Model-Based Diagnostic Frameworks for Fault Detection and System Monitoring in Nuclear Engineering Systems

by

Tat Nghia Nguyen

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Nuclear Engineering and Radiological Sciences)
in the University of Michigan
2020

Doctoral Committee:

Professor Thomas J. Downar, Co-Chair
Dr. Richard Vilim, Co-Chair, Argonne National Laboratory
Associate Professor Eric Johnsen
Professor Annalisa Manera
Associate Professor Necmiye Ozay
Professor Emeritus Neil E. Todreas, Massachusetts Institute of Technology

Tat Nghia Nguyen

nghiant@umich.edu

ORCID iD: 0000-0003-2429-8908

To the teachers in my life.

Acknowledgements

First and foremost, I would like to express my gratitude to my advisor, Prof. Thomas Downar, for his patience and continuous support during my time here at UMich. His passion and vision for multi-physics simulations have provided me with great motivation and guided me through my Ph.D. study. Under his guidance, I have had the opportunity to take part in many interesting projects on various aspects of the field, from the NEUP projects on the fuel performance code BISON, the neutronic code PROTEUS to the current project on the diagnostic code PRO-AID. The experiences from these projects have been invaluable for the development of my career.

I am grateful to my supervisor at Argonne National Laboratory and the co-chair of my dissertation committee, Dr. Richard Vilim, for his trust in me and allowing me the opportunity to work on this project. I know how important it is to him personally. I have benefited from his immense knowledge, particularly on industrial applications of the field. I also want to thank the other members of the committee, Prof. Eric Johnsen, Prof. Annalisa Manera, Prof. Necmiye Ozay and Prof. Neil Todreas for their time reviewing my thesis and for their valuable advices.

I have always considered myself lucky that everywhere I go, from middle school to graduate school, I have always had the chance to meet with teachers and mentors who went beyond their responsibilities to provide me with the support and motivation on my journey. I want to take this opportunity to express my utmost gratitude to Prof. Downar, Prof. Todreas, my teachers in Vietnam, thay Nga, thay Thuc, and all the teachers in my life. Without them I would not be the person I am today, and I am forever in debt. This thesis is dedicated to them.

Finally, I want to thank my sister and brothers for their continuous support and my parents for their lifelong sacrifices in raising us. Many thanks to my friends for their companionship and help during my study. A special mention to Hoang for giving me a place to stay during my time visiting Argonne.

Table of Contents

Acknowledgements iii

List of Tables vii

List of Figures viii

List of Appendices x

Abstract xi

Chapter 1 Introduction and Motivation 1

1.1 Overview of the Diagnostic Problem 4

1.2 Thesis Outline 6

Chapter 2 Fault Detection and Diagnosis Methods 8

2.1 Industrial Applications of FDD Methods 9

2.2 Fundamentals of Model-Based Diagnosis 10

2.3 The Qualitative Approach in PRO-AID 19

2.3.1 Qualitative Physics Based on Confluences 20

2.3.2 Qualitative Models and the Knowledge Base in PRO-AID 22

2.3.3 Limitations of the Qualitative Reasoning Approach in PRO-AID 23

2.4 Quantitative Model-Based Diagnosis Frameworks 24

2.4.1 Quantitative Reasoning Frameworks 24

2.4.2 Model Construction and Residual Generation Approaches 27

Chapter 3 Physics-based Component Models 30

3.1 Physics-based Model Construction 31

3.1.1 Mass Models 33

3.1.2 Momentum Models 33

3.1.3 Energy Models 35

3.2 Residual Generation 39

3.3 Example – Fault Diagnosis of a Single-phase Heat Exchanger 42

3.3.1 Fault Diagnosis Using the MBD Framework.....44
3.3.2 Fault Diagnosis Using the FDI Framework.....46
Chapter 4 Uncertainty Treatment.....49
4.1 Uncertainty Sources and Effects.....50
4.2 Reasoning Under Uncertainty.....51
4.3 Statistical Change Detection.....52
4.4 Deterministic Reasoning Approaches.....54
Chapter 5 Probabilistic Model-based Diagnosis Framework.....58
5.1 Probabilistic Reasoning using Bayesian Network.....59
5.2 Probabilistic Reasoning for Quantitative Model-Based Diagnosis.....62
5.3 Results for the Heat Exchanger Example.....66
Chapter 6 System Level Diagnosis.....71
6.1 Virtual Sensors.....72
6.1.1 Type I Virtual Sensors.....74
6.1.2 Type II Virtual Sensors.....75
6.2 Aggregate Models.....76
6.3 Fault Diagnosis at the System Level.....78
Chapter 7 Results – High Pressure Feedwater System.....82
7.1 System Description.....82
7.2 Virtual Sensors and Balance Equations.....89
7.3 Model Construction and Residual Generation.....91
7.4 Diagnostic Results.....96
7.4.1 Fault Scenarios.....96
7.4.2 Effects of Uncertainty.....100
Chapter 8 Summary, Conclusions, and Future Work.....108
8.1 Summary.....108
8.2 Future Work.....111
Appendix A Regression Methods for Model Calibration.....113
Appendix B Statistical Change Detection Methods.....117
B. 1 Overview.....118
B.2 Generalized Likelihood Ratio Test (GLRT).....119
B.2.1 GLR Test for Sustained Shift in Process Mean (GLR-S).....121

B.2.2 GLR Test for Linear Drift in Process Mean (GLR-D) 122
B.2.3 Examples 123
B.2.4 Decision threshold 125
Appendix C List of Faults for the HP Feedwater System 127
Bibliography 129

List of Tables

Table 2-1. The MBD framework for quantitative model-based diagnosis 25

Table 2-2. The FDI framework for quantitative model-based diagnosis..................................... 26

Table 3-1. Structure of the residuals in the HX Example 43

Table 3-2. Minimal conflicts for a scenario with residuals $(1,1,1,1,0,1,1,1)^T$ 44

Table 3-3. FDI reasoning process for a scenario with residuals $(1,1,1,1,0,1,1,1)^T$ 46

Table 3-4. Fault signatures for the HX example under the notion of exoneration 47

Table 4-1. Reference operating conditions and geometry specifications of the Braidwood CVCS regenerative HX 55

Table 5-1. Conditional probability tables for discrete faults and residuals 64

Table 5-2. The proposed probabilistic reasoning framework for quantitative model-based diagnosis 66

Table 5-3. Results for the two diagnostic scenarios for the single-phase HX 69

Table 7-1. Available diagnostic models for the high-pressure feedwater system. 92

Table 7-2. Model residuals for the high-pressure feedwater system. 95

Table 7-3. Diagnostic result for the case of fouling in FWH 1A.. 98

Table 7-4. Diagnostic result for the case of sensor E2.T fault.. 100

Table 7-5. Mean and standard deviation of each residual when the system is fault-free 101

Table 7-6. Change detection and diagnostic outputs for the case of a sensor fault at E2.T 104

Table 7-7. Fault posterior probabilities by the probabilistic reasoning framework for the case of a sensor fault at E2.T 107

Table A-1. Least squares estimation of the pressure head model parameters 116

Table C-1. List of faults for the high-pressure feedwater system.. 128

List of Figures

Figure 1-1. Simplified P&ID of the Braidwood Nuclear Station CVCS [8] 5

Figure 2-1. Forward mapping from physical states to fault signatures....................................... 13

Figure 2-2. Backward mapping from a set of fault symptoms to possible states in fault diagnosis 14

Figure 4-1. Model-based diagnosis framework using physics-based component models........... 49

Figure 4-2. Application of the GLR-D test to detect non-zero residual 54

Figure 4-3. GLR-D decision functions for the 8 residuals in a sensor fault scenario (left) and a zoomed-in version (right). 56

Figure 5-1. Bayesian network representation of the system with a single fault and single observation.. 61

Figure 5-2. General Bayesian network for quantitative model-based fault diagnosis................. 62

Figure 5-3. A Bayesian network for fault diagnosis in the single-phase heat exchanger example 66

Figure 6-1. Overall scheme of the proposed system level diagnostic framework....................... 80

Figure 6-2. Dependency structure of a model residual in system level diagnosis....................... 81

Figure 7-1. The P&ID of a high-pressure feedwater system 83

Figure 7-2. Steam and water flow paths in a typical vertical high-pressure feedwater heater [98] 85

Figure 7-3. Simplified schematic of the vertical HP feedwater heater. 86

Figure 7-4. Type I virtual sensors for the two first-point FWHs.. 90

Figure 7-5. Fouling in FWH 1A causes two non-zero residuals (right plot) while the other seven remain unaffected (left plot). 96

Figure 7-6. Effect of fouling on the feedwater outlet temperature of FWH 1A 97

Figure 7-7. A simulated fault for the feedwater inlet temperature sensor of FWH 1B 99

Figure 7-8. Sensor fault at E2.T causes four non-zero residuals (right plot)............................... 99

Figure 7-9. The two noisy non-zero residuals under the effect fouling in FWH 1A 102

Figure 7-10. GLR-D decision functions for residual evaluation in the case of fouling in FWH 1A 102

Figure 7-11. The four residuals affected by the temperature sensor fault at the feedwater inlet of FWH 1B 103

Figure 7-12. GLR-D decision functions in residual evaluation for the case of sensor fault at E2.T 103

Figure 7-13. A Bayesian network for the high pressure feedwater system. 105

Figure A-1. Flowrate for a feedwater pump during startup 115

Figure A-2. Pressure head by the feedwater pump during startup.. 115

Figure A-3. Fitted pressure head model for the feedwater pump 116

Figure B-1. A shift in process mean (left) and a drift in process mean (right).......................... 118

Figure B-2. Decision function by GLR-S and GLR-D for a shift in process mean of $\Delta\mu = \sigma$. 124

Figure B-3. Estimation of change point location by GLR tests for a shift of $\Delta\mu = \sigma$ 124

Figure B-4. Probability density function of the GLRS decision function 125

Figure B-5. ARL of the GLRS algorithm as a function of decision threshold. 126

Figure B-6. GLRS false alarm rate as a function of decision threshold. 126

Figure C-1. A Bayesian network for drain pump 1A... 127

List of Appendices

Appendix A Regression Methods for Model Calibration 113

Appendix B Statistical Change Detection Methods.. 117

Appendix C List of Faults for the HP Feedwater System... 127

Abstract

The high operations and maintenance (O&M) cost for nuclear plants is one of the most significant challenges facing the industry today. The research in this thesis is motivated by the ongoing effort to utilize automation and improved operator support technologies to reduce O&M costs in nuclear power plants. A diagnostic framework is first developed for the problem of monitoring equipment health and sensor calibration status in nuclear engineering systems. This is achieved by utilizing real-time data from sensors that are already in place for system monitoring to perform automated diagnostics of equipment degradation. Given the long-time scale over which component degradation typically proceeds, some of the sensors may also inevitably degrade and become unreliable. The need to simultaneously consider equipment and instrument faults is both a technical necessity and a desired capability. The automation of these monitoring tasks contributes to the reduction of the overall O&M cost by reducing the required human resources and by providing better maintenance scheduling.

Early detection of slow degradation over the course of plant operation requires sufficient detection sensitivity from the diagnostic framework. The problem is more complicated in the presence of various sources of uncertainty and possible changes of operating conditions due to plant drifts. To resolve these difficulties and provide the desired capability, the proposed framework is a hybrid integration of quantitative model-based diagnosis, statistical change detection and probabilistic reasoning. Physics-based models are developed to describe the fault-free behavior of system components. Quantitative residuals are generated from the analytical redundancy in each model and serve as fault symptoms for model-based diagnosis. Statistical change detection methods are used to detect changes in the residuals in the presence of uncertainty. Measurement and modelling uncertainty are robustly treated by methods of statistical change detection and probabilistic reasoning. A system level diagnosis framework is proposed to deal with the lack of local sensors to each component.

The overall framework has been implemented and demonstrated with a high-pressure feedwater system whose available sensor set is insufficient for the construction of standalone models for most major components. Results from the demonstration showed that the system level approach can be used to construct models and perform diagnostics for systems with limited instrumentation. Both component faults and sensor faults can be detected, and the effects of uncertainty can be mitigated by the proposed probabilistic reasoning framework. Areas for future work were identified and include the investigation of a dynamic Bayesian network to treat the effects of uncertainty in the diagnosis as well as the investigation of using high fidelity simulation codes to construct simulation-based surrogate models of the basic plant components.

Chapter 1

Introduction and Motivation

The global energy demand is rising rapidly as a result of population and economic growth. A recent report by the U.S. Energy Information Administration projects a 50% increase in world energy consumption between 2018 and 2050 [1]. At the same time, the impact of greenhouse gases emission from energy production is an issue of global concerns, motivating the search for cleaner energy sources. From a technical perspective, nuclear energy can be a crucial part of the solution to both issues. However, to the opposite effect, nuclear power is struggling and its contribution to the energy solution does not live up to its potential.

The main challenges facing the nuclear industry originated from both safety and cost concerns. The economic competition from other lower-cost alternatives makes nuclear energy, even with its numerous benefits, less desirable. To improve the viability of the nuclear power, technology advancements must be made in both safety enhancement and cost reduction [2]. Parallel with progress in the development of advanced fuel cycle and reactor design, advances in operator support technologies are crucial for improving the safety and efficiency of nuclear power plant operations.

The use of operator support technologies is twofold: at the control level, a computerized operator support system can assist plant operators in making timely and informed decisions; at the maintenance level, it can help monitoring the overall equipment condition and system status, improving maintenance scheduling and reducing the need for human resources [3, 4]. The study in this thesis is part of the effort to utilize automation and operator support technologies in reducing the operations and maintenance (O&M) cost in nuclear power plants. Fault diagnosis methods are investigated and developed for application to the problems of monitoring equipment health and instrumentation calibration status.

More specifically, the focus of the diagnostic problem is on equipment performance degradation in the thermal-hydraulic systems over the course of plant operation. Information regarding the status of the system can be collected and analyzed periodically from sensor data. Given the long-time scale over which component degradation typically proceeds, some of the system monitoring sensors may also inevitably degrade and become unreliable. Sensors in nuclear power plants are typically only calibrated once every fuel cycle. The calibration activity requires significant human resources in detecting faulty instruments and recalibrating them. Automated on-line calibration monitoring can be performed during plant operation to detect sensor drifts as they occur and ultimately reduce the O&M cost [5]. It is therefore desirable to have a diagnostic tool with the capability to simultaneously deal with both component faults and sensor faults.

The principal objective of this work is the development of a theoretical framework for performing fault detection and diagnosis in the thermal-hydraulic systems of nuclear power plants. Basic capabilities for this purpose were previously developed at Argonne National Laboratory in the computer code PRO-AID, originally known as PRODIAG [6, 7, 8]. This thesis will build upon PRO-AID by developing a new diagnostic framework to overcome its limitations and improve its capability and applicability. The principal target application for the approach developed here is for immediate implementation in currently operating nuclear power plants, however the methods developed here also would potentially have application for the design and operation of advanced nuclear reactor systems. Practical conditions in terms of available sensor sets and other available data and information are taken into consideration in the development. Afterwards, the diagnostic framework may be applied to the problem of determining the optimal placement for new sensors to improve the monitoring capabilities.

PRO-AID [6] was designed as a rule-based expert system for the detection and diagnosis of upset events in thermal-hydraulic (T-H) systems of nuclear power plants, relying exclusively on T-H instrumentation readings. Faults or malfunctions in any component of a system would result in changing trends of T-H variables. PRO-AID emulates a “human expert” observing individual changing trends in T-H variables and makes diagnostic decisions using a knowledge base provided by a collection of reasoning rules. Changing trends are considered fault symptoms and the knowledge base can provide inferences based on the symptoms allowing the code to deduce the nature of the fault. The knowledge base of PRO-AID is constructed based mainly from the qualitative form of conservation equations. The approach in PRO-AID is function-oriented in the sense that component malfunctions are expected to affect the component's ability to perform its design function. Since each generic component in a T-H system is designed to perform a specific function of either mass, momentum or energy transfer, a component malfunction would lead to an imbalance in at least one of the conservation equations.

The rule-based qualitative reasoning approach in PRO-AID offers several advantages. If the knowledge base is complete and the relevant symptoms are observed, the code can execute the reasoning process efficiently and mistake-free. In this sense, the code can operate as human experts without the human error factors. On the other hand, since the code only observes the system qualitatively, there is a limit on what it can deduce, especially when observations are limited to T-H variable signals. Since some information is lost when the conservation equations are converted into qualitative form, the instrumentation signals available may not be fully utilized. In addition, as the code considers all changing trends as direct fault symptoms, it can only be applicable for situations when that assumption is valid, i.e. for faults that are severe enough to cause direct detectable changing trends and only in a short time window before feedbacks and automated control actions come into effect and begin to overlay the symptoms. Further complications arise when one must consider system noise and measurement uncertainties which make the task of detecting correct changing trends nontrivial. In general, the current version of PRO-AID can only detect abrupt faults and is not applicable for slow degradations in the time frame during which where feedbacks and control system actions may interfere.

Considering the limitations of the qualitative approach in PRO-AID, a new quantitative model-based approach is developed and presented in this thesis. In developing this new approach, similar restrictions as imposed in the previous version of PRO-AID are applied: to rely solely on instrumentation signals and not require any component-specific design parameters. This is due to the business case of the target application. Ultimately, the objective for the development of fault detection and diagnosis tools such as PRO-AID is to reduce the cost of operations and maintenance activities in nuclear power plants. For that purpose, having the approach involve design parameters that require subject matter experts for setup and maintenance would be economically counterproductive.

The most significant original contribution of this work to the field is the development of a physics-based probabilistic framework for system level diagnosis in complex nuclear engineering systems with limited instrumentation. The proposed framework has the capability to deal with both equipment faults and instrument faults. For most systems with limited instrumentation, because of the lack of sensors locally at component level, it is not possible to perform diagnostics for every standalone component which limits the overall detection and monitoring capability. The framework proposed in this thesis resolves this limitation by utilizing the relations between nearby components and sensors available at the system level. Furthermore, one of the challenging issues in engineering applications of model-based diagnosis methods is the difficulty in developing sufficiently accurate models for the diagnostic purpose. High modeling uncertainty may result in unreliable diagnostic results which limit the applicability of model-based methods. In this thesis, a probabilistic reasoning framework using the Bayesian network method has been developed to robustly deal with modeling uncertainty and other inevitable sources of noise and uncertainty in engineering systems.

1.1 Overview of the Diagnostic Problem

Information on the structure of a T-H system will be specified by a piping and instrumentation diagram (P&ID). An example of a P&ID is shown in Figure 1-1 for the chemical and volume control system (CVCS) of the Braidwood Nuclear Generating Station [8]. The P&ID provides a list of all components and sensors as well as specify their locations, interconnections and the fluid flow directions. The basic objective of fault diagnosis in a system is to detect when a fault has occurred and if possible, localize the fault to a specific component or sensor.

A fault is defined to be any change in a component or sensor that affects its performance of the designed function. Generally, faults in a nuclear power plant can be divided into two categories based on the underlying time scale. The first type, occurring abruptly in a short-time scale, is relevant to the control of the plant and thus must be detected and dealt which by human operators in the control room. The second type is performance degradation, occurring slowly in the long-time scale over the course of plant operation. Slow performance degradation may go undetected even with plant staff performing routine maintenance rounds and is typically dealt with during periodic maintenance intervals. The current research will focus on the second category - faults of slow degradation type.

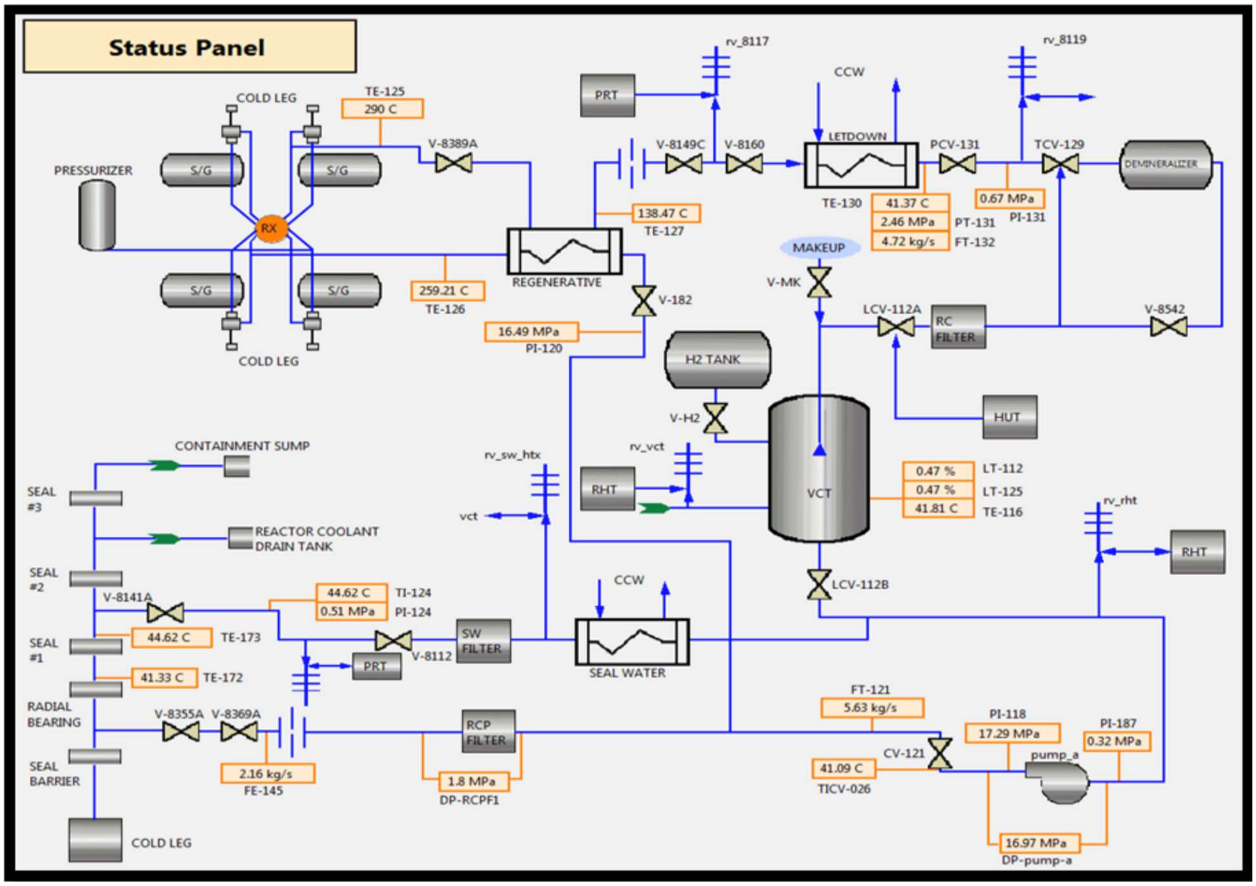

Figure 1-1. Simplified P&ID of the Braidwood Nuclear Station CVCS [8]

Given the long-time scale in consideration, the degradation of the monitoring instruments is also inevitable, e.g. sensors drifting out of calibration. Additionally, the system may undergo changes in operation conditions due to control actions or other changes in boundary conditions such as seasonal temperature changes. The problem is more complicated with the presence of various sources of noise and uncertainty. The desired diagnostic approach should have the capability to deal with both component faults and sensor faults, have high detection sensitivity to detect slow degradations but remain insensitive to the various sources of uncertainty and changes in operating conditions.

The theoretical framework for fault detection and diagnosis (FDD) proposed in this thesis is a hybrid integration of quantitative model-based diagnosis, statistical change detection and probabilistic reasoning. Quantitative physics-based models are constructed to describe the normal, fault-free, behavior of each component in a T-H system. These component models provide the source of analytical redundancy needed to perform fault diagnostics. Discrepancies between fault-free model predictions and measurement data provided by sensor readings are quantified by model residuals. Non-zero residuals are interpreted as fault symptoms in fault detection and diagnosis. With the presence of various sources of uncertainty, a statistical change detection method is needed to reliably evaluate whether a residual has become non-zero statistically. Every time a decision is made on whether a residual is zero or non-zero, there is an associated false detection rate. Taking account of such false detection possibilities in a probabilistic reasoning approach is one of the issues addressed in this work.

The construction of each component model for fault diagnosis requires measured data of a certain number of process variables at the inlet and outlet of the component. In practice, it is rarely the case that there are enough sensors on the boundary of each component for that purpose. Incorporating information from the system level into the model construction process and subsequently in fault diagnosis is one of the original contributions of this thesis. Difficulties due to the lack of sensors are dealt with in system level diagnosis using aggregate models and a new concept of virtual sensor.

1.2 Thesis Outline

Chapter 2 starts with a review on various fault detection and diagnosis methods and their applications with focus on model-based diagnosis methods. The qualitative approach in PRO-AID is discussed next along with its advantage and limitations. The last section of Chapter 2 introduces two quantitative model-based diagnosis frameworks, each provides a reasoning process to obtain fault diagnoses from a set of observed fault symptoms. These two reasoning frameworks will be used as the basis for the work in this thesis.

Chapter 3 deals with the construction of physics-based models for model-based diagnosis and the difficulties involved. From each model, one or several residuals can be computed and utilized in detecting fault symptoms. An example with a single-phase counterflow heat exchanger is discussed in detail to illustrate the model construction and residual generation processes.

In Chapter 4, various sources of uncertainty in a system and their effects on fault detection and diagnosis are discussed. The presence of noise and uncertainty affects not only the process of observing fault symptoms from sensor readings but also the reasoning process going from a set of observed fault symptoms to possible fault diagnoses. Statistical change detection methods to detect non-zero residuals are introduced. If the false detection rates in evaluating model residuals can be considered negligible, the straightforward approach for the reasoning process is to accept the set of observed fault symptoms as definite and proceed with one of the two reasoning frameworks introduced in Chapter 2. Such approach is called ‘deterministic reasoning’ as the possibility that any observed fault symptom could be false is neglected.

Chapter 5 introduces the probabilistic reasoning framework proposed in this study. The possibility of false detection rates in evaluating the model residuals are estimated and accounted for in the reasoning processes. Details on the concept of Bayesian network and the application in probabilistic reasoning are discussed.

Chapter 6 discusses the problem of fault diagnosis at system level. As is often the case, the sensor set available at the boundary of a component is incomplete hence it is not possible to construct a model for the standalone component. One must utilize the relations between nearby components and make use of all available sensors at the system level to maximize the diagnostic capability. The concepts of virtual sensors and aggregate models are introduced for that purpose. Issues with residual generation and fault diagnosis at system level are also discussed.

Chapter 7 provides results for several diagnostic scenarios for a feedwater system in the North Anna Nuclear Generating Station. The overall process from importing and analyzing the P&ID for possible virtual sensors to constructing all possible diagnostic models, generating model residuals and performing diagnosis has been automated in a test implementation. Diagnostic results for various scenarios, including feedwater heat fouling, sensor fault and pump fault are discussed. Possible application of the proposed framework on the problem of determining optimal placements for new sensors is also illustrated.

The final chapter provides a summary of the work and proposes various directions for future work.

Chapter 2

Fault Detection and Diagnosis Methods

Fault detection and diagnosis (FDD) is an essential part of process control and abnormal event management in any engineering and industrial system. A fault is any change to the characteristics of some component in the system that could reduce or disrupt its ability to perform the designed function. Fault detection techniques are usually employed to monitor the system to ensure its efficient working conditions and enhance system safety and reliability. The use of advanced FDD techniques for early detection of faults would allow for timely execution of remedial control actions and for the orderly planning of maintenance activities.

Generally, most FDD methods detect faults by either analyzing inconsistencies in the observed data or by matching the observed data to known fault modes. Inconsistencies in sensor data can be detected by either hardware or analytical redundancy. In a hardware redundancy approach, identical instruments or components are used for the same purpose and their outputs are compared for cross validation. Hardware redundancy is costly and therefore only feasibly applicable for safety-critical applications. The analytical redundancy approaches, on the other hand, rely a priori knowledge of the system for consistency checking. Alternatively, faults can be detected by matching the observed data against known fault models or signal features.

Reviews of FDD methods are available in a great variety of books [9, 10, 11, 12, 13, 14, 15, 16] and journal papers [17, 18, 19, 20, 21, 22, 23, 24, 25, 26]. In general, FDD methods can be classified into two main categories: model-based and process-history-based. Briefly, in model-based approaches, mathematical models of the system or its individual components are constructed to either provide a source of analytical redundancy or describe the system faulty behaviors. Process-history-based approaches include data-driven methods and signal-based methods. For data-driven methods, the availability of a large amount of process history data is required. Various data processing and pattern recognition techniques can be applied to formulate constraints between different sensor readings in the observed data. Signal-based methods on the other hand are univariate and rely on numerical features of individual sensor readings, e.g. vibration or acoustic sensors.

In practice, there is not always a clear distinction between these methods and the advantage of one over the others. For instance, system models can be constructed by data-driven techniques to be used in model-based methods. When considering various FDD techniques for industrial applications, a common consensus is that there is no single method that can satisfy all the desirable features for a specific application. The optimal solution can only be reached by combining different methods in a hybrid form [17, 23].

2.1 Industrial Applications of FDD Methods

In data-driven approaches, multivariate data analysis and supervised learning techniques are applied to large sets of process history data to detect process faults. Data-driven methods can be divided into two categories, namely pattern recognition approaches and data reconstruction approaches [19].

In the first category, the problem of fault diagnosis is formulated as a pattern recognition problem or a classifier. Quantitative features are exacted from the set of process history data and classified by different classes, each of which is associated with a specific type of faults. Faults are detected when the features of the observed data are recognized by a pre-determined class. Feature extraction methods include principle component analysis (PCA) [27, 28, 29, 30, 31], independent component analysis (ICA) [32, 33, 34, 35, 36], partial least square (PLS) [37, 38, 39, 40], linear discriminant analysis (LDA) [41, 42]. Classifier methods include support vector machine (SVM) [43, 44, 45, 46, 47], artificial neural network (ANN) [48, 49, 50, 51].

In the second category of data-driven approaches, quantitative models obtained from process data are used to reconstruct part of the observed data. Faults are detected from the discrepancies between measured data and the model estimations. Methods for data reconstruction include ANN and other supervised machine learning techniques [19] and multivariate state estimation technique (MSET) [52, 53, 54]

Signal-based approaches utilize univariate sensor signals for fault diagnosis, as opposed to multivariate process data in data-driven approaches. Numerical features are extracted from individual sensor readings from which faults can be detected and recognized. Signal features can be classified into three categories: time-domain, frequency domain and joint time-frequency domain. Time-domain features can be extracted from continuous dynamical process variables usually in terms of statistical parameters [55, 23, 56]. Frequency-domain features can be exacted using spectrum analysis tools from relevant sensors, e.g. vibration or acoustic sensors. Frequency-domain signal-based approaches have been widely applied for fault detection using in pump motors and other rotating machineries using vibration sensors [57, 58, 59], in industrial systems using acoustic sensors [60, 61]. In joint time-frequency domain approaches, time-variations of frequency-domain features are monitored and from which faults can be detected [62, 63, 64].

In model-based approaches, models are constructed to provide description of the structure and behavior of a system. Faults are detected by analyzing inconsistencies between observed data and the expected behavior for from matching observations to expected faulty features. Inconsistencies in measurement data are detected from available analytical redundancy relations which are quantified by model residuals in quantitative approaches. Depending on the form of a system model, different techniques are available for the residual generation, including fault detection filter [65, 66], diagnostic observer [67, 13], parity space [68, 17, 67], parameter estimation [69, 17].

2.2 Fundamentals of Model-Based Diagnosis

The characteristics of the diagnostic problem in this thesis is the need to consider complex systems with large numbers of components but limited sensor sets. Under such scenarios, it is not always possible to identify the exact root cause of each upset event and the spatial resolution of the diagnostic results may include multiple components or sensors. We will first investigate possible reasoning frameworks to obtain such diagnostic results from a set of observations. The effects of various sources of uncertainty to the reasoning process are discussed in Chapter 4 and 5. For the current section, we will provide formal definitions of *faults*, *fault symptoms*, *fault diagnoses* and discuss the fundamental reasoning framework for model-based fault diagnosis.

Model-based diagnosis (MBD) is a general framework for fault detection and diagnosis making use of models of the structure and behavior of the system in consideration. The idea of model-based diagnosis as a method that uses analytical redundancy for fault detection can be traced back to the work of Beard [65] and de Kleer [70]. The common approach to model-based diagnosis is to rely on models that describe the normal, fault-free, behavior of a system. Faults can be detected from discrepancies between model predictions and observed data [71, 72]. This approach, formally known as *consistency-based*, will be the focus of development in this thesis. It should be noted that there is an alternative approach, known as *abductive diagnosis*, relying on abnormal models that describe the behavior a system in faulty modes. Faults or malfunctions are detected when their predicted abnormal behaviors match the observed data [73, 74, 75].

The logical framework of *consistency-based* approach for model-based diagnosis was introduced and formalized by Reiter [76] and de Kleer [77]. A detailed description of the approach was provided by de Kleer in [13] for a simplified system of an electronic circuit. In this section, we will generalize that framework to treat systems with the possibility of multiple fault modes in system components as well as sensor faults.

To start, a system in model-based diagnosis is defined by a complete description of its structure, a list of fault-free models for the components in the system and a list of observations obtained from various locations in the system:

Definition 2-1. A system is a triple $(SD, COMPS, OBS)$, where:

- SD - System Description: specifies the structure of the system, including a list of all components, sensors, and their interconnections.
- $COMPS$ - System Components: a list of models that describe the normal (fault-free) behavior of each component
- OBS - Observations: List of data, observed at various locations in the system.

Each component model provides an analytical relation between the process variables on the boundary of the component. The component models together constitute a fault-free system model which imposes various relations and constraints on the observed data. For the formulation in this section, we shall assume that all component models are provided. Furthermore, all sources of uncertainty, including modeling uncertainty and measurement uncertainty, are neglected. Complications in the development of component models and uncertainty treatment will be discussed in Chapter 3 and 5, respectively.

A component *fault* is any change in the characteristics of a component that can cause it to deviate from the expected normal behavior. A sensor *fault* is any change to a sensor that causes its reading value to no longer reflect the true value of the underlying process variable. If the system is fault-free, the observed data provided by sensor readings must satisfy the relations and constraints imposed by the fault-free system model. In that case, we say that the observations are *consistent* with the fault-free system model. Consequently, any inconsistency between the set of observations and the fault-free system model is defined to be a *fault symptom*.

Definition 2-2. A *fault symptom* is an inconsistency detected between the set of observations and the fault-free system model.

Following the underlying physics, each physical state of the system would result in a specific combination of *fault symptoms*. That specific combination of fault symptoms will be referred to as the *fault signature* of the state. An upset event is detected when one or more *fault symptoms* are observed. The objective of fault diagnosis is then to deduce the state of the system from a set of observed *fault symptoms*.

To elaborate, suppose that there are N possible faults in the system, each can be denoted by a label F_i with $1 \le i \le N$. Each physical state of the system can be then identified by a set of faults. Thus, the space of physical states consists of 2^N possible states. Now, suppose that the set of available observations and the fault-free system model allow us to construct n distinct fault symptoms. There are then 2^n possible combinations of fault symptoms.

The causal relations from the underlying physics dictates that to each physical state of the system, there is a specific set of fault symptoms defined to be its *fault signature*. That is, one can theoretically construct a *function* that maps each set of faults to a specific set of fault symptoms. The term *function* is to emphasize that each set of faults is mapped to one and only one set of fault symptoms. Each physical state cannot have more than one signature. On the other hand, it is possible that different physical states can share the same *fault signature*, i.e. different sets of faults can be mapped to the same set of fault symptoms.

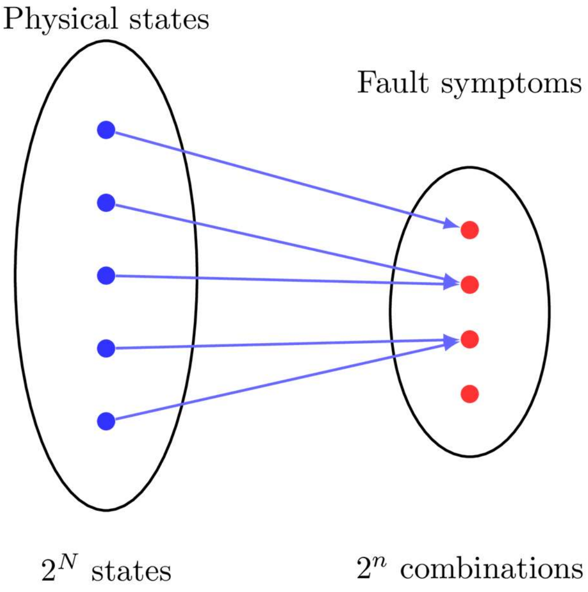

Figure 2-1. Forward mapping from physical states to fault signatures

The forward mapping from physical states to fault signatures in the space of combinations of fault symptoms is illustrated in Figure 2-1. Mathematically, the map is injective but not surjective. Every physical state is mapped into a fault signature but not every combination of symptoms is the signature of a physical state. If a set of fault symptoms is a fault signature to some state, we say the set of fault symptoms *represents* the state. Then among the 2^n possible combinations of fault symptoms, some combinations do not represent any physical states while the others may represent one or multiple states. Usually, the number of distinct fault symptoms in the system is less than the number of possible faults, thus $2^n < 2^N$, and we can see clearly the mapping from 2^N distinct states to 2^n different sets of symptoms cannot be one-to-one.

