FAULT DETECTION AND ROOT CAUSE DIAGNOSIS USING DYNAMIC BAYESIAN NETWORK

by

© Md. Tanjin Amin

A thesis submitted to the

School of Graduate Studies

in partial fulfillment of the requirement for the degree of

Master of Engineering

Faculty of Engineering and Applied Science

Memorial University of Newfoundland

May 2018

St. John's Newfoundland and Labrador

Abstract

This thesis presents two real time process fault detection and diagnosis (FDD) techniques incorporating process data and prior knowledge. Unlike supervised monitoring techniques, both these methods can perform without having any prior information of a fault. In the first part of this research, a hybrid methodology is developed combining principal component analysis (PCA), Bayesian network (BN) and multiple uncertain (likelihood) evidence to improve the diagnostic capacity of PCA and existing PCA-BN schemes with hard evidence based updating. A dynamic BN (DBN) based FDD methodology is proposed in the later part of this work which provides detection and accurate diagnosis by a single tool. Furthermore, fault propagation pathway is analyzed using the predictive feature of a BN and cause-effect relationships among the process variables. Proposed frameworks are successfully validated by applying to several process models.

Acknowledgement

At first, I would like to express profound indebtedness to my supervisors, Dr. Faisal Khan, and Dr. Syed Imtiaz. I got the foremost help from them in any situation. Their motivational feedbacks, research ideas, and patience were the key to reach my goal.

I would also like to remember the sacrifice of my beloved parents, sister, and wife. I was away from them for two years. They always tried to energize me from thousands of miles away with their inspirational words.

It was difficult for me to get adapted in a new condition after reaching St. John’s. My friends, Suvra Chakraborty, Abdul Aziz, Naveel Islam, Jannatul Naeema, and Ahmed Elruby made my life much easier here.

In past two years, I have met many excellent colleagues in the Centre for Risk, Integrity and Safety Engineering (C-RISE). I would like to thank all of them. Specially, I must recognize the contribution of Mohammad Aminul Islam Khan and Hassan Gharahbagheri, who always helped me smilingly in any programming related issue.

I would also like to thank the Natural Sciences and Engineering Research Council (NSERC) for providing the fund.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1. Background

Continuous process monitoring plays a crucial role in ensuring process safety and reliable product quality demand (Severson et al., 2016). Process industries are getting bigger day by day aiming to meet up the ever-increasing end users' demand, resulting in simultaneous interaction of numerous variables, which is making the task of process monitoring arduous. Fault detection and diagnosis (FDD) tools are the central units of any process monitoring technique. A fault can be defined as a deviation of at least one of the variables from the acceptable operational range (Himmelblau, 1978; Isermann, 2005). It may emerge from a malfunctioning actuator or sensor. Fault detection is determining whether a process is in a normal state or in an abnormal state, while fault diagnosis refers to the identification of the root cause(s) of the fault. Early detection helps in taking corrective actions timely, and diagnosis guides the operators to a specific part of the process which is faulty, and needs to be taken care of. FDD is an active area of research in last few decades. Still, process industries suffer from an enormous economic losses due to the occurred abnormal situations initiated by a fault during operation (Nimo, 1995). A fault can spread throughout the entire process if it goes undetected, or corrective measures are not taken timely after detection. It may result in degraded performance or complete system failure depending on the magnitude of the fault, time of detection and activation of the preventive steps. For example, a catastrophic accident was encountered due to a delayed detection in loss of well control at an oil rig in the Macondo field operated by the British Petroleum (BP) L.L.C. (Bea, 2011; Board, 2012).

Abnormal situation management (ASM) mainly comprises of three steps: (1) timely detection of the fault(s), (2) diagnosing the root cause(s), and (3) restoring the system in the normal operating condition. The third part varies from the industry to industry due to the diversified nature of the production processes. Therefore, researchers are mainly focusing on the first two parts of the ASM (Isermann, 2006). Symptoms are the observable effects of a fault. Process data as well as these symptoms contain noise. Moreover, most of the industries are operated under many close loops (Hoo et al., 2003). These make the detection of smaller magnitude faults extremely onerous. Although many FDD methods have been proposed over the years, most of the individual monitoring schemes suffer from accurate diagnostic capacity (Venkatasubramanian et al., 2003b). Combination of two or more methods (popularly known as the hybrid methods) have been proposed by the researchers as an alternative solution to the individual methods. These hybrid methods can provide apt answers to the questions whether a process is being operated in a safe mode or not, and if not where the fault has occurred (Gharahbagheri et al., 2017; Mallick and Imtiaz, 2013; Mylaraswamy and Venkatasubramanian, 1997; Vedam and Venkatasubramanian, 1999; Yu et al., 2015).

1.2. Objectives

The goal of this research is to develop process FDD tools which can detect the fault and diagnose the root cause of the fault precisely to reduce the human error. Prime focus is given on the consistency in accurate diagnosis. Principal component analysis (PCA), Bayesian network (BN) and dynamic Bayesian network (DBN) based FDD methodologies are proposed. The challenges of implementing the existing PCA-BN based method are discussed in Section 3.1. First part of this thesis proposes a comprehensive hybrid methodology which can solve these challenges. Finally, a DBN based scheme is proposed in the second part to facilitate the FDD in a single tool. The main objectives of this thesis are to:

- Detect the fault and diagnose the root cause of the fault without having any prior information of a specific fault type.
- Improve the limited diagnostic capacity PCA integrating with a knowledge based tool: BN to diagnose the root cause more precisely.
- Overcome the challenges faced by existing PCA-BN based methodology.
- Develop a DBN based FDD scheme which can detect the fault as well as diagnose the root cause accurately.

1.3. Thesis Structure

This thesis is a manuscript styled thesis which includes two submitted manuscripts. It is composed of five chapters. Chapter 1 briefly presents the definition of the fault detection and diagnosis as well as the consequences of a fault followed by the motivation and objectives of this research. Chapter 2 provides an extensive literature review on different

FDD methods, mentioning their merits and demerits. Emphasis has been given on the methods relevant to this thesis. In Chapter 3, a hybrid framework combining PCA, BN and multiple likelihood evidence is implemented to detect the process faults, diagnose the root cause of the fault, and identify the propagation pathway. This paper is submitted to the Chemical Engineering Science journal. A DBN based FDD scheme is proposed and validated in Chapter 4. This paper is submitted to the American Institute of Chemical Engineers (AIChE) journal. Finally, the outcomes of this thesis are summarized, and some future directions to improve this research are presented in Chapter 5.

1.4. Software Used

Since the applications of the proposed algorithms have been demonstrated on the simulated data, a well-known and available software, student version of MATLAB Simulink has been used in this thesis. It can be downloaded from https://my.mun.ca/student. All the necessary codes are written in MATLAB as well. GeNIe 2.0 has been used for modeling both the Bayesian network and dynamic Bayesian network for all the process models. It allows performing Bayesian inference for several algorithms. Academic version of this software is available to be downloaded at https://download.bayesfusion.com/files.html?category=Academia without any cost.

1.5. Authorship Statement

I am the primary author of this thesis and, also the two papers, on which this thesis is based on. With the help of the co-authors, Drs. Faisal Khan and Syed Imtiaz, I developed the conceptual understanding of the model requirement and its potential application. Subsequently, I developed detailed PCA-BN based hybrid and DBN based FDD models. I carried out the literature review. I have prepared the first draft of both manuscripts (presented in the thesis), subsequently revised the manuscripts based on the co-authors' feedback. The co-author, Dr. Syed Imtiaz helped in the development, testing and improvement of the models. He also assisted in reviewing and revising the manuscripts. The co-author, Dr. Faisal Khan helped in reviewing, analyzing, and testing the concepts/models, reviewed, and corrected the models and results, and contributed in preparing, reviewing, and revising the manuscript.

Chapter 2

Literature Review

Process FDD is one of the most prevalent areas of research. Over the years, many FDD tools have been proposed by the researchers. These methodologies have been classified into several categories from different perspectives. FDD methods in the existing literatures are broadly categorized into four main groups: (1) model based approaches, (2) process history or data based approaches, (3) knowledge based approaches, and (4) hybrid methods (Chiang et al., 2001; Venkatasubramanian et al., 2003b, 2003c). These approaches are briefly explained in the Sections 2.1, 2.2, 2.3 and 2.4.

2.1. Model based Approaches

Model based approaches mainly depend on the functional or analytical redundancy. Analytic redundancy is obtained from the residuals. These approaches require a residual generator and a residual evaluator. Residual generator provides the residuals by comparing the plant output with the model output, and residual evaluator makes the decision about the faulty or normal state of a process (Frank and Ding, 1997). These approaches highly rely on the process system models developed mainly from the first-principles models. The first-principles models are constructed using the laws of physics (e.g. mass, energy, and momentum balances). Residual is the discrepancy between the plant and model outputs. It can provide an indication to the fact that whether there is any fault or not (Chow and Willsky, 1984). In an ideal condition, the residual should be equal to zero. However, it is usually non-zero in the industrial processes due to noise and modelling uncertainty. Diagnostic observers, Kalman filters, parity relations, parameter estimation etc. are the mostly used tools in this category (Venkatasubramanian et al., 2003c).

Observers are used to estimate the unmeasured state variables. In observer based methods, fault is detected and diagnosed using the residuals generated from the estimated and measured outputs. However, these residuals contain the information of both faults and disturbances (Severson et al., 2016). (Frank, 1990) presented an observer based FDD scheme by decupling the effects of the faults from disturbances. The diagnosis task was performed by combining the analytical and knowledge based redundancy. (Frank and Ding, 1997) proposed a bank of observer based FDD scheme. The key concept was making one residual dedicated to a particular fault and insensitive to the other faults. Consider, four observers are designed to monitor a process that contains four different faults. If observer 1 is dedicated to fault 1, it will not be affected by other faults. In normal operating condition (NOC), all the residuals will be very small. When fault 1 occurs, observer 1 will show large deviation in residual while other observers will continue to provide small residuals, since these are insensitive to this fault. This technique is popularly known as the robust observer based fault diagnosis (Gertler, 2015). Unknown input observer (UIO) is another way of taking out the effect of the disturbances. (Sotomayor and Odloak, 2005) used a bank of UIOs to diagnose the faults in input, output and model parameters under model predictive control (MPC). (Zarei and Poshtan, 2010) proposed a methodology for designing non-linear UIO. Observer gain was estimated using the unscented transformation.