The objective of fault diagnosis is to determine the physical state of a system given a set of observed *fault symptoms*. As we have seen, different states can give the same fault signature and therefore it is not always possible to identify the exact state of a system. One must settle with all possible states that can yield the set of observed symptoms. Each point in the space of physical states can be a guess for the actual state of the system. More formally, we define a *diagnosis* to be any hypothesis on the state of a system in the attempt to explain the observed fault symptoms. A *diagnosis* is *valid* if the corresponding state yields the observed set of *fault symptoms*.

Mathematically, the task of determining all *valid diagnoses* is equivalent to finding the inverse mapping from the space of fault symptoms to the space of physical states.

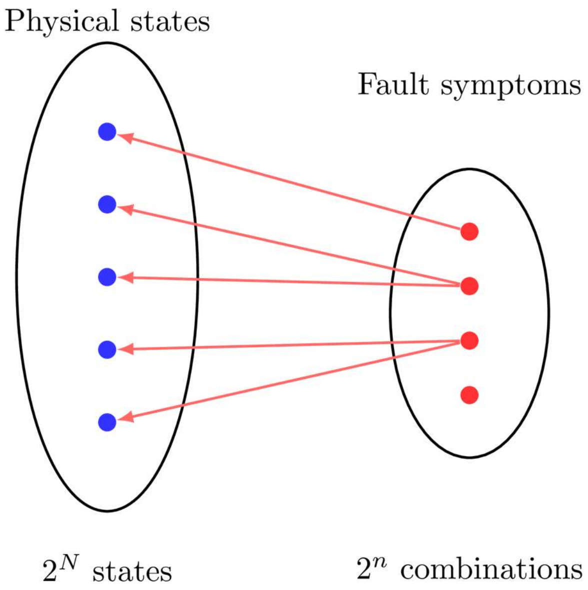

Figure 2-2. Backward mapping from a set of fault symptoms to possible states in fault diagnosis

If one can enumerate all the faults in the systems and can construct the *forward mapping*, as shown in Figure 2-1, from each state to a set of symptoms, then the task of fault diagnosis can be done straightforwardly. All valid diagnoses can be found by a simple search for all physical states whose fault signature matches the observed set of symptoms. However, it is usually not practical to construct the forward mapping, especially for complex systems with high number of continuous fault modes. Furthermore, it is not computationally efficient to perform diagnosis by enumerating all possible states of the system. As shown, the number of possible states increases exponentially with the number of faults.

The objective of the MBD framework formulated here is to determine all *valid diagnoses* for a given set of symptoms without the need to construct the forward mapping by employing a reasoning method known as *backward chaining inference*. *Valid diagnoses* are logically inferred from the implications provided by the observed fault symptoms.

For that, we first need to express each *diagnosis* in the form of a logical statement. Intuitively, each *diagnosis* with a set of faults is a logical statement claiming those faults to be present in the system. Any component or sensor not implicated by this set of faults is implicitly presumed to be fault-free.

Definition 2-3. A *diagnosis* with a set of faults C_f is the hypothesis that the faults included in C_f have occurred and the rest of the system is fault-free.

$$\mathcal{D}(C_f) \equiv \left(\bigwedge_{F_i \in C_f} F_i \right) \wedge \left(\bigwedge_{F_i \notin C_f} \neg F_i \right) \tag{2.1}$$

where $\wedge$ is the logical 'AND' operator; $\neg$ is the logical negation and F_i is the label for a particular fault.

We will use a short-handed notation to write each *diagnosis* by a list of fault labels inside the square brackets '$[\cdot]$'. For instance, $[F_1, F_2]$ is the *diagnosis* claiming that only faults F_1 *and* F_2 have occurred.

Again, at any moment of time, there can be only one *diagnosis* matching the actual state of the system, but we must settle with all the diagnoses that are consistent with the observed symptoms. We defined those to be *valid diagnoses*, or more formally, *consistency-based diagnoses*. More specifically, a diagnosis is called a *consistency-based diagnosis* if its set of faults can account for all the inconsistencies between the observations and the fault-free system model.

For each abnormal event detected by a specific set of observed fault symptoms, we are interested in obtaining the list of all *consistency-based* diagnoses. From a practical point of view, however, it might be helpful to narrow such list down by removing some less useful diagnoses at the tradeoff of some comprehensiveness. That is, it is not practically useful to consider all mathematically valid solutions and one may consider removing some of the less probable diagnoses. In the current framework, we will do this by introducing the concept of *minimal diagnosis* and consider keeping only *minimal diagnoses* in the diagnostic result.

To see the logic behind the use of *minimal diagnoses*, consider a scenario in which we have obtained a list of all *valid* diagnoses for a given set of fault symptoms. Suppose that among the list of *valid* diagnoses, there exists a group of related diagnoses $\mathcal{D}(C_0), \mathcal{D}(C_1), \ldots, \mathcal{D}(C_n)$ such that the first member of the group is a subset of all the other members, i.e. $C_0 \subset C_i$ for all $i \geq 1$. Every diagnosis in this group is just $\mathcal{D}(C_0)$ plus at least one additional fault. In other words, for any diagnosis $\mathcal{D}(C_i)$ in the group with $i \geq 1$, we can remove at least one fault from its set of faults C_i and still have a *valid* diagnosis. The same cannot be said for $\mathcal{D}(C_0)$; we cannot remove any of its faults and still retain a *valid* diagnosis. In this case, $\mathcal{D}(C_0)$ is called a *minimal diagnosis*.

From a practical point of view, for the purpose of reducing the number of possibilities one needs to consider in the diagnostic result, it is a reasonable choice to focus only on the *minimal diagnosis* $\mathcal{D}(C_0)$ and ignore the other *non-minimal* members of the group. In terms of actionable information, every diagnosis in the group implies the set of faults C_0 in $\mathcal{D}(C_0)$. In terms of prior probability, assuming the faults are independent, it is clear that the *minimal diagnosis* $\mathcal{D}(C_0)$ is the most likely candidate among the group, regardless of what the prior probability of each fault might be. The prior probability of each diagnosis $\mathcal{D}(C_i)$ in the group with at least one additional fault compared to $\mathcal{D}(C_0)$ is smaller than the prior probability of $\mathcal{D}(C_0)$. (To be precise, the requirement for this statement to hold is that the prior probability of each fault is less that 50% which we can safely assume to always be the case for all practical purposes).

Formally, a valid *consistency-based* diagnosis $\mathcal{D}(C_0)$ is called a *minimal diagnosis* if there is no proper subset C'_0 of C_0 such that $\mathcal{D}(C'_0)$ is also a *consistency-based* diagnosis. Given a list of all valid diagnoses, we can divide it to separate groups of related diagnoses as described and from that obtain all *minimal diagnoses*. It should be noted that the term "*minimal*" here does not necessarily imply a minimum number of faults. For instance it could be the case that both $[F_1, F_2]$ and $[F_3]$, for some fault labels F_1, F_2, F_3, are valid minimal diagnoses and without considering the prior probabilities, one has no basis to prefer one minimal diagnosis to another.

Definition 2-4. A *minimal diagnosis* is a valid consistency-based diagnosis such that none of its proper subsets is also a valid diagnosis.

The use of the minimal diagnosis concept allows one to logically remove some of the less useful diagnoses, thus reduces the number of diagnoses one needs to consider. The tradeoff is that in the unlikely event that multiple faults occur in a system and the actual state of the system corresponds to a non-minimal diagnosis then by only considering minimal diagnoses, one may miss some of the multiple faults.

The objective of model-based diagnosis is then to obtain all *minimal diagnoses* for a given set of observed fault symptoms. In framework of MBD, *minimal diagnoses* are deduced by logical inference using the information obtained from each fault symptom. Each fault symptom represents an inconsistency between the observations and the fault-free system model. From each fault symptom, one can conclude that at least one of the involved components or sensors must be faulty. Such statement is known as a *conflict*. Formally, a *conflict*, identified by a set of faults C_f , is defined to be the logical statement claiming at least one fault in C_f must be true.

Definition 2-5. A conflict for a set of faults C_f is the statement that at least one of the faults included in C_f must have occurred.

$$\mathcal{C}(C_f) \equiv \bigvee_{F_i \in C_f} F_i \tag{2.2}$$

where $\vee$ is the logical 'OR' operator.

In that context, a conflict among some set of fault C_f is valid if the observed fault symptoms cannot be explained without at least one fault in C_f. Parallel to the definition of *minimal diagnosis*, a *minimal conflict* is defined to be a valid *conflict* such that none of its proper subsets is also a valid *conflict*. One cannot remove any fault from the set of faults in a *minimal conflict* without invalidating the associated logical statement.

As an example, suppose that from a certain fault symptom one can logically conclude that at least one fault among the three faults, labeled by F_1, F_2, and F_3, must be true. Then we have a valid conflict among the set of these three faults, conventionally written as $\langle F_1, F_2, F_3 \rangle$. Notice that if one were to pick from the space of all possible conflicts, any conflict with the same three faults plus some additional faults, e.g. $\langle F_1, F_2, F_3, F_4 \rangle$, is also valid as the associated logical statement immediately follows. However, $\langle F_1, F_2, F_3, F_4 \rangle$ is not a *minimal conflict* since one can remove F_4 and still retain a valid conflict.

In summary, each fault symptom provides a logical statement in the form of a conflict. Thus, for each abnormal event, a collection of conflicts can be derived from the observed fault symptoms. A fault diagnosis is valid if it is consistent with every observed conflict. By logical inference, one can then obtain all valid fault diagnoses for the abnormal event.

Alternatively, in the language of set theory, given the set of all *minimal conflicts* derived from the observed fault symptoms, the list of all *minimal diagnoses* can be obtained using the following proposition:

Proposition 2-1. Let Π denote the set of all *minimal conflicts* in a system, a diagnosis $\mathcal{D}(\Delta)$ is a valid *minimal diagnosis* if and only if its set of faults Δ is a *minimal* set to have a non-empty intersection with the set of faults expressed by every *conflict* in Π:

$$\Delta \cap C_f \neq \varnothing, \quad \forall C_f : \mathcal{C}(C_f) \in \Pi \tag{2.3}$$

Δ being a minimal implies that no proper subset of Δ satisfies the same condition.

To summarize, if the logical implication from each fault symptom is known, the model-based diagnosis framework formulated in this section provides an algorithm to obtain all *valid minimal diagnoses* without the need to construct the fault signatures for all possible combination of faults.

The algorithm consists of the following steps:

1) Search for inconsistencies between observed data and fault-free system model. Each inconsistency serves as a *fault symptom,* which gives rise to a *conflict*.

2) Obtain a set of *minimal conflicts* from the observed fault symptoms.
3) Search for all *minimal diagnoses* from the set of *minimal conflicts*, using the proposition given by Eqn. (2.3).

Up until now, we have not discussed how component models and subsequently fault symptoms can be constructed. For the original applications of MBD in electronic circuits, component models are usually known as part of the system specifications. For applications in engineering systems, however, components models are generally not provided and need to be developed for diagnostic purposes. The construction of component models is one of the challenging issues one needs to address in order to apply the MBD framework to complex engineering systems. For our current application for TH systems, the development of physics-based component models will be discussed in Chapter 3.

2.3 The Qualitative Approach in PRO-AID

The idea of qualitative physics based on confluences was introduced by de Kleer and Brown in [78]. Intuitively, when we observe the behavior of a physical system, the instantaneous values of the process variables are often not of interest. The changing trends of the process variables convey most of the information regarding the status of the system. From the qualitative understanding of the system, one can make sense of what the observed trends indicate without performing any calculations or knowing the exact quantitative details of the system.

Wei and Reifman [6] applied this concept to the FDD problem of TH systems. More specifically, the concept of qualitative physics provided an approach to construct qualitative models for each generic type of component in a TH system without the need to know component-specific design parameters. Since each generic component in a T-H system is designed to perform a specific function of either mass, momentum or energy transfer, faults in the component would result in the violation of at least one of the balance equations formulated under normal working conditions. The qualitative form of each balance equation serves as a *qualitative model* for the component.

Each qualitative component model provides several relations between individual variable trends at the inlet and outlet of the component and an imbalance indicator, referred to as a Q value. Each component fault would lead to changing trends in one or several related Q values and thus, changing trends in Q values serve as *fault symptoms* for the reasoning process in fault diagnosis.

Qualitative rules based on the qualitative component models form the knowledge base of PRO-AID. Using the knowledge base, *fault symptoms*, i.e. Q trends, can be inferred from observed process variable trends and from that possible faults can be deduced by applying the MBD reasoning framework.

2.3.1 Qualitative Physics Based on Confluences

We will use the square brackets "$[\cdot]$" to denote the qualitative property of a variable, which could take three possible values: positive (+), zero (0) or negative (-).The trend of a variable Q can be determined by the qualitative value of its differential dQ:

$$[dQ] = + \quad \leftrightarrow \quad Q \text{ is increasing } (\uparrow)$$

$$[dQ] = 0 \quad \leftrightarrow \quad Q \text{ is unchanging } (-)$$

$$[dQ] = - \quad \leftrightarrow \quad Q \text{ is decreasing } (\downarrow)$$

Confluence equations are simple qualitative forms of differential equations. For our application, we are interested in obtaining the relations between qualitative trends of various T-H variables and the imbalance indicators in generic components of the system. The confluence equations in that case can be derived from the corresponding conservation equations of either mass, momentum or energy.

To illustrate the transformation of a quantitative equation into qualitative form, consider an example given by the following arbitrary equation:

$$Q = ax - by \tag{2.4}$$

where x and y are some variables of interest, say some sensor readings; a and b are some positive but otherwise unknown parameters; Q is an indicator that we are interested in and cannot measure directly. Again, this is just a hypothetical arbitrary equation which does not necessarily represent any actual physical phenomenon.

Under these conditions, we can only observe the qualitative trends of x and y separately and not the value of $ax - by$. To transform the equation into the qualitative form, we can simply differentiate both sides and consider the sign of each term:

$$\begin{aligned} & dQ = a\ dx - b\ dy \\ \Rightarrow & [dQ] = [a\ dx - b\ dy] \\ \Rightarrow & [dQ] = [a\ dx] - [b\ dy] \\ \Rightarrow & [dQ] = [dx] - [dy] \end{aligned} \tag{2.5}$$

For the third step, $[a\ dx - b\ dy]$ is set equal to $[a\ dx] - [b\ dy]$. Such operation is only valid if the two terms, $a\ dx$ and $b\ dy$ have *opposite signs*. The fourth step immediately follows since both a and b are assumed to be positive. $[dQ] = [dx] - [dy]$ is the qualitative form we want. The trend of the imbalance indicator Q can then be inferred from the trend of the individual variables x and y.

Notice the loss of information in the third step. The sign $[a\ dx - b\ dy]$ can always be determined if one knows the values of a and b but the difference $[a\ dx] - [b\ dy]$ is mathematically ill defined. For example, if $a\ dx$ and $b\ dy$ are both positive, their qualitative difference is undetermined:

$$\text{positive} - \text{positive} = \text{unknown}$$

Such loss of information is inevitable in the transformation of an equation from quantitative form to qualitative form. The emphasis here is that when applicable, the qualitative form $[dQ] = [dx] - [dy]$ allows us to infer the trend of the variable Q even though the exact value of the "design parameters" are a and b unknown.

Formally, the set of rules to formulate qualitative equations and manipulate qualitative variables are summarized below, as described in [78]:

1. $[0][x] \rightarrow [0]$
2. $[0] + [x] \rightarrow [x]$
3. $[+][x] \rightarrow [x]$
4. $[-][x] \rightarrow -[x]$
5. $[xy] \rightarrow [x][y]$
6. $[x + y] \rightarrow [x] + [y]$

Most of these transformation rules are intuitive and straightforward, except for the last rule which may cause some loss of information as discussed.

2.3.2 Qualitative Models and the Knowledge Base in PRO-AID

For PRO-AID, qualitative component models are constructed using confluence equations derived from the conservation equations of mass, momentum and energy. Consider, for example, a single inlet/outlet component under quasi-static conditions, the mass conservation equation can be written as:

$$w_{\mathrm{out}} - w_{\mathrm{in}} = Q_{\mathrm{mass}} \tag{2.6}$$

where w_{in} and w_{out} are the inlet and outlet flow rates, assumed to be available by sensor readings. Q_{mass} is a source/sink term in the mass balance and will be used as the imbalance indicator. For this component under normal working conditions we can expect to have $Q_{mass} = 0$, i.e. there is no source or sink of mass.

Following the transformation rules established in the last section, it is straightforward to obtain the following confluence equation from Eqn. (2.6):

$$[dQ_{\mathrm{mass}}] = [dw_{\mathrm{out}}] - [dw_{\mathrm{in}}] \tag{2.7}$$

This serves as a *qualitative mass balance model* for the component. The trend of the imbalance indicator Q_{mass} can be obtained from the inlet and outlet flowrate trends using the following reasoning rules:

$$\boxed{\begin{array}{l} \text{If } w_{in} \uparrow \text{ and } w_{\mathrm{out}} \downarrow \text{ then } Q_{\mathrm{mass}} \downarrow \\ \text{If } w_{in} \downarrow \text{ and } w_{\mathrm{out}} \uparrow \text{ then } Q_{\mathrm{mass}} \uparrow \end{array}} \tag{2.8}$$

Changing trends in Q_{mass} serve as *fault symptoms* indicating there is a leak in or out of the component.

Similarly, the static momentum conservation equation for the component can be written as:

$$P_{\mathrm{in}} - P_{\mathrm{out}} - \frac{kw^2}{\rho A^2} = 0 \tag{2.9}$$

where P_{in} and P_{out} are inlet and outlet pressures, respectively; w is the flow rate through the component. k, ρ, and A respectively denote the loss coefficient, fluid density and effective cross-sectional area. k and A are component-specific parameters and can be expected to remain the same if the component is fault-free. We can lump these parameters into a status indicator Q_{mom}, thus:

$$P_{\text{in}} - P_{\text{out}} - \frac{w^2}{Q_{\text{mom}}} = 0 \tag{2.10}$$

Under normal working conditions we can expect to have $Q_{\text{mom}} = \text{const.}$ Using the transformation rules described in the last section, we can obtain the qualitative form:

$$\begin{aligned} &dP_{\text{in}} - dP_{\text{out}} - 2\frac{w}{Q_{\text{mom}}}dw + \frac{w}{Q^2{}_{\text{mom}}}dQ_{\text{mom}} = 0 \\ &\rightarrow [dQ_{\text{mom}}] = [dw] - [dP_{\text{in}}] + [dP_{\text{out}}] \end{aligned} \tag{2.11}$$

This qualitative equation serves as a *qualitative momentum model*, from which we can construct reasoning rules to infer the trend of Q_{mom} for the component. For instance:

$$\boxed{\text{If } w\downarrow \text{ and } P_{\text{in}}\uparrow \text{ and } P_{\text{out}}\downarrow \text{ then } Q_{\text{mom}}\downarrow} \tag{2.12}$$

Q_{mom} decreasing would serve as a *fault symptom* indicating a blockage-type fault, i.e. increased loss coefficient, in the component.

This procedure of constructing qualitative models can be generalized for each generic component type in TH systems. The reasoning rules defined using such qualitative models form the knowledge base for PRO-AID. The knowledge base allows the code to infer changing trends of various Q values from individual process variable trends. Changing trends of the Q values serve as *fault symptoms* and the logical MBD framework as described in Section 2.2 can be applied to obtain possible diagnoses by logical inference.

2.3.3 Limitations of the Qualitative Reasoning Approach in PRO-AID

The main advantage of the qualitative approach in PRO-AID is that qualitative models can be constructed for each type of generic component in a TH system without the need to know component-specific design parameters. The approach was applied successfully in detection of abrupt faults in various systems [79, 80, 81]. However, the applicability of this qualitative approach in detecting long-time scale slow degradations is subject to several limitations. Furthermore, its capability to deal with sensor faults is limited.

The limitations in PRO-AID are intrinsic to the qualitative approach. The qualitative model developed for each component as a specific set of if-then rules which does not account for the possibility of sensor faults generally cannot handle multiple-fault events. The approach requires continuous sensor readings to be transformed into discrete qualitative values, which limits the detection sensitivity. Faults need to be severe enough to trigger the reasoning rules within a relatively short timeframe before feedback and control action responses obscure the underlying variable trends. Some loss of information is inevitable when a quantitative equation is transformed into qualitative form, which could lead to scenarios where no reasoning rule is applicable. The issue is even more significant if number of qualitative variables involved in a confluence equation is high in which case the effectiveness of the approach is very limited.

For the current application, the need for detection of slow degradations in both system components and sensors motivates the search for an alternative quantitative approach.

2.4 Quantitative Model-Based Diagnosis Frameworks

Quantitative models can be constructed to either describe the normal fault-free behavior of a system or to provide a description of its different fault modes. For a consistency-based fault diagnosis approach, we need models of the fault-free behavior to check for inconsistencies in the observed data. The construction of fault-free quantitative models and their usage in fault diagnosis are the focus of this section.

We will first discuss the reasoning process in quantitative approaches, assuming that quantitative models and subsequently model residuals can be constructed for each component in a system. Furthermore, model residuals are constructed in a way that would allow us to infer which faults are implicated when a residual is non-zero. Various model construction and residual generation approaches are then discussed in Section 2.4.2 and explored more in detail in Chapter 3.

2.4.1 Quantitative Reasoning Frameworks

From the formulation in Section 2.2, the key step in the MBD framework as a consistency-based approach is to identify the set of all minimal conflicts. Conflicts can be derived from discrepancies between the observations and the fault-free system, defined as *fault symptoms*. The constraints between observations and the expected fault-free behavior are formally defined as analytical redundancy relations (ARRs). Each analytical redundancy relation provides an equation governing the relation between certain observations and components in the system. All the ARR equations must hold when the system is fault-free.

In a quantitative approach, analytical redundancy relations can be expressed by quantitative equations. If a certain ARR equation does not hold then at least one of the involved components or sensors must be faulty and the degree to which the relation is violated is quantified by a *residual*, defined as the difference between the two sides of the ARR equation. Thus, each non-zero *residual* from each independent ARR equation serves as a *fault symptom* from which a *minimal conflict* can be derived.

Following the formulation in Section 2.2, the MBD framework for the quantitative approach can be summarized in three main steps as listed in Table 2-1.

Table 2-1. The MBD framework for quantitative model-based diagnosis

Step 1	Identify available ARRs from component models, from which define possible *residuals*.
Step 2	Obtain *minimal conflicts* from *non-zero* residuals
Step 3	Obtain all valid *minimal diagnoses* from the set of all *minimal conflicts*.

The MBD framework as introduced above was originally developed in the computer science and artificial intelligence community [72, 82]. An independent but closely related framework was developed in parallel by the fault detection and isolation (FDI) community [82]. We will refer to this alternative approach as the FDI framework just to distinguish it from the former although both serve as a framework for model-based diagnosis.

The two frameworks are analogous in the sense that both are consistency-based and rely on system models and the same redundancy information for consistency-checking. More specifically for both frameworks in the context of a quantitative approach, the occurrence of a fault is detected by non-zero *residuals* indicating discrepancies between the observed behavior and the normal operation behavior described by the system model. The distinction is on how diagnoses are obtained from the observed residuals.

As described above, in the MBD framework, only non-zero residuals are used as fault symptoms to formulate the required minimal conflicts while zero residuals are ignored. In the FDI framework, a significant but commonly employed assumption is the *notion of exoneration* using *zero* residuals. In addition to the fact that non-zero residuals indicate possible faults, it is often assumed *zero* residuals indicate 'no-faults', thus all relevant faults in the involved components and sensors are exonerated [72, 67].

The notion of exoneration is an approximation and not mathematically exact. It is possible that multiple faults can occur simultaneously and compensate one another to give a zero residual. In such scenario, the invalid exonerations may lead to false diagnoses. However, for most cases in practice, such scenarios can be considered statistically insignificant. The notion of exoneration can help simplify the reasoning process and provide more detailed diagnoses.

Using the notion of exoneration, relevant faults from zero residuals can be removed from the observed *conflicts*. The reasoning process for fault diagnosis in the FDI framework is summarized in Table 2-2.

Table 2-2. The FDI framework for quantitative model-based diagnosis

Step 1	Identify available ARRs from component models, from which define possible *residuals*.
Step 2	Obtain *minimal conflicts* from *non-zero* residuals
Step 3	Use *zero* residuals to exonerate relevant faults from the conflicts
Step 4	Obtain all valid diagnoses from the reduced conflicts.

So far, we have assumed that all component models are known and there is no uncertainty in the model predictions or observed data. That is not the case for most applications in engineering systems. Modeling uncertainty is inevitable, and all observations are subject to noise and measurement uncertainty. The effects of modeling and measurement uncertainty add another layer of complications one needs to account for.

In the next chapter, we will discuss the development of component models in TH systems of nuclear power plants. As dictated by the underlying physics, the models for most components in the thermal-hydraulic systems are non-linear, even in steady-state operation. A crucial part of the model development process is to account for the possibility that the plant may undergo various controlled changes in operating conditions and the validity of the component models should be insensitive to such changes.

2.4.2 Model Construction and Residual Generation Approaches

Generally, quantitative models of technical processes in engineering systems can either be derived analytically from understanding of the underlying physics or constructed empirically as black-box models using past data. In practice, the two approaches may be combined to create gray-box models using both physical laws and operational data [17].

The dynamics of a technical process in a system can be characterized by its response outputs $\mathbf{y}(t)$ to inputs $\mathbf{u}(t)$, where $\mathbf{y}(t)$ and $\mathbf{u}(t)$ are column vectors. For *linear time-invariant* (LTI) systems, the response is more conveniently expressed by an input-output model under Laplace transform:

$$\mathbf{y}(s) = G_u(s)\mathbf{u}(s) \tag{2.13}$$

where $G_u(s)$ is known as the transfer matrix; the label s denotes the complex variable of Laplace transform. For any time-dependent function $f(t)$, Laplace transform is a generalization of Fourier transform to express the function in frequency domain, defined by:

$$F(s) = \int_{-\infty}^{\infty} f(t)e^{-st}dt \tag{2.14}$$

where s is a complex variable with $s = \sigma + i\omega$ for a weighting constant σ and frequency variable ω. For $\sigma = 0$ we get the Fourier transform. The inverse transform is given by:

$$f(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}F(s)e^{st}d\omega \tag{2.15}$$

Thus, to model the system, one must either obtained the transfer matrix $G_u(s)$ either analytically or empirically. Both the inputs $\mathbf{u}(t)$ and outputs $\mathbf{y}(t)$ are assumed to be measurable. At any given time t, the state of the system can defined by a set of unmeasurable *state variables* $\mathbf{x}(t)$. The model of the LTI system can be expressed in the state-space representation as:

$$\begin{aligned}\dot{\mathbf{x}}(t)&=A\mathbf{x}(t)+B\mathbf{u}(t)\\ \mathbf{y}(t)&=C\mathbf{x}(t)+D\mathbf{u}(t)\end{aligned} \tag{2.16}$$

where A, B, C and D are parameter matrices. The state-space model (2.16) can be obtained directly by modeling or derived from the minimum state-space realization of the transfer matrix $G_u(s)$ using the relation $G_u(s)=D+C\left(sI-A\right)^{-1}B$ [67].

In general, the fault-free behavior of a dynamic LTI system can be described by either an input-output model or a state-space model. Given the fault-free model, one can then generate model residuals for fault diagnosis. Ideally, we want the residuals to be mostly sensitive to the effects of various faults and insensitive to noise and model uncertainty. A straightforward approach to residual generation is to use the difference between the measured outputs $\mathbf{y}(s)$ and fault-free model prediction $\hat{\mathbf{y}}(s)$:

$$\mathbf{r}(s)=\mathbf{y}(s)-\hat{\mathbf{y}}(s)=\mathbf{y}(s)-G_u(s)\mathbf{u}(s) \tag{2.17}$$

However, in the presence of model uncertainty and possible sources of noise and disturbances, the performance of such residual generator is often poor [67]. In the effort of reducing the sensitivity to noise and model uncertainty, residuals are defined as functions of the measured inputs and outputs using filter and factorization techniques. Ding and Frank introduced the following form for residuals of LTI systems [67]:

$$\mathbf{r}(s)=R(s)\left[M_u(s)\mathbf{y}(s)-N_u(s)\mathbf{u}(s)\right] \tag{2.18}$$

where $R(s)$ is a *parametrization matrix*, and:

$$M_u(s)=I-C(sI-A+LC)^{-1}L \tag{2.19}$$

$$N_u(s) = D + C(sI - A + LC)^{-1}(B - LD) \tag{2.20}$$

for a matrix L known as the *observer gain matrix*. The forms of $R(s)$ and L depend on the particular residual generator chosen for a system. Residual generation methods for LTI systems include: fault detection filter [65, 66], diagnostic observer [67, 13], parity space [68, 17, 67].

In the above analysis, the time variable is continuous. For discrete time, the Laplace transform is replaced by z-transform and the analyses hold with s replaced by the variable z of the z-transform. It should be noted that in general, the effects of different faults may combine in the residual vector $\mathbf{r}(s)$. Thus, additional techniques in analyzing the residuals are needed to differentiate between different faults in a system.

In practice, most technical processes are non-linear, and it is generally not possible to analytically model non-linear dynamic processes. Linear assumptions or linearization techniques are required to represent such systems by linear input-output or state-space models. For instance, the behavior of each non-linear system in operating conditions close to a reference point can be approximated by a linear model. Alternatively, data-driven methods can be used to directly construct the relation between input and output variables, provided that a large set of operational data is available. Data-driven methods to construct quantitative models include principle component analysis (PCA), artificial neural network (ANN) [23].

Besides input-output and state-space models, one can also construct models to express certain process parameters $\theta(t)$ as functions of both the input and output variables. The parameters $\theta(t)$ are in turn related to other physical parameters and variables relevant to different faults in the system. Using the models for $\theta(t)$, one can generate different residuals relevant to various faults in the system. Such approach to residual generation is known as the *parameter estimation* approach [69, 17].

For the current application in this thesis, we will use a physics-based approach to construct models for each performance-related parameter in a component and generate model residuals via the *parameter estimation approach*. The approach will be discussed in more detail in Chapter 3.

Chapter 3

Physics-based Component Models

Considering the characteristics of the diagnostic problem and the objectives for the target application, the quantitative model-based approach was selected as the basis for the fault detection and diagnosis framework in this thesis. The key factors that shaped the development of the proposed FDD approach include the need for sufficiently high detection sensitivity to detect faults of slow degradation type, the capability to simultaneously deal with both component faults and sensors faults, and the desired insensitivity to various sources of uncertainty and possible changes in operating conditions.

The proposed theoretical framework consists of a quantitative model-based approach to quantify slow performance degradations and uncertainty treatments, including statistical change detection and probabilistic reasoning, to robustly deal with modeling uncertainty and measurement error. As discussed in Chapter 2, developing component models adequate for diagnostic purpose is one of the challenging issues one needs to address in order to apply the model-based diagnosis frameworks to complex engineering systems.

There are generally two approaches to constructing quantitative component models, namely data-driven approach and physics-based approach. For the data-driven approach, each component model is constructed by machine learning techniques relying solely on a set of data describing past behavior of the component. As such, the quality of the model depends on the coverage of the data set used as training data and the capability to extrapolate is limited. In applications for nuclear power plants, the data available are often not ideal for purely data-driven approaches.

The physics-based approach, on the other hand, relies on understanding of the underlying physics to formulate the functional form of the component model with only a few parameters left to be determined for each specific component. The quality of physics-based models is less data-dependent and generally one can expect the physics-based approach to be less sensitive to changes in operation conditions. The physics-based approach for model construction is preferred for the framework developed in this thesis.

3.1 Physics-based Model Construction

In this section, we will discuss the physics-based approach to construct quantitative models for the model-based diagnosis frameworks. Although some of the component-specific design parameters may be available, it is often not practical to use them directly in deriving analytical models to describe the underlying physics. Constructing models solely from geometric design parameters often requires computational simulations and cannot be done analytically. The general approach proposed here is to construct parametric models to describe the behavior of the component by simplifying the underlying physics. Effectively, all geometric characteristics of the component are lumped into a few unknown parameters of the parametric models. The unknown model parameters are to be determined using past data.

Each TH system in a nuclear power plant can be decomposed into separate components of known generic types, e.g. valve, pump, heat exchanger. Each component of a generic type is designed to perform a basic function of either mass, momentum or energy transfer. The behavior or performance of a component in normal working conditions can be described by separate models constructed for each of these three processes.

A fault is defined to be any change in the characteristics of a component that affect its ability to perform its designed function. Any fault or malfunction in the component would alter its characteristic which will be reflected by an inconsistency between the actual observed behaviors and a model prediction. More specifically, since the models for each component are constructed based on the three conservation equations, any fault in the component would result in an unaccounted imbalance in at least one of the conservation equations which leads to a non-zero residual for the corresponding model. For the current application, we are considering faults of the slow degradation type in the long-time scale during which the operation of each component can be considered quasi-static.

To clarify, consider a one-dimensional incompressible flow through a single inlet/outlet component. The integral equations for the conservation of mass, momentum and energy for the control volume are respectively given by [83]:

$$\frac{dm}{dt} = w_{\text{in}} - w_{\text{out}} \tag{3.1}$$

$$\left(\frac{l}{A}\right)_T \frac{dw}{dt} = P_{\text{in}} - P_{\text{out}} + \rho g(z_{\text{in}} - z_{\text{out}}) + \frac{w^2}{2\rho}\left(\frac{1}{A_{\text{in}}^2} - \frac{1}{A_{\text{out}}^2}\right) - \Delta P_{\text{loss}} \tag{3.2}$$

$$\frac{d}{dt} H = w_{\text{in}} h_{\text{in}} - w_{\text{out}} h_{\text{out}} + Q_{\text{eng}} \tag{3.3}$$

where the indices ‘in’ and ‘out’ denote the inlet and outlet location; m is the total fluid mass enclosed in the control volume; w denotes the mass flow rate; P denotes pressure; A denotes cross-sectional area; $\left(l/A\right)_T$ is the equivalent inertia length for the component defined by its geometric dimensions; ρ is the fluid density; g is the gravitational acceleration constant; z denotes the relative elevation at each location of the flow; ΔP_{loss} denotes the total pressure loss; H denotes the total energy enclosed in the control volume; h denotes the specific enthalpy and Q_{eng} is the combined energy source/sink term.

By assuming quasi-static conditions, we are setting the time derivative terms, i.e. the left sides of Eqns. (3.1-3.3) to zero. Explicitly, we are assuming that even though the TH process variables at the inlet and outlet of each component may vary over time, the contribution of the time derivatives to the conservation equations are negligible.

In general, for each component, one can construct three models to describe its fault-free behavior with regard to the conservation of mass, momentum and energy. For brevity, we will refer to these models for each component as the mass, momentum and energy models. As discussed, we will formulate physics-based parametric models for each generic type of generic component with a few component-specific parameters left to be determined using past data.

3.1.1 Mass Models

Using Eqn. (3.1) and setting the derivative term to zero, we obtain a simple relation enforcing the balance of mass flow rates between the inlet and outlet. This serves as a mass model for the component with no unknown parameters involved. For a generic component with multiple inlet or outlet, the mass model is given by:

$$\boxed{\sum w_{\text{in}} = \sum w_{\text{out}}} \tag{3.4}$$

To evaluate the performance of a component regarding the mass model, a full set of flowrate sensors at every inlet and outlet location is required.

3.1.2 Momentum Models

Setting the left side to zero and rearranging the terms in Eqn. (3.2), we have:

$$P_{\text{in}} - P_{\text{out}} = -\rho g(z_{\text{in}} - z_{\text{out}}) - \frac{w^2}{2\rho}\left(\frac{1}{A_{\text{in}}^2} - \frac{1}{A_{\text{out}}^2}\right) + \Delta P_{\text{loss}} \tag{3.5}$$

This relation is valid for single-phase incompressible flows in components with no source of momentum, i.e. not a pump. The first term is the gravity effect which is fixed for each component. The second term, account for the acceleration of the fluid, vanishes if the cross-sectional areas at the inlet and outlet are the same. The total pressure loss ΔP_{loss} can be further decomposed into pressure losses due to friction ($\Delta P_{\text{friction}}$) and form losses (ΔP_{form}) due to any abrupt changes of flow direction or geometry.

Generally, for incompressible flow, the form loss term ΔP_{form} is proportional to the square of the flow rate. The friction loss $\Delta P_{\text{friction}}$ can be written as:

$$\Delta P_{\text{friction}} = \bar{f}\frac{L}{D_e}\frac{\rho v_m^2}{2} \tag{3.6}$$

where $\bar{f}$ is the frictional pressure drop coefficient, L is the length of the flow channel, D_e the equivalent diameter and v_m is the bulk velocity of the flow. The pressure drop coefficient $\bar{f}$ is usually given as function of the *Reynolds number* Re, with:

$$\mathrm{Re} = \frac{\rho v_m D_e}{\mu} \tag{3.7}$$

where μ is the viscosity of the fluid. The exact form of $\bar{f}$ depends on the flow regime, identified by the magnitude of Re, and other characteristics of the flow surface. For example, with laminar flow:

$$\bar{f} = \frac{64}{\mathrm{Re}} \tag{3.8}$$

Or for turbulent flow inside a smooth tube using the McAdams correlation with $10^4 < \mathrm{Re} < 10^6$ [83]:

$$\bar{f} = 0.184\,\mathrm{Re}^{-0.20} \tag{3.9}$$

Notice that the bulk velocity v_m is generally proportional to the flow rate. Thus, the right side of Eqn. (3.5) only depends on a single process variable, the flow rate w. All other parameters are either constant or are geometric characteristics of the component which should remain the same if the component is fault-free. The dependence of fluid properties on slight change of pressure or temperature can be neglected.

Therefore, to monitor the performance of the component regarding the conservation of momentum, we need to construct a parametric model for the pressure difference between the inlet and outlet as a function of the flow rate:

$$P_{\mathrm{in}} - P_{\mathrm{out}} = f(w) \tag{3.10}$$

The parametric form of the function $f(w)$ is to be determined. Expressed as a polynomial of w, it can have a zero-order term from the gravity effect, a second-order term from the form losses term (ΔP_{form}). The friction loss contribution can be lumped into a term of order between first and second order, depending on the flow regime. Overall, it suffices to use a quadratic form for the parametric function $f(w)$. Thus, the momentum model for a generic component with incompressible fluid is given by:

$$\boxed{P_{\mathrm{in}} - P_{\mathrm{out}} = \theta_0 + \theta_1\, w + \theta_2\, w^2} \tag{3.11}$$

where θ_0, θ_1, and θ_2 are the three unknown model parameters to be determined for each specific component using sensor data in a calibration process. Effectively, the geometric and design characteristics of the component are lumped into these three model parameters. Any momentum-related faults, e.g. leakage or blockage, would result the change of at least one parameter.

The calibration and subsequent use of this momentum model requires reading data from three sensors: inlet pressure P_{in}, outlet pressure P_{out} and flowrate w at either the inlet or outlet.

For pumps, we have an additional term for momentum gain (ΔP_{gain}) to Eqn. (3.5) due to the pump motor shaft power. In general, ΔP_{gain} may depend on not only the flow rate but also pump speed and other operating conditions. Momentum model for pumps, if not provided, need to be developed for each specific design type. For a simple constant speed pump, we can use a model similar to Eqn. (3.11) with the inlet and outlet pressure interchanged to quantify the pump head:

$$\Delta P_{\text{head}} = P_{\text{out}} - P_{\text{in}} = \theta_0 + \theta_1 \; w + \theta_2 \; w^2 \tag{3.12}$$

3.1.3 Energy Models

The quasi-static conservation of energy equation for a generic component is given by Eqn. (3.3) with the left side set to zero:

$$w_{\text{out}} h_{\text{out}} - w_{\text{in}} h_{\text{in}} = Q_{\text{eng}} \tag{3.13}$$

The energy source term Q_{eng} is component- and situation-dependent. Thus, to model the component behavior in energy-related process, we need a model for Q_{eng} in terms of other process variables.

For a pump, Q_{eng} is the effective shaft power provided by the pump motor which will be modeled as a function of the flowrate and pump speed. For other non-heat-exchanging components, Q_{eng} is a small loss term, including external heat loss, which is often not of interest and can be neglected. For heat exchangers, e.g. heaters or condensers, Q_{eng} depends on the component type and operating conditions. In general, we only need to construct energy models for heat-exchanging components with the parametric forms depend on the component type.

As an example, consider the case of a single-phase counterflow shell-tube heat exchanger. Following the above analysis, the heat exchanger is a composite of two single inlet/outlet components, one for each side. The loss term Q_{eng} from the hot side is approximately the gain term Q_{eng} for the cold side. Thus, it is more convenient to model both sides together.

Assuming negligible external heat loss, the total heat transfer rate is related to the inlet and outlet enthalpies on each side by:

$$Q_{\text{eng}} = w^h (h^h_{\text{in}} - h^h_{\text{out}}) = w^c (h^c_{\text{out}} - h^c_{\text{in}}) \tag{3.14}$$

where the superscripts h and c denote the hot side and cold side of the heat exchanger, respectively. Each enthalpy value can be obtained from the corresponding temperature sensor given the operating pressure.

This is the heat balance equation that holds when there is no leakage or significant external heat loss in the heat exchanger. Besides the heat balance, we are more interested in monitoring the heat-exchanging capability of the heat exchanger. Faults related to the heat-exchanging capability, like fouling, would not affect the heat balance. Briefly, fouling in a heat exchanger is the accumulation of unwanted materials on the heat-exchanging surfaces which may affect both the heat transfer process as well as the momentum transfer along the axial direction.

To monitor the heat-exchanging capability, the overall heat transfer coefficient is defined via the log-mean temperature difference (LMTD) model:

$$Q_{\text{eng}} = UA \frac{\Delta T_o - \Delta T_i}{\ln \dfrac{\Delta T_o}{\Delta T_i}} \tag{3.15}$$

where $\Delta T_o = T^h_{\text{out}} - T^c_{\text{in}}$ and $\Delta T_i = T^h_{\text{in}} - T^c_{\text{out}}$ are the temperature differences between the two sides at the outlet and inlet of the hot side. U is defined to be the overall heat transfer coefficient and A is the effective heat transfer area.

As A can be considered a constant for each heat exchanger, we can combined U and A into a single parameter UA which can be used to monitor the heat transfer performance of the heat exchanger. For brevity, we will also refer to UA as the overall heat transfer coefficient.

For the exchanger, even in fault-free state, UA is not a constant but depends on other process variables. The problem of constructing a model for the heat transfer process reduces to constructing a model for UA. At each axial location, the local heat transfer coefficient of the shell-tube geometry is given by [84]:

$$\frac{1}{U_i} = \frac{1}{h_h} + \frac{d_i}{d_o}\frac{1}{h_c} + R_w \tag{3.16}$$

where d_i and d_o are the inner and outer tube diameters; h_h and h_c are the film heat transfer coefficients for the hot side and cold side, respectively; R_w is the heat resistance by the tube wall per unit area on the inner side. The film heat transfer coefficient on each side can be expressed in terms of the Nusselt number by:

$$h = \frac{k_f Nu}{D_h} \tag{3.17}$$

where k_f is the fluid conductivity and the hydraulic diameter D_h for each side is used as the characteristic length.

To obtain a parametric model for UA as a function of the process variables at the inlet and outlet of the heat exchanger, we first need an expression for U_i. For that purpose, we need to investigate the functional form of the heat transfer coefficients h_h and h_c, which in general depend on the fluid properties, heat exchanging surface conditions and flow conditions. For instance, for fully developed turbulent flow of nonmetallic fluids, the Nusselt number is given by a generic expression [83]:

$$Nu = C\,\mathrm{Re}^{\alpha}\,\mathrm{Pr}^{\beta}\left(\frac{\mu_w}{\mu}\right)^{\kappa} \tag{3.18}$$

where μ_w is the fluid viscosity at wall temperature; μ is the fluid viscosity at bulk temperature; C, α, β, and κ are constants that depend on the fluid properties and geometry of the flow channel; and Pr is the Prandtl number of the fluid.

For single-phase water with $\mu \approx \mu_w$, the Dittus-Boelter correlation is commonly used [83, 84]:

$$Nu = 0.023\,\text{Re}^{0.8}\,\text{Pr}^{n} \tag{3.19}$$

with $n = 0.3$ for cooling and $n = 0.4$ for heating.

Notice that for each side of the heat exchanger, the Reynolds number Re is proportional to the flow rate w hence in this case using Eqn. (3.19) the Nusselt number Nu is proportional to a power of the mass flowrate, $w^{0.8}$. If the dependence of other fluid properties on slight changes of on temperature can be considered negligible, we have the following expression for the local heat transfer coefficient U_i under changes of flow rate on each side:

$$\frac{1}{U_i} = \frac{1}{h_{h0}}\left(\frac{w_{h0}}{w_h}\right)^{0.8} + \frac{1}{h_{c0}}\left(\frac{w_{c0}}{w_c}\right)^{0.8} + R_{w0} \tag{3.20}$$

where the subscript 0 the values evaluated at a reference operating point.

Note that the overall heat transfer coefficient U as defined by Eqn. (3.15) is a weighted average of the local U_i along the axial direction. Thus, using the expression for U_i, we can obtain a parametric model for UA that explicitly only depends on the two flowrates w_h and w_c:

$$\boxed{\frac{1}{UA} = \theta_h w_h^{-0.8} + \theta_c w_c^{-0.8} + \theta_0} \tag{3.21}$$

with three model parameters θ_h, θ_c and θ_0 to be determined in the calibration process. Physically, these three parameters encode the fluid properties and geometric characteristics of the heat exchanger. Heat-exchanging-related faults would cause the heat exchanger to deviate from this expected model.

In practice, the flow characteristic on the shell side is not the same as the tube side. For instance, in the presence of the baffles on the shell side (usually the cold side), a better approximation is to adjust the power of the flow rate term from -0.8 to -0.6 [85]. Thus, for a heat exchanger with the baffles in the cold shell side, the parametric model for UA becomes:

$$\frac{1}{UA} = \theta_h w_h^{-0.8} + \theta_c w_c^{-0.6} + \theta_0 \tag{3.22}$$

In practice, an extra correction factor, to account for the temperature profile along the axial direction especially for the case with multiple shell or tube passes, is often included in the definition of UA in Eqn. (3.15) [85]. For the current application as we only consider small deviations around a steady state operating point, the variation of such correction factor due to changes in temperature is assumed to be negligible.

To summarize, the energy model for a single-phase counterflow shell-tube heat exchanger include the heat balance given by Eqn. (3.14), the overall heat transfer coefficient UA computed using the LMTD model in Eqn. (3.15) and a parametric model for UA in terms of the two flow rates given by Eqn. (3.21) or (3.22). The parametric model for UA has three unknown model parameters left to be determined for each specific heat exchanger by using measurement data. Monitoring both the heat balance and the heat transfer performance via UA requires six sensors at the inlet and outlet of the heat exchanger: the flow rate, inlet and outlet temperatures on each side. If the heat balance can be ensured by other means, the construction of the parametric model for UA requires five of the six sensors. In that case the sixth sensor can be computed from the other five using the heat balance equation.

In this section, we have discussed the approach to construct physics-based models for components in TH systems. Only some of the most common components have been discussed. The procedure will be generalized to apply for a larger class of components of each design type. Each parametric model for a component in general may contain several model parameters. These parameters are to be determined for each specific component by fitting the model against measurement data in a process referred to as model calibration. Depending on the parametric form of the model, model calibration can be performed using linear or polynomial regression [86]. Details on the relevant regression methods will be provided in Appendix A.

3.2 Residual Generation

After quantitative models have been constructed for all possible components in a system, the next step is to obtain all available analytical redundancy relations from each model. Each ARR can be used to generate one residual whose non-zero value serves as a fault symptom in quantitative model-based diagnosis as discussed in Section 2.4. From Section 3.1, the physics-based models for each component are constructed separately for each of the mass, momentum and energy processes. As such, ARRs from each model will only be sensitive to specific type of faults. We have developed component models and subsequently model residuals in a way that would allow us to differentiate between different types of faults in a component.

By definition, each ARR must remain valid even under changes of boundary or operating conditions if the related components and sensors are fault-free. As an example, for a single-phase heat exchanger, the overall heat transfer coefficient UA is a performance-related parameter but it may not stay the same under changes in operating conditions. Thus, setting UA to a reference value does not produce a valid ARR. Fouling in the heat exchanger would affect UA but so do possible changes in operating conditions like the flow rate on either side. An ARR to detect fouling must be obtained from a model that can account for the possible changes of operating conditions.

Residuals from Mass Models

The mass balance model for a generic component does not contain any model parameters. For each mass model, we have a single ARR enforcing the flow rate balance between the inlets and outlets the residual can be computed as:

$$r_{\text{mass}} = \sum w_{\text{in}} - \sum w_{\text{out}} \tag{3.23}$$

The calculation of r_{mass} involves a set of flow rate sensors, one for each inlet or outlet. A non-zero value for r_{mass} would indicate either a sensor fault or a leakage in or out of the component.

Residuals from Momentum Models

For a generic momentum model given by Eqn. (3.11), we have a single ARR relating the measured pressure loss provided by the two pressure readings and the model pressure loss model prediction. The residual from that ARR can be computed as:

$$r_{\text{press}} = \left(P_{\text{in}} - P_{\text{out}}\right) - \left(\theta_0 + \theta_1\, w + \theta_2\, w^2\right) \tag{3.24}$$

The calculation of r_{press} requires three sensor readings for P_{in}, P_{out} and w. A non-zero value for r_{press} would indicate either one of the sensor faults or a blockage or leakage in the component.

Residuals from Energy Models

Consider energy model for the heat exchanger described in 3.1.3 which includes a heat balance equation and a parametric model for UA, we have two independent ARRs:

$$w^h(h_{\text{in}}^h - h_{\text{out}}^h) = w^c(h_{\text{out}}^c - h_{\text{in}}^c) \qquad \text{(Heat Balance)} \tag{3.25}$$

$$\frac{\text{LMTD}(T_{\text{in}}^c, T_{\text{out}}^c, T_{\text{in}}^h, T_{\text{out}}^h)}{w^h(h_{\text{in}}^h - h_{\text{out}}^h)} = \theta_h w_h^{-0.8} + \theta_c w_c^{-0.8} + \theta_0 \qquad \text{(HX Performance)} \tag{3.26}$$

Each of these ARRs involves six sensors: a flow rate sensor, inlet temperature and outlet temperature for each side. For a given operating pressure on each side, the enthalpy values can be obtained directly from temperature sensor readings.

If all six sensors are available, we can straightforwardly compute the following two residuals:

$$r_L = w^h(h_{\text{in}}^h - h_{\text{out}}^h) - w^c(h_{\text{out}}^c - h_{\text{in}}^c) \tag{3.27}$$

$$r_0 = \frac{\text{LMTD}(T_{\text{in}}^c, T_{\text{out}}^c, T_{\text{in}}^h, T_{\text{out}}^h)}{w^h(h_{\text{in}}^h - h_{\text{out}}^h)} - \left(\theta_h w_h^{-0.8} + \theta_c w_c^{-0.8} + \theta_0\right) \tag{3.28}$$

Note that for each side of the heat exchanger, the flow rate could be measured at either the inlet or outlet. A leakage causing a loss of mass would violate both ARRs and cause both residuals to be non-zero. The calculation of each residual involves six sensors. In general, a non-zero r_L would indicate a leakage or one of the sensor faults. A non-zero r_0 would indicate either leakage, fouling or one of the sensor faults.

To detect and differentiate sensor faults, we can combine the two ARRs and compute one residual without using one of the sensors. That is, when all six sensors are available, we have the option to leave out one sensor. Using the other five sensors and assuming the first ARR holds, we can estimate the sixth sensor and use the result with the second ARR to generate a new residual. That way, we can have up to six residuals using each combination of five sensors. Each residual serves as a possible fault symptom and as discussed in Chapter 2, we would like to maximize the number of fault symptoms in order to improve the resolution of the diagnostic result.

If only five sensors are available, we cannot enforce the two ARRs. In this case, one must rely on other methods to detect possible leakage in the heat exchanger. Then assuming no leakage, the heat balance equation can be used to estimate the missing sixth sensor. The result can be used with the second ARR to generate one residual.

3.3 Example – Fault Diagnosis of a Single-phase Heat Exchanger

After all residuals have been constructed from the component models, one can apply one of the two frameworks listed in Table 2-1 and 2-2 to perform fault diagnosis.

To demonstrate this process, consider an example with the single-phase counterflow heat exchanger described in Section 3.1 and 3.2. We will assume six sensors are available. Let S_i with $i = 1, 2, ...6$ to denote the sensors for the flow rate, inlet temperature, outlet temperature of the cold side and then those for the hot side, respectively.

We will consider both component faults and sensor faults:

- Component faults: leakage (denoted by F_L) and fouling (F_0)
- Sensors faults: fault in sensor S_i , denoted by F_i for $i = 1, 2, ...6$

A leakage fault is the loss of mass from tube side to the shell side or the shell side to the external environment. Fouling is the accumulation of unwanted materials on the tube inner and outer surfaces affecting both the heat transfer process between the tube and shell sides as well as the momentum transfer along the axial direction. A sensor is said to be out of calibration if there is a significant bias between its reading value and the true value.

Since we only have a flow rate sensor for each side of the heat exchanger, it is possible to construct any mass or momentum model. We can only construct the energy model as described in 3.1.3. It follows that we can construct eight residuals in total. One from the heat balance relation given by Eqn. (3.27), a second residual from the HX performance relation given by Eqn. (3.28), and the other six residuals by combining the two ARRs and use each combination of five sensors, leaving out one sensor.

For instance, without using the sensor for the cold side inlet temperature, i.e. sensor S_2, the value for that variable can be estimated from the heat balance equation:

$$h_{\text{in}}^{c}(T_{\text{in,p}}^{c}) = h_{\text{out}}^{c} - \frac{w^{h}}{w^{c}}(h_{\text{in}}^{h} - h_{\text{out}}^{h}) \tag{3.29}$$

where the subscript 'p' indicates that the temperature $T_{\text{in,p}}^{c}$ is to be predicted from the compute enthalpy h_{in}^{c}. Using $T_{\text{in,p}}^{c}$ with the second ARRs, we can compute a residual, denoted by r_2:

$$r_2 = \frac{\text{LMTD}(T_{\text{in,p}}^{c}, T_{\text{out}}^{c}, T_{\text{in}}^{h}, T_{\text{out}}^{h})}{w^{h}(h_{\text{in}}^{h} - h_{\text{out}}^{h})} - \left(\theta_h w_h^{-0.8} + \theta_c w_c^{-0.8} + \theta_0\right) \tag{3.30}$$

The subscript '2' in r_2 is to emphasize that this residual is computed without using sensor S_2. Similarly, we have six residuals, denoted by r_i for $i = 1, 2, ...6$, each computed without using sensor S_i.

Overall, this system has eight independent faults, denoted by F_L, F_0, …, F_6 and we have constructed eight different eight residuals r_L, r_0,…, r_6. The dependencies of each residual are summarized in Table 3-1.

Table 3-1. Structure of the residuals in the HX Example

Residual	*Relevant Component Fault Types*	*Sensors Involved*
r_L	Leakage	S_1, S_2, S_3, S_4, S_5, S_6
r_0	Leakage, Fouling	S_1, S_2, S_3, S_4, S_5, S_6
r_1	Leakage, Fouling	S_2, S_3, S_4, S_5, S_6
r_2	Leakage, Fouling	S_1, S_3, S_4, S_5, S_6
r_3	Leakage, Fouling	S_1, S_2, S_4, S_5, S_6
r_4	Leakage, Fouling	S_1, S_2, S_3, S_5, S_6
r_5	Leakage, Fouling	S_1, S_2, S_3, S_4, S_6
r_6	Leakage, Fouling	S_1, S_2, S_3, S_4, S_5

The calculation of each residual involves a certain number of sensors as listed in the third column. For each residual, the relevant components faults are those that can affect the underlying analytical redundancy relation and cause the residual to become non-zero. With the structure of each residual known, one can then apply a quantitative model-based diagnosis framework to obtain possible diagnoses for a given set of fault symptoms.

As mentioned in Section 2, we will use a short-handed notation to write diagnoses and conflicts:

- Each *diagnosis* is denoted by the square brackets "$[]$" containing a list of faults. For example, $[F_0, F_3]$ is a diagnosis claiming both F_0 *and* F_3 must be true.
- Each *conflict* relation is denoted by the angle brackets "$\langle\rangle$" containing a list of faults. For instance, $\langle F_0, F_3\rangle$ is a conflict relation claiming that either F_3 *or* F_3 must be true.

3.3.1 Fault Diagnosis Using the MBD Framework

Consider a scenario in which all residuals are observed to be non-zero, except for r_3. The set of residual values can be written in a column vector as $(1,1,1,1,0,1,1,1)^T$, with each index list the value for a residual in the order of appearance in Table 3-1. The binary value 1 indicates a residual is non-zero.

We will now apply the MBD framework for quantitative model-based diagnosis, as detailed in Table 2-1. Step 1 has already been done. The next step is to obtain a conflict from each non-zero residual. The logical basis to construct a conflict is the claim: if a residual is non-zero then either the component is faulty with a fault of the specified relevant types or one of the involved sensors is faulty. Thus, using the structure of the residuals listed in Table 3-1, we can construct the following seven *minimal conflicts* from the seven non-zero residuals:

Table 3-2. Minimal conflicts for a scenario with residuals $(1,1,1,1,0,1,1,1)^T$

Symptom	Minimal conflict
$r_L \neq 0$	$\langle F_L, F_1, F_2, F_3, F_4, F_5, F_6\rangle$
$r_0 \neq 0$	$\langle F_L, F_0, F_1, F_2, F_3, F_4, F_5, F_6\rangle$
$r_1 \neq 0$	$\langle F_L, F_0, F_2, F_3, F_4, F_5, F_6\rangle$
$r_2 \neq 0$	$\langle F_L, F_0, F_1, F_3, F_4, F_5, F_6\rangle$
$r_4 \neq 0$	$\langle F_L, F_0, F_1, F_2, F_3, F_5, F_6\rangle$
$r_5 \neq 0$	$\langle F_L, F_0, F_1, F_2, F_3, F_4, F_6\rangle$
$r_6 \neq 0$	$\langle F_L, F_0, F_1, F_2, F_3, F_4, F_5\rangle$

Thus, we have a set of seven *minimal conflicts*. Each conflict is the logical statement claiming one of the faults as listed must be true. Following the logical framework, a *diagnosis* is valid if and only if it can simultaneously satisfy all these seven logical statements.

Using the proposition given by Eqn. (2.3), we can search for all valid *minimal diagnoses* $\mathcal{D}(\Delta)$ under the condition that Δ is a minimal set with at least one element from the set of faults in each conflict. In this case, the list of all valid *minimal diagnoses* can be found to be:

$$\boxed{[F_L] \text{ or } [F_3] \text{ or } [F_i, F_j] \text{ for any } i, j \neq L, 3} \tag{3.31}$$

As shown in Table 3-2, F_L and F_3 are two common elements for all seven conflicts. It follows that $[F_L]$ and $[F_3]$ are two valid single-fault minimal diagnoses. Any combination of two or more faults is also a *valid* diagnosis but since we are considering only *minimal diagnoses,* the other possibilities are limited to two-fault diagnoses $[F_i, F_j]$ for the indices $i, j \neq L, 3$. All other possibilities are either not valid or not minimal.

Therefore, in this scenario the diagnostic result contains two single-fault diagnoses ($[F_L]$, $[F_3]$) and 15 two-fault diagnoses ($[F_i, F_j]$ with $i, j \neq L, 3$). Mathematically, all these 17 diagnoses are equally valid. The only way one can narrow the list down further is by considering the prior probability of each fault. For instance, if all eight faults can be assumed to be equally likely with a small probability, then multiple-fault events can be considered significantly less likely than single-fault events. In that case, one can consider the single-fault assumption and eliminate all multiple-fault diagnoses.

In this framework, non-zero residuals are used to construct conflicts while zero residuals are not utilized. More specifically, the approach only makes a backward reasoning claim that when a residual is non-zero then at least one of the involved components or sensors must be faulty. It does not make any assumption in the forward cause-effect direction or any claim on how an ARR is affected when a fault occurs. In particular, it does not eliminate the possibility that some faults may not be detected by some ARRs in which they are involved. In the context of the current example, such possibility means a sensor may involve in the calculation of multiple residuals, but a fault in that sensor, depending on the magnitude, may affect only some of those residuals. Without neglecting such possibility, no definite logical statement can be drawn from a zero residual.

Overall, the MDB framework is logically sound and mathematically exact. But since no assumption or simplification is made, the diagnostic results in some cases may be too generic. As shown for the current example with the residuals observed to be $(1,1,1,1,0,1,1,1)^T$, there are 17 valid *minimal diagnoses* one need to consider.

3.3.2 Fault Diagnosis Using the FDI Framework

We will now proceed to the FDI framework, following the steps as described in Table 2-2. The first two steps are common to the MBD framework. For step 3, zero residuals are used to reduce the observed conflicts by the *notion of exoneration*.

Consider again the scenario with the residuals observed to be $(1,1,1,1,0,1,1,1)^T$. From the seven non-zero residuals, we can obtain seven minimal conflicts as listed in Table 3-2. The residual $r_3 = 0$ allows us to exonerate all of its relevant faults from the seven conflicts. From the structure of r_3 listed in Table 3-1, the relevant faults are $\{F_L, F_0, F_1, F_2, F_4, F_5, F_6\}$. The process of reducing the observed conflicts is summarized in Table 3-3.

Table 3-3. FDI reasoning process for a scenario with residuals $(1,1,1,1,0,1,1,1)^T$

Symptom	Minimal conflicts	Reduced conflicts
$r_L \neq 0$	$\langle F_L, F_1, F_2, F_3, F_4, F_5, F_6 \rangle$	$\langle F_3 \rangle$
$r_0 \neq 0$	$\langle F_L, F_0, F_1, F_2, F_3, F_4, F_5, F_6 \rangle$	$\langle F_3 \rangle$
$r_1 \neq 0$	$\langle F_L, F_0, F_2, F_3, F_4, F_5, F_6 \rangle$	$\langle F_3 \rangle$
$r_2 \neq 0$	$\langle F_L, F_0, F_1, F_3, F_4, F_5, F_6 \rangle$	$\langle F_3 \rangle$
$r_4 \neq 0$	$\langle F_L, F_0, F_1, F_2, F_3, F_5, F_6 \rangle$	$\langle F_3 \rangle$
$r_5 \neq 0$	$\langle F_L, F_0, F_1, F_2, F_3, F_4, F_6 \rangle$	$\langle F_3 \rangle$
$r_6 \neq 0$	$\langle F_L, F_0, F_1, F_2, F_3, F_4, F_5 \rangle$	$\langle F_3 \rangle$

Thus, after exoneration, all conflicts are reduced to $\langle F_3 \rangle$ and we are left with only one possibility, a fault in sensor S_3. The final diagnostic result in this case is $[F_3]$.

It is clear that in this case, the FDI framework produced a more detailed result. That is possible because the additional information from zero residuals was utilized by the use of the notion of exoneration. Effectively by using the notion of exoneration for all zero residual, the FDI framework makes two claims:

- A fault would result in a non-zero residual for all ARRs it is involved in.
- Multiple faults do not counteract one another to give a zero residual.

Both of these claims are only approximations and not mathematically exact. Nevertheless, the notion of exoneration allows us to simplify the reasoning process and obtain more detailed results.

The two claims in the notion of exoneration can be combined into a single *forward* statement that if one or more faults relevant to a residual occur then the residual is non-zero. Using such statement, one can construct the *forward* mapping from each set of faults to a set of *fault symptoms,* known as the fault signature as discussed in Section 2.2 and illustrated in Figure 2-1. For the current example, the fault signatures for all possible scenarios are listed in Table 3-4.

Table 3-4. Fault signatures for the HX example under the notion of exoneration

	F_L	F_0	F_1	F_2	F_3	F_4	F_5	F_6	*Multiple Faults*	*No Fault*
r_L	1	0	1	1	1	1	1	1	1	0
r_0	1	1	1	1	1	1	1	1	1	0
r_1	1	1	0	1	1	1	1	1	1	0
r_2	1	1	1	0	1	1	1	1	1	0
r_3	1	1	1	1	0	1	1	1	1	0
r_4	1	1	1	1	1	0	1	1	1	0
r_5	1	1	1	1	1	1	0	1	1	0
r_6	1	1	1	1	1	1	1	0	1	0

F_L *: leakage,* F_0 *: fouling,* F_i *for* $i \geq 1$ *: fault in sensor* S_i
1 *indicates a non-zero residual,* 0 *indicates zero residual*

Given the fault signature table, fault diagnosis can be done by simply matching the fault signatures to the observed set of fault symptoms. For the current example, the observed set of symptoms is $(1,1,1,1,0,1,1,1)^T$ which matches the signature for $[F_3]$. The final diagnostic result is $[F_3]$, in agreement with the result we have obtained previously by backward inference.

In this simple example for a single component, we were able to construct the forward mapping represented by a fault signature table. All the multiple-fault states were shown to have the same signature but that is not the case in general. Again, it should be emphasized that the premise of the model-based diagnosis frameworks formulated in Chapter 2 is to perform diagnosis without the need to construct the forward mapping as that task may not be practically possible, especially for complex system with a large number of faults.

Chapter 4

Uncertainty Treatment

The physics-based approach developed in Chapter 3 can be summarized by the following flowchart.

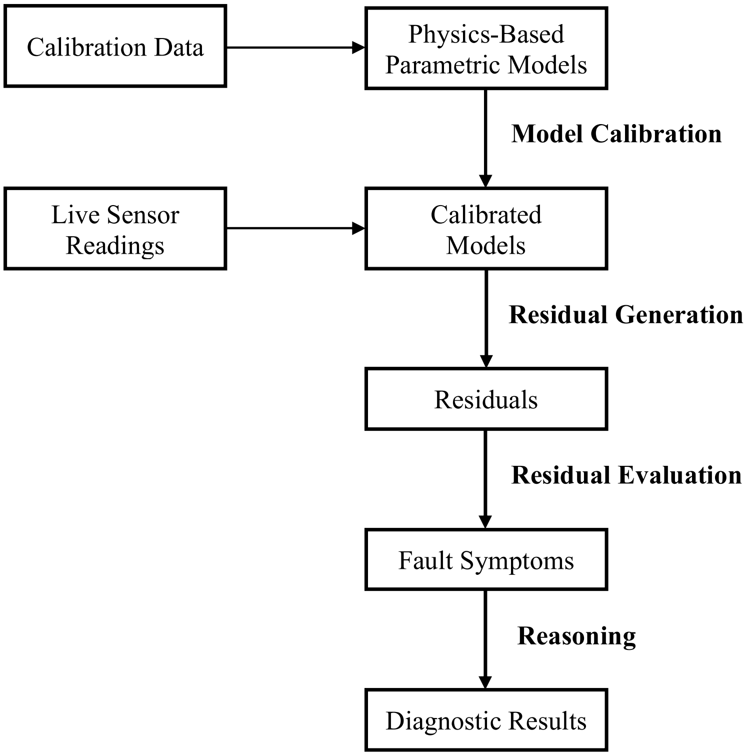

Figure 4-1. Model-based diagnosis framework using physics-based component models

Physics-based parametric models are obtained for each generic type of component from simplifying the underlying physics. Each parametric model may contain several model parameters which are then determined for each specific component in a calibration process using a set of training data. Calibration data can be obtained from the startup data of the system. The calibrated component models provide the source of analytical redundancy relations for fault diagnosis. Residuals are computed from each component model using live sensor readings. Residuals are then evaluated with each non-zero residual serving as a fault symptom. The reasoning process going from fault symptoms to diagnostic results is performed using one of the diagnosis frameworks as discussed in Section 2.4. Such reasoning processes were demonstrated in Section 3.3 for an example of a single-phase heat exchanger.

Up until now, we have not considered the effects of noise and uncertainty. In practice, various sources of uncertainty could be present in a system and affect most of the steps of the flowchart in Figure 4-1. Different approaches to uncertainty treatment in model-based diagnosis are discussed in this chapter.

4.1 Uncertainty Sources and Effects

The two main sources of uncertainty in the current diagnostic problem are measurement uncertainty and modeling uncertainty. Measurement uncertainty, present in both the calibration data and live sensor data, originates from the uncertainty in the reading value of each sensor. Modeling uncertainty comes from both the parametric form for each model and the calibration process. Parametric models are obtained by simplifying the underlying physics and thus inevitably cannot describe the physical phenomena exactly. Furthermore, in the calibration process to determine the model parameters, the presence of measurement uncertainty in the calibration data leads to uncertainty in model parameters. The two sources of uncertainty in model parameters and the parametric form of each model combine into modeling uncertainty in the calibrated models.

In the *residual generation* step, the measurement uncertainty in live sensor readings combines with modeling uncertainty into the uncertainty of each residual computed at each time step. The immediate effect is that all residuals appear ‘noisy’. Subsequently, for the *residual evaluation* step, a statistical treatment is needed to decide at each given time if a residual is statistically zero or non-zero. Such statistical treatment is generally known as *change detection*. A residual is observed to be *non-zero* when its distribution is detected to have deviated from the original expected distribution.

In statistical change detection, there are two types of detection error:

- Type I error (false positive): A residual is zero but detected to be non-zero.
- Type II error (false negative): A residual is non-zero but detected to be zero.

Consequently, the uncertainty in the value of each residual inevitably leads to uncertainty in the observed fault symptoms. There is an associated false alarm rate whenever a fault symptom is observed, and there is a possibility that some fault symptoms may not be detected.

In the reasoning frameworks described in Chapter 2, one proceeds from a definite set of fault symptoms and cannot consider the possibility that the observed fault symptoms could be wrong. As discussed, in the presence of uncertainty, some of the observed symptoms could be false and some of the fault symptoms may be undetected. Performing reasoning on a set of incorrect fault symptoms would inevitably lead to false diagnoses.

Furthermore, for practical applications, not all type of component faults and sensor faults are equally likely to occur. Such information, provided as the prior probability of each fault, is relevant to fault diagnosis. Dealing with the possibility of false alarms and the prior probability of different types of faults in the reasoning process is the topic of *reasoning under uncertainty.*

4.2 Reasoning Under Uncertainty

The two types of error in change detection lead to the possibility that an observed set of fault symptoms could be false. In general, such false alarm rates depend on the statistical change detection tool being used and the magnitude of the changes relatively to the uncertainty of the residuals. A straightforward approach in dealing with the possibility of false observations is to rely on a statistical change detection tool to minimize the false detection rates. Then if the false alarm rates can be reduced to a tolerable level, they can be neglected, and the two reasoning frameworks developed in Chapter 2 can be directly. Such treatment is referred to as *deterministic reasoning*. The term “deterministic” is used to emphasize that these reasoning approaches take the inputs as definite fact. The two deterministic reasoning approaches correspond to the two model-based diagnosis frameworks described in Section 2.4 are:

- Deterministic I: Use statistical change detection to detect non-zero residuals then apply the MBD framework (summarized in Table 2-1) for fault diagnosis.
- Deterministic II: Use statistical change detection to detect non-zero residuals then apply the FDI framework (summarized in Table 2-2) for fault diagnosis.

Information on prior probability of the faults can be used to eliminate some of the less probable diagnoses from the results obtained using Deterministic I or II. Various statistical change detection methods are summarized in the next section with the details to be provided in Appendix B.

In general, however, it is not always possible to reduce the false alarm rates to a negligible level. For a given change detection method, the rate of type I error depends on the detection threshold whereas the rate of type II error depends on both the detection threshold and the magnitude of the change relatively to the uncertainty of the residual. Raising the detection threshold would reduce the rate of type I error but increase the rate of type II error. Also, the rate of type II error depends on the ratio between the change in mean value and the standard deviation representing the uncertainty. Recall that the residual uncertainty originates from both measurement uncertainty and modeling uncertainty. Thus, the rate of type II error depends on the quality of the component models used for diagnosis.

Therefore, depending on the quality of the component models, it may not be possible to choose a detection threshold such that both type I and type II errors can be neglected. In that case, one must account for the possibility of false alarms in the reasoning process. This is the motivation for the probabilistic reasoning framework developed in this study and will be discussed in Chapter 5.