Kalman filter (KF) is the optimal state estimator/observer when process is linear, and noise follows a Gaussian distribution. Although KF based observers perform well in the steadily shifting faults, it may be difficult to identify the jump failures (Venkatasubramanian et al., 2003c). Classical KF do not provide optimal solution when the system is non-linear. Extended KF (EFK) and unscented KF (UKF) are the two derivatives of conventional KF to tackle the non-linearity. EKF linearizes the model using the Taylor series expansion, while UKF determines a set of sigma points and transform each of these points through the non-linear function to compute the Gaussian from the transformed and weighted points (LaViola, 2003). UKF provides the best performance compared to the KF and EKF when the system is highly non-linear (Wan and Van Der Merwe, 2000). Many observers have been developed in recent years to handle different issues such as disturbances, faults, optimum estimation, non-linearity etc. (Ali et al., 2015) presented an exhaustive review on different types of observers recently used in the chemical process systems.

Examining parity equation relation is one of the earliest and popular model based approaches. The key idea is to check the parity or consistency of the monitoring models with the measurements and known inputs, since any model based residual generator can be expressed in terms of parity relation (Gertler, 1991; Venkatasubramanian et al., 2003c). Parity vector is constructed from a linear combination of sensor outputs and applied inputs in the open loop operation, and the outputs are reconstructed from the measurements using the observers. Residuals need to be zero in absence of any fault and process uncertainty

(Patton and Chen, 1991). However, it is difficult to build an explicit model, and uncertainty always prevails in the measurements due to noise, which may make the residuals non-zero. (Willsky, 1976) first proposed the dynamic parity relations. (Chow and Willsky, 1984) developed a method to construct the parity equations from the state-space model. (Staroswiecki et al., 1993) improved the robustness of fault detection and isolation (FDI) using the optimal structured residuals generated from the parity relations. (Magni and Mouyon, 1994) presented a methodology which needed to estimate only those state vectors, required for residual generation using the parity space approach. Other notable works using the parity relation are presented by (Gertler, 1997; Odendaal and Jones, 2014; J. Wang et al., 2017; Zhong et al., 2015). More information can be found in (Patton and Chen, 1991). One of the main assumptions of this approach is that the model is linear. Hence, parity space approach becomes less suitable for monitoring batch processes since operating condition continuously changes, and often non-linearity exists in large scale production facilities.

The underlying assumption of the parameter estimation based FDD technique is that the process faults also affect the parameters. These parameters are not directly measurable. On-line measurements are obtained using the assumption that the process parameters have relationships with the state variables. (Isermann, 1982; Young, 1981) contemplated different parameter estimation techniques such as least-squares estimation, determination of time derivatives, instrumental variables parameter estimation, and parameter estimation via discrete-time models. (Isermann, 1985) proposed a FDD method using the parameter estimation technique. The threshold of the process parameters is determined in normal operating condition, and on-line measurements through the state variables are compared. Any violation from the threshold is considered as a fault. However, diagnosis becomes more strenuous due to the complex interaction of the process parameters and fault symptoms. (Isermann, 1997, 1993) incorporated expert opinion with the parameter estimation technique to improve the diagnosis capacity. A fault influences specific parameters and symptoms. These symptoms can be analyzed, and if-then rule can be applied to diagnose the fault. (Höfling and Pfeufer, 1994) proposed a methodology combining the parameter estimation and parity relation for optimal fault detection. (Che Mid and Dua, 2017) applied a parameter estimation based fault detection technique to a single stage evaporator and the four tank systems. This approach requires precise dynamic models which are difficult to obtain in large scale industrial processes. Diagnostic performance of the parametric estimation technique is complex and often misleading (Venkatasubramanian et al., 2003c).

The salient review papers in the field of model based FDD are presented by (Frank, 1996; Gao et al., 2015; Gertler, 2015, 1991; Himmelblau, 1978; Isermann, 2006, 1997; Isermann and Balle, 1997; Katipamula and Brambley, 2005; Simani et al., 2003; Venkatasubramanian et al., 2003c).

The biggest advantage of using the model based approaches is the elimination of costly hardware redundancy. However, the models need to be accurate for better FDD performance. Most of the model based FDD tools are constructed considering the process to be linear, which is often invalid in practical cases. Moreover, it is costly and time consuming to build a model. Another crucial fact is that the model may vary from plant to plant for same product depending on the process flow diagrams (PFDs). In batch processes, it is often difficult to obtain the details of the physical relationships among the process variables which results in modelling error and uncertainty. Thus, model based approaches become less suitable in large scale process monitoring (Venkatasubramanian et al., 2003c).

2.2. Data based Approaches

Data based methods exploits the correlation between the variables for detecting process fault and thus eliminates the need for explicit models. Data based methods train the monitoring scheme using the historical process data. Hence, these approaches are also known as the process history based methods. A large amount of historical data is transformed to construct the monitoring scheme for FDD. This transformation technique is known as the feature extraction. Data based methods can be divided into two categories depending on the extraction process: qualitative and quantitative methods. Qualitative methods include the expert systems and qualitative trend analysis (QTA), while quantitative methods include the statistical tools (e.g. Principal Component Analysis, Partial Least Squares), artificial neural network (ANN) and support vector machine (SVM) (Venkatasubramanian et al., 2003b).

Statistical methods contain both univariate and multivariate techniques. The univariate techniques include the Shewhart control chart, exponentially weighted moving average (EWMA) control chart and cumulative sum (CUSUM) control chart. The Shewhart chart developed by (Shewhart, 1930) is the most widely used univariate control chart. It is also known as the $\bar{x}$ chart. The upper control limit (UCL) and lower control limit (LCL) of the Shewhart control chart are calculated from the mean (μ) and standard deviation (σ). A popular choice of UCL and LCL are $\mu+3\sigma$ and $\mu-3\sigma$ respectively. When a sample exceeds any of these control limits, fault is detected. Unlike the Shewhart chart, both EWMA and CUSUM control charts have some memory, since they use the first order linear filter. Both these charts are computationally more expensive, and their performance is largely affected by the proper selection of the tuning parameter (Montgomery and Runger, 2010). The main advantage of the univariate monitoring techniques is the ease of implementation. However, these charts do not consider the change in operating condition while taking decision, resulting in unnecessary false alarms. Another drawback of the univariate monitoring is that an individual chart is required to monitor a variable which makes the monitoring more complex for the operators. Hence, only a few quality variables are possible to monitor using the univariate techniques (Kourti and MacGregor, 1995).

Multivariate statistical process monitoring (MSPM) tools removes some of the limitations of the univariate monitoring techniques using different statistics such as Hotelling's T^2, squared prediction error (SPE) and I^2 statistics. These tools can represent the process data in a lower dimensional space which essentially reduce the monitoring cost. Another robust feature of the MSPM tools is that they can capture the correlation among the process variables. As a result, false alarms can be minimized when operating condition changes. Principal component analysis (PCA), partial least square (PLS), independent component analysis (ICA) and Fisher discriminant analysis (FDA) are the notable MSPM tools.

PCA is a dimensionality reduction tool which has a wide range of application in the fields of medical science (Price et al., 2006), dynamic risk assessment (Adedigba et al., 2017), fault detection (Zadakbar et al., 2012), quality monitoring (Jackson, 2005) and so on. PCA became popular in the process industries after the works of (Kresta et al., 1991; Wise et al., 1988). Reduced dimension in PCA is obtained by rotating the axes. PCA projects the variables onto the principal component space and extracts a new set of variables. These extracted variables are called the principal components (PCs) (Dunia et al., 1996). PCs are linear combination of the original variables. PCs are orthogonal to each other and the first PC explains the most variance in the data, then the second PC, and so on. PCs can be obtained sequentially or through eigenvalue decomposition/singular value decomposition (SVD). SVD decomposes the covariance or correlation matrix to a set of vectors which has the same properties as PCs. Selection of appropriate number of PCs largely affects the monitoring performance. Details of selection procedure of number of PCs can be found in Section 3.2.1. Figure 2.1 shows the geometric representation of PCA.

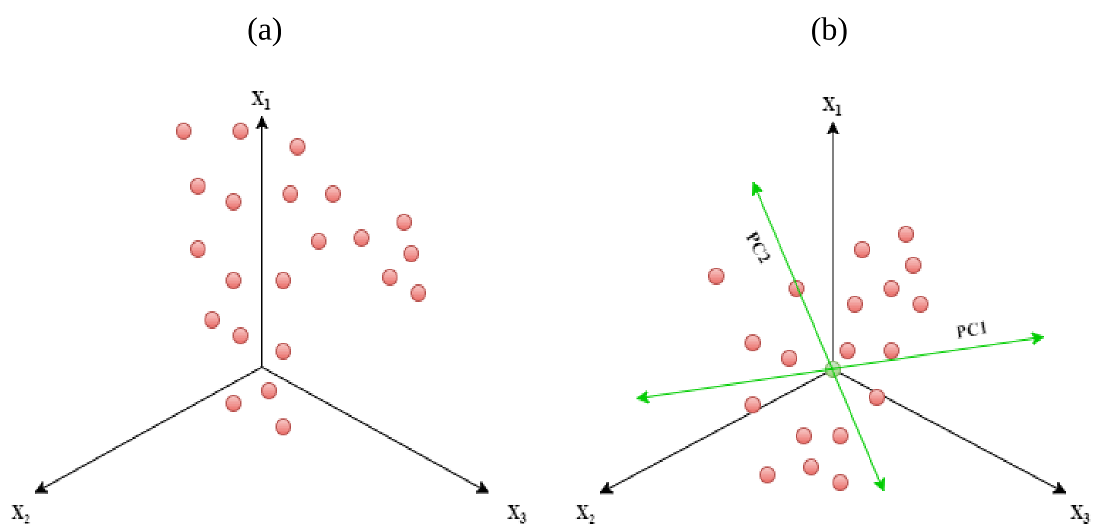


Figure 2.1: Geometric representation of PCA (a) raw data (b) axis rotation by PCA

Hotelling's T^2 and squared prediction error (SPE) are the two most common statistics associated with PCA. T^2 measures the distance between the sample space and center of the feature space. SPE indicates the lack of goodness of fit of sample data from the residual space. Whenever an on-line sample violates the threshold of T^2 or SPE, fault is detected. Another excellent feature of T^2 and SPE statistics is that they can provide the multivariate contribution plots which can help to identify the root cause of the fault. However, the variable with the highest contribution is not always the root cause of the fault which makes the diagnosis incomplete and complex (Joe Qin, 2003).