4.3 Statistical Change Detection

In the context of the current application, the problem of statistical change detection is to detect whether the mean value of a noisy residual has deviated from its normal value, i.e. changing from zero to non-zero. We will assume that the mean μ_0 and standard deviation σ_0 in the zero state are known or can be estimated. The new mean value after a change is unknown but the standard deviation is assumed to remain the same. The change in mean value can occur either as an abrupt shift or a slow drift.

The most straightforward approach to detect a change in mean value is by using the limit-checking method, formally known as the Shewhart control chart [87, 88]. A change is detected when the difference between the current value and the normal mean exceeds a pre-defined threshold. This approach is easy to implement but it was shown to be less effective for detecting small changes [89].

Other approaches include the exponentially weighted moving average (EWMA) control chart and the generalized likelihood ratio (GLR) test [89, 90]. The criteria to assess the performance of a change detection method includes detection delay, false positive and false negative rates. Comparison between various change detection methods can be found in [89]. We will be using the GLR test for the current application. The methodology of the GLR method is summarized in the remainder of this section.

Consider a noisy variable that can be described by a Gaussian distribution with known mean μ_0 and standard deviation σ_0. After a change, the mean value of the variable shifts to an unknown value μ . We would like to detect the change, estimate the time step at which the change started and the new mean μ .

The values of the variable, collected at discrete time steps, can be put into a time series $\{y_k\}$. For the GLR test, at a time step k , a decision function can be evaluated using past values of the variable and a change is detected when the decision function exceeds a pre-defined threshold. The GLR decision function to detect a shift in mean value is given by:

$$g_k = \frac{1}{2\sigma_0{}^2} \max_{1 \le j \le k} \frac{1}{k-j+1} \sum_{i=j}^{k} (y_i - \mu_0)^2 \qquad (4.1)$$

If a change is detected, the location of the change is the index j that maximizes the above expression. The detection threshold can be defined based on a pre-defined tolerable false detection rate.

For the case of slow drift in mean value, as opposed to an abrupt shift, a slight modification is needed, as discussed in [91, 92, 93]. The GLR-D decision function for detecting small drift is given by:

$$g_k = \frac{1}{2{\sigma_0}^2} \max_{1 \le j < k} \frac{\left[\sum_{i=j}^{k} (i-j)(y_i - \mu_0) \right]^2}{\sum_{i=j}^{k} (i-j)^2} \tag{4.2}$$

To demonstrate, consider the heat exchanger example with the residual r_0 computed for the heat transfer model given by Eqn. (3.28). Gaussian noise was added to simulate measurement uncertainty. The mean and standard deviation of the residual was estimated during normal operation when the residual can be considered statistically zero. The plot of the residual at each time step and the corresponding GLR-D decision function is shown in Figure 4-2. A change is detected when the decision function exceeds the detection threshold, shown by the red line on the right plot.

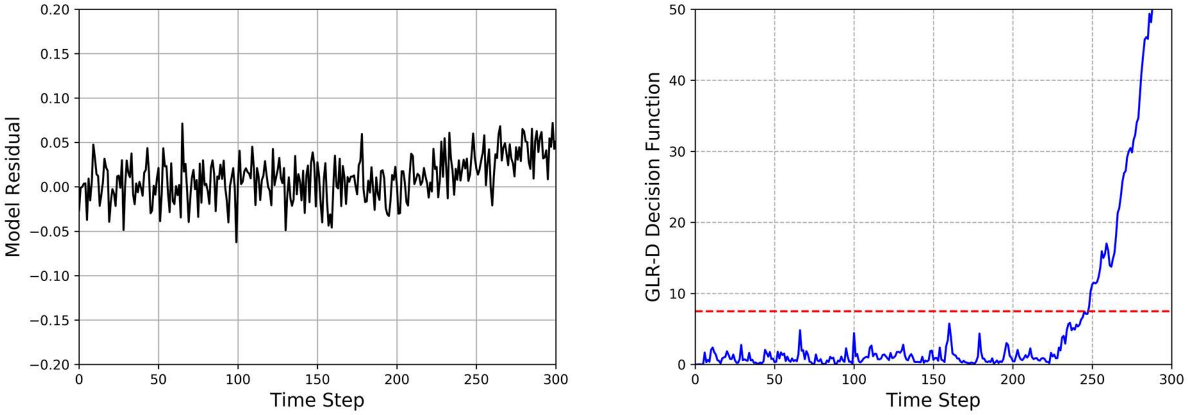

Figure 4-2. Application of the GLR-D test to detect non-zero residual

4.4 Deterministic Reasoning Approaches

To demonstrate the two deterministic reasoning approaches, we consider again the example with a single-phase counter flow heat exchanger as analyzed in Chapter3. Simulation data was obtained for the heat exchanger using design parameters from those of a regenerative heat exchanger in the chemical and volume control system (CVCS) of the Braidwood Nuclear Generating Station [3]. For reference, the operating conditions and geometry specifications are listed in Table 4-1, as modeled in the GPASS 1-D system code [94].

Table 4-1. Reference operating conditions and geometry specifications of the Braidwood CVCS regenerative HX

Parameter	*Value*
Hotside mass flow rate (kg/s)	4.724
Hotside inlet temperature (^{o}C)	290
Hotside pressure (MPa)	15.0
Coldside mass flow rate (kg/s)	3.467
Coldside inlet temperature (^{o}C)	40
Coldside pressure (MPa)	14.65
Configuration Type	Shell-Tube
Number of tubes	256
Tube inner diameter (mm)	9.525
Tube outer diameter (mm)	12.633
Shell inner diameter (m)	0.254
Total length (m)	5.0
Tube wall thermal conductivity (W/m.K)	25
Tube wall roughness (m)	0.00001

For the calibration process, GPASS simulation data with noise added for uncertainty was used to calibrate the UA model given by Eqn. (3.21). The eight residuals were generated as described in Section 3.3. Afterwards, the mean μ_0 and standard deviation σ_0 of each residual was estimated using data sampled around the reference operating point. These distribution parameters (μ_0, σ_0) are needed for the GLR test to evaluate each residual. These steps are straightforward and are omitted here.

Consider a scenario with sensor S_3 drifting out of calibration starting at certain time step. The sensor fault was simulated by an increasing bias added to its reading value. More specifically, the simulated fault started at $t_{cp} = 200$ with an increasing bias rate of 5% per 100 time-steps. Applying the GLR-D test to each of the 8 residuals, the results are plotted on Figure 4-3.

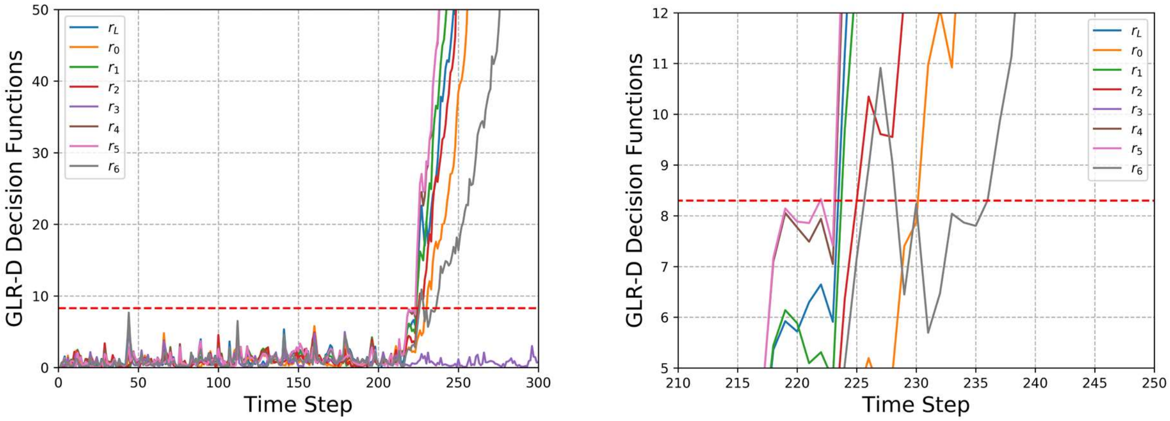

Figure 4-3. GLR-D decision functions for the 8 residuals in a sensor fault scenario (left) and a zoomed-in version (right).

The detection threshold, as shown by the dashed red line in Figure 4-3, was set to $h = 8.3$ for a false positive rate of 0.1%. After sufficient wait-time, all the residuals except r_3 can be observed to be non-zero as their decision functions eventually exceed the threshold. The observed set of fault symptoms is $(1,1,1,1,0,1,1,1)^T$ after around timestep 236. We can now apply the two deterministic reasoning approaches to perform fault diagnosis. As already discussed in Section 3.3, the diagnostic result by Deterministic I for this particular set of fault symptoms is given by Eqn. (3.31) consisting of 17 minimal diagnoses. The result by Deterministic II is $[F_3]$, which is the correct diagnosis.

The notion of exoneration using zero residuals from the FDI framework allows Deterministic II to provide more detailed diagnostic results. However, in situations where that notion does not hold, the Deterministic II approach may fail to produce a valid diagnosis. In the presence of uncertainty, such situations arise more often due to the possibility of false negative in change detection, i.e. a residual with changed mean value detected to be zero.

To elaborate on the limitation of the Deterministic II approach, notice in Figure 4-3 that for a period prior to time step 236, the GLR-D decision function for residual r_6 dropped below the detection threshold and the change detection algorithm failed to detect the change. During that period, the observed fault symptoms are $(1,1,1,1,0,1,1,0)^T$. For Deterministic I, we have a set of 5 minimal conflicts (see Table 3-2, excluding the last row) which lead to the following minimal diagnoses:

$$\boxed{[F_L] \text{ or } [F_3] \text{ or } [F_6] \text{ or } [F_i, F_j] \text{ for any } i, j \neq L, 3, 6} \quad (4.3)$$

In comparison with Eqn. (3.31), the consequence of the false negative, r_6 incorrectly observed to be zero, is that we now have one additional single-fault diagnosis F_6. On the other hand, the deterministic II approach failed to produce a valid diagnosis since the combination of $r_3 = 0$ and $r_6 = 0$ exonerates all 8 faults.

Generally, in the presence of uncertainty, Deterministic II is more sensitive to errors in change detection. In Deterministic I, the effect of change detection false alarms often results in an increased number of possible diagnoses. On the other hand, in Deterministic II, the set of observed fault symptoms is required to exactly match the fault signature defined under the notion of exoneration. Thus, false alarms that invalidate the notion of exoneration often result in no valid diagnosis being found.

An alternative approach to deal with such issue in Deterministic II is to use a concept of distance when matching the observed fault symptoms to fault signatures. Instead of requiring a fault signature to exactly match the observed symptoms, one can use a definition of distance, e.g. the norm of the difference, to find a fault signature closest to the observed set of fault symptoms. This way, the Deterministic II approach will always produce at least one diagnosis. However, in doing so, all false detection rates are implicitly assumed to be equal. Furthermore, in such process, one cannot incorporate information on the prior probability of each fault. Thus, in general, the shortest distance may not necessarily imply the most likely diagnosis.

To overcome these issues, one must take account of both the fault prior probabilities and the possibility of false alarms in the reasoning process. In the next chapter, we will discuss the probabilistic reasoning framework for quantitative model-based diagnosis in which diagnostic results are determined based on the fault posterior probabilities computed using both prior probabilities and the conditional probabilities between related variables.

Chapter 5

Probabilistic Model-based Diagnosis Framework

The characteristics of the two quantitative reasoning frameworks formulated in Chapter 2, namely the MBD and FDI frameworks, were discussed in Chapter 3. We have demonstrated that the MBD framework, inherently a logically sound framework, cannot fully utilize the information available from zero and non-zero residuals. The FDI framework relying on the *notion of exoneration* can generally provide more detailed diagnostic results. By applying the notion of exoneration, one must assume that a residual is only zero if none of the relevant faults in the underlying analytical redundancy relation is present. While the notion of exoneration is not mathematically exact, it can be considered a good approximation for practical use in the absence of measurement and modeling uncertainty.

When considering various sources of uncertainty in a system, the straightforward approach is to neglect the possibility of false detections in evaluating residuals and directly apply the two quantitative reasoning frameworks. We referred to such approach as *deterministic reasoning*. The deterministic reasoning approaches are justifiable when the false alarm rates in evaluating noisy residuals can be minimized to an insignificant level by a statistical change detection tool. However, that is not always possible, especially when modeling uncertainty is significant. In particular, when modeling uncertainty is high relatively to the amplitude of the effects of the faults we are trying to detect, it may not be possible to reduce the possibility of false negative to a negligible level. When the false negative rate in residual evaluation is significant, non-zero residuals can be falsely detected to be zero and the notion of exoneration for zero residuals may longer be a good approximation. We have demonstrated the limitations of the two deterministic reasoning frameworks in Chapter 4.

In the two quantitative reasoning frameworks, diagnostic results are obtained from a definite set of observed fault symptoms. In the presence of uncertainty, false detections in evaluating noisy residuals could result in uncertain observation of fault symptoms. For a probabilistic reasoning framework, one must account for the possibility that some of the observations could be false in the reasoning process going from observed fault symptoms to fault diagnoses. Furthermore, for a general system, some faults are more likely to occur to others. Information on the prior probability of each fault is relevant to fault diagnosis, especially when there could be multiple valid diagnoses for the same set of fault symptoms. Incorporating both the prior probability of faults and the possibility of false observations into the reasoning process of fault diagnosis is the focus of this chapter on probabilistic reasoning.

5.1 Probabilistic Reasoning using Bayesian Network

As discussed in Chapter 2, each state of a system in a fault diagnosis problem can be identified by a set of faults. The physical cause-effect relations dictate that each set of faults result in a certain set of fault symptoms. In the reasoning process, we are interested in inferring the state of the system from an observed set of fault symptoms. In a probabilistic setting, the reasoning process becomes the task of computing the probability of each physical state given the set of observed symptoms, known as the *posterior probability*.

To elaborate, let us consider the simplest case: a system with a single fault and a single observation denoted by binary variables F and O, respectively. $F = 0$ if the system is fault-free and $F = 1$ if the fault is present. Suppose that we cannot measure F directly and can only make observations through O. From a model of the system, we know that $F = 1$ would cause O to be 1 and thus, O can be considered a fault symptom. The cause-effect relation is $F = 1$ leads to $O = 1$. In the reasoning process for fault diagnosis, one can infer $F = 1$ if O is observed to be 1.

In the presence of noise and uncertainty, one cannot observe O exactly and in general cannot establish deterministic cause-effect relations. Statements such that $F = 1$ leads to $O = 1$ must now be provided in terms of conditional probabilities, e.g. $P(O = 1 | F = 1)$. In the forward cause-effect direction, one must provide the probability $P(F)$ for each value of F known as the *prior probability* of the fault; and $P(O | F)$ for each combination of F and O known as the *likelihood* of the observations. Intuitively, the *prior probability* indicates how likely it is for the fault to occur and the *likelihood* specifies how often each value of O is observed in each state of the system.

In performing probabilistic reasoning, one must compute the *posterior probabilities* $P(F \mid O)$ for the observed value of O. If O is observed to be 1, the *posterior probability* $P(F=1 \mid O=1)$ indicates how probable it is that the fault is present. One can then conclude that the fault has occurred if $P(F=1 \mid O=1)$ is close to 100%. The *posterior probability* can be computed using Bayes' theorem:

$$P(F=F_i \mid O=O_i) = \frac{P(O_i \mid F_i) \times P(F_i)}{P(O_i)} \tag{5.1}$$

Since the *prior probability* $P(F_i)$ and *likelihood* $P(O_i \mid F_i)$ are provided, the *posterior probability* can be computed if the *marginal probability* $P(O_i)$ is known. $P(O_i)$ is the probability to observe O_i regardless of whether the fault is present. $P(O_i)$ can be computed by summing over all possibilities in a process known as *marignalization*:

$$\begin{aligned} P(O_i) &= \sum_{F_j} P(O_i \mid F_j) F_j \\ &= P(O_i \mid F=0) \times P(F=0) + P(O_i \mid F=1) \times P(F=1) \end{aligned} \tag{5.2}$$

In fact, to decide if the fault has occurred, one only need to compare the relative magnitude of the *posterior probability* of the fault, $P(F=1 \mid O_i)$, with that of the other possibility, $P(F=0 \mid O_i)$. In that case, since $P(O_i)$ is common to the two posibilities, it may not be necessary to compute $P(O_i)$ explicitly.

In summary, to perform probabilistic reasoning for this simple example, one need to provide the *prior probability* of the fault and the *likelihood* of the observations. Afterwards, the *posterior probability* of the fault can be computed.

To apply this probabilistic reasoning process to a general problem of fault diagnosis, there are several complications. For a system in general, there are various faults and observations with each observation only affected by some of the faults. Furthermore, the faults may not directly affect the observations, but the causal effects may propagate through various intermediate variables. Therefore, it is generally not possible to pre-determine the likelihood $P(O \mid F)$ for complex systems. Instead, one must provide the conditional probability distributions of directly related variables. For instance, if the effects of fault F propagate to O through an intermediate variable r, one must provide the distributions $P(r \mid F)$ and $P(O \mid r)$ from which the *likelihood* $P(O \mid F)$ can be computed.

The conditional dependencies between different variables in a system can be represented graphically by a Bayesian network. Formally, a Bayesian network is a *directed acyclic graph* whose *nodes* represent random variables and directed *edges* represent conditional dependent relations [95]. The arrow directions of the directed edges represent cause-effect directions. As an example, the simple system with a single fault F and single observation O we have considered above can be represented by the Bayesian network shown in Figure 5-1.

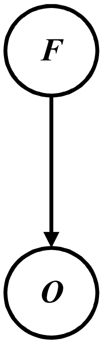

Figure 5-1. Bayesian network representation of the system with a single fault and single observation

In this case, we only have two nodes, representing the two variables F and O. The direct effect of F on O is represented by the arrow between the two nodes. As discussed, for a general system, we may have any number of nodes and edges. To completely define a system for probabilistic reasoning, one need to define all the nodes to represent relevant variables, directed edges between the nodes to represent the cause-effect relations and provide the conditional probability distribution of each node on its parent nodes. Generally, the first layer, consisting of nodes with no parent nodes, represent physical states and the last layer represents observations.

The structure of the Bayesian network allows one to compute the *likelihood* of the observations and from that the *posterior probability* of the physical states.

5.2 Probabilistic Reasoning for Quantitative Model-Based Diagnosis

In the framework of quantitative model-based diagnosis, the state of a system at any given time can be described by a set of faults. Faults, including component faults and sensor faults, can lead to changes in the mean values of certain model residuals which will be observed by change detection tools as non-zero residuals. The observations of zero and non-zero residuals in the evaluation of model residuals are the inputs for the reasoning process for fault diagnosis.

More specifically, a fault can directly lead to changes in the mean value of a model residual. At the same time, the computed value of the residual is also affected by measurement and modeling uncertainty. The observation on whether the residual is statistically zero or non-zero is obtained from performing statistical change detection on the computed residual. Consider a system with a single fault mode and a single model residual, the dependencies between the variables can be represented by the following Bayesian network.

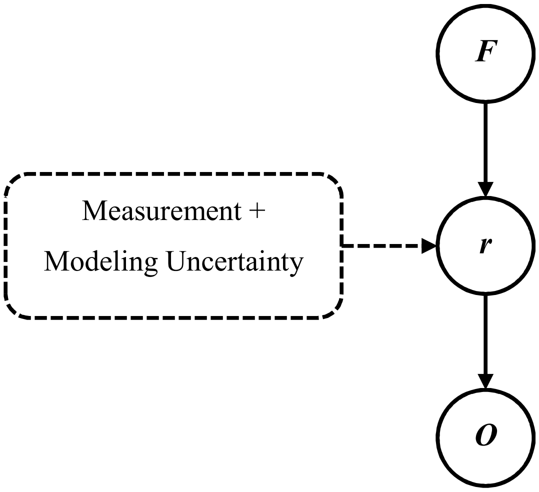

Figure 5-2. General Bayesian network for quantitative model-based fault diagnosis

The change detection output O only depends on the time series of the computed values of residual r. At any given time, the value of r depends on the various uncertainty sources and on whether the fault F has occurred. Usually, the combined measurement and modeling uncertainty can be considered constant and thus can be omitted from the graph. To completely define the structure of the network, one need to specify the probability distributions for nodes, which in this case are $P(F)$, $P(r|F)$, and $P(O|r)$.

$P(F)$ the *prior probability* of the fault, indicating how likely it is for the fault to occur. The conditional probability distribution $P(r|F)$ of residual r on fault F is directly related to the sensitivity of the residual on the fault. The conditional probability distribution $P(O|r)$ of O on r is related to the sensitivity of the change detection tool.

Note that the observation O can be considered binary, i.e. 0 if the detection output is zero and 1 if *non-zero*, but the residual r and fault F are natively continuous variables. One may take the magnitude of the fault as the continuous value for F. The change in mean value of residual r, as the result of a fault, is continuous and depends on the magnitude of the fault. In this case, one need to provide the conditional distributions $P(r|F)$ and $P(O|r)$ in forms of probability density functions. Alternatively, one may choose to discretize the fault F and residual r into binary variables:

- $F=1$ if the fault has occurred with magnitude exceeding a certain threshold; $F=0$ otherwise.
- $r=1$ if there is a change in the mean value of the residual that exceeds a certain threshold; $r=0$ otherwise.

The discretization of faults into binary variables comes naturally from the practical use. As is the case with deterministic reasoning, at any given time, we are mostly interested in whether a fault has occurred. Thus, each fault is practically a binary variable. Considering the residuals as binary variables, the CPTs $P(O|r)$ can be computed directly from the false positive and false negative rates of the change detection tool. Such discretization process of the faults and residuals allows one to simplify the calculation and reduce the computational cost. Although the use of a finer discretization can theoretically provide more accurate computation, it is not necessary for the current practical application.

The discretization of sensor faults is straightforward. For each sensor in practice, there are pre-defined thresholds to determine when the sensor is considered out of calibration. One can then simply define the binary sensor fault based on such thresholds. For component faults, the threshold may be set based on the effect of the fault on performance-related parameters. For example, one can set the threshold for the definition of fouling based on the change of the overall heat transfer coefficient in a heat exchanger. After the threshold for each fault is defined, the *prior probability* of the fault, $P(F)$, may be assumed or estimated from past data.

After the threshold for each residual is defined, one can estimate the distributions $P(r \mid F)$ and $P(O \mid r)$ by based on the sensitivity of the residual on the fault and the characteristics of the change detection tool being used. The distributions $P(r \mid F)$ and $P(O \mid r)$ will be provided as conditional probability tables (CPTs) for each combination of the discrete variables:

Table 5-1. Conditional probability tables for discrete faults and residuals

	$F=0$	$F=1$
$r=0$	$P_{r\mid F}(0\mid 0)$	$P_{r\mid F}(0\mid 1)$
$r=1$	$P_{r\mid F}(1\mid 0)$	$P_{r\mid F}(1\mid 1)$

	$r=0$	$r=1$
$O=0$	$P_{O\mid r}(0\mid 0)$	$P_{O\mid r}(0\mid 1)$
$O=1$	$P_{O\mid r}(1\mid 0)$	$P_{O\mid r}(1\mid 1)$

The entries for the $P(r \mid F)$ CPT can be estimated by sampling the model being used to compute the residual. Let $\Delta\mu$ be the shift in mean value that was used as the threshold to discretize the residual r. Then, for example, $P_{r\mid F}(1\mid 1)$ is the probability that the change in residual r as the effect of fault F is larger than $\Delta\mu$. Each entry of the CPT for $P(r \mid F)$ depends on the chosen threshold $\Delta\mu$ and the sensitivity of the model to the fault F.

In the CPT for $P(O \mid r)$, $P_{O\mid r}(1\mid 0)$ is the false positive rate, i.e. the probability to detect a change when $r=0$. $P_{O\mid r}(0\mid 1)$ is the false negative rate, the probability that a change is undetected. Note that the computed value for each residual is subject to measurement and modeling uncertainty. Let σ denote the standard deviation of the residual as the result of the combined uncertainty. The ratio $\alpha = \Delta\mu/\sigma$ is known as the *signal-to-noise ratio*. For each change detection tool, the false detection rates can be pre-computed as functions of the *signal-to-noise ratio* α.

The estimation process for the CPTs $P(r \mid F)$ and $P(O \mid r)$ depends on how the threshold $\Delta\mu$ is defined. Physically, it is reasonable to set $\Delta\mu$ based on the sensitivity of the fault, allowing

$P_{r|F}(1|1)$ to be approximately set to 1.0. In general, each residual may depend on multiple faults in which case one can set $\Delta\mu$ based on the sensitivity of the least sensitive faults but will need to sample each combination of faults to compute the entries of the CPT. In this case, the ratio $\alpha = \Delta\mu/\sigma$ is different for each residual and one will need to compute a $P(O|r)$ CPT for each residual. An alternative, and somewhat more convenient, approach is to set $\Delta\mu$ based on the 'noise level' σ, i.e. choose the same $\alpha = \Delta\mu/\sigma$ for all residuals. In that case, the same $P(O|r)$ table is applicable for all residuals and one just need to sample the model to compute the $P(r|F)$ table for each residual.

In summary, a Bayesian network for probabilistic reasoning in quantitative model-based diagnosis consists of three layers: the nodes on the first layers represent different faults in the system; the second layer represents model residuals and the third layer represents change detection observations. Each residual may be connected to various faults as dictated by the underlying model. On the other hand, each change detection output only depends on the residual on which change detection is performed. To define the structure of the network, one needs to provide the prior probability $P(F)$ and the conditional probability tables $P(r|F)$ and $P(O|r)$.

After the diagnostic problem has been formulated in form of a Bayesian network with its structure defined, the existing methods of Bayesian inference can be applied to compute the posterior probability of each fault given a set of observation. The structure of the Bayesian network can be used as input for a generic Bayesian network tool or probabilistic reasoning engine to perform the posterior probability calculation. Efficient algorithms for such calculations, which fall outside the scope of this thesis, are omitted here. Overall, the probabilistic reasoning framework for quantitative model-based diagnosis is summarized to the four main steps listed in Table 5-2.

Table 5-2. The proposed probabilistic reasoning framework for quantitative model-based diagnosis

Step 1	Identify available ARRs from component models, from which define possible *residuals*.
Step 2	Construct a Bayesian network representing the dependencies between *faults* and *residuals*
Step 3	Compute the *posterior probability* of each fault given a set of observed zero and non-zero residuals.
Step 4	Obtain the final diagnosis from faults with significant *posterior probability*.

5.3 Results for the Heat Exchanger Example

As an example, let us consider the case with a single-phase counterflow heat exchanger as discussed in Chapter 3 and 4. The available residuals and related component and sensor faults have been identified in Chapter 3 and listed in Table 3-1. The diagnostic problem in this case can be represented by the Bayesian network in Figure 5-3.

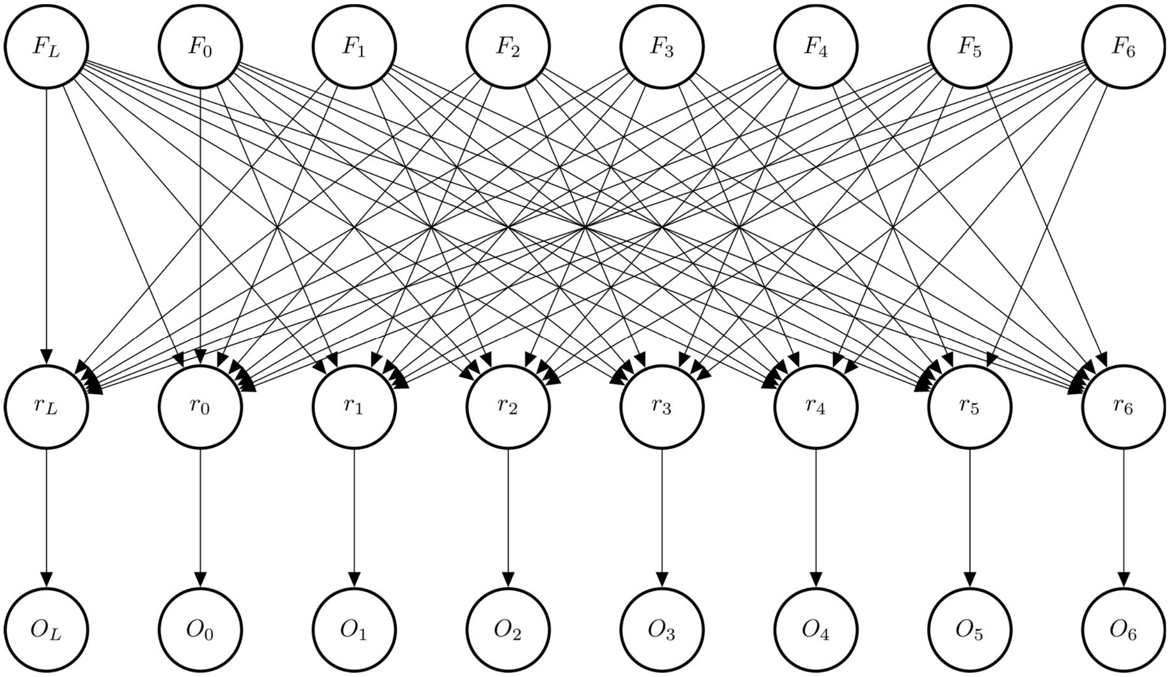

Figure 5-3. A Bayesian network for fault diagnosis in the single-phase heat exchanger example

The first layer consists of eight nodes representing the eight component and sensor faults. The second layer represents the eight model residuals. Each relevant fault to a residual is connected to the residual by an arrow as shown. Effects of measurement and modeling uncertainty are considered constant and omitted in the drawing. Overall, the structure of the network represents the causality relations between faults, residuals and the observations. For instance, the true mean value of each residual depends on certain faults as indicated by the arrows: fault F_0 does not affect r_0 as there is no edge connecting the two.

To completely define the Bayesian network, we must specify the conditional probability distribution at each node. To demonstrate the calculation of posterior probabilities in probabilistic reasoning, for this section we will assume the faults and residuals have been discretized into binary variables. For the first layer, we need to provide the prior probability of each fault, which usually depends on the fault type and time into the operation cycle. Prior probabilities can be estimated from past experience with the system. In this example, we will arbitrarily assume that at the time of consideration, sensors have a 5% chance to be out of calibration; the prior probability for fouling in the heat exchanger is 10% chance and that for leakage is 1% , making it the least likely fault:

$$P(F_L = 1) = 0.01 \tag{5.3}$$

$$P(F_0 = 1) = 0.10 \tag{5.4}$$

$$P(F_i = 1) = 0.05 \text{ for } i \geq 1 \tag{5.5}$$

The conditional probability of each residual on its relevant faults depends on the sensitivity of the underlying model to each fault. Furthermore, in multiple-fault scenarios, it is possible that faults can counteract one another. In general, the conditional probability distribution of each residual can be computed by sampling the underlying model. For the current demonstration, we will apply the two assumptions that were used for the notion of exoneration in the FDI framework:

- A fault would result in a non-zero residual for all ARRs it involves in.
- Multiple faults do not counteract one another to give a zero residual.

Under these assumptions, a residual is non-zero if at least one of its parent nodes takes value 1, thus:

$$P(r_i = 0 \mid \text{all parent nodes} = 0) = 1.0 \tag{5.6}$$

$$P(r_i = 1 \mid \text{any parent node} = 1) = 1.0 \tag{5.7}$$

The conditional dependence of each observation O_i, i.e. change detection output, on the true mean value of the residual r_i depends on the false positive and false negative rates of the change detection algorithm. These false detection rates depend on the specific change detection method being used, the change magnitude relatively to the variance and the chosen detection threshold. For the demonstration, suppose that the detection threshold has been chosen such that the false positive rate is 0.1% and the false negative rate is 1.0%. Thus, by definition:

$$P(O_i = 1 \mid r_i = 0) = 0.001 \tag{5.8}$$

$$P(O_i = 0 \mid r_i = 1) = 0.01 \tag{5.9}$$

In Eqns. (5.3) - (5.9), we have provided the conditional probability for every node in the Bayesian network. Thus, the structure of the network is completely defined. Existing methods of Bayesian network can then be employed to compute the *marginal probability* and *likelihood* of each set of observations and from that the *posterior probability* for each fault or each diagnosis.

For example, consider the case with the fault symptoms observed to be $O = (1,1,1,1,0,1,1,1)^T$, as discussed in Sections 3.3 and 4.4, the *posterior probability* for $F_3 = 1$, i.e. for sensor S_i to be faulty, can be calculated to be:

$$P(F_3 = 1 \mid O = (1,1,1,1,0,1,1,1)^T) = 0.986 \tag{5.10}$$

The posterior probability of all other faults is negligible. Thus, we can conclude that F_3 is the most likely fault. Here $P(F_3 = 1 \mid O)$ represents the probability for fault F_3, i.e. sensor S_3 out of calibration, regardless of the status of the other faults. The probability for the single-fault diagnosis $P(F_3 = 1, F_i = 0 \text{ for } i \neq 3 \mid O)$ can be computed similarly and is slightly lower.

For the case with the fault symptoms $O = (1,1,1,1,0,1,1,0)^T$, which the Deterministic II approach failed to produce a valid diagnosis due to the invalid notion of exoneration, the two dominated possibilities are:

$$P(F_3 = 1 \mid O = (1,1,1,1,0,1,1,0)^T) = 0.498 \tag{5.11}$$

$$P(F_6 = 1 \mid O = (1,1,1,1,0,1,1,0)^T) = 0.498 \tag{5.12}$$

The probability for other faults is less than 1%. Hence, we can conclude that the two most likely faults are F_3 and F_6. In comparison, recall the diagnosis produced by the Deterministic I approach, given by Eqn. (4.3), which includes $[F_L]$ and multiple two-fault diagnoses, in addition to F_3 and F_6. Here we were able to use the *posterior probability* to eliminate $[F_L]$ and the multiple-fault diagnoses.

The proposed probabilistic framework requires posterior probability calculation for each fault in the system. This task could be computationally expensive if the number of possible faults is large. Additionally, it may be helpful to consider specific diagnosis instead of each fault independently, e.g. $[F_3]$ which implies $F_3 = 1, F_i = 0$ for $i \neq 3$ instead of $F_3 = 1$ regardless of the other faults. The alternative approach is by combining the Deterministic I approach with the probabilistic framework: One can use the Deterministic I approach to produce a list of minimal diagnoses and then use the probabilistic framework to compute the posterior probability for each minimal diagnosis. Diagnoses with insignificant posterior probability can then be eliminated.

The diagnostic results by the three reasoning approaches for the two scenarios of the single-phase the heat exchanger example is summarized in Table 5-3.

Table 5-3. Results for the two diagnostic scenarios for the single-phase HX

Symptoms	Deterministic I	Deterministic II	Probabilistic
$(1,1,1,1,0,1,1,1)^T$	$[F_L]$ or $[F_3]$ or $[F_i, F_j]$ for any $i, j \neq L, 3$	$[F_3]$	$[F_3]$
$(1,1,1,1,0,1,1,0)^T$	$[F_L]$ or $[F_3]$ or $[F_6]$ or $[F_i, F_j]$ for any $i, j \neq L, 3, 6$	None found	$[F_3]$ or $[F_6]$

Effectively in this probabilistic framework, information from zero residuals can be utilized in a similar manner to the notion of exoneration but without suffering its limitation due to false detections in residual evaluations. Overall, by considering the possibility of false alarms and the prior probability of each fault in the reasoning process, the probabilistic approach can provide improved diagnostic results in the presence of modelling and measurement uncertainty.

Chapter 6

System Level Diagnosis

The methods we have discussed in the last four chapters constitute the overall quantitative diagnostic framework. In Chapter 3, physics-based models are constructed to describe the fault-free behavior of T-H components. Quantitative residuals can be generated from each component model to serve as possible fault symptoms. Chapter 2 provides the reasoning frameworks to obtain valid fault diagnoses from a set of observed fault symptoms. The presence of measurement and modeling uncertainty affects all the steps from model construction, residual generation to diagnostic reasoning. Treatments of the effects of uncertainty are discussed in Chapter 4 and 5. The overall process has been demonstrated for the example of a single-phase heat exchanger.

To apply this physics-based diagnostic framework to complex T-H systems, the general strategy is to decompose each system into separate components of known generic types whose physical behavior can be described by pre-defined models. Fault-free models for each generic component type can be formulated from the underlying physical laws in forms of parametric models. Each parametric model may contain a few unknown parameters which are to be determined in a process, referred to as *model calibration*, for each specific component by using training data obtained from measured data of various process variables on the boundary of the component.

To provide the required measured data for the model calibration process, a certain number of sensors available to the component is needed. For the discussion in the previous chapters, we have assumed that there are enough sensors on the boundary of each component for that purpose. That is not usually the case in practice. In fact, most T-H systems in nuclear power plants are not fully instrumented and it is rarely the case that one has sufficient sensors to allow the calibration of the parametric models for each standalone component. To improve the diagnostic capability, one needs to utilize information available not just locally to each component but also from other components and sensors in the system.

In case the sensor set available at the inlet and outlet of a component is insufficient for the calibration process, there are two possible directions one can follow for a solution:

- Construct parametric models for a combination of multiple components that can then be calibrated by the available sensor set.
- Compute the missing sensors by utilizing other sensors at system level and the relations between the component with nearby component on its upstream and downstream.

The first solution produces *aggregate models* as permitted by the available sensors, i.e. physics-based models covering multiple physical components. The second option gives rise to the concept of *virtual sensors*. Virtual sensors are created in place of missing physical sensors by solving system balance equations. The calculation of each virtual sensors involves certain physical sensors and components. For the purpose of model construction, during which the system can be considered fault-free, each virtual sensor acts as a physical sensor to provide calibration data for component models. For fault diagnosis, special care must be taken since the validity of each virtual sensor depends on the status of the involved components and physical sensors. For the framework developed in this thesis, both concepts of aggregate models and virtual sensors are utilized to maximize the diagnostic capabilities for T-H systems with limited sensor sets.

6.1 Virtual Sensors

For each T-H component of a known generic type, one can formulate models to describe its performance in the processes of mass, momentum and energy transport. Physics-based models for the component are generally expressed by parametric models with a few unknown parameters. For each component model, the process of determining the model parameters and the subsequent use of the model in monitoring the component require measurement data of certain process variables. For example, as discussed in Section 3.1, the mass balance model for a single inlet/outlet component requires two flowrate sensors whereas the momentum model require two pressure sensors and a flowrate sensor.

Ideally, measured data for the required variables are provided by sensors at the inlet and outlet of the component. When a required sensor is missing, one can try to solve for the underlying process variables by using reading values of other physical sensors and the conservation laws available at system level. Balance equations among components at system level may allow one to solve for certain process variables on the boundary of each component. This is the main idea for the concept of virtual sensor. We will refer to this concept of virtual sensor, obtained from solving balance equations to be used in both model construction and subsequently in fault diagnosis, as *type I virtual sensor*.

> *Definition 6-1.* A *type I virtual sensor* is the analytical solution of a process variable that is required for the construction of certain component models but there is no available physical sensor for that purpose.

Additionally, we have demonstrated in Section 3.3 for the example of a single-phase heat exchanger that even with a full set of sensors, one may generate additional residuals by using different combinations of sensors. In that case, for each of the additional residuals, one sensor was left out and the underlying process variable was computed from the other sensors. The additional residuals helped differentiate sensor faults from component faults. As each residual could potentially serve as a fault symptom, maximizing the number of residuals would help improving the resolution of diagnostic results. This is the motivation for a second type of *virtual sensors,* henceforth referred to as *type II virtual sensors*. Even if a physical sensor is already available, it may be helpful to create a type II virtual sensor for the underlying variable to be used as an alternative for the purpose of maximizing the number of model residuals. Type II virtual sensors are utilized for residual generation and not needed during model construction.

> *Definition 6-2.* A *type II virtual sensor* is the analytical solution of a process variable that can be used as the alternative of a physical sensor or type I virtual sensor in the process of generating model residuals.

6.1.1 Type I Virtual Sensors

The process variables of interest in a T-H system can be characterized to flowrate, pressure and enthalpy. Under the quasi-static condition, the conservation laws for a control volume, given by Eqns. (3.1) – (3.3) in Chapter 3, can be written as simple balance equations providing constraints between T-H process variables at various locations in a system. When the system is fault-free, there is no loss of mass and external heat loss can be considered negligible. In that case, collections of system level mass and heat balance equations may be used to solve for *virtual sensors* of flowrate and enthalpy type. On the other hand, since pressure loss is a crucial part of the momentum equation, it is generally not possible to create virtual sensors for pressure.

Type I virtual sensors will be constructed solely from solving system balance equations. If a virtual sensor is solvable, its value is valid if and only if all the involved balance equations are valid, i.e. the involved sensors and components must be free of the related faults. Such conditions can be assumed in the model construction process during which the system can be considered fault-free. Type I virtual sensors are constructed firstly for model calibration purposes. In the subsequent use of such virtual sensor for residual generation, one must keep track of the relevant faults that can invalidate the virtual sensor.

More specifically, the mass balance equation for a block of a system with an arbitrary number of inlet and outlet points is given by:

$$\sum w_{\mathrm{in}} = \sum w_{\mathrm{out}} \tag{6.1}$$

where w denotes a flowrate variable and the subscripts indicate either inlet or outlet locations. The validity conditions for this equation include no fault among the flowrate sensors and no leakage in any component in the block.

Similarly, the heat balance equation for a system block with no heat-exchanging component is given by:

$$\sum w_{in} \cdot h_{in} = \sum w_{out} \cdot h_{out} \tag{6.2}$$

where the h 's denote enthalpy values. For heat exchangers, the heat balance equation is given by:

$$\left[w\cdot\left(h_{out}-h_{in}\right)\right]_{\text{cold side}}=\left[w\cdot\left(h_{in}-h_{out}\right)\right]_{\text{hot side}} \tag{6.3}$$

The validity conditions for heat balance equations in general include no sensor faults, no leakage for the involved component and no faults that can cause significant external heat loss, if applicable.

Given the structure of a system, mass and heat balance equations can be generated and collected, each involves certain virtual sensors. One can then put the equations into groups of common unknowns and search for solvable virtual sensors. If a virtual sensor can be solved from a certain group of balance equations, its computed value is valid only if all equations in the group are valid. One can keep track of the relevant faults that can invalidate the virtual sensor from the involved components and sensors in each equation.

6.1.2 Type II Virtual Sensors

In the previous section, type I virtual sensors are obtained from solving system balance equations. Type I virtual sensors are created to provide the required data for the model construction process as opposed to type II virtual sensors which are only created and used in the subsequent residual generation process after all available models have been constructed.

Having various component models constructed using physical sensors and type I virtual sensors, one can then utilize both model predictions and system balance equations to create additional virtual sensors of type II. Type II virtual sensors are created for the purpose of increasing the number of independent model residuals one can generated for the system in order to improve the diagnostic resolution.

The availability of type II virtual sensors is system- and situation-specific. We recall from Section 3.3 that for a single-phase heat exchanger, when a full set of six physical sensors are available, the six sensors are constrained by the heat balance equation:

$$w^h\left[h(T^h_{\text{in}})-h(T^h_{\text{out}})\right]=w^c\left[h(T^c_{\text{out}})-h(T^c_{\text{in}})\right] \tag{6.4}$$

where enthalpy values are evaluated as functions of temperature readings. Under the assumption that this equation holds, one can *predict* the value of each sensor from the other five sensors. For example, the prediction for inlet temperature on the cold side is $T^c_{\mathrm{in,p}}$ such that $h(T^c_{\mathrm{in,p}}) = h^c_{\mathrm{in,p}}$ with:

$$h^c_{\mathrm{in,p}} = h(T^c_{\mathrm{out}}) - \frac{w^h}{w^c}\left[h(T^h_{\mathrm{in}}) - h(T^h_{\mathrm{out}}) \right] \tag{6.5}$$

In this case, $T^c_{\mathrm{in,p}}$ is a type II virtual sensor in additional to the available physical sensor T^c_{in}. To generate residuals from the heat transfer model, one has the option to use either the physical sensor or its alternative - the type II virtual sensor.

As with type I virtual sensors, when using a type II virtual sensor for residual generation in fault diagnosis, one must keep track of the relevant faults that can invalidate its value. In this case, it is clear that $T^c_{\mathrm{in,p}}$ is only valid if there is no fault among the five sensors being used and the balance equation holds. Thus, the relevant faults include five sensor faults, leakage in the heat exchanger and any other faults that can violate the balance equation.

In principle, type II virtual sensors can be created from any model predictions. However, in the presence of uncertainty, both measurement and modeling uncertainty can combine in the model predictions use for type II virtual sensors. When such type II virtual sensors are used to compute a residual, uncertainty from multiple models can combine resulting in high uncertainty for the residual. To utilize such residuals in fault diagnosis, the residual evaluation tool and subsequently the reasoning framework must have the capability to tolerate such scenarios. In this thesis, the probabilistic reasoning framework developed to deal with cases when measurement and modeling uncertainty are significant, thus can be expected to fulfil that role.

6.2 Aggregate Models

After all type I virtual sensors have been identified for a system, one can proceed to construct component models from the combined set of physical sensors and virtual sensors. It is often the case that even with the addition of type I virtual sensors, it is not possible to construct models for every separate component. In that case, one must resort to models of multiple nearby components, which we refer to as *aggregate models*.

Recall from the discussion previously that type I virtual sensors are limited to flowrate and temperature/enthalpy variables only. It is not possible to obtain virtual pressure sensors from balance equations. In addition to the condition of no sensor faults, virtual flowrate sensors obtained from solving balance equations are only valid if there is no *leakage* in the involved components. Similar conditions apply for virtual temperature or enthalpy sensors.

Aggregate Mass Models

Mass balance models are created for the purpose of detecting leakage and therefore should only involve physical flowrate sensors. As a generalization of the discussion in Section 3.1, an aggregate mass balance model can be constructed for any block of components if a physical flowrate sensor is available at every inlet and outlet point. The aggregate mass model is expressed by:

$$\sum w_{\text{in}} = \sum w_{\text{out}} \tag{6.6}$$

where each inlet w_{in} or outlet w_{out} flowrate must be available by a physical sensor.

Aggregate Momentum Models

In Section 3.1.2, we have developed a momentum model for a component with a single inlet and single outlet. For the generalization to aggregate momentum models, we limit our consideration to a block of *isolated* components between single inlet and outlet points. The term '*isolated'* is to emphasize that there is no external mass exchange in between the inlet and outlet. More specifically, for a part of the system under single-phase flow between a single inlet and outlet, the overall pressure loss can be expressed as a parametric model of the flowrate as:

$$P_{\text{in}} - P_{\text{out}} = f(w) \tag{6.7}$$

where P_{in} and P_{out} are the inlet and outlet pressure readings and w is a flowrate reading at either the inlet and outlet; $f(w)$ is the parametric form of the model. Generally, we can use the quadratic form as provided by Eqn. (3.11). As mentioned, there is no virtual sensors for pressure variables thus both P_{in} and P_{out} must be provided by physical sensors. On the other hand, the flowrate reading may be obtained from a type I virtual sensors. Given the structure of a system, one can search for all available pressure models satisfying these conditions.

Aggregate Energy Models

Unlike mass end momentum models, the energy process is only relevant to specific component types, e.g. pumps or heat exchangers. Thus, energy models are only constructed for separate components of relevant types. The sensor requirement is model-specific. In general, flowrate and enthalpy variables at the inlet and outlet of each component are of interest which must be provided by either available physical sensors or type I virtual sensors.

For the example of the single-phase heat exchanger described in Section 3.1.3, the full set of six sensors for the construction of the heat transfer model include the two flowrates, inlet and outlet temperatures on each side. If only five of those six are available, the missing sensor can be replaced by the virtual sensor computed using the heat balance equation.

6.3 Fault Diagnosis at the System Level

Following the framework developed in this thesis, the overall process to perform fault diagnosis at system level for arbitrary T-H systems is summarized by the flowchart in Figure 6-1.

A model library is developed with specifications on possible models for each generic T-H component type. The model specifications include information on sensor requirement and possible generalization to include multiple components in aggregate models.

After the P&ID specifying the structure of the system is imported, all virtual sensors that could be helpful for model construction is spawned by comparing the available sensor set against the sensor requirement provided by the model library. Balance equations throughout the system are collected to determine solvable virtual sensors. We referred to these solvable variables as type I virtual sensors. Unsolvable virtual sensors are then discarded.

The next step is to search for available models, including component-specific models and aggregate models as allowed by the available physical sensors and active type I virtual sensors. From the available models, we can then define additional type II virtual sensors for residual generation.

For each virtual sensor, including both types, information on the involved sensors and component faults is stored. These are the faults that can invalidate the computed value of the virtual sensor. From the available models, different model residuals can be defined using both real sensors and virtual sensors. The structure of each residual, i.e. the sensors and components involved in the underlying ARR, is identified and stored.

The next step is to determine the unknown parameters in each model in a model calibration process using a set of training data. Afterwards, model uncertainty and the uncertainty in each model residual are quantified. Information on the distribution of each model residual, i.e. its mean and variance, will be used by the statistical change detection tool selected for the residual evaluation step.

For each new time step, live sensor readings are collected and imported. System balance equations are then solved to update the value of available virtual sensors. The value of each model residual is then updated for the new time step. Afterwards, each residual is evaluated by a change detection tool to determine if the residual is statistically zero or non-zero.

After the residual evaluation step, the list of zero and non-zero residual will be used as input for the reasoning process for fault diagnosis. To perform diagnosis, we have the option to use either one of the deterministic reasoning frameworks as discussed in Section 4.2 or the probabilistic reasoning framework developed in Chapter 5. Diagnostic result is obtained from the reasoning engine and updated after each time step.

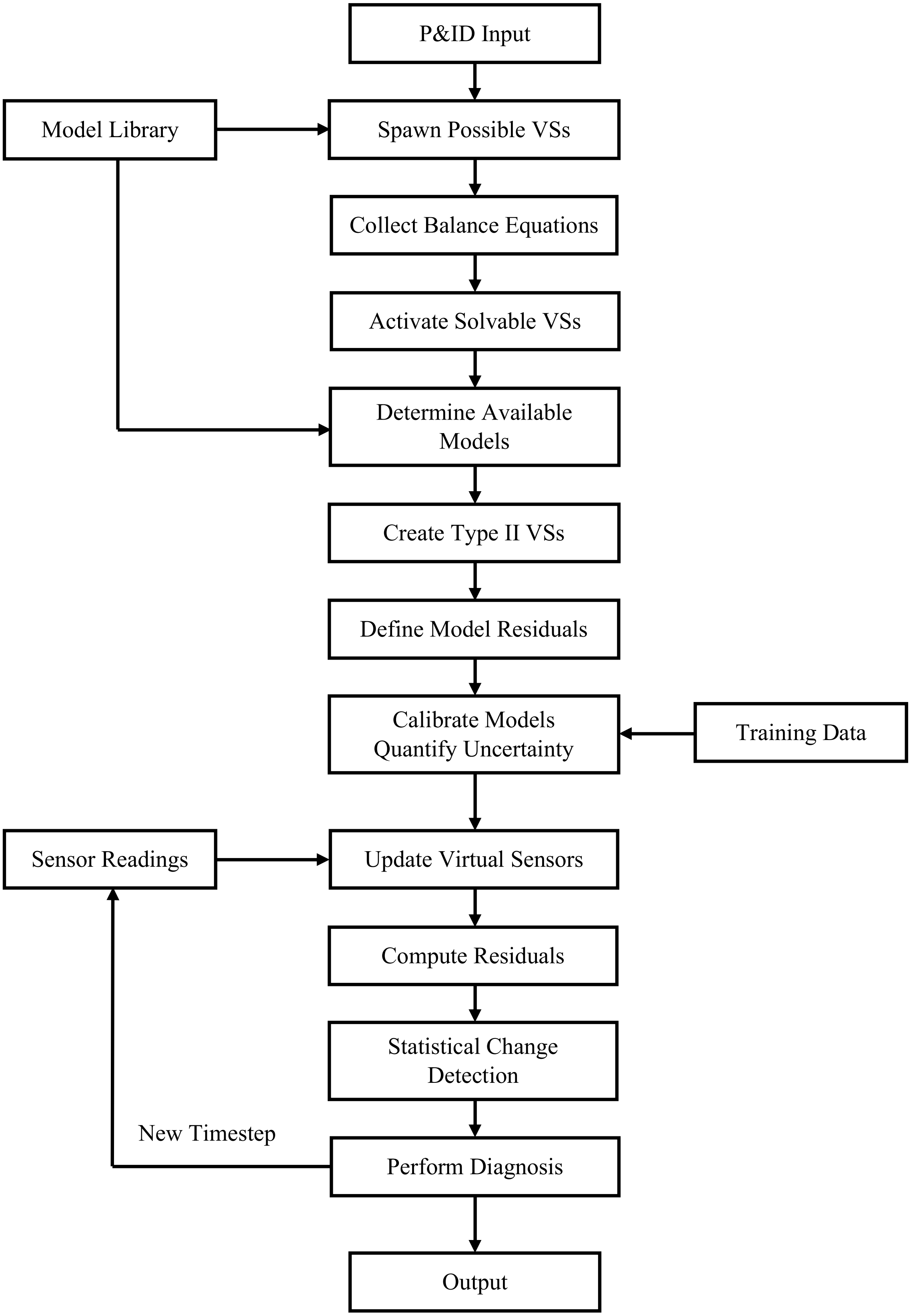

Figure 6-1. Overall scheme of the proposed system level diagnostic framework

Recall that to perform the reasoning process in quantitative model-based diagnosis, one needs to know the dependency structure of each residual. For the deterministic reasoning frameworks, the list of relevant faults that can cause a residual to become non-zero is needed to construct a *conflict* relation from each non-zero residual. For the probabilistic framework, the structure of each residual is needed to construct a Bayesian network for the system.

To clarify the reasoning process in system level diagnosis, it should be emphasized that each residual, even if computed from a component-specific model, may involve other components and sensors at system level. More specifically, the calculation of each model residual may involve one or more virtual sensors whose validity depends on other system components and sensors. The dependency structure of a model residual in general is given by the flowchart in Figure 6-2. A fault can cause a non-zero model residual in two ways: either by directly affecting the model used for residual generation or by affecting the balance equations or model predictions used in the calculation of the involved virtual sensors.

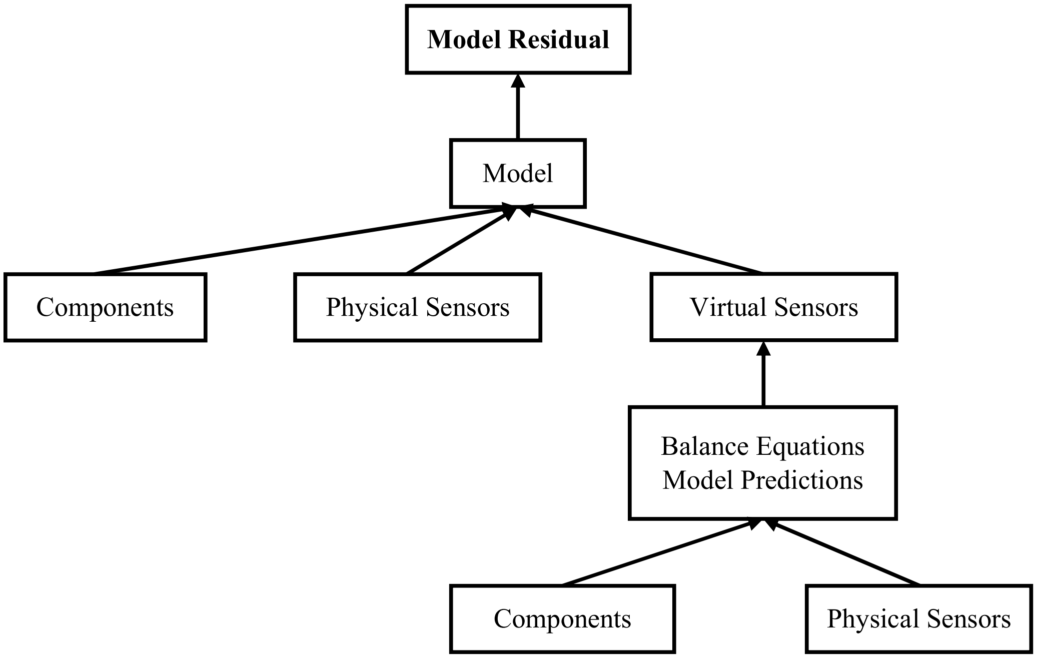

Figure 6-2. Dependency structure of a model residual in system level diagnosis

In general, the list the relevant faults to each residual may include not just faults from the components and sensors directly involved in the underlying model but also the faults that can invalidate the involved virtual sensors. Such list of relevant faults will be generated and stored when each model residual is defined.

Chapter 7

Results – High Pressure Feedwater System

7.1 System Description

In this chapter, to demonstrate the proposed diagnostic framework, we will consider various diagnostic scenarios in a high-pressure feedwater system of a typical PWR plant. The high-pressure feedwater system is part of the condensate and feedwater system that is responsible for the supply of pre-heated feedwater to the steam generators. Exhausts from the turbines turn into condensate and get heated up by feedwater heaters in multiple heating stages before re-entering the steam generators. The high-pressure feedwater system we are considering here consists of the two heating stages closest to the inlets of the steam generators, referred to as the first-point and second-point stages. Detailed description of such system can be found, for example, in the final safety analysis report for Unit 1 and 2 of the North Anna Power Station [96]. The structure of the system in consideration is illustrated by the P&ID in Figure 7-1.

The feedwater heater in each stage is of the closed two-shell type and thus, each heating stage effectively consists of two feedwater heaters in parallel piping lines [96]. Therefore, for this example we have four feedwater heaters: two first-point heaters, labeled by 1-FW-E-1A and 1-FW-E-1B, and two second-point heaters, labeled by 1-FW-E-2A and 1-FW-E-2B, as shown by the P&ID in Figure 7-1. The system as shown also include three steam generator feed pumps, labeled by 1-FW-P-1A to -P-1C, and three drain pumps, labeled by 1-SD-P-1A to -P-1C.

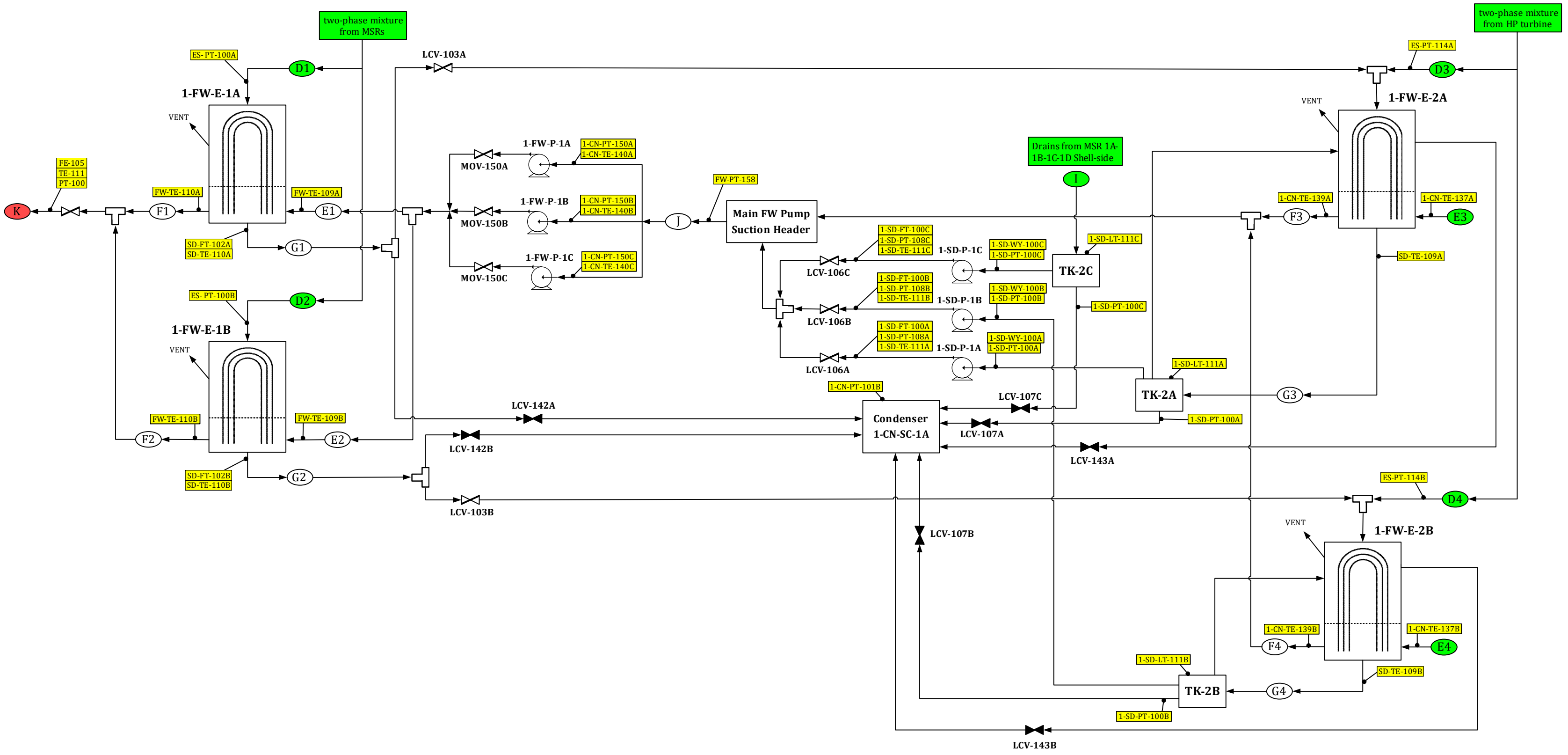

Figure 7-1. The P&ID of a high-pressure feedwater system

From the top of the P&ID, two-phase mixtures from the high-pressure turbine and moisture separator reheaters flow into the shells of the first and second-point heaters. Drains from the first-point heaters flow into the shells of the second-point heaters. On the right of the P&ID, feedwater from later heating stages flows through the two second-point heaters to the suction header of the feed pumps. Drains from the two-point heater shells are collected by the high-pressure heater drain receivers, TK-2A and TK-2B. Drains from the moisture separators (not shown in Figure 7-1) are collected by a third drain receiver, TK-2C. The three drain pumps pump condensate from the drain receivers to the suction header of the feed pumps. During normal operation, only two of the three feed pumps operate to pump feedwater through the two first-point heaters to a discharge header to supply the steam generators. In emergency situations, excessive drains from the first-point heaters and the drain receivers are collected by a condenser. All the valves to the condenser are otherwise closed off during normal operation.

The yellow tags in Figure 7-1 indicate the sensors typically available for such system. Each label containing PT denotes a pressure sensor, FE denotes a flowrate sensor and TE denotes a temperature sensor. For brevity, we will use short-handed labels when referring sensors. For example, E2.T, with E2 being the location label and T the variable type, refers to the temperature sensor at the inlet of FWH 1B as shown in Figure 7-1. A Dymola simulation model for this system has been developed at Argonne National Laboratory [97]. Simulation data from the model will be used for the analysis in this chapter.

For this demonstration, we will exclude the condenser from the P&ID as all its incoming piping lines are normally closed. The components in the system can be characterize by known generic types. Each drain receiver will be treated as coolant tank. Model development for each generic component type is discussed in the remainder of this section.

Vertical Feedwater Heaters

The feedwater heaters in this example are of the vertical channel down shell-tube design. The steam and water flow paths in a typical heater of this design are illustrated by the diagram in Figure 7-2 [98]. Steam flows in on the shell side, exchanges heat with the feedwater on the tube side and turns into condensate collected in the drain pool. In general, if the inlet steam is superheated, the shell side of the heater consists of three sections: a de-superheating section, a condensing section and a drain cooling section. The feedwater on the tube side remains in the subcooled regime throughout the heating process.

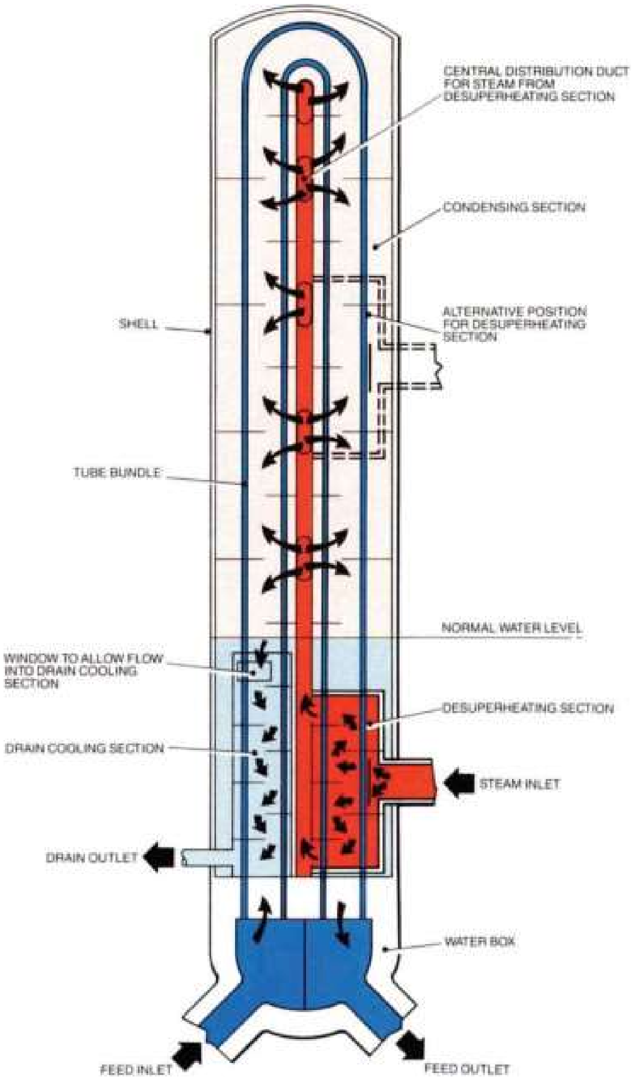
CENTRAL DISTRIBUTION DUCT FOR STEAM FROM DESUPERHEATING SECTION
CONDENSING SECTION
SHELL
ALTERNATIVE POSITION FOR DESUPERHEATING SECTION
TUBE BUNDLE
NORMAL WATER LEVEL
WINDOW TO ALLOW FLOW INTO DRAIN COOLING SECTION
DESUPERHEATING SECTION
DRAIN COOLING SECTION
STEAM INLET
DRAIN OUTLET
WATER BOX
FEED INLET
FEED OUTLET

Figure 7-2. Steam and water flow paths in a typical vertical high-pressure feedwater heater [98]

For the system in consideration, we have two-phase mixtures, i.e. wet steam, coming into the shell side of the feedwater heaters. Thus, the shell of each heater consists of only two sections, condensing and drain cooling, without the de-superheating section. For normal operating conditions, the water level of the drain cooling section is maintained relatively constant by a level control system.

For the diagnostic problem, we are interested in using sensor readings at the inlet and outlet of the heater to monitor its performance. As with the example of the single-phase counterflow heat exchanger analyzed in Section 3.1.3, the performance-related criteria include the heat balance and overall heat transfer capability.

Recall that with the single-phase heat exchanger, to establish the heat balance equation, we need *six* sensors: mass flowrate, inlet temperature and outlet temperature for each side. Temperature readings are used to compute the corresponding enthalpy values. From these six sensors, the heat transfer performance of the single-phase heat exchanger can be evaluated in terms of the overall heat transfer coefficient UA. If only *five* out of those six sensors are available, one can make the assumption that the heat balance equation, Eqn. 3.14, is valid and from that compute the missing sensor. Thus, a set of five sensors is the minimum requirement for one to evaluate UA and from that construct a performance model for the single-phase heat exchanger.

To apply the same procedure to the feedwater heater, note that we have wet steam at the inlet of the shell side whose enthalpy depends strongly on two variables: pressure and steam quality. Thus, to check for the heat balance, a set of *seven* sensors is required: feedwater flow rate, feedwater inlet and outlet temperatures, steam inlet pressure, steam inlet quality, drain flowrate and drain temperature. Locations of these sensors on the steam and water flow paths are shown by the diagram in Figure 7-3. T_3 and T_4 are the intermediate feedwater temperatures leaving and re-entering the drain cooling section.

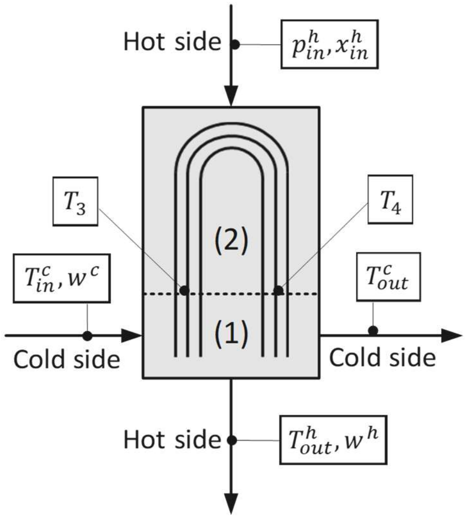

Figure 7-3. Simplified schematic of the vertical HP feedwater heater.

T_3 and T_4 can be computed from the heat balance equations. The other variable labels as shown in Figure 7-3 denote the full set of seven sensors: each label T denotes a temperature sensor, w denotes mass flowrate, p denotes pressure and x denotes steam quality.

Steam quality sensors are generally not available and in that case one must assume the heat balance in order to compute the steam inlet enthalpy. Thus, by a similar analysis as with the single-phase heat exchanger, the minimum sensor requirement to evaluate the performance of the feedwater heater is a set of six sensors, as shown in Figure 7-3 excluding steam quality.

The dashed line in Figure 7-3 representing the water level separating the condensing zone and the drain cooling zone. The condensing zone is responsible for most of the heat transfer rate in the heater. Wet steam from the shell side inlet can be considered to completely condense to saturated liquid as it reaches the boundary of the drain cooling zone.

The heat transfer rate between the outgoing feed water and the drain pool can be considered negligible and thus $T_4 \approx T_{\text{out}}^c$. The intermediate temperature T_3 can be computed from the heat balance equation for the condensing zone:

$$w^c \left[h(T_{\text{out}}^c) - h(T_3) \right] = w^h \left[h(p_{\text{in}}^h, x_{\text{in}}^h) - h_{\text{sat}}(p_{\text{in}}^h) \right] \tag{7.1}$$

where h_{sat} denoted the water saturation enthalpy at the given pressure.

The overall heat transfer coefficient of the condensing zone can be evaluated from a LMTD model:

$$Q_c = UA_c \times \frac{\left(T_{\text{sat}} - T_3\right) - \left(T_{\text{sat}} - T_{\text{out}}^c\right)}{\ln \dfrac{T_{\text{sat}} - T_3}{T_{\text{sat}} - T_{\text{out}}^c}} \tag{7.2}$$

where $T_{\text{sat}} = T_{\text{sat}}(p_{\text{in}}^h)$ is the saturation temperature at the given pressure p_{in}^h. The total heat transfer rate of the condensing zone, Q_c, is given by either side of the heat balance equation in Eqn. (7.1). It should be emphasized that the flow path of the feedwater in the condensing zone is a two-pass U-tube. To evaluate the overall heat transfer coefficient in practice, one usually needs to include a correction factor F_t into the LMTD model to account for the two-pass geometry [85]. The correction factor F_t is typically a function of the temperature profile along the flow path. For diagnostic purpose, here we are effectively lumping F_t into the overall coefficient UA_c and assume the dependence on small variation of temperature is negligible.

As with the single-phase heat exchanger, we will now construct a parametric model for UA_c as a function of the two flow rates, assuming the water level of the drain cooling section can be considered constant. The functional form of the parametric model depends on the flow and heat transfer condition on each side. For the tube side, the feedwater remains subcooled and thus the dependence of UA_c on the tube side flow rate w^c can be taken to be the same as in the single-phase heat exchanger, given by Eqn. (3.21). For the shell side, the heat transfer process can be described as vertical film condensation. For turbulent flow outside vertical tubes, the recommended correlation for the heat transfer coefficient is [99]:

$$h_{\text{cond}} = 0.0076\left(\frac{\rho_f^2 g k_f^3}{\mu_f^2}\right)^{\frac{1}{3}} \text{Re}_b^{0.4} \tag{7.3}$$

where ρ_f, k_f and μ_f are respectively the density, heat conductivity and viscosity of the condensate, evaluated at the tube wall temperature; Re_b is the Reynold's number of the condensate flow.

Following the same procedure as with the single-phase heat exchanger, we have the following parametric model for the overall heat transfer coefficient in the condensing zone:

$$\frac{1}{UA_c} = \theta_{10} + \theta_{11}\left(w^c\right)^{-0.8} + \theta_{12}\left(w^h\right)^{-0.4} \tag{7.4}$$

where θ_{10}, θ_{11} and θ_{12} are the three model parameters to be determined using training data.