One of the limitations of PCA is that the PCs are time invariant. When process data are highly time dependent, classical PCA may provide higher false alarm rate or miss detection rate. (Ku et al., 1995) proposed dynamic PCA (DPCA) to address this issue. (Westerhuis et al., 1998) applied multi-block PCA to the processes that contain several stages. Multi-block PCA provides better diagnostic information than single PCA (Qin et al., 2001).

Process data often experience gradual drift which creates significant false alarms unless the monitoring model is updated with time. (Li et al., 2000) proposed a recursive PCA (RPCA) algorithm to recursively update the control limits and PCs for adaptive monitoring. Moving window PCA is another way to handle the gradual drift (Jeng, 2010). (Nomikos and MacGregor, 1994) successfully applied multi-way PCA to handle the multidimensionality of data in batch processes. Another extension of PCA to monitor batch processes is combining it with wavelet filtering, which is commonly known as the multi-scale PCA (Bakshi, 1998; Misra et al., 2002; Nielsen and Jensen, 2009; Zhang et al., 1999). It mainly works by removing the time-scales with slower variations.

PCA provides the optimal solution when process data follow a Gaussian distribution (Rhoads and Montgomery, 1996). ICA provides robust performance in a non-Gaussian feature using higher order statistics such as kurtosis and negentropy (Kano et al., 2003; Lee et al., 2004b). The major weakness of the conventional ICA is that it cannot distinguish the dominant independent components (ICs). Modified ICA (MICA) was proposed by (Lee et al., 2006; Zhang and Zhang, 2010) to tackle this issue by preserving the ranking of PCs in the PCA whitening step.

Linear PCA cannot provide the best performance in case of process non-linearity. (Kramer, 1991) used an auto-associative neural network based non-linear PCA (NLPCA) to monitor non-linear data. In recent days, kernel tricks are becoming popular to handle the non-linearity, and many researchers have proposed kernel PCA (KPCA) based FDD (Choi et al., 2005; Choi and Lee, 2004; Lee et al., 2004a). Selection of proper kernel function (e.g. Gaussian, polynomial, radial, and sigmoidal) along with the associated parameters is a huge challenge in constructing the KPCA model. It is also computationally expensive when sample data are large since the lower dimensional data are projected onto the higher dimensional space.

Both ANN and SVM have robustness in case of process non-linearity (Mahadevan and Shah, 2009; Sorsa and Koivo, 1993). Nevertheless, these methods are suitable when in-depth information about the process is available, since both normal and faulty data are needed to train the monitoring scheme to ensure accurate diagnosis.

(Chiang et al., 2001; Yin et al., 2012) compared the detection performance of different multivariate techniques using the benchmark Tennessee Eastman (TE) chemical process. (Ge et al., 2013) presented the merits and demerits of different data based FDD tools in their review paper. Other informative review articles on the process history based FDD tools include (Ding, 2014; Qin, 2012; Russell et al., 2012; Venkatasubramanian et al., 2003b; Yin et al., 2014; Yoon and MacGregor, 2001).

Data based FDD tools are very popular in the process industries due to ease of application. Although these tools provide quick detection performance, they lack in accurate diagnostic capacity. The multivariate contribution plots often fail to diagnose the root cause when the fault magnitude is very low. Another issue is more than one variable are shown as faulty due to the smearing effect, which eventually makes the diagnostic task more complex (Liu, 2012).

2.3. Knowledge based Approaches

Knowledge based tools reflect the human knowledge in terms of computer programs to improve the diagnostic task. Prior knowledge is utilized to build the model using if-then-else clauses (Venkatasubramanian et al., 2003a). Experts systems, case based reasoning (CBR), fault tree analysis (FTA), signed digraph (SDG), possible cause and effect graph (PCEG) and Bayesian network (BN) are the most popular knowledge based FDD tools.

Expert systems are the earliest knowledge based tools which attempted to reflect the expert knowledge in terms artificial intelligence (AI) using computer aided programs. The objectives of the expert systems are to: (1) infer process anomalies from the observed information, (2) detect potential hazards due to these malfunctions, and (3) recommend the remedies (Quantrille and Liu, 2012). MODEX2 is one of the primeval expert systems developed by (Venkatasubramanian and Rich, 1988) which used abductive reasoning for fault diagnosis. (Venkatasubramanian and Chan, 1989) developed an expert system to diagnose the faults in the fluid catalytic cracking unit (FCCU). This is well known as the

CATDEX. It uses simple AND-OR decomposition strategy to diagnose the root cause. FAX was built by (Chen and Modarres, 1992) using the hierarchical system decomposition technique. It utilized the Bayes' theorem (BT) as the inference algorithm. (Nan et al., 2008) introduced a computer aided tool which can integrate both expert knowledge and sensor data. Fuzzy logic was used for inference purpose. The advantages of expert systems lie on the facts that they are easier to develop and can provide viable solutions to take corrective actions. However, they lack in adaptability to new fault conditions.

CBR is another AI based approach which tries to solve the emerged problems using the knowledge of past. CBR has mainly three important properties: (1) store all the formerly occurred faults and the logical reasons behind these incidents, (2) robustly diagnose the cause of a fault which has already been encountered, and (3) try to present the reasons for the new faults based on the experience of past (Kolodner, 1992). Adaptability to new situations make it more user friendly (Lee, 2017). (Grant et al., 1996) applied CBR in fault diagnosis of industrial printers using CheckMate software and showed that CBR performs more efficiently than the rule based expert systems. (Olivier-Maget et al., 2009) used the outputs of the EKF as the input evidence in CBR to improve the diagnosis capacity of the FDD tool. (Zhao et al., 2017) proposed an improved CBR algorithm to reduce the computational time. Different case studies in the Tennessee Eastman (TE) chemical process were used to validate the developed model. More information on CBR can be found in (Aamodt and Plaza, 1994; Ashley, 2003; Watson and Marir, 1994). Although it has adaptability to new fault conditions, required domain expert knowledge to build the diagnostic model is essentially higher than the rule based expert systems. It also takes longer time to present the diagnostic result.

FTA is a robust tool in safety, risk and reliability analysis (Lee et al., 1985). It is a is a top-down deductive failure analysis technique, where a top event represents the failure of a variable due to failure of one or more variables, located at the bottom of the FT. These variables are causally related. Variable at the top of the FT is dependent on the other variables. FT assumes the bottom variables to be independent of each other (Vesely et al., 1981). Boolean logical 'AND' and 'OR' gates are used to define the relationships among the variables. Root cause can be diagnosed by analyzing the minimal cut sets, once a fault is detected. Although FTA is easy to be implemented, it is generic in nature and cannot show the interdependency among the variables. It is difficult to build an exact FT model for large scale processes due to complex interdependency among the variables. Thus, it is seldom used for process fault diagnosis.

SDG is the most popular knowledge based FDD tool. SDG reflects the causal relations among the process variables in the graphical form. Each node in an SDG represents a random process variable. The nodes take values of 0, + and -; implying steady state, higher than steady state and lower than steady state respectively. The arcs are directed from the cause nodes to the effect nodes. SDG can be derived from algebraic or differential equations. SDG was introduced in process fault diagnosis by (Iri et al., 1979). (Shiozaki et al., 1985) used conditional arcs in SDG to improve the diagnosis. (Yang et al., 2012) proposed an SDG modelling technique based on cross-correlation analysis of the process data and transfer entropy, and validated the constructed model using prior knowledge.

Different SDG modelling techniques can be found in (Maurya et al., 2003a, 2003b; Oyeleye and Kramer, 1988).

The PCEG analysis is an extension of SDG as it possesses certain characteristics of SDG (Venkatasubramanian et al., 2003a). The main difference between the SDG and PCEG is that the number of states at a node in the PCEG is unrestricted which provides more explicit information about the states of a variable. Hence, improved root cause diagnosis is achieved. (Wilcox and Himmelblau, 1994a) discussed the construction of the PCEG in detail, and its application to a counter flow heat exchanger and the Syschem plant is demonstrated in (Wilcox and Himmelblau, 1994b).

Although the aforementioned knowledge based tools can diagnose the root cause of the faults, uncertainty affects their performance. Since process measurements are extremely noisy and diagnosis is a process of reaching to a certain conclusion compiling several noisy uncertain evidences, the diagnostic tool needs to have robustness to uncertainty. Bayesian network (BN) is an emerging tool which can handle uncertainty. It is widely used in the fields of medical science (Friedman et al., 2000), safety, risk and reliability engineering (Abimbola et al., 2015; Musharraf et al., 2013), dependability and maintenance engineering (Weber et al., 2012). The power of BN has not been fully exploited in the area of process fault detection and diagnosis.

A BN is a probabilistic framework, which shows the complex interaction among the variables of a system in the pictorial view (Neapolitan, 2004). It provides an incisive representation of the joint probability distribution (JPD) of a set of random variables (Van Der Gaag, 1996). The difference between the classical and Bayesian probability is that classical probability does not put any weightage to the observation or evidence while Bayesian probability always contains certain degree of belief from the evidence (Heckerman, 1998). The most robust feature of a BN is that it can be constructed with limited data or even in absence of data integrating expert knowledge (Heckerman et al., 1995; Martin et al., 2012). A new evidence can be used to update the BN which helps in drawing certain conclusion (Madsen, 2008).

A BN has four structural components: nodes, acyclic arcs, prior and conditional probabilities (Lauritzen and Spiegelhalter, 1988). Nodes and arcs are the qualitative part of a BN. Prior and conditional probabilities enable a BN to perform the quantitative analysis (Bobbio et al., 2001). A node represents a random variable, while an arc indicates the nature of dependency between the connected nodes. An arc is generated from a parent node (cause) and directed to a child node (effect) (Yu et al., 2015). The arcs are irreversible. A parent node can have several child nodes and vice versa. Prior is the initial information about a node. Conditional probabilities determine the degree of influence among the variables as well as uncertainty (Lemmer and Kanal, 2014). Two types of conditional probabilities exist in a BN: likelihood probabilities and posterior probabilities. Likelihood probability reflects the probability of a child node given a prior of its parent node, and posterior probability is developed based on an evidence in a child node (mathematical explanation presented in Section 3.2.2).