For the drain cooling zone, most of the heat transfer takes place between the incoming leg of the feedwater and the drain pool. Similarly, the overall heat transfer coefficient of the drain cooling zone is given by a LMTD model:

$$Q_d = UA_d \times \frac{\left(T_{\text{out}}^h - T_{\text{in}}^c\right) - \left(T_{\text{sat}} - T_3\right)}{\ln \frac{T_{\text{out}}^h - T_{\text{in}}^c}{T_{\text{sat}} - T_3}} \tag{7.5}$$

$$Q_d = w^c\left[h(T_3) - h(T_{\text{in}}^c)\right] = w^h\left[h_{\text{sat}}(p_{\text{in}}^h) - h(T_{\text{out}}^h)\right] \tag{7.6}$$

The parametric model for the overall heat transfer coefficient of the drain cooling zone can be taken to be that of a single-phase exchanger by:

$$\frac{1}{UA_d} = \theta_{20} + \theta_{21}\left(w^c\right)^{-0.8} + \theta_{22}\left(w^h\right)^{-0.8} \tag{7.7}$$

with three model parameters θ_{20}, θ_{21} and θ_{22}.

Eqn. (7.4) and (7.7) provide two parametric models for the heat transfer coefficients that can be used to monitor the performance of the feedwater heater. To calibrate these models, one need to sensor data to evaluate the heat transfer coefficients from Eqn. (7.2) and (7.5). For that, a minimum of six sensors is required: w^c, w_h, T^c_{in}, T^c_{out}, p^h_{in} and T^h_{out}. Provided enough sensors, these models can be calibrated and used to generate residuals for model-based diagnosis.

Other Components

The other major components of the system as shown in Figure 7-1 include the feed pumps and drain pumps. For each pump, the two performance-related parameters of interest are the pressure head, $P_{\text{out}} - P_{\text{in}}$, and consumed power, W. As discussed in Chapter 3, for a general variable-speed pump, both the pump head and power can be described by parametric models of the rotational speed n and flow rate w:

$$P_{\text{out}} - P_{\text{in}} = f(n, w) \tag{7.8}$$

$$W = g(n, w) \tag{7.9}$$

The functional forms $f(n, w)$ and $g(n, w)$ can be obtained from the pump specifications or analytically formulated. For constant speed pumps, both the head and power reduce to function of only the total flow rate w. Unless a specific performance curve is provided for $f(w)$, one can take the quadratic parametric form as described in Section 3.1.2. To calibrate the head model for a constant-speed pump, the required sensors are flowrate w, inlet pressure P_{in} and outlet pressure P_{out}. To calibrate the power model, a sensor reading for the consumed power is required. For variable-speed pumps, an additional sensor for the pump speed is required.

For the rest of the system, the drain receivers, without being fully instrumented, will be treated as a mass source/sink. The valves, including motor-operated valves and pressure-operated valves, will be treated as generic pressure loss components as no reading on the opening of each valve is available.

7.2 Virtual Sensors and Balance Equations

By comparing the available sensors at the inlet and outlet of each component to the sensor requirement for the construction of its models, it is straightforward to check for the missing sensors and from that create all possible virtual sensors. One must then collect all the available system balance equations and from that determine which of the virtual sensors created previously are solvable. Only solvable virtual sensors are kept, which we referred to as type I virtual sensors, the rest are discarded. For the system in consideration, among the missing sensors, the most notable virtual sensors that are solvable and of interest to us are the feedwater flowrate for the two first-point feedwater heaters (FWHs).

More specifically, as described in Section 7.1, the full sensors set for each FWH consists of seven sensors, at least six of which are required for the construction of its heat transfer coefficient models. As shown in Figure 7-1, only five of those sensors are available for each of the first-point FWH. Thus, the available sensor set for each FWH is insufficient for the construction of the FWH model.

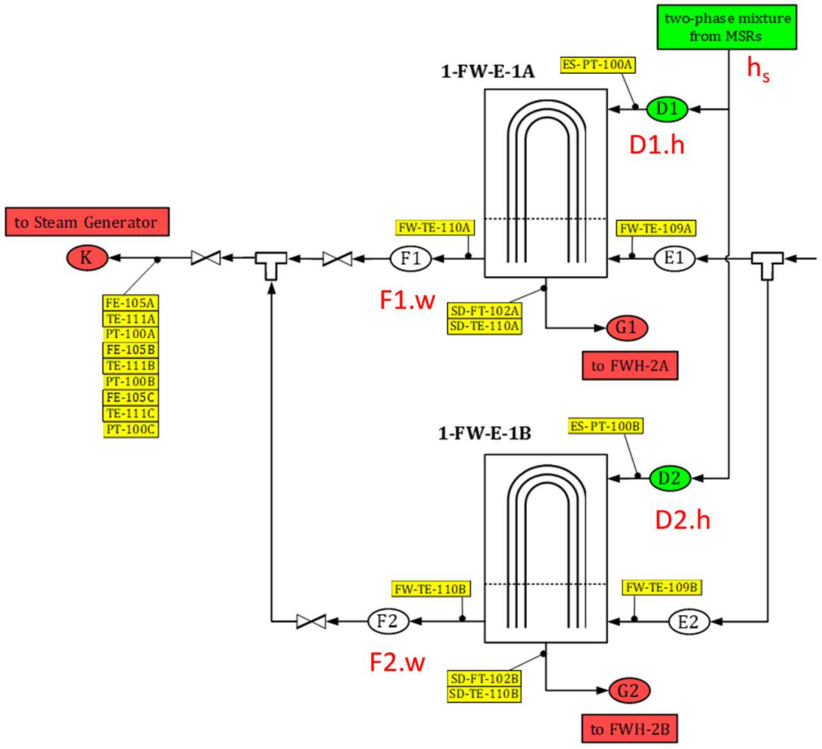

Figure 7-4. Type I virtual sensors for the two first-point FWHs

The two missing sensors are for each first-point FWH are: feedwater flowrate and steam inlet quality, or equivalently steam inlet enthalpy. The locations of the missing sensors are shown by the red labels, $F1.w$, $D1.w$, $F2.w$, and $D2.w$, in Figure 7-4. Additionally, notice that the two first-point FWHs share the same steam inlet condition whose enthalpy is denoted by h_s.

The balance equations involving these virtual sensors are:

$$\begin{aligned}
&1)\quad F1.w + F2.w - K.w = 0 \\
&2)\quad D1.h - h_s = 0 \\
&3)\quad D2.h - h_s = 0 \\
&4)\quad F1.w\left[F1.h - E1.h\right] = G1.w\left[D1.h - G1.h\right] \\
&5)\quad F2.w\left[F2.h - E2.h\right] = G2.w\left[D2.h - G2.h\right]
\end{aligned} \tag{7.10}$$

where the capitalized labels denote measurement locations for brevity. The first equation establishes the mass balance between the feedwater flowrate from the two first-point FWHs with the total flowrate to the steam generators. The second and third equations enforce the same steam inlet condition for the two FWHs. The fourth and fifth equations are the heat balance equations for the two FWHs. With five equations and five unknowns, the virtual sensors are solvable.

Other solvable virtual sensors include to total mass flowrate from the feed pump suction header, flowrate coming out of each drain receiver, total flowrate from the two second-point FWHs. For the second-point FWHs, each has three missing sensors, none of which is solvable. The missing flow rate for each feed pump is also unsolvable.

7.3 Model Construction and Residual Generation

The calibration and subsequent use of each component or aggregate model in diagnostics require measured data of a certain number of process variables. Such required measurements can be provided by either physical sensors or type I virtual sensors. After all solvable type I virtual sensors have been identified, the next step is to determine all possible component and aggregate models as allowed by the set of available physical sensors and type I virtual sensors.

For the current system, no mass balance model is available as each mass model requires a full set of physical flowrate sensors for all inlets and outlets.

The generic pressure model for each single inlet/outlet block of components require two pressure readings and one flowrate reading. One model of this type is available between the feed pump suction header and the discharge header. For each drain pump, the available sensor set allows a model for the pressure head. No model is possible for the individual feed pump. We will use a quadratic form for the parametric models of the generic pressure model and the pump head, as described in Eqn. (3.11) and (3.12).

Notice that the two first-point FWHs are on parallel piping lines. From the constraint of equal pressure loss on each line, one can construct a flow ratio model to monitor the flowrate ratio between the two line. There are no physical flow rate sensors available for these two piping lines, but the virtual flow rate sensors constructed for the two FWHs can be used for this purpose. If the system is fault-free, one can expect the flow ratio to remain constant.

Since no power meter is available for the pumps, it is not possible to construct pump power models. For the FWHs, the addition of the virtual flowrate sensor for the feedwater side allows us to construct a heat transfer model for each first-point FWH. No model is possible for the second-point FWHs as the available sensor set is insufficient.

A summary of the available models is provided in Table 7-1.

Table 7-1. Available diagnostic models for the high-pressure feedwater system.

ID	*Name*	*Model Type*	*Components*	*Relevant Fault Types*
1	DP-1	Generic pressure difference	Feed pumps 1A and 1B, valves, FWH 1A and 1B, pipes	Leakage, Blockage
2	FR-1	Flow ratio	FWHs 1A and 1B, pipes	Leakage, Blockage
3	SDP-1A	Pump head	Drain pump 1A	Pump fault
4	SDP-1B	Pump head	Drain pump 1B	Pump fault
5	SDP-1C	Pump head	Drain pump 1C	Pump fault
6	FWH-1A	HX performance	FWH 1A	Leakage, Fouling
7	FWH-1B	HX performance	FWH 1B	Leakage, Fouling

DP-1 is the generic pressure difference between the feed pump suction header and discharge header near the inlet of the steam generators. The pressure difference between those two points depends on the pressure gain provided by the pumps and the pressure loss along the piping lines. Recall that for normal operation, only two of the feed pumps are running.

From these models, one can generate model residuals for diagnostics. In Chapter 3, we have discussed possible residuals for each model type. For the model DP-1, we can generate a residual from the model prediction of the pressure difference:

$$r_1 = \Delta P_p(w) - (P_{\text{out}} - P_{\text{in}}) \tag{7.11}$$

where $\Delta P_p(w)$ is the model prediction of the pressure difference, as a function of the flowrate reading w; P_{in} and P_{out} are the inlet and outlet pressure measured at the feed pump suction header and discharge header, respectively. The calculation of this residual involves three sensors for w, P_{in}, and P_{out}.

From the model FR-1, we can generate a residual to monitor changes in the flow rate ratio. The fault-free ratio is computed from calibration data.

$$r_2 = \frac{F1.w}{F2.w} - \left(\frac{F1.w}{F2.w}\right)_{\text{fault free}} \qquad (7.12)$$

The calculation of r_2 involves two type I virtual sensors $F1.w$ and $F2.w$. Thus, in addition to the faults that can cause the actual flow ratio to deviate, any faults that can invalidate these two virtual sensors would also cause the residual to be non-zero.

For each pump head model, we can generate a residual for the pressure head prediction, similar to that of the pressure difference model:

$$r_3 = \Delta P_p(w) - (P_{\text{out}} - P_{\text{in}}) \qquad \text{(Drain pump 1A)} \qquad (7.13)$$

$$r_4 = \Delta P_p(w) - (P_{\text{out}} - P_{\text{in}}) \qquad \text{(Drain pump 1B)} \qquad (7.14)$$

$$r_5 = \Delta P_p(w) - (P_{\text{out}} - P_{\text{in}}) \qquad \text{(Drain pump 1C)} \qquad (7.15)$$

The calculation of each of these residuals require three sensors for the flow rate w, suction pressure P_{in} and discharge pressure P_{out}. By an abuse of notations, we have used the same variable labels for all three drain pumps. Each drain pump has a distinct set of sensors, as shown by the P&ID in Figure 7-1.

For each of the two FWH models, in this demonstration we will focus on the condensing zone only as it is responsible for most of the heat transfer rate in the FWH. From the model prediction for the overall heat transfer coefficient of the condensing zone, we can generate one residual from each model:

$$r_6 = \frac{1}{UA_{c,p}(w^c, w^h)} - \frac{1}{UA_c(w^c, T_{\text{in}}^c, T_{\text{out}}^c, w^h, p_{\text{in}}^h, T_{\text{out}}^h)} \quad \text{(for FWH 1A)} \tag{7.16}$$

$$r_7 = \frac{1}{UA_{c,p}(w^c, w^h)} - \frac{1}{UA_c(w^c, T_{\text{in}}^c, T_{\text{out}}^c, w^h, p_{\text{in}}^h, T_{\text{out}}^h)} \quad \text{(for FWH 1B)} \tag{7.17}$$

where the notations for the T-H variables are as described for Eqns. (7.1) and (7.2). The label for each variable is based on the variable type and its location relatively to the attached FWH. It should be noted that although the variables in Eqns. (7.16) and (7.17) have the same labels, they are provided by a different set of sensors for each FWH. Here $UA_{c,p}(w^c, w^h)$ is the model prediction for the overall heat transfer coefficient of the condensing zone; UA_c is the value computed directly from its definition. In this demonstration, we will use the parametric model as expressed by Eqn. (7.4) for $UA_{c,p}(w^c, w^h)$. Considerations of more sophisticated models will be part of the future work. The calculation of each of these residuals requires six readings: feedwater flowrate w^c, feedwater inlet temperature T_{in}^c, outlet temperature T_{out}^c, drain flowrate w^h, steam inlet pressure p_{in}^h, and drain temperature T_{out}^h. Five of these six variables can be obtained from physical sensors while the flowrate w^c will be provided by a *virtual sensor*. Recall that the feedwater flowrate virtual sensors for FWHs 1A and 1B, $F1.w$ and $F2.w$, can be obtained from solving the system of five equations in (7.10). These are type I virtual sensors.

For the purpose of residual generation, we can also use the prediction of the flow ratio model FR-1 to compute these feedwater flowrates thus obtain two *type II virtual sensors* that can be used as alternative:

$$\begin{gathered} \frac{F1.w}{F2.w} = \left(\frac{F1.w}{F2.w}\right)_{\text{fault free}} \\ F1.w + F2.w = K.w \end{gathered} \tag{7.18}$$

where the fault-free ratio is the prediction of the flow ratio model. The solution of these two equations give us the two type II virtual sensors that can be used as alternative to the two type I virtual sensors obtained from solving (7.10). When the system is fault-free, the solutions of (7.10) and (7.18) agree, i.e. the value of each type II virtual sensor is equivalent to its corresponding type I virtual sensor.

However, notice that the validity conditions of the balance equations in (7.18) are not the same as those for the equations in (7.10). It follows that the validity conditions of the type II virtual sensors are different from those of the type I virtual sensors. A sensor fault for the inlet temperature of FWH 1B would violate the fifth equation in (7.10) thus invalidate the solutions for the two type I virtual sensors but it would not affect the equations in (7.18).

Using the type II virtual sensors for the feedwater flowrate, denoted by w_{II}^{c}, we can generate two additional residuals for FWHs 1A and 1B:

$$r_8 = \frac{1}{UA_{c,p}(w_{II}^{c}, w^{h})} - \frac{1}{UA_{c}(w_{II}^{c}, T_{\text{in}}^{c}, T_{\text{out}}^{c}, w^{h}, p_{\text{in}}^{h}, T_{\text{out}}^{h})} \quad \text{(for FWH 1A)} \tag{7.19}$$

$$r_9 = \frac{1}{UA_{c,p}(w_{II}^{c}, w^{h})} - \frac{1}{UA_{c}(w_{II}^{c}, T_{\text{in}}^{c}, T_{\text{out}}^{c}, w^{h}, p_{\text{in}}^{h}, T_{\text{out}}^{h})} \quad \text{(for FWH 1B)} \tag{7.20}$$

Note the distinction between the residual r_6 in Eqn. (7.16) where a type I virtual sensor is used for the flowrate w^{c} and r_8 where a type II virtual sensor is used instead. Recall the dependency structure of each residual as shown by the flowchart in Figure 6-2. For this example, these two residuals, r_6 and r_8, have different dependency structures although they were computed from the same model. Similarly, the same applies for r_7 and r_9.

Table 7-2. Model residuals for the high-pressure feedwater system.

Residual	*Model*	*Sensors*	*Formula*
r_1	DP-1	Inlet, outlet pressure, flowrate	(7.11)
r_2	FR-1	Two virtual flowrates	(7.12)
r_3	SDP-1A	Inlet, outlet pressure, flowrate	(7.13)
r_4	SDP-1B	Inlet, outlet pressure, flowrate	(7.14)
r_5	SDP-1C	Inlet, outlet pressure, flowrate	(7.15)
r_6	FWH-1A	Five physical sensors, 1 type I VS	(7.16)
r_7	FWH-1B	Five physical sensors, 1 type I VS	(7.17)
r_8	FWH-1A	Five physical sensors, 1 type II VS	(7.19)
r_9	FWH-1B	Five physical sensors, 1 type II VS	(7.20)

7.4 Diagnostic Results

To demonstrate the diagnostic process, we will use simulation data obtained from a Dymola model developed at Argonne National Laboratory [97]. A start-up procedure was simulated, and the simulation data was used for the calibration of the seven models listed in Table 7-1. Afterwards, various scenarios of faults were simulated and investigated. In this section, we will consider a case of fouling in one of the FWHs and a case of sensor faults.

7.4.1 Fault Scenarios

To clarify the diagnostic process, let us first consider the fault scenarios without measurement uncertainty.

Fouling in FWH 1A

For a case of fouling in FWH 1A, the residuals were computed and plotted in Figure 7-5.

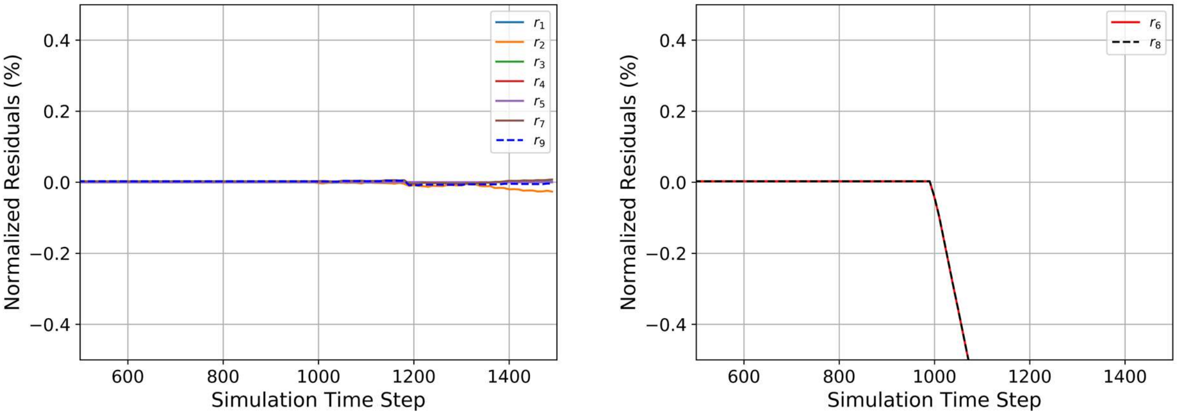

Figure 7-5. Fouling in FWH 1A causes two non-zero residuals (right plot) while the other seven remain unaffected (left plot).

The fouling fault, started at time step 1000, causes two non-zero residuals, r_6 and r_8, which were computed from the heat transfer model for FWH 1A. Notice that when the system is fault-free, all nine residuals are all approximately zero, indicating that the modeling uncertainty in each model is negligible. This is because the simulation was performed in Dymola which, for its 1-D T-H component models, uses similar pressure drop and heat transfer correlations to the ones that were based on for the development of the physics-based parametric models in this thesis. The simulation results are therefore in good agreement with the diagnostic models.

Furthermore, on the right plot of Figure 7-5, the residual r_6 is shown to be nearly identical to r_8. This is expected since the fouling fault does not affect any balance equations. The solutions of (7.10) for the type I virtual sensors are equivalent to the solution of (7.18) for the type II virtual sensors and thus, the value of r_6 and r_8 are equivalent.

To show the effect of the fouling fault, the outlet feedwater temperature, F1.T, from FWH 1A is plotted on Figure 7-6 in comparison with the temperature, F2.T, from FWH 1B.

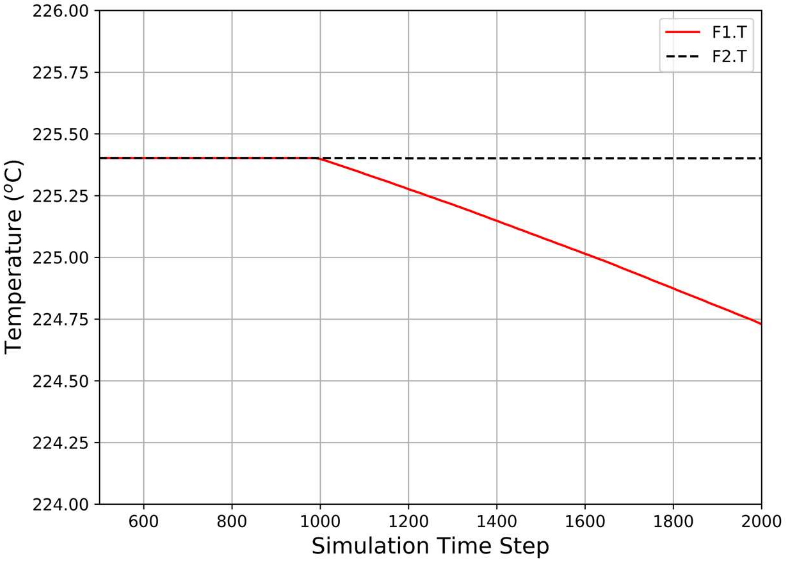

Figure 7-6. Effect of fouling on the feedwater outlet temperature of FWH 1A

To perform the reasoning process, we have the option to use either the *probabilistic reasoning framework* in Chapter 5 or one of the *deterministic reasoning approaches* described in Chapter 4. Without measurement and modeling uncertainty, there is no benefit in using probabilistic reasoning. In the deterministic reasoning approaches, we first need to detect non-zero residuals and then apply either the MBD reasoning framework in Table 2-1 (Deterministic I) or the FDI reasoning framework in Table 2-2 (Deterministic II).

The task of change detection in this case is trivial. The observed fault symptoms are the two non-zero residuals (r_6 and r_8). All other residuals are zero. With the structure of all residuals known, we can apply the FDI reasoning framework similarly to the process described in detail for the single-phase heat exchanger example. The valid *minimal diagnoses* are fouling in FWH 1A or sensor fault for the steam pressure of FWH 1A.

Table 7-3. Diagnostic result for the case of fouling in FWH 1A

Fault symptoms	*Diagnoses*
r_6 and r_8 non-zero	[Fouling in FWH 1A] *or* [Pressure sensor fault at FWH 1A steam inlet]

More specifically, recall the dependency of each residual listed in Table 7-2. From residual r_6 being non-zero, one can obtain a *conflict* implicating either fouling or leakage in FWH 1A, a sensor fault among the five involved physical sensors or invalidity of the type I virtual sensor for the flowrate F1.w. Similarly, r_6 being non-zero provides a *conflict* implicating either fouling or leakage in FWH 1A, a sensor fault among the five involved physical sensors or invalidity of the type II virtual sensor for F1.w.

Using the *notion of exoneration* with the zero residuals, one can then exonerate most of the faults from the two *conflicts* and is left with only two possibilities of either fouling in FWH 1A or a sensor fault in the steam inlet pressure of FWH 1A. Both these faults can directly cause r_6 and r_8 to be non-zero and in this case, we cannot differentiate between the two faults.

For this system, the number of possible faults is much higher than the number of fault symptoms. Under such condition, as discussed in Chapter 2, without the notion of exoneration, the diagnostic result using the MBD reasoning framework (Deterministic I) often consists of too many possibilities, thus is not suitable for practice uses. For most cases in practice when considering both component faults and sensor faults, we will use only either Deterministic II or probabilistic reasoning.

Sensor Fault at E2.T

The second fault scenario we will consider is a temperature sensor fault at the feedwater inlet of FWH 1B. The sensor fault is simulated by a bias, increasing over time, added to its reading value. Reading values from the faulted sensor is plotted in Figure 7-7. The residuals in this scenario were computed and plotted in Figure 7-8. The sensor fault causes four non-zero residuals: r_2, r_6, r_7 and r_9.

A fault in sensor E2.T directly affect the calculation of the heat transfer model in FWH 1B, causing both r_7 and r_9 to be non-zero. Additionally, the sensor fault violates the fifth balance equations in Eqn. (7.10). Thus, the two type I virtual sensors obtained by solving 7.10 are invalid which results in r_2, residual for the flow rate ratio, and r_6, residual for FWH 1A, as both are computed using at least one type I virtual sensor.

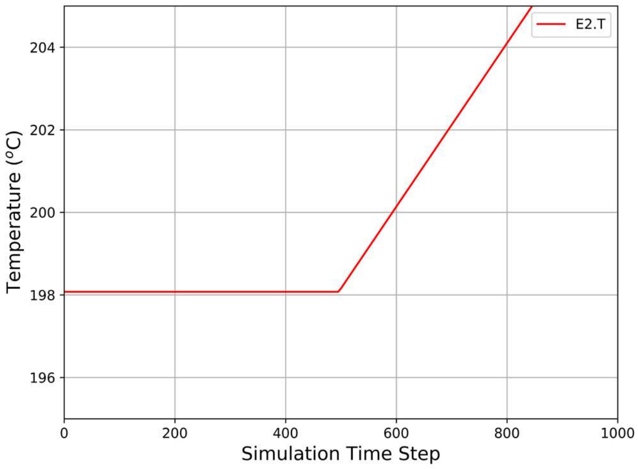

Figure 7-7. A simulated fault for the feedwater inlet temperature sensor of FWH 1B

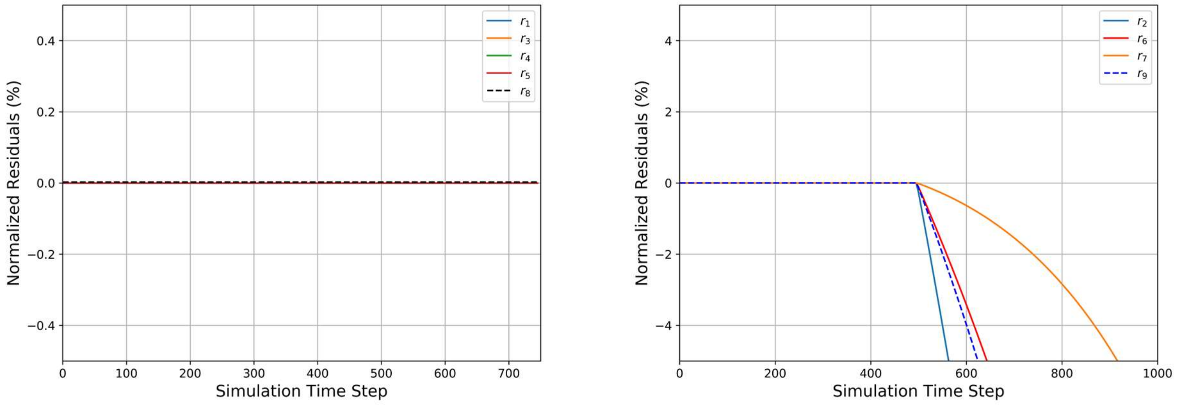

Figure 7-8. Sensor fault at E2.T causes four non-zero residuals (right plot).

Applying the FDI reasoning framework for the Deterministic II approach, the diagnostic result for this set of fault symptoms are listed in Table 7-4. In this case, the diagnostic result includes one component fault, leakage from the shell of FWH 1B, or one of the four sensor faults. The available sensor set does not allow us to differentiate between the true fault, which is sensor E2.T, from the other four faults.

Table 7-4. Diagnostic result for the case of sensor E2.T fault

Fault symptoms	*Diagnoses*
r_2, r_6, r_7 and r_9 non-zero	[Leakage from the FWH 1B shell] *or* [Sensor fault E2.T] *or* [Sensor fault F2.T] *or* [Sensor fault G2.w] *or* [Sensor fault G2.T]

7.4.2 Effects of Uncertainty

To investigate the effect of measurement uncertainty, we will add Gaussian noise to each variable of the simulation output. The measurement uncertainty of each sensor depends on the sensor type. The two common types of temperature sensors in nuclear systems are resistance temperature detectors (RTDs) and thermocouples. Typical RTDs have an accuracy of around $\pm 0.3^o C$ while thermocouples have a lower accuracy of up to $\pm 2.2^o C$ or 0.75%. Flowrate and pressure readings are provided by pressure transmitters typically with an accuracy of $\pm 0.25\%$ for high precision sensors and up to $\pm 1.25\%$ for others [100]. Sensors can be sampled every second or minute.

For the current application, we are interested in faults of slow-degradation type and thus, only need to run diagnostics on a longer timescale, e.g. once every hour or day. In that case, it is not necessary to process every data point as collected from the sensors. The general practice would be to take the moving average of a certain number of sensor data points as one data point for the diagnostic tool. Doing so would effectively reduce the measurement uncertainty in each data point. For this demonstration, we will assume an effective standard deviation of $\pm 0.1^o C$ for each temperature reading and $\pm 0.1\%$ for pressure and flowrate readings. It should be noted that some sensors may have measurement errors in forms of bias, as opposed to white noise, which do not get canceled or reduced by averaging. We will be assuming all sensors have been calibrated initially such that the biases are negligible. The increasing of sensor bias over time will be recognized as a sensor fault.

The distribution parameters of each residual computed from 500 fault-free data points are listed in Table 7-5.

Table 7-5. Mean and standard deviation of each residual when the system is fault-free

Residual	Mean	Std. Dev.
r_1	-8.5×10^{-5}	2.63×10^{-3}
r_2	4.7×10^{-4}	7.35×10^{-3}
r_3	2.5×10^{-6}	1.99×10^{-3}
r_4	1.6×10^{-5}	1.94×10^{-3}
r_5	-1.6×10^{-4}	1.82×10^{-3}
r_6	-1.0×10^{-4}	1.28×10^{-2}
r_7	-3.4×10^{-4}	1.27×10^{-2}
r_8	-3.2×10^{-4}	1.41×10^{-2}
r_9	5.4×10^{-4}	1.37×10^{-2}

The mean values of all nine residuals are close to zero when the system is fault-free. Notice that the standard deviations of the four FWH performance residuals r_6 to r_9 are significantly higher than the other five residuals. That is expected since the calculation of those residuals involve more sensors.

For the case of fouling in FWH 1A, from the results in Section 7.4.1, we expect the two residuals r_6 and r_8 to be affected. In the presence of uncertainty, the plots of r_6 and r_8 are shown in Figure 7-9. Measurement uncertainty from the involved sensors combined in the uncertainty of each residual.

To detect non-zero residual, we will be using the GLR-D change detection tool whose decision function was provided in Eqn. (4.2). The decision functions for all nine residuals in this case are shown in Figure 7-10. The detection threshold was set to $h = 8.3$ for a false positive rate of 0.1%.

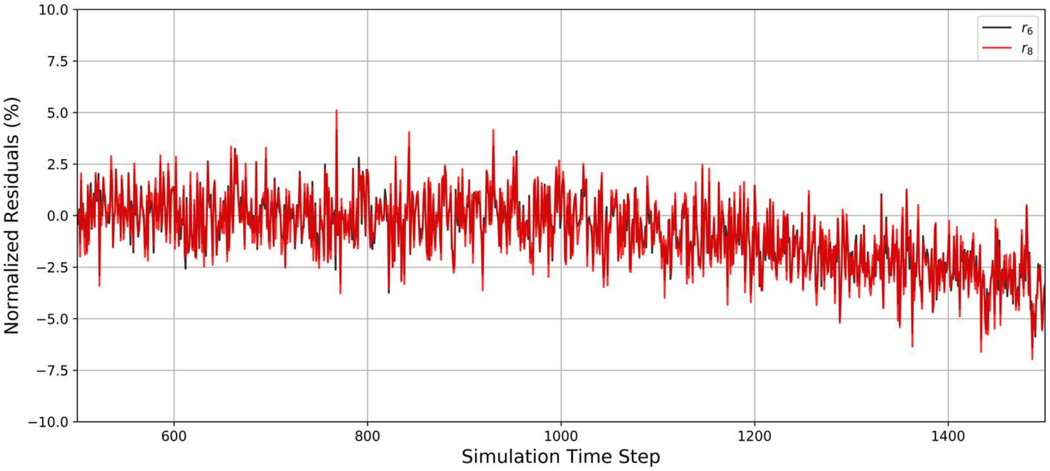

Figure 7-9. The two noisy non-zero residuals under the effect fouling in FWH 1A

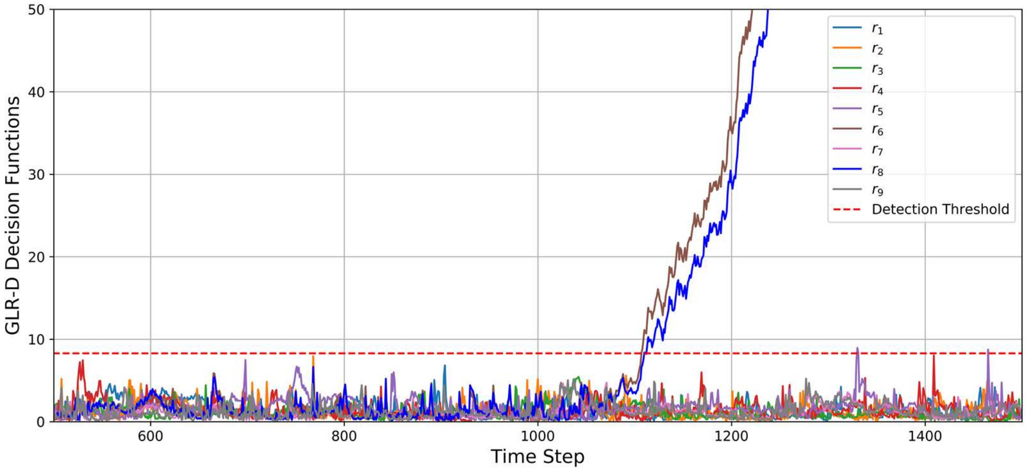

Figure 7-10. GLR-D decision functions for residual evaluation in the case of fouling in FWH 1A

It is clear that after both r_6 and r_8 can be observed to be non-zero given enough wait-time while the other residuals are observed to be zero. For this set of observed fault symptoms, the diagnostic result by the Deterministc II approach was discussed in the last section and summarized in Table 7-3. The two possibilities are fouling in FWH 1A and pressure sensor fault at the inlet of FWH 1A.

For the case of a sensor fault at E2.T as discussed in the last section, plots of the residuals affected by the fault are shown in Figure 7-11.

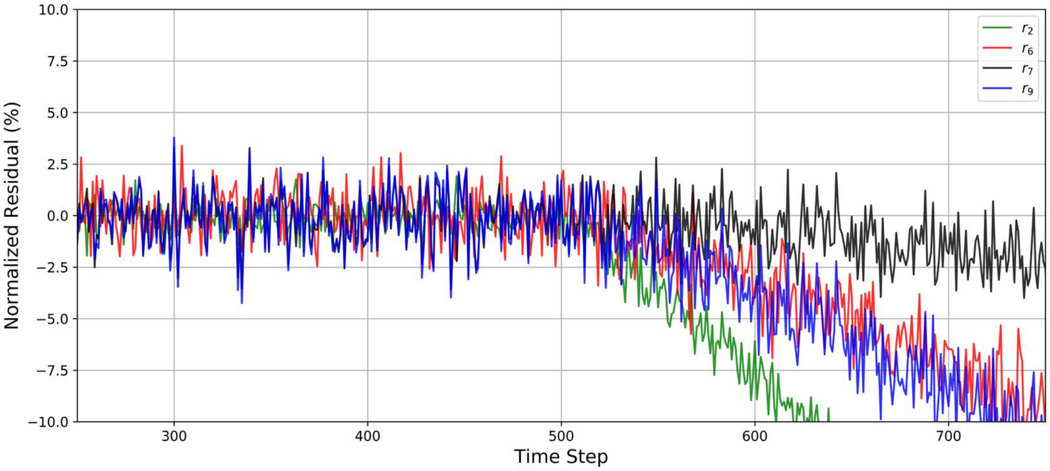

Figure 7-11. The four residuals affected by the temperature sensor fault at the feedwater inlet of FWH 1B

It can be observed from Figure 7-11 that the sensitivity of residual r_7 to the sensor fault is significantly lower compared to the other three residuals. The GLR-D decision functions for this case are plotted in Figure 7-12.

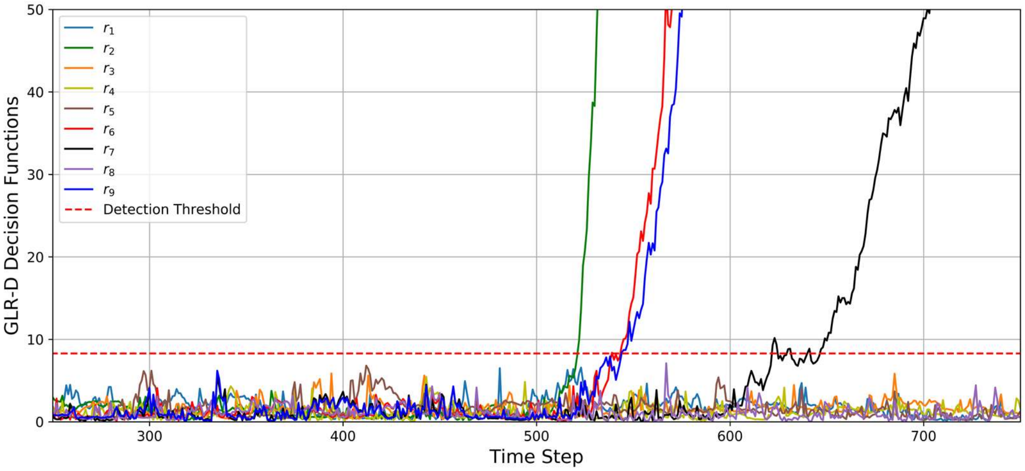

Figure 7-12. GLR-D decision functions in residual evaluation for the case of sensor fault at E2.T

Note that the sensor drift starts at time step 500. From Figure 7-12, r_2 can be observed to be non-zero starting at time step 520, r_6 and r_9 starting at step 540 while the change in residual r_7 goes undetected until around time step 650. Prior to timestep 650, because of the lower sensitivity of

r_7 to the fault, the change detection tool fails to detect the non-zero residual – a scenario we referred to as a false negative.

In the Deterministic II, the FDI reasoning framework is applied using the observations of non-zero and zero residuals from the change detection tool. The change detection results and the corresponding diagnostic results are summarized in Table 7-6.

Table 7-6. Change detection and diagnostic outputs for the case of a sensor fault at E2.T

Time Step	*Change Detection Output*	*Diagnostic Output*
$t < 520$	No non-zero residual	No fault.
$520 \leq t < 543$	r_2	No valid diagnosis.
$543 \leq t < 647$	r_2, r_6, r_9	No valid diagnosis.
$t > 647$	r_2, r_6, r_7, r_9	[Leakage from the FWH 1B shell] *or* [Sensor fault E2.T] *or* [Sensor fault F2.T] *or* [Sensor fault G2.w] *or* [Sensor fault G2.T]

Prior to time step 647, one or more affected residuals are undetected and the Deterministic II fails to find a valid diagnosis. It should be noted that in this case, as shown in Figure 7-7, the sensor bias increases over time, thus the effect in r_7 is eventually detected. For a small shift of the bias, the change in r_7 may not be detected because of its lower sensitivity to the fault.

This example showed that the Deterministic II approach is susceptible to false detections, a limitation we also discussed in Section 4.4. The deterministic reasoning framework may fail to find a valid diagnosis in case of one or more false positive or false negative by the change detection tool. In the presence of uncertainty, some of the residuals with lower sensitivity may not be detected until the fault becomes sufficiently significant. This issue effectively limits the diagnostic sensitivity. In Chapter 5, the probabilistic reasoning framework was developed to deal with this difficulty, i.e. to account for the possibility of false positive and false negative.

To apply the probabilistic framework, as described in Chapter 5, the first step is to construct a Bayesian network to represent the dependency between faults and residuals. The Bayesian network for this case is shown in Figure 7-13. Each number on the first layer represents a distinct fault. The list of faults is provided in Appendix C. In particular, number 1 is fouling in FWH 1A and 15 is the sensor fault at E2.T. The residual for the head model of each drain pump, r_3, r_4, and r_5, is independent from the rest the of the system. Thus, each drain pump can be represented by a separated Bayesian network with four independent faults: a pump fault and three sensor faults.

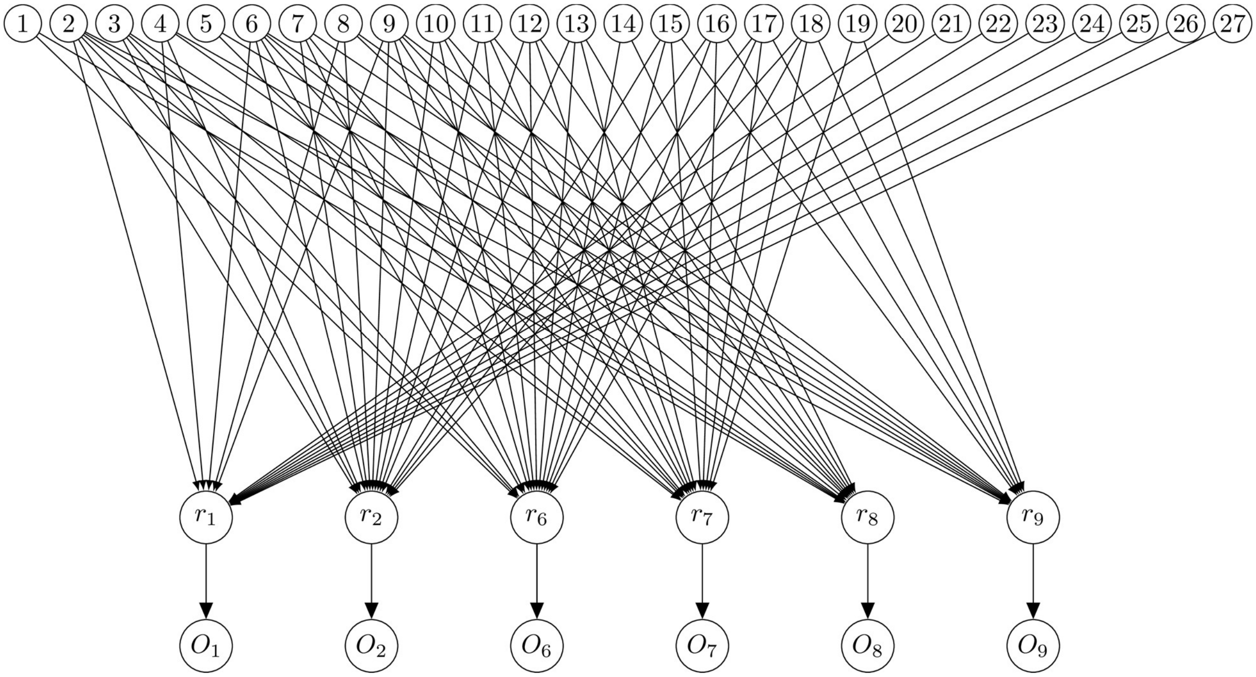

Figure 7-13. A Bayesian network for the high pressure feedwater system.

To define the structure of the network, we need to provide the prior probability of each fault and the conditional probability tables for the residuals and change detection observations. We will assume a prior probability of 10% for fouling, 1% for leakage, 5% blockage and 5% for sensor faults.

As with the example of the single-phase heat exchanger, to simplify the calculation, we will apply the same assumptions used for the notion of exoneration: each fault affects every residual it is involved in. The conditional probability $P(r \mid F)$ is then simply given by:

$$P(r_i = 0 \mid \text{all parents node} = 1) = 1.0 \tag{7.21}$$

$$P(r_i = 1 \mid \text{any parent node} = 1) = 1.0 \tag{7.22}$$

Suppose that the detection threshold has been chosen such that the false positive rate is 0.1% and the false negative rate is 1.0%, thus:

$$P(O_i = 1 \mid r_i = 0) = 0.001 \tag{7.23}$$

$$P(O_i = 0 \mid r_i = 1) = 0.01 \tag{7.24}$$

After the structure of the network has been defined, we can then compute the posterior probability of each fault given a set of observations. Prior to time step 520, all residuals are observed to be zero and no fault is detected.

Between time step 520 and 543, only r_2 is observed to be non-zero. The posterior probability of every fault is approximately zero and no fault is detected. The non-zero value of r_2 is interpreted as a false positive by the change detection tool.

Between time step 543 to 647, r_2, r_6, and r_9 are observed to be non-zero. Results of fault posterior probabilities for this case are listed in the third column of Table 7-7. Notice that the values found for fouling in FWH 1B (10.1%) and sensor fault D2.P (5.0%) are just the prior probability provided for the faults. The evidence, r_2, r_6, r_9 being non-zero, neither implicates nor exonerates these faults. The faults with significant posterior probability are the four sensors faults, in E2.T, F2.T, G2.T and G2.T, and leakage from the shell side of FWH 1B. This is the result found by the Deterministic II approach after time step 647 as listed in Table 7-6. The distinction between the sensor faults and the leakage is because of the difference in prior probability. The sensor faults are more likely than the leakage. Here, by using the probabilistic reasoning framework, we obtain the correct diagnostic result even when the change in residual r_7 is undetected.

After time step 647, r_7 is observed to be non-zero in addition to r_2, r_6, r_9. The results are listed in the fourth column in Table 7-7. The addition of r_7 to the evidence does not significantly change the posterior probabilities. The faults with significant posterior probability are the four sensor faults and FWH 1B shell leakage, in agreement with the result in Table 7-4 for the ideal case of no uncertainty. One can conclude the sensor faults are more likely than the leakage but, because of the limitation of the current sensor set, cannot differentiate between the four sensor faults.

Table 7-7. Fault posterior probabilities by the probabilistic reasoning framework for the case of a sensor fault at E2.T

Fault ID	*Fault*	*Symptoms*	
		r_2 , r_6 , r_9	r_2 , r_6 , r_7 , r_9
15	Sensor E2.T	25.7%	25.8%
16	Sensor F2.T	25.7%	25.8%
17	Sensor G2.w	25.7%	25.8%
18	Sensor G2.T	25.7%	25.8%
5	FWH 1B, Fouling	10.1%	10.1%
7	FWH 1B, Shell leak.	5.1%	5.2%
19	Sensor D2.P	5.0%	5.0%
Other faults		< 0.1%	< 0.1%

Chapter 8

Summary, Conclusions, and Future Work

8.1 Summary

The study in this thesis focused on the application of diagnostic methods to the problem of monitoring equipment health and sensor calibration status in nuclear engineering systems. The research is motivated by the ongoing effort to utilize automation and operator support technologies for cost reduction in nuclear power plants. The task of detecting equipment performance degradation can be automated by a diagnostic framework which make use of measurement data from instruments that are already in place for system monitoring. Given the long-time scale over which component degradation typically proceeds, some of the system monitoring sensors may also inevitably degrade and become unreliable. The human resources required to detect and recalibrate faulty sensors contribute a significant fraction to the overall O&M cost.

The objective of this thesis was the development of a diagnostic framework for thermal-hydraulic systems in commercial nuclear power plants, capable of dealing with both equipment faults and instrument faults. The principal target application for the approach developed here is for immediate implementation in currently operating nuclear power plants, however the methods developed here also would potentially have application for the design and operation of advanced nuclear reactor systems. In order to detect slow performance degradation and sensor drift, a high detection sensitivity is needed while the plant may undergo changes in operating conditions and the sensor data are subject to noise and uncertainty. For that purpose, the diagnostic framework needs to be insensitive in changes of boundary conditions and the various sources of uncertainty. Other challenging issues for the research problem include the lack of sensors that can be used for the specific diagnostic purposes.

The theoretical framework proposed in this thesis is a hybrid of quantitative model-based diagnosis, statistical change detection and probabilistic reasoning. Given the need for the diagnostic tool to be insensitive to operating condition changes under plant drifts, a physics-based approach was developed. In the framework of model-based diagnosis, physics-based models are used to describe the fault-free behavior of T-H components. Quantitative model residuals can be generated from the analytical redundancy relations provided by the fault-free component models. Non-zero model residuals serve as fault symptoms for the reasoning process in the model-based diagnosis framework developed here.

The presence of measurement and modeling uncertainty affects both the residual evaluation as well as the reasoning processes. A statistical change detection tool is necessary to detect whether a model residual is statistically zero or non-zero. Consequently, there is an associated false detection rate when each residual is detected to be zero or non-zero. If the false detection rates can be considered negligible, conventional deterministic reasoning frameworks can be applied to obtain possible fault diagnoses from a set observed fault symptoms. However, when false detection rates are significant, e.g. in the case of large modeling uncertainty, the deterministic reasoning frameworks may fail to produce valid diagnoses. The probabilistic reasoning framework using the method of Bayesian network was proposed to deal with such scenarios by considering the possibility of false observations in the reasoning process for fault diagnosis.

The construction of physics-based models requires the decomposition of each T-H system into separate components of known generic types. Physics-based models are developed for each generic component type in the form of parametric models. Each model may contain some unknown parameters which are determined for each specific component during the model calibration process. A training data set from measurement data of the process variables on the boundary of the component is required for this process. Thus, each component model has a minimum sensor requirement for the calibration process that must be performed. For most T-H systems in currently operating nuclear power plants, the available sensor set is limited and insufficient for model construction of standalone components. In this thesis, the lack of sufficient sensors is mitigated by the introduction of the concept of virtual sensors. Relations between different components and sensors at the system level are utilized to solve for the missing variables required for model construction. At the same time, the search for available models is expanded to cover multiple nearby components as allowed by the sensor set.

The proposed diagnostic framework was automated in a Python test implementation. In Chapter 8, the implementation was applied to a typical high-pressure feedwater system to demonstrate its capability. With the exception of a pressure head model for each drain pump, it was not possible to construct standalone component models for the major components in the system with the current sensor set. Most notably this was the case for the feedwater heaters (FWH) and steam generator feed pumps. By utilizing system balance equations, the missing feedwater flowrate sensors for the two first-point FWHs were computed from the other available sensors. The solutions of the balance equations, which we referred to as virtual sensors, was used in the construction of two FWH models. Furthermore, various aggregate models for pressure loss, flow rate ratio were created as listed in Table 7-1. Residuals were then generated from the models. In order to perform diagnosis at the system level it was necessary to keep track of the dependency of each residual on the involved components, on the sensors and on underlying assumptions for the system balance equations.

Results for the case of feedwater heater fouling and for the case of a sensor fault were investigated and the results were shown in Section 7.4.1. Useful results were obtained for each scenario, however, due to the limitation of the available sensor set, there were multiple valid diagnoses for each case, and it was not possible to always identify a unique diagnosis.

The effects of measurement uncertainty were demonstrated in Section 7.4.2. In the presence of measurement and modeling uncertainty, there is an inevitable delay in the time it takes for a non-zero residual to be detected. Depending on the sensitivity of each residual to each fault, some of the mathematically affected residuals were not always detected. In Table 7-6, an example was shown of a case in which it was not possible to provide a valid diagnosis when the change detection tool failed to detect one of the non-zero residuals. Such issues limit the diagnostic sensitivity and reliability of deterministic approaches in the case of significant uncertainty. The limitation can be overcome by the use of probabilistic reasoning approaches and results using probabilistic reasoning based on a Bayesian network were shown in Table 7-7. The correct diagnostic result was obtained even when one of the affected residuals was undetected.

The most significant original contribution of this thesis to the field is the development of a physics-based probabilistic framework for fault diagnostics of T-H systems with limited instrumentation in nuclear power plants. The integration of model-based diagnosis with probabilistic reasoning at system level is possible due to the physics-based approach in model construction and residual generation. From the underlying analytical relations, each residual can be directly linked to various component and sensors faults at system level. The use of physics-based models for quantitative model-based diagnosis allows the proposed framework to be less sensitive to possible changes of operating conditions and capable of dealing with both component faults and sensor faults. The relations between different system components, represented by conservation laws, are utilized through the concept of virtual sensors and aggregate models to effectively reduce the number of sensors required for each component and provide a more detailed diagnosis. Modeling and measurement uncertainty are robustly dealt with by statistical change detection and probabilistic reasoning.

8.2 Future Work

The focus of this thesis has been the development of a diagnostic framework applicable to complex engineering systems with a limited sensor set. Diagnostic results for each system were produced consistent with the best spatial resolution permitted by the available sensor set. The diagnostic benefit from each possible new sensor can be systematically analyzed from the effects of the addition to each step of the framework by using the model construction to residual generation and reasoning developed in this work. Thus, the physics-based framework as formulated provides a straightforward transition to the ‘inverse problem’ of determining optimal placement for new sensors for a given monitoring need. Solving the inverse problem considering both technical and economic aspects will be an important part of the future application of this work.

The reliability of model-based diagnosis depends on the quality of the models being used. High modeling uncertainty may lead to unreliable or false diagnoses. The difficulty in developing models with tolerable uncertainty for complex technical processes is one of the limitations of model-based diagnosis. In this work, models of T-H components were constructed in a physics-based approach and thus are less data-dependent.   The effects of uncertainty were robustly treated using a probabilistic framework. Nevertheless, the model construction process requires simplifications of the underlying physics which inevitably lead to intrinsic modeling uncertainty. As possible direction for future work in the modeling aspect of the work here is to utilize the design parameters available for each component and investigate the use of high-fidelity simulation codes to construct simulation-based surrogate models.

The proposed framework has been formulated to consist of separated steps and allow various methods in each step to be interchangeable. Specifically, for the reasoning process, one has the option to use one of the deterministic reasoning approaches or the probabilistic reasoning approach. Residuals are evaluated by statistical change detection tools whose outputs in the form of zero and non-zero residuals are usable by both deterministic reasoning and probabilistic reasoning approaches. In the presented probabilistic results for the heat exchanger example and the feedwater system, several assumptions were applied to simplify the estimation of the conditional probability tables. Such simplifications are justifiable for practical use if one is not particularly interested in the exact posterior probability of each fault but only needs rough estimation of the relative magnitude between faults. However, as discussed in Section 5.2, a more accurate estimation of the conditional probability tables can be obtained by sampling the underlying model. This will be part of the future work for the application of the probabilistic framework in predictive and preventative maintenance for nuclear systems.

Finally, from a statistical point of view, potential conflicts arise in the factorization of the uncertainty treatment into two separated steps, a change detection step using the GLR test and a probabilistic reasoning step using Bayesian network. The former is a *frequentist* approach using likelihood ratio test while the latter is a *Bayesian* approach. As discussed, such separation is necessary to allow the reasoning approaches to be interchangeable. However, there are several Bayesian change detection methods available with comparable performance to the GLR tests [101, 102] and for future work, methods can be investigated to combine both steps of the change detection and reasoning into a dynamic Bayesian network to treat the effects of uncertainty in diagnosis.

Appendix A

Regression Methods for Model Calibration

As described in Chapter 3, physics-based models can be developed for each generic type of T-H components in form of parametric models. Each parametric model may contain several unknown parameters to be determined for each specific component in a process we referred to as model calibration. Details of the model calibration process are discussed in this appendix.

Each of the parametric models as formulated can be expressed as a linear model:

$$y = \beta_0 + \sum_{j=1}^{p} X_j \beta_j \tag{A.1}$$

where y is the model output; β_0 and β_j are the model parameters; X_j are the input variables for the model; and p is the number of distinct input variables. Each X_j could be from a different physical quantity or just a different power of the same physical variable. For example, for the model of the overall heat transfer coefficient in Eqn. (3.21), the X_j 's are $w_c^{-0.8}$ and $w_h^{-0.8}$. The objective of the model calibration process is to determine the model parameters using a training data set.

Each data point of the training set provides a value of y for certain measured input X_j. Let $\mathbf{y}$ be the column vector containing N output from the training set; $\mathbf{X}$ denotes the $N \times (p+1)$ input matrix with each row representing the p input variables for each training data point and a 1 in the first position. The model can be written in linear form as:

$$\hat{\mathbf{y}} = \mathbf{X}\beta \tag{A.2}$$

where $\beta = (\beta_0, \beta_1, ..., \beta_p)^T$ is the parameter vector and $\hat{\mathbf{y}}$ is the model prediction for the N training data points. We need to estimate the parameter vector β by fitting the model prediction $\hat{\mathbf{y}}$ against the measured output $\mathbf{y}$. The most common estimation method is least squares, in which the parameters are selected to minimize the sum of squares of the differences between $\hat{\mathbf{y}}$ and $\mathbf{y}$, known as the *residual sum of squares* [86]*:*

$$\begin{aligned} \mathrm{RSS}(\beta) &= \left(\mathbf{y} - \hat{\mathbf{y}}\right)^T \left(\mathbf{y} - \hat{\mathbf{y}}\right) \\ &= \left(\mathbf{y} - \mathbf{X}\beta\right)^T \left(\mathbf{y} - \mathbf{X}\beta\right) \end{aligned} \tag{A.3}$$

To minimize the residual sum of squares, we can now take the derivative w.r.t β and set it to zero to obtain a least square estimation of the model parameters:

$$\hat{\beta} = \left(\mathbf{X}^T \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{y} \tag{A.4}$$

Under the assumption that the observations y_i are uncorrelated and have a constant variance σ^2, the variance-covariance matrix of the estimated parameters is given by:

$$\mathrm{Var}(\hat{\beta}) = \left(\mathbf{X}^T \mathbf{X}\right)^{-1} \sigma^2 \tag{A.5}$$

where the variance σ^2 can be estimated from the observations by:

$$\hat{\sigma}^2 = \frac{1}{N - p - 1} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \tag{A.6}$$

$\hat{\sigma}$ is known as the residual standard error (RSE). The residual standard error could be used as an indicator for the accuracy of the model. However, as it is estimated from training data, the RSE could underestimate the model prediction uncertainty. The standard practice when there are sufficient data is to split a data set into three parts: a *training set* used to fit the models; a *validation set* used to estimate prediction error for model selection; and a *test set* used to access the prediction error of the final chosen model [86]. In the current application, a physics-based approach is used for model selection, thus, a *validation set* is not required but a *test set* is still preferred to estimate the prediction uncertainty.

As an example, consider the calibration process for a pressure head model for a pump. We will use the quadratic form from Eqn. (3.12):

$$\Delta P_{\text{head}} = \theta_0 + \theta_1\ w + \theta_2\ w^2 \tag{A.7}$$

A set of real plant startup data for a feedwater pump is shown in Figure A-1 for the measured flowrate and Figure A-2 for the pressure head.

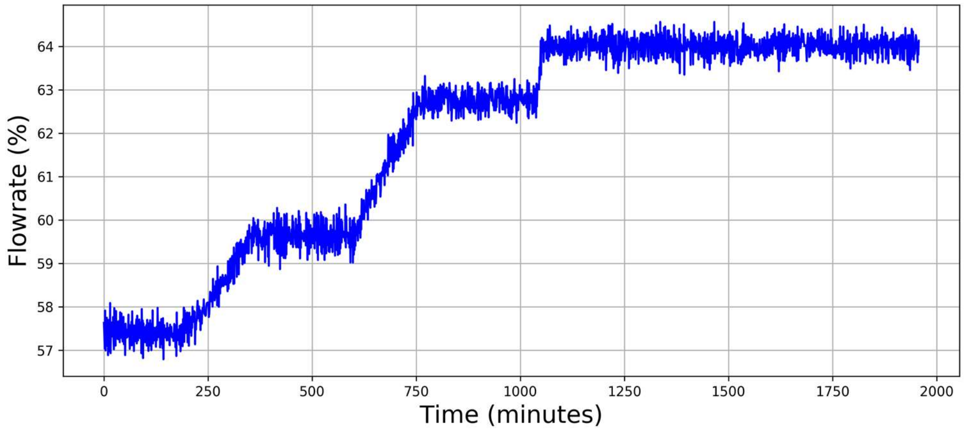

Figure A-1. Flowrate for a feedwater pump during startup

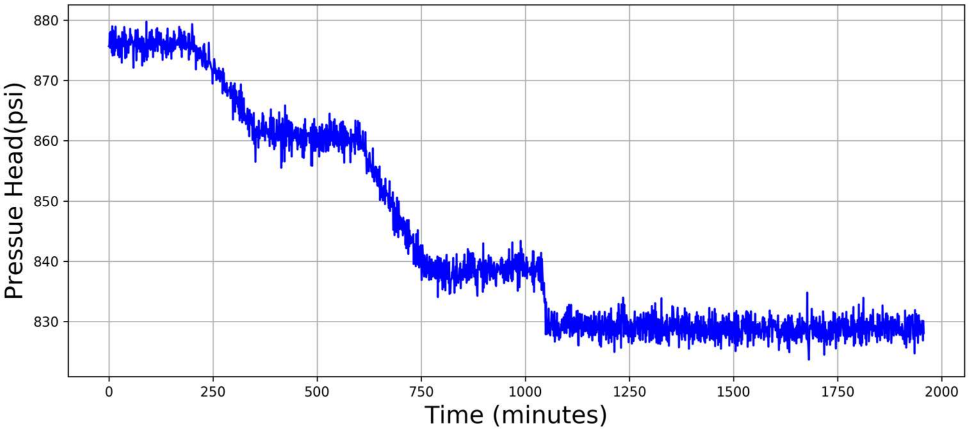

Figure A-2. Pressure head by the feedwater pump during startup

The startup data set is split in to two data sets: a training set with 70% of the data and a test set from the rest. By fitting the model in Eqn. (A.7) against the training set, the estimated model parameters are listed in Table A-1.

Table A-8-1. Least squares estimation of the pressure head model parameters

Model parameters		Estimated	Std. Err.	95% confidence interval	
				Lower limit	Upper limit
$\Delta P(w)$	θ_0	1068.0	31.1□	1007.1	1129.0
	θ_1	0.025	1.019	–1.974	2.025
	θ_2	–0.059	0.008	–0.075	–0.042

The standard prediction error can be estimated from the test data set to be 1.30 (psi). A plot of the fitted model is shown in Figure A-3.

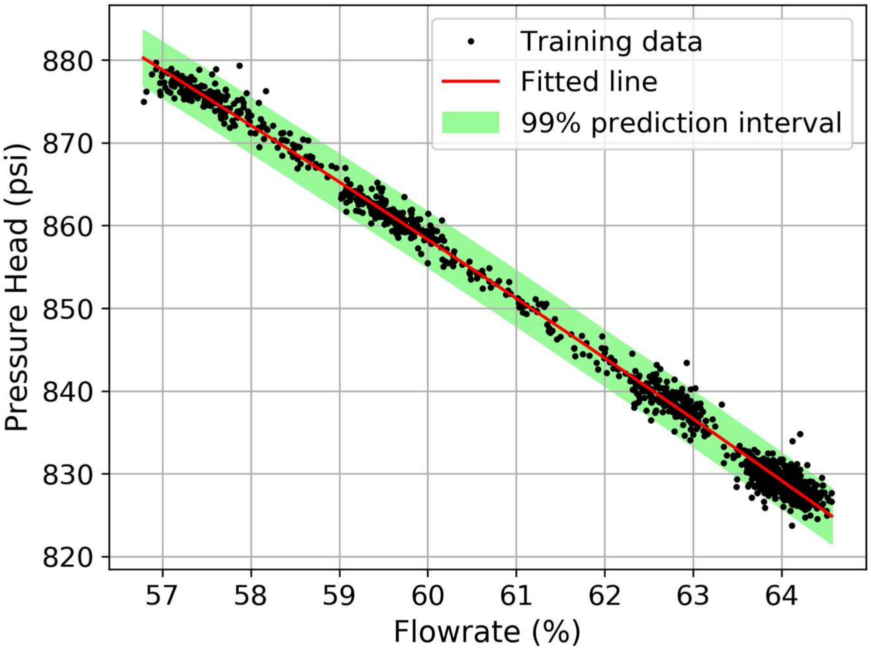

Figure A-3. Fitted pressure head model for the feedwater pump

Appendix B

Statistical Change Detection Methods

In the presence of noise and uncertainty, a statistical treatment is needed to evaluate the residuals in order to determine if a residual has become non-zero. In this appendix, statistical change point detection methods are discussed, and the most suitable method is selected for the current application.

In the problem of statistical change point detection (CPD), one aims to detect whether a process or a random variable has deviated from its normal behavior, i.e. if there is a change in the underlying distribution. Based on the conditions of the target application, we will assume distribution parameters of a process variable in its normal state (in control) are known or can be estimated but the distribution after a change (out of control) is unknown.

We are interested in detecting a change of mean value. The variance or standard deviation is assumed to remain the same. In general, a change in process mean can occur in two ways:

- Shift: At the changepoint, the process mean abruptly shifts to a different value and stays there after the change.
- Drift: At the changepoint, the process mean starts drifting gradually from the original value. For the problem of on-line change detection, one aims to minimize the detection delay, i.e. the time between the changepoint and detection point. In that period, it will be assumed that the drift is linear.

The two change modes are illustrated in Figure B-1. The primary objective for applying a CPD method is to detect a change as soon as possible, i.e. minimizing the delay between change time and detection time. A secondary objective, which may or may not be possible depending on the method, is to estimate the time and amplitude of the change.

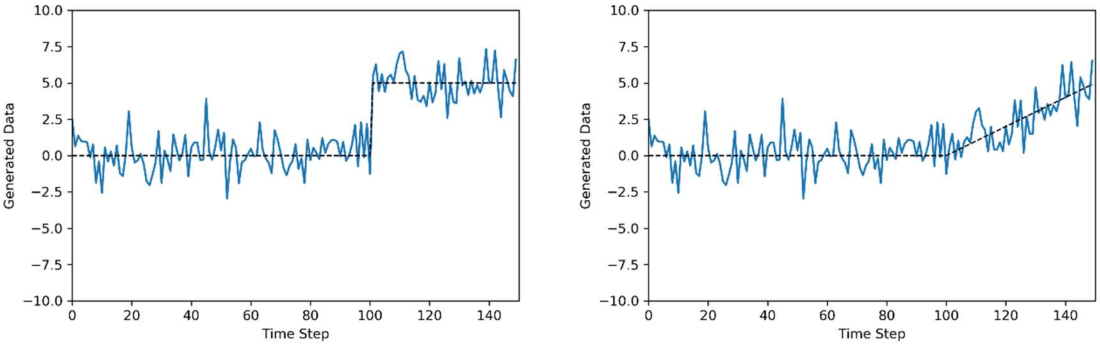

Figure B-1. A shift in process mean (left) and a drift in process mean (right)

B. 1 Overview

Since the distribution parameters after a change are unknown, some of the most well-developed CPD methods, including the CUSUM method, are not applicable [15]. Some of the possible options are:

- Shewhart control chart
- Exponentially weighted moving average (EWMA) control chart
- Generalized likelihood ratio (GLR) tests

The Shewhart control chart [89, 88] is a statistical process control tool designed to determine if a process has gone out of control, i.e. has deviated from its expected behavior. It is formulated as a limit-checking detector: a process is considered out of control when its deviation from the expected mean value exceeds the limits set based on the expected standard deviation. The Shewhart control chart is easier to implement and useful for detection of large shift in mean value but it offers no way of estimating the change magnitude or locating the change point. Also, it has proved to be less effective for detecting small shift or drift.

In a similar manner, the EWMA control chart operates by setting upper and lower limits based on the standard deviation but like the name suggested, it use exponentially weighted moving average values of the process instead of directly measured values in the limit checking process. Recent observations have higher weights in the considerations. The EWMA method can be used to detect both shift and drift changes in mean value [90].

Another method which can be formulated specifically for the detection of both shift and drift changes is the GLR tests. Furthermore, it also provides estimations of the change magnitude and change time. Previous studies have shown that the GLR method when set up properly can provide the best overall performance in term of detection delay and parameter estimations, i.e. change time and change magnitude [89, 91]. The methodology of the GLR method will be summarized in the remaining of this section.

Consider a statistical process, i.e. a series of sequentially sampled data, which is expected to have constant mean and standard deviation values under normal behavior (in-control). For each time step when a new data point is sampled, we are interested in detecting whether the mean value of the process has changed. We assume the process can be described by a normal (Gaussian) distribution whose mean and standard deviation are known or can be estimated. Under this assumption, the distribution for a new data point is given by:

$$p(y \mid \mu_0, \sigma_0) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{(y-\mu_0)^2}{2\sigma_0^2} \right) \tag{B.1}$$

Given a sequence of observed data and a statistical model of the distribution, the likelihood function represents the plausibility of the model. In other words, the higher the likelihood function, the more likely that the statistical model is correct. Given two hypotheses, or two different models, of a distribution, the ratio of the likelihood functions can be used to decide on which hypothesis is more likely to be true. For a sequence of independently observed data, the likelihood function is the product of the likelihood on each data point. For the Gaussian distribution, the likelihood function given each data point is:

$$l(\mu, \sigma \mid y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2} \right) \tag{B.2}$$

B.2 Generalized Likelihood Ratio Test (GLRT)

Consider a series of sequentially sampled data $\{y_k\}$. For each time step when a new data point is sampled, a decision rule is computed to test between two hypotheses:

- H_0: There has been no change. The distribution parameter is given by defined by $\theta_0 = (\mu_0, \sigma_0)$
- H_1: There was a change point happened at some time step j, after the change point, distribution parameter has changed to θ_1

We have assumed that the model parameters (μ_0, σ_0) for the null hypothesis H_0, i.e the distribution before any change occurs, are known. If the parameter after change, θ_1, is also known, the decision can be made based on the likelihood ratio between the two distributions.

If the parameter after change is unknown, however, the ratio test cannot be performed. One solution as proposed by Wald [19,20] is to replace the unknown parameter θ_1 by its maximum likelihood estimate. The likelihood ratio test is then based on the ratio of likelihoods defined by:

$$\hat{\Lambda}_N = \frac{\sup_{\theta_1} L(H_1)}{L(H_0)} \tag{B.3}$$

If $\hat{\Lambda}_N < 1$, the likelihood of the null hypothesis H_0 is larger than the maximum likelihood of the hypothesis H_1 and therefore one can select H_0 and conclude that there has been no change. Otherwise, a change can be reported with its location determined by the maximum of the likelihood of H_1.

In H_1, we assume a change point happened at time index j after which the model parameter changed to θ_1. For time index *1* to $j-1$, both H_0 and H_1 assume the same model parameter (θ_0) and thus the likelihood ratio during this period cancels out. The log-likelihood ratio given observations from time index up to time index $k > j$ is therefore given by:

$$S_1^k(\theta_1) = S_j^k(\theta_1) = \ln\left(\prod_{i=j}^{k} \frac{p_{\theta_1}(y_i)}{p_{\theta_0}(y_i)}\right) = \sum_{i=j}^{k} \ln\left(\frac{p_{\theta_1}(y_i)}{p_{\theta_0}(y_i)}\right) \tag{B.4}$$

The log likelihood ratio, $S_j^k(\theta_1)$, is a function of two variables: the change time j and the model parameter value after change to θ_1. These two parameters are selected to maximize the log likelihood ratio. The maximum log-likelihood ratio is defined by:

$$g_k = \max_{1 \le j \le k} \sup_{\theta_1} S_j^k(\theta_1) \tag{B.5}$$

The maximum log-likelihood ratio g_k is used as the decision function for the GLR test at time step k. Hypothesis H_1, with optimal parameters $\hat{j}_k$ and $\hat{\theta}_{1,k}$ to maximize $S_j^k(\theta_1)$, is accepted if the decision function g_k exceeds a pre-defined threshold h. Otherwise, H_0 is accepted and no change is reported.

For detection of changes in process mean, the GLR test has been commonly used to detect sustained shift [89]. For better estimations of the change point and change magnitude, the GLR test can be formulated specifically for each change model, i.e. sustained shift and linear drift. GLR change detection formulation for linear drifts were presented by Fahmy and Elsayed [92] and Wang et al. [93].

B.2.1 GLR Test for Sustained Shift in Process Mean (GLR-S)

For a Gaussian process, the model parameters are mean value μ and standard deviation σ. We will assume the standard deviation to remain the same even after a shift in process mean. The distribution before change is given by:

$$p_{\mu_0}(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu_0)^2}{2\sigma^2} \right) \tag{B.6}$$

After a shift in process mean $\theta_1 = (\mu_1, \sigma)$, the distribution is given by:

$$p_{\mu_1}(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu_1)^2}{2\sigma^2} \right) \tag{B.7}$$

The log-likelihood ratio for observations up to time k after a changepoint at j is:

$$S_1^k(\mu_1) = \sum_{i=j}^{k} \ln\left(\frac{p_{\mu_1}(y_i)}{p_{\mu_0}(y_i)} \right) = \sum_{i=j}^{k} \frac{(y_i-\mu_0)^2 - (y_i-\mu_1)^2}{2\sigma^2} \tag{B.8}$$

Take derivative with respect to μ_1 and set it to zero to obtain the optimal value of μ_1 that maximizes the likelihood ratio:

$$\hat{\mu}_{1,j} = \frac{1}{k-j+1}\sum_{i=j}^{k} y_i \tag{B.9}$$

Substitute into Eqn. (B.8) and (B.5), the decision function for a shift in process mean is given by:

$$g_k = \frac{1}{2\sigma^2}\max_{1\le j\le k}\frac{1}{k-j+1}\sum_{i=j}^{k}(y_i-\mu_0)^2 \tag{B.10}$$

A change is detected when decision function g_k exceeds a pre-defined threshold h. The process mean after the change can then be estimated using Eqn. (B.9) with the change point j determined by maximizing Eqn. (B.10).