A BN can be constructed using prior knowledge and process flow diagrams (PFDs) as well as available data. Learning a BN from data is known as NP-hard problem (Chickering et al., 2004). Many score-based learning algorithms are available for structural learning of a

BN from data such as K2 algorithm (Cooper and Herskovits, 1992), three phase dependency algorithm (TPDA) (Cheng et al., 1997), bootstrap approach (Cheng et al., 1997) and so on. Conditional probabilities can also be estimated from expert judgment and historical data. Maximum likelihood estimation (MLE) and Bayesian estimation (BE) are the most popular techniques for defining conditional probabilities from data (Grossman and Domingos, 2004; Kuhner, 2006).

(Mehranbod et al., 2003) proposed a BN based fault detection and identification method for three types of sensor faults: noise, bias and drift in steady operating condition. (Mehranbod et al., 2005) extended the methodology for transient operating condition. Fault detection was done by comparing the normal state probabilities with the updated operating state probabilities. Rule based methods were used to identify the fault type by analyzing the maximum probable states (MPSs). (Dey and Stori, 2005) utilized a BN to diagnose the root cause of variation of a machine tool. Multiple sensor data in operational mode were used as the evidence to update the network and identify the type of variation (e.g. stock size, workpiece hardness, tool wear by drilling and tool wear due to face milling). (Verron et al., 2008) proposed a Condensed Semi Naïve Bayesian Network (CSNBN) based fault diagnosis technique. The basic idea of a CSNBN is that a node can represent some variables, and these variables follow a multivariate distribution. Supervised classification strategy was used for diagnosis. (Azhdari and Mehranbod, 2010) demonstrated the application of a BN in detecting and diagnosing the faults in the TE chemical process. The detection procedure is analogues to the methods proposed by (Mehranbod et al., 2005, 2003). They used the pattern recognition technique to diagnose the fault. One of the interesting application of a BN was presented by (Gonzalez et al., 2015). They used the BN for fault diagnosis as well as dimensionality reduction. Knowledge of a fault was utilized to reduce the network size. (Atoui et al., 2016) proposed another CSNBN based fault diagnosis approach. They collected the structured residuals in an incident matrix which provided the evidence to the BN to diagnose the root cause. Besides these, many hybrid methods have been proposed where a BN is used for identifying the root cause of the fault. These methods are discussed in Section 2.4.

The BNs used in most of the current literatures are static. Static BN is not suitable for monitoring the dynamic processes, since it cannot capture the temporal relationships among the process variables. (Yu and Rashid, 2013) introduced a dynamic BN (DBN) based process monitoring technique. They used time variant JPD to detect the fault. Dynamic Bayesian contribution index (DBCI) was used to diagnose the root cause of the fault. The advantage of using a DBN is that it enables comprehensive FDD using a single tool as well as it can identify the fault propagation pathway. They used a duplicate dummy node to represent the recycling variables in the network, which enable the acyclic BN to represent recycling variables. (Zhang and Dong, 2014) integrated the output of a Gaussian mixture model (GMM) with a DBN to detect and diagnose the fault. They showed the methodology can handle some missing data. A review on the application of a BN in chemical industries can be found in (Zerrouki and Smadi, 2017).

2.4. Hybrid Methods

According to the comprehensive review by (Venkatasubramanian et al., 2003a, 2003b, 2003c), it is apparent that no individual method can capture all the expected aspects of

FDD. Each method has some advantages as well as some pitfalls. Hybrid methods have been proposed by many researchers to overcome these handicaps of the individual methods. These methods are constructed using two or more independent FDD methods and enable to capture the complete feature of FDD more strongly, integrating the strength of different individual methods (Das et al., 2012). For example, the MSPM tools can detect the fault early and diagnose the probable causes by generating the multivariate contribution plots. Reaching to a certain conclusion from these contributions is often tedious. Although knowledge based tools have poor detection capacity, they can be utilized to diagnose the root cause. The incomplete diagnosis report can be provided to the KB tools to complete the diagnosis. In this way, the MSPM tools are overcoming the limited detection capacity of the KB tools, and the KB tools are improving the diagnosis capacity of the MSPM tools. Thus, it is possible to ensure higher performance from both detection and diagnosis perspective which is not possible by any of these individual tools.

(Yu et al., 1996) proposed a hybrid method incorporating the parity equations and neural network. Parity equations generated residuals at the first stage and sent the incomplete diagnostic information to the neural network. A bank of neural nets was employed to diagnose the fault at the second stage. Signal to noise ratio of the residuals was used as the fault isolation index. (Mylaraswamy and Venkatasubramanian, 1997) combined the neural network with SDG to develop a hybrid FDD scheme for large scale industrial processes. They named the methodology as the DKit. It was successfully applied to the Amoco FCCU for 13 fault scenarios. In the DKit, a neural network was used to detect the fault, and SDG diagnosed the root cause. Combination of PCA and SDG was proposed by (Vedam and

Venkatasubramanian, 1999). SDG enabled diagnosis more precisely which was not possible by PCA. (Wang et al., 2012) developed a PCA and SymCure reasoning based hybrid FDD tool. SymCure is an expert system which can diagnose the fault. A lab scale distillation column was used to test the performance of this methodology. (Guo and Kang, 2015) proposed a hybrid methodology combining hazard and operability study (HAZOP), KPCA, wavelet neural network (WNN) and FTA. HAZOP was first used to identify the fault mode and process variables to be monitored. KPCA was used to extract the features from these variables. A bunch of WNN models were trained to detect the fault mode from on-line samples. Finally, FTA was used to identify the root cause of the fault. Although this method can provide comprehensive monitoring scheme, it is costly and time consuming to build the monitoring models. Furthermore, it is only applicable to the known faults stored in the trained models. (Jung et al., 2016) combined a model based and a data based tool to detect and diagnose the fault. Residuals generated from the parity equations was used to detect the fault. Fault information was provided to a one-class SVM where the variables were ranked based on the previously trained fault signatures. This method is suitable for diagnosing the known fault types only.

Most of the above-mentioned hybrid methods can improve the performance of individual FDD tools by capturing some of the features. Still, diagnosis is either complicated for SDG based methods or computationally expensive for neural network and SVM based methods, since detailed fault information is required. Using a BN for fault diagnosis can overcome these issues to some extents. (Mallick and Imtiaz, 2013) proposed a hybrid methodology comprised of PCA and BN. (Yu et al., 2015) developed an MICA and BN based framework which can capture the non-gaussian feature of process data. They showed that a BN can be used to diagnose the faults that originated from an unmonitored variable. (Wang et al., 2017) used a semi-parametric PCA and BN based methodology. Semi-parametric PCA enables capturing the non-linear, non-Gaussian and non-monotonic natures of the process data. They also demonstrated the capacity of a BN to diagnose the unmonitored root cause variable. (Gharahbagheri et al., 2017) presented a hybrid framework combining KPCA and BN. KPCA can handle non-linear process data using the kernel mapping function. Limited diagnostic information from KPCA was used to update the BN and diagnose the root cause of the fault. Granger causality and transfer entropy were used to construct the network, and prior knowledge was utilized to validate the developed BN structure. They used PCA residuals to define the CPTs. In all these works, a BN was used as a diagnostic tool in the second stage of the framework. These methods do not require any in-depth fault information. Hence, it is possible to diagnose the unknown faults. The variable which has the highest contribution in the multivariate contribution plot is considered as the fault symptom, and root cause is diagnosed among the other variables. A fault usually exhibits multiple symptoms and the contribution of the variables to the fault is uncertain in nature due to the smearing effect. One cannot guarantee that the selected variable to update the BN is the only fault symptom. Another issue is that the root cause is often accurately diagnosed by the MSPM tools. Using this evidence to update the BN will lead to false diagnosis. These issues are yet to be addressed in the BN based hybrid frameworks.

2.5. Conclusion

The following conclusions can be drawn from the above literature review of different FDD approaches:

i. No individual method can provide robust detection and diagnosis performance alone.

ii. MSPM tools can detect the fault early, and PCA is the most popular MSPM tool in the process industries. It does not require enormous data to build the monitoring model like ANN and SVM.

iii. A BN is capable of diagnosing the root cause, and diagnosis can be done in an unsupervised classification technique.

iv. Hybrid methods are becoming popular in the process industries, since they can provide a comprehensive solution.

v. Smearing effect and fault magnitude can affect the current BN based hybrid frameworks.

vi. Considering above facts, a hybrid method comprising of PCA, BN and multiple likelihood evidence is proposed. This method utilizes more diagnostic information from PCA while updating the BN and provides more comprehensive solution than existing hard evidence based diagnosing techniques.

vii. A DBN can represent the dynamic nature of a process, and it can detect and diagnose the fault. Considering this fact, a DBN based FDD methodology is also proposed which provides more comprehensive solution than the Shewhart control chart, PCA and static BN.

Chapter 3

Process System Fault Detection and Diagnosis using a Hybrid Technique

Abstract: This paper presents a hybrid methodology to detect and diagnose the faults in dynamic processes based on principal component analysis (PCA with T^2 statistics) and a Bayesian network (BN). It deals with the uncertainty generated by the multivariate contribution plots and improves the diagnostic capacity by updating the BN with multiple likelihood evidence. It can diagnose the root cause of the process fault precisely as well as identify the fault propagation pathway. This methodology has been applied to the continuous stirred tank heater and the Tennessee Eastman chemical process for twelve fault scenarios. The result shows that it provides better diagnostic performance over conventional principal component analysis with hard evidence based approaches.

Keywords: Process monitoring, hybrid methodology, principal component analysis, Bayesian network, likelihood evidence.

3.1. Introduction

Monitoring is important in modern process industries due to their complexity, increased safety requirements and product quality demands (Chiang et al., 2000; Dong et al., 2015). Abnormal situations often occur in the process industries, resulting in huge economic losses (Nimmo, 1995). An abnormal situation initiates with a fault during operation. A fault can be defined as the deviation of a process variable from an acceptable operational range (Venkatasubramanian et al., 2003c). Fault detection and diagnosis (FDD) is the first step in abnormal situation management (ASM) (Kresta et al., 1991). Data-driven multivariate statistical process monitoring (MSPM) techniques are widely used in process industries due to their effectiveness and ease of development. These techniques can extract features from highly correlated-high dimensional data to detect and diagnose the fault (Bakshi, 1998; Joe Qin, 2003; Kresta et al., 1991).

Principal component analysis (PCA) and partial least square (PLS) are the most widely used data-driven MSPM techniques, and are optimal for monitoring the process variables, following a multivariate Gaussian distribution (Kano et al., 2001; Rhoads and Montgomery, 1996). Independent component analysis (ICA) can capture a non-Gaussian feature by using higher order statistics like kurtosis and negentropy (Kano et al., 2003; Lee et al., 2004b). Modified ICA (MICA) can extract some dominant independent components (ICs), and improves the performance of conventional ICA (Lee et al., 2006; Zhang and Zhang, 2010). Kernel PCA (KPCA) has also been used to handle non-linearity (Cho et al., 2005; Lee et al., 2004a). All these tools need only a few historical data in normal operating condition (NOC) to estimate the control limit (CL), and information of faulty data behavior is not required for a successful performance. The artificial neural network (ANN) and support vector machine (SVM) have also been applied to process monitoring (Chiang et al., 2004; Mahadevan and Shah, 2009; Sorsa and Koivo, 1993; Weerasinghe et al., 1998).

These techniques require pre-classified training data of both normal and faulty samples, and are suitable where in-depth fault information is available.

Despite many advancements in the field of MSPM, diagnosis of the root cause of a fault is still a challenge. A BN is an emerging tool in FDD which is becoming popular due to its ability to incorporate process data with expert opinion, and it has many successful applications in root cause diagnosis. (Liu and Chen, 2009) proposed a Bayesian classification based PCA approach, which can successfully detect and isolate faults. (Weidl et al., 2005) applied an object-oriented BN (OOBN) to digester fiber line to diagnose the root cause. (Yu and Rashid, 2013) used a dynamic Bayesian network (DBN) based process monitoring approach for detecting the fault, diagnosing the root cause of the fault, and identifying the fault propagation pathway. They proposed the abnormality likelihood index (ALI) and dynamic Bayesian probability index (DBPI) to detect the fault. (Zhang and Dong, 2014) applied three time-slice DBN with a mixture of Gaussian output (3TDBN-MG) to handle some missing data and non-Gaussian process data.

Two or more techniques have been combined by many researchers to overcome the limitations of an individual method, which are popularly known as hybrid methods (Mylaraswamy and Venkatasubramanian, 1997; Venkatasubramanian et al., 2003b). Fault detection is performed at the first stage using MSPM tools (e.g. PCA, ICA etc.), and diagnosis is performed in the second stage by knowledge based tools (e.g. BN, causality analysis etc.) utilizing the evidence generated by the first stage detection tool in terms of the multivariate contribution plot. Thus, higher monitoring accuracy is achieved. (Mallick and Imtiaz, 2013) integrated the BN with PCA to improve its diagnosis capacity. (Yu et al.,

2015) used MICA and the BN to diagnose the root cause in an unmonitored variable. They determined the causal relationships among process variables and the conditional probability tables (CPTs) from prior knowledge and historical data. (Gharahbagheri et al., 2017) applied KPCA with a BN to capture the non-Gaussian feature of process data, and to diagnose the root cause of a fault. Transfer entropy and Granger causality were used to identify the causal relationships, and prior knowledge was utilized to validate the network. CPTs were estimated using maximum likelihood estimation (MLE). A BN is acyclic in nature. But, chemical processes often have recycling variables. (Yu and Rashid, 2013) and (Gharahbagheri et al., 2017) used duplicate dummy node to represent recycling variables in a BN.

In all the developed hybrid methods, first a data based method (e.g. PCA, KPCA etc.) was used for fault detection and diagnosis. Often, the diagnosis information is incomplete, and points to a group of variables as the probable cause. The current practice is to use heuristic rules to reduce the information to one faulty variable. Usually, the variable with the highest contribution is taken as the faulty variable. Accordingly, a 100% faulty state is assigned (commonly known as hard evidence) to the highest contributing variable from the multivariate contribution plot. In this approach, the contribution of other variables to the fault is ignored. This has several limitations. Firstly, the evidence generated from the multivariate contribution plots is uncertain in nature. The BN is updated considering the observations certain. Secondly, two or more variables can have very close contributions to the fault; selecting a particular variable as faulty is challenging in those cases. Finally, the root cause variable can be an intermediate node, and it can have highest contribution in the multivariate contribution plots. This means that the statistical tool already accurately does diagnosis. When a statistical tool diagnoses a root node as the root cause, a BN is not required. A BN is used to diagnose the root cause, when a child node has the highest contribution in the multivariate contribution plot. If hard evidence is used in this case, it causes diagnostic error. These limitations can be overcome by updating the BN with multiple uncertain evidence. However, this type of evidence has very limited use in fault diagnosis due to being complex in nature. Uncertain evidence in a BN has not yet been used in process fault diagnosis to authors' best knowledge. There are three main types of uncertain evidence- likelihood or virtual evidence, fixed probabilistic evidence and not-fixed probabilistic evidence (Mrad et al., 2015). In this work, we will focus only on the likelihood or virtual evidence.

The focus of this work is to improve the diagnostic capability of a BN based model using the information received from the multivariate contribution plots. A new hybrid methodology has been proposed integrating, principal component analysis (PCA) with T^2 statistics with the BN model. It is henceforth referred to as PCA-T^2-BN. This methodology considers multiple likelihood evidence to improve the diagnostic capacity of conventional PCA and PCA with BN (using a hard evidence based approach). The methodology has been applied to two benchmark processes – the continuous stirred tank heater (CSTH) and Tennessee Eastman (TE) chemical process. It is proven that the proposed PCA-T^2-BN methodology performs better than convectional PCA and its derivatives.

This paper is organized as follows: Section 3.2 presents preliminaries on PCA and BN. Section 3.3 will discuss the methodology in detail. Section 3.4 will show the application and suitability of the proposed methodology. Results and discussion will be summarized in Section 3.5. The contribution, advantages, and future work scope will be discussed in Section 3.6.

3.2. Preliminaries

3.2.1. Principal Component Analysis (PCA)

PCA is a dimensionality reduction technique, and is also used for process fault detection and diagnosis (Bakshi, 1998). PCA projects data from a high dimensional data space to a lower dimensional subspace, which preserves the maximum variation of the original space in reduced dimensions (Mallick and Imtiaz, 2013). It provides a new set of uncorrelated variables from a set of correlated variables using linear transformation. If a process contains m variable and n samples, the data matrix can be represented as $X \in \Re^{n \times m}$. The covariance matrix, R, is given by:

$$R = \mathrm{cov}(X) = \frac{X^T X}{n-1} \tag{3.1}$$

Singular value decomposition (SVD) of R is performed in such a way so that $R=V\Lambda V^T$. Λ is a diagonal matrix, which contains the eigenvalues in a descending order ($\lambda_1 > \lambda_2 > \ldots\ldots > \lambda_m$). V are the eigen vectors.

Selection of the number of principal components (PCs) can be done either by SCREE plot or the cumulative percent variance (CPV) approach (Jackson, 2005). SCREE is a graphical method, where eigenvalues are shown on X axis in descending order, and their correspondence variances are shown on Y axis. CPV is a simple approach. Equation 3.2 represents it:

$$\mathrm{CPV(b)} = \frac{\sum_{i=1}^{b} \lambda_i}{\mathrm{trace(R)}} \times 100\% \tag{3.2}$$

b is the selected number of PCs. Usually, b is selected when CPV(b)≥90%. The transformation matrix, P, is generated based on b. P contains m number of rows and b number of columns. The columns of P are called loadings. Scores are the values of the original space in a reduced feature space. Score, T:

$$\mathrm{T} = \mathrm{XP} \tag{3.3}$$

T can be transformed into original space using Equation 3.4:

$$\hat{\mathrm{X}} = \mathrm{TP}^{\mathrm{T}} \tag{3.4}$$

The residual matrix, E can be calculated as:

$$\mathrm{E} = \mathrm{X} - \hat{\mathrm{X}} \tag{3.5}$$

Two types of statistics: Hotelling's T^2 and squared prediction error (SPE) statistics are typically used in PCA based process monitoring. T^2 measures the correlated distance between the center of the feature space and projected data samples, while SPE measures the Euclidean distance between the PC feature space and the residual space (Yu et al., 2015).

For PCA:

$$\mathrm{T}_i^2 = \mathrm{t}_i \mathrm{t}_i^{\mathrm{T}} \tag{3.6}$$

where $t_i = x_i P \wedge_b^{-1/2} P^T$ is the contribution of the i^{th} monitored sample.

The control limit of Hotelling's T^2 can be calculated with a level of significance, α, using Equation 3.7:

$$T^2_{critical} = \frac{(n^2-1)b}{n(n-b)} \times F_\alpha(b, n-b) \tag{3.7}$$

where F_α (b, n-b) is the probability obtained from F distribution with (b, n-b) degrees of freedom with 1-α level of confidence.

The SPE or Q value is calculated as:

$$Q_i = e_i x_i^T \tag{3.8}$$

where $e_i = x_i(I - PP^T)$ is the contribution of the i^{th} monitored sample.

I is an m by m diagonal matrix with diagonal values as 1. The upper control limit of Q statistics with a level of significance, α, can be computed as:

$$Q_\alpha = \Theta_1 \left[\frac{h_o c_\alpha \sqrt{2\Theta_2}}{\Theta_1} + 1 + \frac{\Theta_2 h_o (h_o - 1)}{\Theta_1^2} \right]^{\frac{1}{h_o}} \tag{3.9}$$

where $\Theta_i = \sum_{j=b+1}^{m} \lambda_j^i$ and $h_o = 1 - \frac{2\Theta_1\Theta_3}{3\Theta^2}$

c_α is obtained from normal distribution for α level of confidence (Jackson and Mudholkar, 1979).

3.2.2. Bayesian Network (BN)

A BN is a strong tool for probabilistic reasoning, and helps to reach a certain conclusion using uncertain observations (Neapolitan, 2004). It is the representation of knowledge in graphical form. It provides the logical relations between variables through graphical representation in terms of conditional probabilities. Three terms are frequently used in a BN: prior, conditional, and posterior probability. ‘Prior’ reflects the initial information about a variable. ‘Conditional probability’ represents the mutual information shared among the variables. ‘Posterior’ is the degree of belief a variable contains based on an evidence. The priors can be renewed based on the collected information. This information is also known as evidence. Bayes’ theorem (BT), shown in Equation 3.10, is the heart of the network.

$$P(\theta/X) = \frac{P(X/\theta) \times P(\theta)}{P(X)} \tag{3.10}$$

where P(Θ) is the prior belief, and P(X) is the probability of an observation or evidence. P(X/Θ) is the conditional probability of X given Θ. It is also called the likelihood. P(Θ/X) is the conditional probability of Θ given X. It can also be referred to as posterior probability, since it contains the degree of belief relying on observation.