B.2.2 GLR Test for Linear Drift in Process Mean (GLR-D)

For the case with a drift in process mean, we assume that the change starts exactly at the end a time step. Wang et al. [93] considered a more general case in which the change could occur during a time step but the difference in performance was shown to be negligible. With the change in process mean modeled by a linear drift of rate β per time step, starting at time index j, the process mean and the probability density at time index $i \ge j$ are given by:

$$\mu_i = \beta(i-j)+\mu_0 \tag{B.11}$$

$$p_{\theta_1}(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{\left(y-\beta(i-j)-\mu_0\right)^2}{2\sigma^2}\right) \tag{B.12}$$

The log-likelihood ratio at time step k after a drift starting at j:

$$S_j^k(\beta) = \sum_{i=j}^{k}\ln\left(\frac{p_{\theta_1}(y_i)}{p_{\theta_0}(y_i)}\right) = \sum_{i=j}^{k}\frac{\left(y_i-\mu_0\right)^2-\left(y_i-\mu_0-\beta(i-j)\right)^2}{2\sigma^2} \tag{B.13}$$

Maximize the right-hand side of Eqn. (B.13) with respect to β to obtain:

$$\hat{\beta}_j = \frac{\sum_{i=j}^{k}(i-j)(y_i-\mu_0)}{\sum_{i=j}^{k}(i-j)^2} \tag{B.14}$$

Substitute this result into Eqn. (B.13) and (B.5), the decision function for GLR-D is given by:

$$g_k = \frac{1}{2\sigma^2} \max_{1 \le j \le k} \frac{\left[\sum_{i=j}^{k} (i-j)(y_i - \mu_0) \right]^2}{\sum_{i=j}^{k} (i-j)^2} \tag{B.15}$$

A change is detected when decision function g_k exceeds a pre-defined threshold h. The drift rate can then be estimated using Eqn. (B.14) with the change point j determined by maximizing Eqn. (B.15).

B.2.3 Examples

The GLR-S and GLR-D tests are formulated specifically for the detection and estimation of changes by shift and drift in process mean, respectively. Previous studies have shown that both tests perform equally well in term of detection delay when applied to either type of change in process mean [93]. The advantage of specializing a test for each type of change is realized in change magnitude estimation. It should be expected that the GLR-S, formulated for predicting shift change, will not perform well in estimating the drift change and vice versa. To demonstrate the difference between the two formulations, consider an example with a series of observations which can be described by a Gaussian distribution with standard deviation σ_0. Starting at $t = 500s$, the process mean is subjected to a shift of $\Delta\mu = \sigma_0$. The decision functions for the GLR-S and GLR-D tests are plotted in Figure B-2 with the decision threshold selected albitrary at $h = 7.5$. It can be observed from Figure B-22 that both tests have similar performance in terms of detection delay.

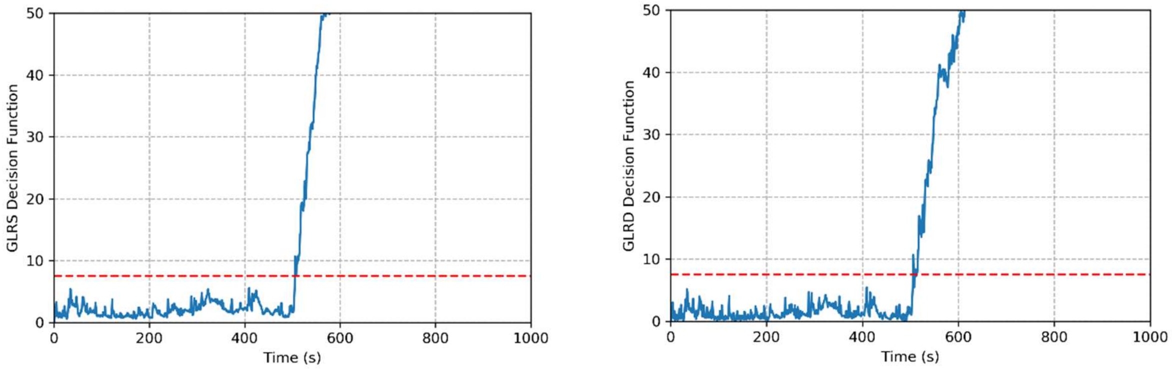

Figure B-2. Decision function by GLR-S and GLR-D for a shift in process mean of $\Delta\mu = \sigma$

After a change is detected, that is when the decision function exceeds the threshold, the GLR tests can also provide estimated values for the change point location and the change magnitude. The estimations of change point location by GLR-S and GLR-D for the case above are plotted in Figure B-3. Note that the change point estimation is to be disregarded until a change is reported when the decision function reaches its threshold. Figure B-3 shows that for the case of a shift in mean value, the GLR-S can provide a good estimation of the change point location while the GLR-D test cannot, which is expected.

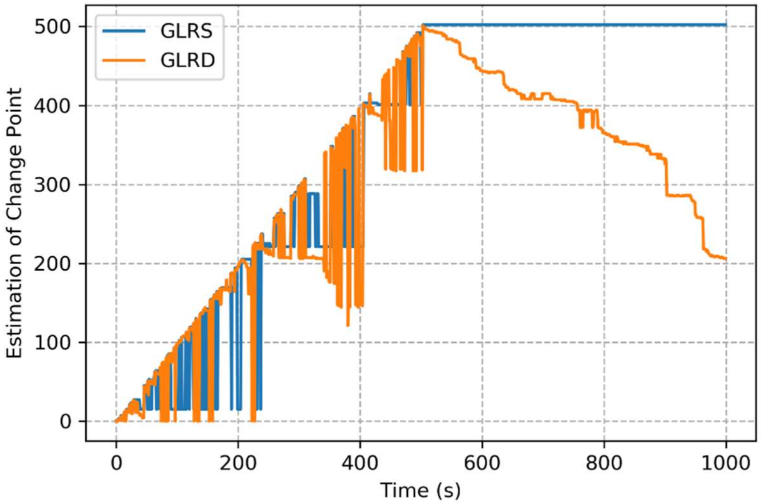

Figure B-3. Estimation of change point location by GLR tests for a shift of $\Delta\mu = \sigma$

B.2.4 Decision threshold

The decision threshold h for the change detection algorithm can be set based on a pre-defined false alarm rate, i.e the probability for the change detection algorithm to incorrectly detect a change even though the process variable is still in its normal distribution

For a given threshold for the decision function, the false alarm rate can be estimated by the inverse of the average run length (ARL). When the process variable is still in control, the average run length is defined as the average number of samples evaluated (the number of times the decision function is checked against the detection threshold) before a change is detected, which count as a false alarm since the variable is still in the distribution for its normal behavior.

For the GLR tests, it may be too computationally costly to use all the history data to compute the decision function. Instead, the algorithm can be run using only a number of m most recent data points. The average run length, and the false alarm rate, depend slightly on m .

The average run length or false alarm rate can be estimate by sampling the decision function using the normal behavior distribution of the process variable. Figure B-4 show the distribution of the decision function for the GLRS algorithm with $m = 200$, obtained by sampling 10^6 data points of a normalized normal distribution.

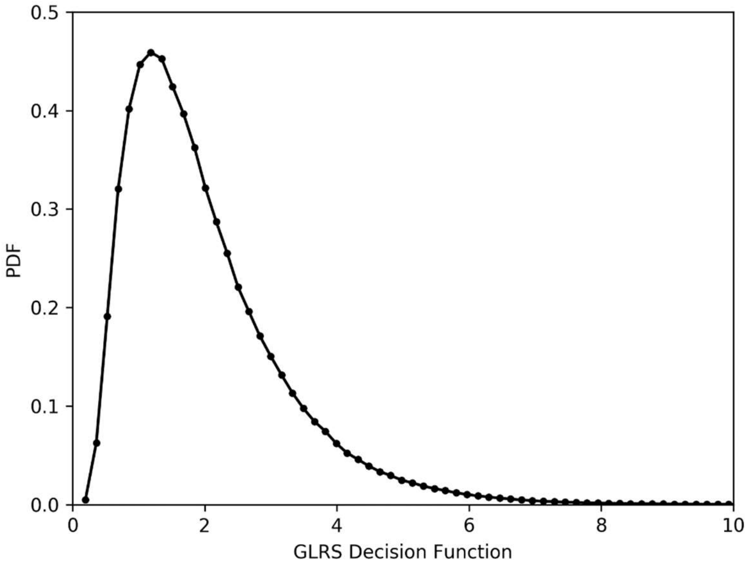

Figure B-4. Probability density function of the GLRS decision function

From the probability density function of the decision function, the average run length and the false alarm rate can be estimated, which are plotted on Figures B-5 and B-6, respectively. For example, the ARL corresponding to a 0.1% false alarm rate is 1000 and the threshold needed is approximately 8.36.

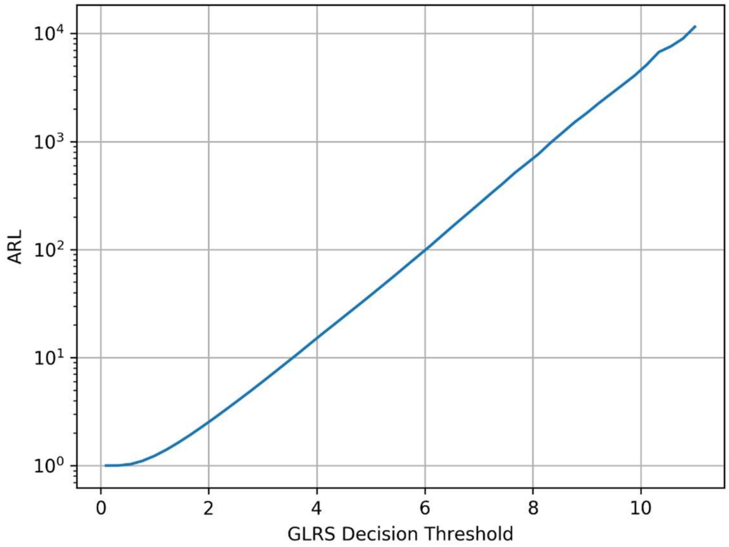

Figure B-5. ARL of the GLRS algorithm as a function of decision threshold.

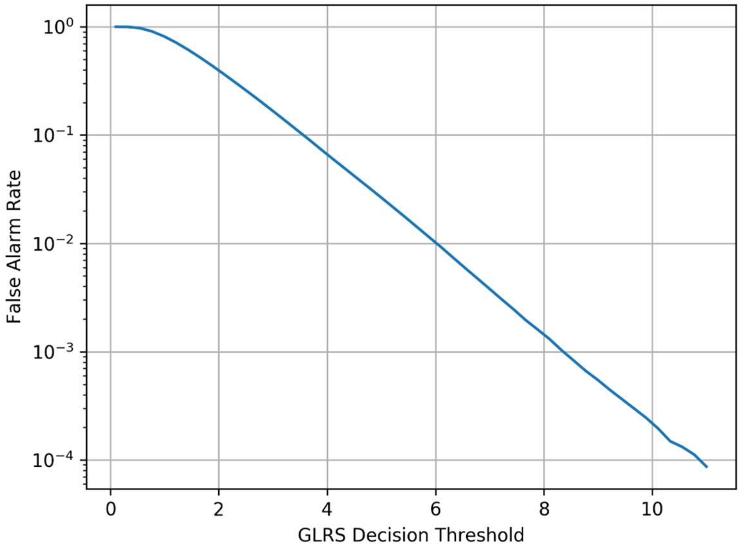

Figure B-6. GLRS false alarm rate as a function of decision threshold.

Appendix C

List of Faults for the HP Feedwater System

A list of faults for the high-pressure feedwater system considered in Chapter 7 and identified in Section 7.4.2 is provided in Table C-1. Note that here we have neglected the possibilities of piping leakage and blockage but the addition of such faults into the framework is straightforward.

For the probabilistic reasoning results in Section 7.4.2, we have only considered the first 27 faults that are relevant to the six residuals r_1, r_2, r_6, r_7, r_8, and r_9. The residuals computed from the pressure head models for the drain pumps, r_7, r_8, and r_9, are independent from those six residuals and only depend on the pump fault and sensor faults at the inlet and outlet of each pump.

More specifically, since a standalone pressure head model can be constructed for each of the drain pumps, the residual for each drain pump is independent from the rest of the system. To perform probabilistic reasoning for each drain pump, we can construct a Bayesian network that consist of only the pump fault and three sensor faults around the pump. An example is shown in Figure C-1 for the drain pump 1A.

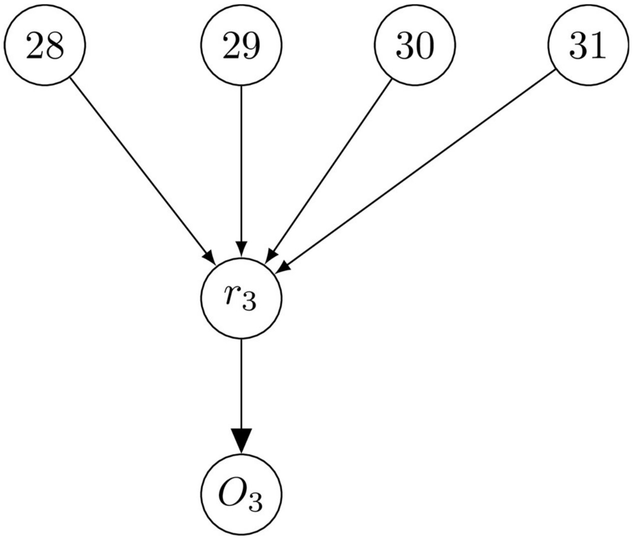

Figure C-1. A Bayesian network for drain pump 1A

Table C-1. List of faults for the high-pressure feedwater system

ID	Comp. Label	Comp. Type	Fault
1	1-FW-E-1A	FWH	Fouling
2	1-FW-E-1A	FWH	Tube leak
3	1-FW-E-1A	FWH	Shell leak
4	1-FW-E-1A	FWH	Tube block
5	1-FW-E-1B	FWH	Fouling
6	1-FW-E-1B	FWH	Tube leak
7	1-FW-E-1B	FWH	Shell leak
8	1-FW-E-1B	FWH	Tube block
9	FE-105	Flow sensor	Sensor fault
10	FW-TE-109A	Temp. sensor	Sensor fault
11	FW-TE-110A	Temp. sensor	Sensor fault
12	SD-FT-102A	Flow sensor	Sensor fault
13	SD-TE-110A	Temp. sensor	Sensor fault
14	ES-PT-100A	Press. sensor	Sensor fault
15	FW-TE-109B	Temp. sensor	Sensor fault
16	FW-TE-110B	Temp. sensor	Sensor fault
17	SD-FT-102B	Flow sensor	Sensor fault
18	SD-TE-110B	Temp. sensor	Sensor fault
19	ES-PT-100B	Press. sensor	Sensor fault
20	FW-PT-158	Press. sensor	Sensor fault
21	PT-100	Press. sensor	Sensor fault
22	1-FW-P-1A	Feed pump	Pump fault
23	1-FW-P-1B	Feed pump	Pump fault
24	MOV-150A	Valve	Leakage
25	MOV-150A	Valve	Blockage
26	MOV-150B	Valve	Leakage
27	MOV-150B	Valve	Blockage
28	1-SD-P-1A	Drain pump	Pump fault
29	1-SD-PT-100A	Press. sensor	Sensor fault
30	1-SD-FT-100A	Flow sensor	Sensor fault
31	1-SD-PT-108A	Press. sensor	Sensor fault
32	1-SD-P-1B	Drain pump	Pump fault
33	1-SD-PT-100B	Press. sensor	Sensor fault
34	1-SD-FT-100B	Flow sensor	Sensor fault
35	1-SD-PT-108B	Press. sensor	Sensor fault
36	1-SD-P-1C	Drain pump	Pump fault
37	1-SD-PT-100C	Press. sensor	Sensor fault
38	1-SD-FT-100C	Flow sensor	Sensor fault
39	1-SD-PT-108C	Press. sensor	Sensor fault

Bibliography

[1] U.S. Energy Information Administration, "International Energy Outlook," 2019.

[2] OECD/NEA, "Reduction of Capital Costs of Nuclear Power Plants," OECD Publishing, Paris, 2000.

[3] R. Vilim, A. Grelle, R. Lew, T. Ulrich, R. Boring and K. Thomas, "Computerized Operator Support System and Human Performance in the Control Room," in *10th International Topical Meeting on Nuclear Plant Instrumentation, Control, and Human-Machine Interface Technologies (NPIC & HMIT 2017)*, San Francisco, CA, 2017.

[4] R. Vilim, "Automating O&M Monitoring Using Physics-Based Qualitative and Quantitative Reasoning," in *14th Pacific Basin Nuclear Conference and Technology Exhibition*, San Francisco, CA US, 2018.

[5] EPRI, "Requirements for On-Line Monitoring in Nuclear Power Plants," Palo Alto, CA, 2008.

[6] T. Y. C. Wei and J. Reifman, " PRODIAG: A Process-Independent Transient Diagnostic System - I: Theoretical Concepts," *Nuclear Science and Engineering,* vol. 131, no. 3, pp. 329-347, 1999.

[7] J. Reifman and T. Y. C. Wei, "PRODIAG: A Process-Independent Transient Diagnostic System—II: Validation Tests," *Nuclear Science and Engineering,* vol. 131, no. 3, pp. 348-369, 1999.

[8] R. Vilim, Y. Park and A. Grelle, "Parameter-Free Conservation-Based Equipment Fault Diagnosis," in *9th International Conference on Nuclear Plant Instrumentation, Control and Human - Machine Interface Technologies*, Charlotte, NC, 2015.

[9] J. Korbicz, J. M. Koscielny, Z. Kowalczuk and W. Cholewa, Fault Diagnosis: Models, artificial intelligence, applications, Berlin: Springer, 2004.

[10] R. Isermann, Fault-diagnosis Systems: An introduction from fault detection to fault tolerance, Berlin: Springer, 2006.

[11] V. Palade and C. D. Bocaniala, Computational Intelligence in Fault Diagnosis, London: Springer, 2006.

[12] R. Isermann, Fault-Diagnosis Applications - Model-based condition monitoring: Actuators, drives, machinery, plants, sensors, and fault-tolerant Systems, Berlin: Springer, 2011.

[13] S. X. Ding, Model-Based Fault Diagnosis Techniques: Design schemes, algorithms and tools, London: Springer, 2012.

[14] C. Aldrich and L. Auret, Unsupervised Process Monitoring and Fault Diagnosis with Machine Learning Methods, London: Springer, 2013.

[15] S. X. Ding, Data-driven Design of Fault Diagnosis and Fault-tolerant Control Systems, London: Springer, 2014.

[16] Z. Chen, Data-Driven Fault Detection for Industrial Processes: Canonical correlation analysis and projection based methods, Berlin: Springer, 2017.

[17] V. Venkatasubramanian, R. Rengaswamy, K. Yin and S. N. Kavuri, "A Review of Process Fault Detection and Diagnosis Part I: Quantitative Model-Based Methods," *Computers and Chemical Engineering,* vol. 27, pp. 293-311, 2003.

[18] V. Venkatasubramanian, R. Rengaswamy and S. N. Kavuri, "A Review of Process Fault Detection and Diagnosis Part II: Qualitative Models and Search Strategies," *Computers and Chemical Engineering,* vol. 27, pp. 313-326, 2003.

[19] V. Venkatasubramanian, R. Rengaswamy, S. N. Kavuri and K. Yin, "A Review of Process Fault Detection and Diagnosis Part III: Process History-Based Methods," *Computers and Chemical Engineering,* vol. 27, pp. 327-346, 2003.

[20] J. Ma and J. Jiang, "Applications of fault detection and diagnosis methods in nuclear power plants: A review," *Progress in Nuclear Energy,* vol. 53, pp. 255-266, 2011.

[21] S. J. Qin, "Survey on data-driven industrial process monitoring and diagnosis," *Annual Reviews in Control,* vol. 36, pp. 220-234, 2012.

[22] A. Mouzakitis, "Classification of Fault Diagnosis Methods for Control Systems," *Measurement and Control,* vol. 46, no. 10, pp. 303-308, 2013.

[23] Z. Gao, C. Cecati and S. X. Ding, "A Survey of Fault Diagnosis and Fault-Tolerant Techniques—Part I: Fault Diagnosis With Model-Based and Signal-Based Approaches," *IEEE Transactions on Industrial Electronics,* vol. 62, no. 6, pp. 3757-3767, 2015.

[24] Z. Gao, C. Cecati and S. X. Ding, "A Survey of Fault Diagnosis and Fault-Tolerant Techniques—Part II: Fault Diagnosis With Knowledge-Based and Hybrid/Active Approaches," *IEEE Transactions on Industrial Electronics,* vol. 62, no. 6, pp. 3768-3774, 2015.

[25] Y. Liu and A. M. Bazzi, "A review and comparison of fault detection and diagnosis methods for squirrel-cage induction motors: State of the art," *ISA Transactions,* vol. 70, pp. 400-409, 2017.

[26] N. Md Nor, C. Che Hassan and M. Hussain, "A review of data-driven fault detection and diagnosis methods: applications in chemical process systems," *Reviews in Chemical Engineering,* pp. 1-40, 2019.

[27] R. Dunia, J. S. Qin, F. T. Edgar and J. T. McAvoy, "Identification of faulty sensors using principal component analysis," *AIChE Journal,* vol. 42, pp. 2797-2812, 1996.

[28] B. M. Wise and N. B. Gallagher, "The process chemometrics approach to process monitoring and fault detection," *Journal of Process Control,* vol. 6, pp. 329-469, 1996.

[29] X. Wang, U. Kruger, G. Irwin, G. McCullough and N. McDowell, "Nonlinear PCA with the local approach for diesel engine fault detection and diagnosis," *IEEE Transactions on Control Systems Technology,* vol. 16, no. 1, pp. 122-129, 2008.

[30] B. Jiang, J. Xiang and Y. Wang, "Rolling bearing fault diagnosis approach using probabilistic principal component analysis denoising and cyclic bispectrum," *Journal of Vibration and Control,* vol. 22, no. 10, pp. 2420-2433, 2014.

[31] Y. Zhang, C. Bingham and M. Gallimore, "Fault detection and diagnosis based on extensions of PCA," *Advances in Military Technology,* vol. 8, no. 2, pp. 27-41, 2013.

[32] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural Networks,* vol. 13, pp. 411-430, 2000.

[33] J. Ding, V. A. Gribok, W. J. Hines and B. Rasmusse, "Redundant sensor calibration monitoring using independent component analysis and principal component analysis," *Real-Time Systems,* vol. 27, pp. 27-47, 2004.

[34] D. Tsai, S. Wu and W. Chiu, "Defect detection in solar modules using ICA basis images," *IEEE Transactions on Industrial Informatics,* vol. 9, no. 1, pp. 122-131, 2013.

[35] Y. Zhang, N. Yang and S. Li, "Fault isolation of nonlinear processes based on fault direction and features," *IEEE Transactions on Control Systems Technology,* vol. 22, no. 4, pp. 1567-1572, 2014.

[36] Y. Guo, J. Na, B. Li and R. Fung, "Envelope extraction based dimension reduction for independent component analysis in fault diagnosis of rolling element bearing," *Journal of Vibration and Control,* vol. 333, no. 13, pp. 2983-2994, 2014.

[37] S. J. Qin, "Recursive PLS algorithms for adaptive data modeling," *Computers and Chemical Engineering,* vol. 22, pp. 503-514, 1998.

[38] S. Ding, S. Yin, K. Peng, H. Hao and B. Shen, "A novel scheme for key performance indicator prediction and diagnosis with application to an industrial hot strip hill," *IEEE Transactions on Industrial Informatics,* vol. 9, no. 4, pp. 2239-22247, 201.

[39] R. Vitalea, O. De-Noordb and A. Ferrera, "A kernel-based approach for fault diagnosis in batch processes," *Journal of Chemometrics,* vol. 28, no. 8, pp. 697-707, 2014.

[40] X. Zhao, Y. Xue and T. Wang, "Fault detection of batch process based on multi-way Kernel T-PLS," *Journal of Chemical and Pharmaceutical Research,* vol. 6, no. 7, pp. 338-346, 2014.

[41] L. H. Chiang, E. L. Russell and R. D. Braatz, "Fault diagnosis in chemical processes using Fisher discriminant analysis, discriminant partial least squares, and principal component analysis," *Chemometrics and Intelligent Laboratory Systems,* vol. 50, pp. 243-252, 2000.

[42] Q. P. He, S. J. Qin and J. Wang, "A new fault diagnosis method using fault directions in Fisher discriminant analysis," *AIChE Journal,* vol. 51, pp. 555-571, 2005.

[43] D. Tax, A. Ypma and R. Duin, "Pump failure determination using support vector data description," *Lecture Notes in Computer Science,* vol. 1642, pp. 415-420, 1999.

[44] A. Widodo and B. Yang, "Support vector machine in machine condition monitoring and fault diagnosis," *Mechanical Systems and Signal Processing,* vol. 21, no. 6, pp. 2560-2574, 2007.

[45] S. Yin, X. Gao, H. Karimi and X. Zhu, "Study on support vector machine based fault detection in Tennessee Eastman Process," *Abstract and Applied Analysis,* vol. 2014, 2014.

[46] M. Namdari, H. Jazayeri-Rad and S. Hashemi, "Process fault diagnosis using support vector machines with a genetic algorithm based parameter tuning," *Journal of Automation and Control,* vol. 2, no. 1, pp. 1-7, 2014.

[47] Z. B. Sahri and R. B. Yusof, "Support vector machine-based fault diagnosis of power transformer using k-nearest-neighbor imputed DGA dataset," *Journal of Communications and Computer Engineering,* vol. 2, no. 9, pp. 22-31, 2014.

[48] Y. Shatnawi and M. Al-Khassaweneh, "Fault diagnosis in internal combustion engines using extension neural network," *IEEE Transactions on Industrial Electronics,* vol. 61, no. 3, pp. 1434-1443, 2014.

[49] C. Yan, H. Zhang and L. Xu, "A novel real-time fault diagnosis system for steam turbine generator set by using strata hierarchical artificial neural network," *Energy and Power Engineering,* vol. 1, no. 1, pp. 7-16, 2009.

[50] O. Elnokity, I. Mahmoud, M. Refai and H. Farahat, "ANN based sensor faults detection, isolation and reading estimates-SFDIRE: Applied in a nuclear process," *Annals of Nuclear Energy,* vol. 49, no. 11, pp. 131-142, 2012.

[51] M. Valtierra-Rodriguez, R. Romero-Troncoso, R. Osornio-Rios and A. Garcia-Perez, "Detection and classification of single and combined power quality disturbances using neural networks," *IEEE Transactions on Industrial Electronics,* vol. 61, no. 5, pp. 2473-2482, 2014.

[52] J. P. Herzog, S. W. Wegerich and K. C. Gross, "MSET modeling of Crystal River-3 venturi flow meters," in *The 6th International Conference on Nuclear Engineering*, San Diego, CA, 1998.

[53] N. Zavaljevski and K. C. Gross, "Sensor fault detection in nuclear power plants using multivariate state estimation technique and support vector machines," Argonne National Laboratory, Argonne, Illinois, 2000.

[54] J. W. Hines and A. Usynin, "MSET performance optimization through regularization," *Nuclear Engineering and Technology,* vol. 37, pp. 177-184, 2005.

[55] M. Ueda, K. Tomobe, K. Setoguchi and A. Endou, "Application of autoregressive models to in-service estimation of transient response for LMFBR process instrumentation," *Nuclear Technology,* vol. 137, pp. 163-168, 2002.

[56] L. Hong and J. Dhupia, "A time domain approach to diagnose gearbox fault based on measured vibration signals," *Journal of Sound and Vibration,* vol. 333, no. 7, pp. 2164-2180, 2014.

[57] M. E. H. Benbouzid, "A review of induction motors signature analysis as a medium for faults detection," *IEEE Transactions on Industrial Electronics,* vol. 47, pp. 984-993, 2000.

[58] P. Tavner, L. Ran, J. Penman and H. Sedding, Condition monitoring of rotating electrical machines, Stylus Publishing, 2000.

[59] G. Joksimovic, J. Riger, T. Wolbank, N. Peric and M. Vasak, "Stator-current spectrum signature of healthy cage rotor induction machines," *IEEE Transactions on Industrial Electronics,* vol. 60, no. 9, pp. 4025-4033, 2013.

[60] C. J. Li and S. Y. Li, "Acoustic emission analysis for bearing condition monitoring," *Wear,* vol. 185, pp. 67-74, 1995.

[61] U. Kunze, "Experience with the acoustic leakage monitoring system ALUES in 17 VVER plants," *Progress in Nuclear Energy,* vol. 34, pp. 213-220, 1999.

[62] Z. Feng, M. Liang and F. Chu, "Recent advances in time–frequency analysis methods for machinery fault diagnosis: a review with application examples," *Mechanical Systems and Signal Processing,* vol. 38, pp. 165-205, 2013.

[63] D. Yu, Y. Yang and J. Cheng, "Application of time–frequency entropy method based on Hilbert–Huang transform to gear fault diagnosis," *Measurement,* vol. 40, pp. 823-830, 2007.

[64] E. Cabal-Yepez, A. Garcia-Ramirez and R. Romero-Troncoso, "Reconfigurable monitoring system for time–frequency analysis on industrial equipment through STFT and DWT," *IEEE Transactions on Industrial Informatics,* vol. 9, no. 2, pp. 760-771, 2013.

[65] R. Beard, "Failure Accommodation in Linear System Through Self-Reorganization (Doctoral Dissertation)," MIT, Cambridge, MA, 1971.

[66] L. Jones, "Failure Detection in Linear Systems (Doctoral Dissertation)," MIT, Cambridge, MA, 1973.

[67] P. M. Frank, S. X. Ding and T. Marcu, "Model-based fault diagnosis in technical processes," *Transactions of the Institute of Measurement and Control,* vol. 22, no. 1, pp. 57-101, 2000.

[68] E. Chow and A. Willsky, "Analytical redundancy and the design of robust failure detection systems.," *IEEE Transactions on Control,* vol. 29, pp. 603-614, 1984.

[69] R. Isermann, "Process fauls detection based on modelling and estimation methods - a survey," *Automatica,* vol. 20, no. 4, pp. 387-404, 1984.

[70] J. d. Kleer, "Local Methods for Localizing Faults in Electronic Circuits," MIT, Cambridge, MA, 1976.

[71] J. d. Kleer and B. Williams, "Diagnosing multiple faults," *Artificial Intelligence,* vol. 32, pp. 97-130, 1987.

[72] J. d. Kleer and J. Kurien, "Fundamentals of Model-Based Diagnosis," *IFAC Proceedings Volumes,* vol. 36, no. 5, pp. 25-36, 2003.

[73] L. Console, D. T. Dupre and P. Torasso, "A Theory of Diagnosis for Incomplete Causal Models," in *The 11th International Joint Conference on Artificial Intelligence*, Detroit, MI, 1989.

[74] Y. Peng and J. A. Reggia, Abductive Inference Models for Diagnostic Problem Solving, New York: Springer, 1990.

[75] D. Poole, "A Methodology for Using a Default and Abductive Reasoning System," *International Journal of Intelligent Systems,* vol. 5, no. 5, pp. 521-548, 1990.

[76] R. Reiter, "A theory of diagnosis from first principles," *Artificial Intelligence,* vol. 32, pp. 57-95, 1987.

[77] J. d. Kleer, A. Mackworth and R. Reiter, "Characterizing Diagnoses and Systems," *Artificial Intelligence,* vol. 52, pp. 197-222, 1992.

[78] J. S. Brown and J. d. Kleer., "A qualitative physics based on confluences," *Artifical Intellegence,* vol. 24, no. 1-3, pp. 7-83, 1984.

[79] J. Reifman and T. Y. C. Wei, "PRODIAG - Dynamic Qualitative Analysis for Process Fault Diagnosis," in *9th Power Plant Dynamics, Control & Testing Symposium*, Knoxville, Tennessee, 1995.

[80] J. Reifman, T. Y. C. Wei, J. E. Vitela, C. A. Applequist and T. M. Chasensky, "PRODIAG - A Hybrid Artificial Intelligence Based Reactor Diagnostic System for Process Faults," in *4th International Conference on Nuclear Engineering*, New Orleans, Louisiana, 1996.

[81] A. L. Grelle, Y. S. Park and R. B. Vilim, "Development and Testing of Fault-Diagnosis Algorithms for Reactor Plant Systems," in *24th International Conference on Nuclear Engineering*, Charlotte, North Carolina, 2016.

[82] M.-O. Cordier, P. Dague, M. Dumas, F. Lévy, J. Montmain, M. Staroswiecki and L. Travé-massuyès, "AI and Automatic Control Approaches of Model-Based Diagnosis: Links and Underlying Hypotheses," *IFAC Proceedings Volumes,* vol. 33, no. 11, pp. 279-284, 2000.

[83] N. E. Todreas and M. S. Kazimi, Nuclear Systems Volume I: Thermal Hydraulic Fundamentals, London: CRC Press, 2011.

[84] T. L. Bergman, A. S. Lavine, F. P. Incropera and D. P. DeWitt, Fundamentals of Heat and Mass Transfer, New York: Wiley, 2011.

[85] J. Edwards, "Design and rating shell and tube heat exchangers," P&I Design Ltd., Teesside, UK, 2008.

[86] T. Hastie, R. Tibshirani and J. Friedman, The Elements of Statistical Learning - Data Mining, Inference, and Prediction, New York: Springer-Verlag, 2009.

[87] W. A. Shewhart, Economic Control of Quality Manufactured Product, New York: D. Van Nostrand Company, 1931.

[88] E. S. Page, "Control charts for the mean of a normal population," *Journal of the Royal Statistical Society,* vol. 16, no. 1, pp. 131-135, 1954.

[89] M. Basseville and I. V. Nikiforov, Detection of Abrupt Changes - Theory and Application, Prentice Hall, In., 1993.

[90] F. Gan, "EWMA control chart under linear drift," *Journal of Statistical Computation and Simulation,* vol. 38, pp. 181-200, 1991.

[91] Y. Liu, C. Zou and Z. Wang, "Comparisons of control schemes for monitoring the means of processes subject to drifts," *Metrika,* vol. 70, no. 2, pp. 141-163, 2009.

[92] H. M. Fahmy and E. A. Elsayed, "Detection of linear trends in process mean," *International Journal of Production Research,* vol. 43, no. 3, pp. 487-504, 2006.

[93] S. Wang, L. Xu and M. R. R. Jr, "A generalized likelihood ratio control chart for monitoring the process mean subject to linear drifts," *Quality and Reliability Engineering International,* vol. 29, pp. 545-553, 2013.

[94] R. Vilim, T. Lee and S. Passerini, "GPASS – A Code for Design and Analysis of Power Plant Control and Protection Systems," in *2016 International Congress on Advances in Nuclear Power Plants*, San Francisco, CA, 2016.

[95] J. Pearl, Causality: Models, Reasoning, and Inference, New York, NY: Cambridge University Press, 2009.

[96] Dominion Energy Virginia, "Revision 54 to Updated Final Safety Analysis Report, Chapter 10, Steam and Power Conversion System," U.S. NRC, 2018.

[97] R. Ponciroli and R. Vilim, *Personal Communication*. July 2019.

[98] British Electricity International, Turbines, Generators and Associated Plant, London: Pergamon, 1991.

[99] W. McCabe, J. Smith and P. Harriott, Unit Operations of Chemical Engineering, McGraw-Hill, 1993.

[100] J. B. Coble, R. M. Meyer, P. Ramuhalli, L. J. Bond, H. M. Hashemian, B. D. Shumaker and D. S. Cummins, "A Review of Sensor Calibration Monitoring for Calibration Interval Extension in Nuclear Power Plants," Pacific Northwest National Laboratory, Richland, Washington, 2012.

[101] R. P. Adams and D. J. MacKay, "Bayesian online changepoint detection," *arXiv:0710.3742,* 2007.

[102] A. B. Downey, "A novel changepoint detection algorithm," *arXiv:0812.1237,* 2008.