For a fixed value of X, Equation 3.10 can be written as,

$$P(\theta/X) \propto P(X/\theta) \times P(\theta) \tag{3.11}$$

So, posterior is proportional to the product of likelihood and prior probabilities. Dividing the right-hand side of Equation 3.11 by a normalizing constant will give the posterior, P(Θ/X). This normalizing constant depends on the achieved evidence. For a certain evidence of X:

$$P(\theta/X) = P(X/\theta) \times P(\theta) \tag{3.12}$$

Equation 3.12 is called the Bayesian belief updating equation (Gharahbagheri et al., 2017). A BN is a directed acyclic graph (DAG). It is actually a combination of nodes, arcs, prior and conditional probabilities. Each node represents a random variable, and arcs represent the causal relationships among the random variables. The direction of the arc depicts the dependency between the nodes. The node from which the arc is generated is called the parent node, and the node to which the arc is directed is called the child node. The nodes which do not have any child node are called leaf nodes, and the nodes which do not have any parent node are called the root nodes. An intermediate node acts as both a parent and child node in a BN (Y. Wang et al., 2017; Yu et al., 2015). The state of any node can be updated depending on the observed evidence, and decisions can be made based on renewed probability (Nielsen and Jensen, 2009). Figure 3.1 shows a simple BN of four nodes to illustrate the types of nodes in a BN.

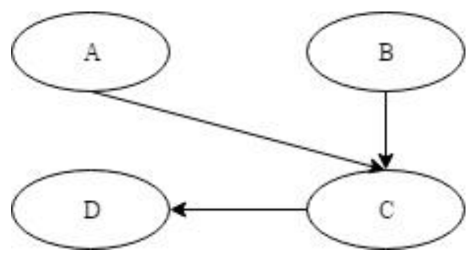


Figure 3.1: A simple BN

Both A and B are parent nodes of C. C is the parent node of D. D has no child node, while A and B have no parent node, so both A and B are root nodes, C is an intermediate node, and D is a leaf node. It is worth noting that no arch can come back from D to A or B, as it will make the network cyclic. In essence, a BN works in five simple steps in fault diagnosis by:

1. Collecting the information about the variables.
2. Determining the causal relationships among variables.
3. Estimating the prior and conditional probabilities from collected information.
4. Updating the belief based on the observed information.
5. Making decision from renewed prior beliefs.

3.3. PCA-BN with Multiple Likelihood Evidence

Figure 3.2 shows the flow diagram of the proposed methodology. This methodology is a combination of PCA-T^2-BN with multiple likelihood evidence. PCA is the primary detection tool and the observed information provider to the BN, which is the second stage root cause diagnosis tool. PCA has been used because of the simplicity in its application and reliable performance. Only a sufficient number of historical data is needed to determine the PCA loading vectors and threshold values of T^2 and SPE statistics. In this work, only T^2 based monitoring is shown; since the diagnostic performance of SPE is not consistent for the tested fault scenarios.

BN has been selected for its suitability in representing the relationships among process variables in pictorial view, and its ability to incorporate the expert opinion with in-depth process knowledge. The causal relationships of the BN have been determined based on prior knowledge, and process flow diagrams (PFDs). CPTs have been calculated based on maximum likelihood estimation (MLE) from the PCA residuals of faulty samples. (Gharahbagheri et al., 2017) showed that it is more suitable to estimate the CPTs from PCA residuals than from normal process data, since residuals contain causal variations, mitigating the variations due to noise and process abnormality. The residual data which fall outside one standard deviation around the mean value are considered to be faulty samples.

The off-line application of the methodology consists of four steps:

Step 1: A sufficient number of samples in NOC is collected. These samples are auto-standardized to zero median and unit variance.

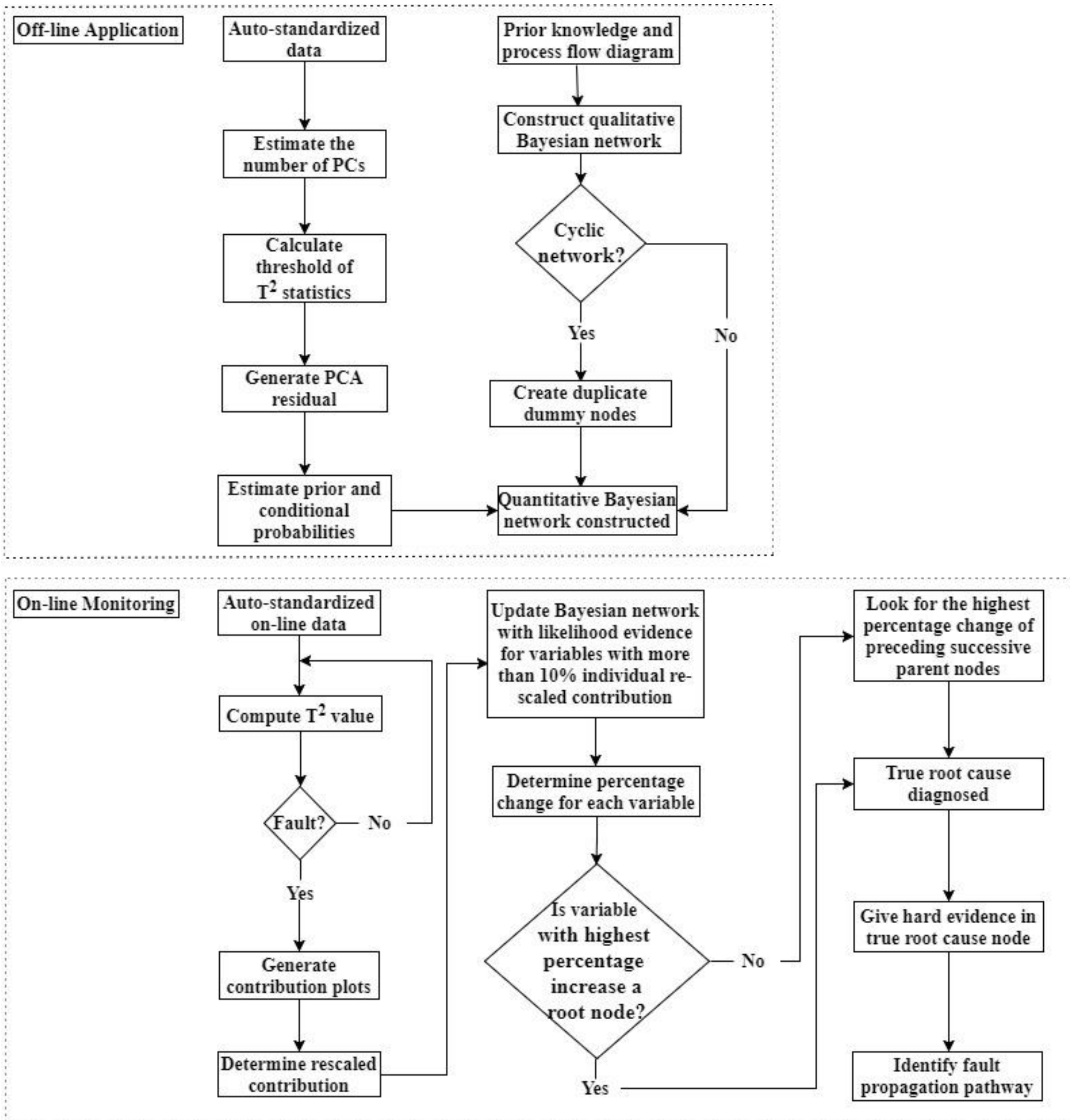


Figure 3.2: Proposed PCA-BN with multiple likelihood based methodology for FDD

Step 2: The required number of PCs are selected using Equation 3.2. The threshold value of the T^2 control chart is calculated from Equation 3.7.

Step 3: Prior knowledge and PFD are utilized to construct the qualitative BN. If the network is cyclic, the duplicate dummy nodes are created.

Step 4: CPTs are estimated from PCA residuals and the BN in NOC is constructed.

On-line monitoring is done in five steps:

Step 1: The T^2 value of each sample is computed using Equation 3.6 and compared with the threshold value. It should be noted that on-line samples have been auto-standardized based on the same median and standard deviation value as are calculated in the off-line mode.

Step 2: The T^2 contribution plot has been generated. These contributions are rescaled from 0-80% according to their relative weight. The highest contributing variable has been considered as having an 80% probability of being in a faulty state, and other variables have been assigned a probability of fault relative to the contribution of the highest contributing variable. Suppose A, B and C are three variables with 45%, 40% and 15% contribution to the fault, respectively, in the multivariate contribution plot. After rescaling, the relative contributions to the fault will be $P(A)_{fault}=0.80$, $P(B)_{fault}=0.71$ and $P(C)_{fault}=0.27$.

Step 3: The variables which have more than 10% contribution to the fault after rescaling have been used to update the BN. The advantages of steps 2 and 3 in updating the BN are:

- In a contribution plot, all variables may have individual contribution less than 50%. If this evidence is used to update the BN, it will show all variables have increased probability in the normal state; which is impractical, since the process is in the faulty state. Rescaling will allow the BN to show an approximately accurate picture of the state of process variables.
- Two variables may have contributions of 55% and 42%. If the BN is updated with this evidence, it will show a decision in favor of the variable with 55% contribution.

Rescaling will allow the BN to consider the variable with 42% contribution, since its relative contribution will be more than 61%.

- If variables with small a contribution have been used to update the BN, it will create diagnostic error.

Step 4: Percentage change in probability for all the variables is observed after updating the BN. The root cause is the variable; which has the highest percentage increase in the faulty state if it is a root node. If it is a child node, the search for the highest percentage change in the faulty state among its parent nodes is carried out to diagnose the root cause.

Step 5: After diagnosing the root cause, hard evidence (P(faulty state)=100%) is given to the corresponding node, and an increasing tendency of other variables towards the faulty state has been observed. The variables which have an increasing inclination to the faulty state, are included in the fault propagation pathway.

3.3.1. Pearl's Belief Propagation (BP) Algorithm

Belief propagation (BP) in a BN is known as BN inference or message passing. Several inference algorithms are used in BNs depending on the structure and type of the network. These algorithms can be divided into two types: exact inference and approximate inference. Exact inference includes the forward-backward algorithm (also known as Pearl's Belief Propagation algorithm), the variable elimination, and the junction tree algorithm. Approximate inference includes the loopy belief propagation, the factor frontier algorithm, the Rao-Blackwellized particle filtering, and the Boyen Koller algorithm (Łupińska-Dubicka, 2012; Murphy, 2002). Despite having several algorithms, sometimes it is difficult to infer in a BN if the network is too complex and a large number of states exist for each node. The forward-backward algorithm proposed by (Pearl, 1988) is the most widely used message passing algorithm for discrete BN. This algorithm has been used in this work, as it allows to incorporate both hard and likelihood evidence. According to this algorithm, the messages going from the parent to child nodes are called "π" message and the messages going from the child to parent nodes are called "λ" message. This algorithm assumes the BN as to be a singly connected DAG. Figure 3.3 shows how a message is passed from one node to another node in a singly connected network.

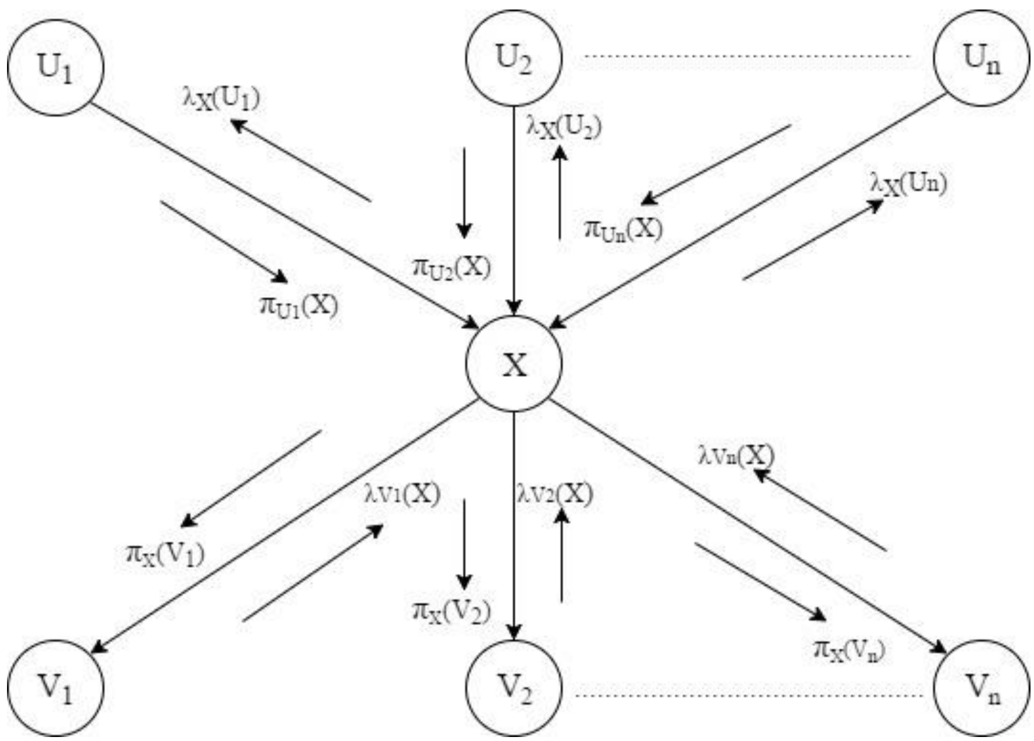


Figure 3.3: Message passing in a singly connected network

The prior probability of a node, π(x), equals to the product of the conditional probabilities of x given to the values of all possible combinations of its parents and the π message is passed down from its parents.

$$\pi(x) = \sum_{u} P\,(x/u) \prod_{k=1}^{n} \pi_x(u_k) \tag{3.13}$$

The likelihood of a node, $\lambda(x)$, equals the product of the entire λ messages from its child nodes (shown in Equation 3.14).

$$\lambda(x) = \prod_{j=1}^{n} \lambda_{v_j}(x) \tag{3.14}$$

The belief of any node, BEL(x), will be the product of likelihood and prior probability.

$$\mathrm{BEL}(x) = \alpha \left[\prod_{j=1}^{n} \lambda_{v_j}(x)\right]\left[\sum_{u} P\,(x/u) \prod_{k=1}^{n} \pi_x(u_k)\right] \tag{3.15}$$

When evidence is provided after receiving certain information about an event, λ and π messages are updated according to the Equations 3.16 and 3.17.

$$\lambda_x(u_i) = \beta \sum_{x} \lambda(x) \sum_{u_k\ k\neq i} P\,(x/u) \prod_{k\neq i} \pi_x(u_k) \tag{3.16}$$

$$\pi_{v_j}(x) = \alpha \prod_{k\neq j} \lambda_{Y_k}(x)\pi(x)$$

$$\text{So, } \pi_{v_j}(x) = \alpha\,[\mathrm{BEL}(x)/\lambda_{Y_j}(x)] \tag{3.17}$$

Therefore, the message going to a parent node, X, from a child node, Y, is equal to the belief of that node divided by the message sent by the child. α and β are normalized constants.

The boundary conditions for X will vary depending on the type of node. Table 3.1 shows the initial values of X for three distinct types of nodes: root node, child node and evidence node.

Table 3.1: Boundary condition setting criteria for different node types

Node Type	**π (Node)**	**λ (Node)**
Root Node	Prior Value	(1,1)
Child Node	BEL (Node)	(1,1)
Evidence Node	Prior Belief of Node	Evidence Belief of Node

3.3.2. Explanation of Pearl's BP Algorithm with Example

To understand the Pearl's BP algorithm, let us consider a simple BN, shown in Figure 3.4. A and B are the root nodes and parent nodes of D. A is the parent node of C.

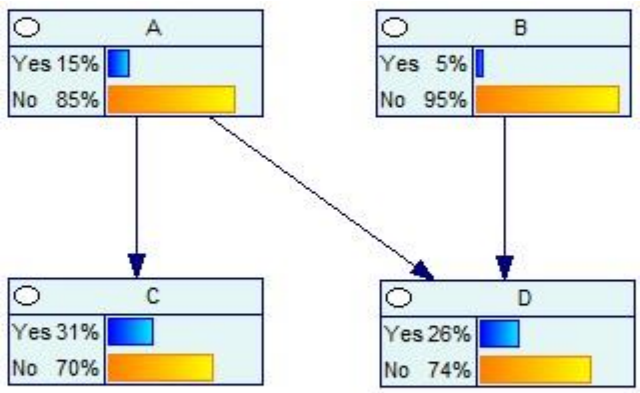


Figure 3.4: A simple BN of four nodes

The priors are P(A) = 0.15 and P(B)= 0.05. The conditional probability of C and D are shown in Table 3.2 and 3.3 respectively. The initial boundary condition is shown in Table 3.4. The π messages of C and D can be calculated using Equation 3.13 from the prior values of A and B.

π(C) = [(P(C=Yes/A=Yes)*P(A=Yes) + P(C=Yes/A=No)*P(A=No)),

(P(C=No/A=No)*P(A=No) + P(C=No/A=Yes)*P(A=Yes))]

= (0.90*0.15 + 0.20*0.85, 0.80*0.85 + 0.10*0.15)

= (0.305,0.695)

Table 3.2: CPT for node C

A		Yes	No
C	Yes	0.90	0.20
	No	0.10	0.80

Table 3.3: CPT for node D

A		Yes		No	
B		Yes	No	Yes	No
D	Yes	0.98	0.95	0.85	0.10
	No	0.02	0.05	0.15	0.90

According to Equation 3.15, BEL(C) = (0.305,0.695)

Similarly, π(D) = [(P (D=Yes/A=Yes,B=Yes)*P (A=Yes)*P(B=Yes) + P (D=Yes/A=Yes, B=No)*P (A=Yes)*P(B=No)+P (D=Yes/A=No,B=Yes)*P(A=No)*P(B=Yes)+ P(D=Yes/ A=No,B=No)*P(A=No)*P(B=No)), (P(D=No/A=No, B=No)*P(A=No)*P(B=No) + P(D=No/A=Yes, B=No)* P(A=Yes) * P(B=No) + P(D=No/A=No,B=Yes)*P(A=No)* P(B=Yes) + P(D=No/A=Yes, B=Yes)*P(A=Yes)*P(B= Yes))]

= (0.98*0.15*0.05+0.95*0.15*0.95+0.85*0.85*0.05+0.10*0.85*0.95,

0.90*0.85*0.95+0.05*0.15*0.95+ 0.15*0.85*0.05+0.02*0.15*0.05)

= (0.26,0.74)

And, BEL(D)= (0.26,0.74)

Table 3.4: Initial boundary condition for the BN of Figure 3.4

Node	**BEL (X)**	**π(X)**	**λ(X)**
A	(0.15,0.85)	(0.15,0.85)	(1,1)
B	(0.05,0.95)	(0.05,0.95)	(1,1)
C			(1,1)
D			(1,1)

GeNIe 2.0 has been used for simulation based on stated data. Figure 3.4 supports the result obtained in Table 3.5.

If any certain evidence shows that D has occurred, λ(D)= (1,0). This evidence will propagate throughout the network and update the node states. λ messages of A and B can be calculated using Equation 3.16.

Table 3.5 shows the updated condition of the BN after belief propagation throughout the network.

Table 3.5: Updated node condition after belief propagation

Node	**BEL (X)**	**π(X)**	**λ(X)**
A	(0.15,0.85)	(0.15,0.85)	(1,1)
B	(0.05,0.95)	(0.05,0.95)	(1,1)
C	(0.305,0.695)	(0.305,0.695)	(1,1)
D	(0.26,0.74)	(0.26,0.74)	(1,1)

λ(A) = [(e=1*P(D=Yes/P(A=Yes,B=Yes) * P(B=Yes)+ e=1 * P(D=Yes/P(A=Yes,B=No)* P(B=No) + e=0 * P(D=No/P(A=Yes,B=Yes)*P(B=Yes)+ e=0*P(D=No/P(A=Yes,B=No)* P(B=No)), (e=1*P(D=Yes/P(A=No,B=Yes)*P(B=Yes)+e=1*P(D=Yes/P(A=No,B=No)* P(B=No)+e=0 * P(D=No/P(A=No,B=Yes) * P(B=Yes)+e=0 * P(D=No/P(A=No,B=No) * P(B=No))]

= (1*0.98*0.05+1*0.95*0.95+0*0.02*0.05+0*0.05*0.95,

1*0.85*0.05+1*0.10*0.95+0*0.15*0.05 +0*0.90*0.95)

= (0.952,0.138)

A has been marginalized considering D has occurred due to A and P(A=Yes)=1.

Similarly, λ(B) = (0.87,0.228)

Table 3.6: Updated node condition after providing evidence

Node	**BEL (X)**	**π(X)**	**λ(X)**
A	(0.549,0.451)	(0.15,0.85)	(0.952,0.138)
B	(0.167,0.833)	(0.05,0.95)	(0.87,0.228)
C	(0.585,0.415)	(0.585,0.415)	(1,1)
D	(1,0)	(0.26,0.74)	(1,0)

π messages of A and B will not change. However, the beliefs of A and B will be changed, and can be calculated using Equation 3.15.

BEL(A) = (0.15*0.952/0.26,1- 0.15*0.952/0.26) = (0.549,0.451)

And, BEL(B) = (0.05*0.87/0.26,1- 0.05*0.87/0.26) = (0.167,0.833)

The new π message of C can be calculated using Equation 3.13. The π message has changed due to the change in belief of node A.

π (C) = (0.90*0.55+0.20*0.45,0.10*0.55+0.80*0.45)

= (0.585,0.415)

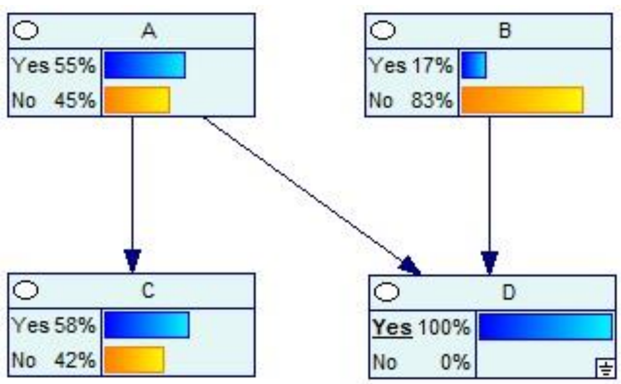


Figure 3.5: Node status after giving evidence to node D and belief propagation

Table 3.6 shows the updated status of nodes after giving evidence and the propagation of the message. Figure 3.5 shows the updated BN after giving evidence and belief propagation, which supports the beliefs of different nodes, stated in Table 3.6.

3.3.3. Hard Evidence vs Uncertain Evidence

When certain information has been obtained about an observation, usually the corresponding node is updated with a 100% true value for that state. This type of evidence is called hard evidence. However, often 100% certainty is not achievable from an observation, especially when another statistical tool has been assigned to determine the contribution of each variable to an incident. In these cases, the reliability of the obtained evidence is doubtful. This type of evidence is called uncertain evidence. A BN can be updated with multiple uncertain evidence to reach a certain conclusion. This type of evidence can play a significant role in root cause diagnosis.

The dissolution tank model described in (Mallick and Imtiaz, 2013) has been considered to illustrate how updating a BN with multiple likelihood evidence can improve diagnosis. The dissolution tank model has four variables: water flow, solid flow, water level, and density. Both water and solid flow have an impact on the state variables- water level and density. Hence, a fault in water flow or solid flow will affect these two state variables significantly. For the sake of simplicity, the same prior failure probability has been assumed for both water and solid flow. The CPTs for water level and density are shown in Table 3.7 and 3.8 respectively.

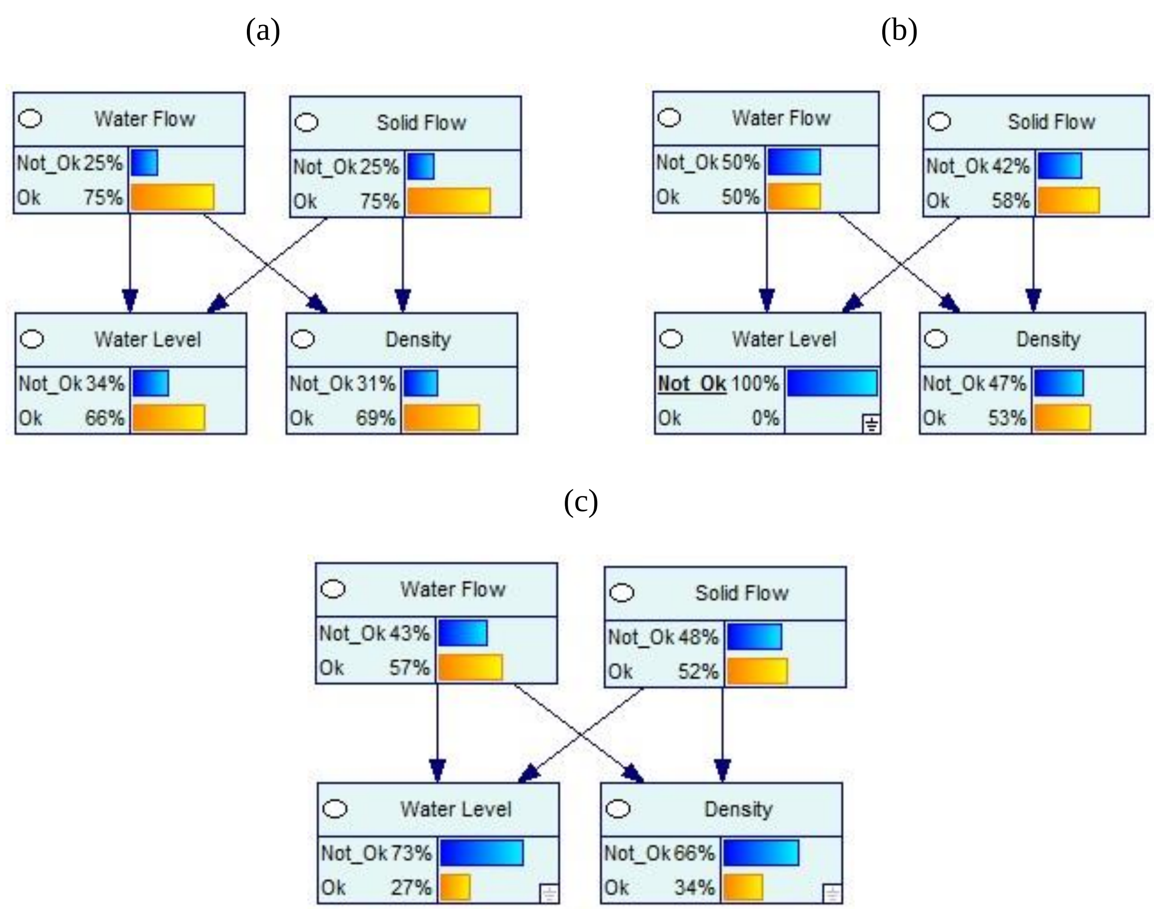
Table 3.7: CPT for water level

Water Flow		Not_Ok		Ok	
Solid Flow		Not_Ok	Ok	Not_Ok	Ok
Water Level	Not_Ok	0.90	0.60	0.45	0.15
	Ok	0.10	0.40	0.55	0.85

Table 3.8: CPT for density

Water Flow		Not_Ok		Ok	
Solid Flow		Not_Ok	Ok	Not_Ok	Ok
Water Level	Not_Ok	0.92	0.35	0.70	0.10
	Ok	0.08	0.65	0.30	0.90

Consider that a fault has occurred in solid flow and that the T^2contributions of water level and density to the fault have been obtained as 48% and 45% respectively. According to the rescaling technique, the water level and density will possess 80% and 75% rescaled contributions respectively.



(d) (e)

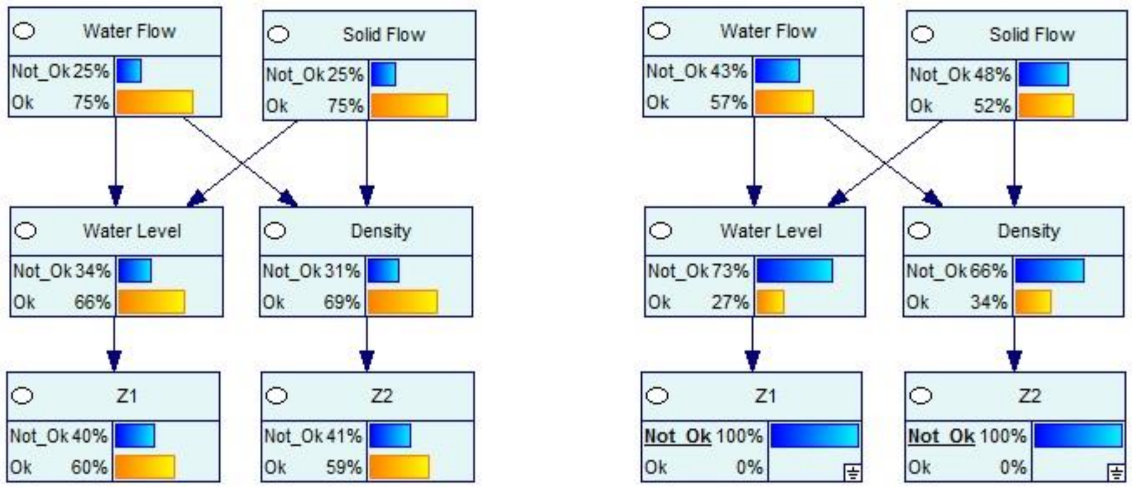


Figure 3.6: Comparison of performance of likelihood and hard evidence in dissolution tank, and working principle of likelihood evidence (a) BN in NOC (b) updated BN with hard evidence (c) updated BN with two likelihood evidence (d) creation of likelihood nodes Z1 and Z2 (e) providing hard evidence to Z1 and Z2

For root cause diagnosis, hard evidence will be set to the Not_Ok state of the water level node in the conventional approach, and water flow will be falsely diagnosed as the root cause (Figure 3.6(b)). However, Figure 3.6(c) shows that if the BN is updated with two likelihood evidence, density is accurately identified as the root cause of the process abnormality. Hence, it can be inferred that multiple likelihood evidence can improve the diagnostic capacity of a BN depending on the observed information and degree of dependency among variables.

Figure 3.6(d) and 3.6(e) explain how likelihood evidence works in a BN. First, two likelihood nodes, Z1 and Z2 are created, which possess the likelihood ratio of the observed information from the T^2 contribution plot for water level and density respectively (Pearl, 1988; Peng et al., 2010). Sufficient information indicates that density has 75% probability to be in the Not_Ok state. This means the likelihood ratio for density in the Not_Ok to Ok state is 3:1. This is reflected in the CPT of Z2. Then, hard evidence is given in the Not_Ok state of Z2 to compute the updated probabilities of other nodes. The likelihood node Z1 is created using a likelihood ratio of 4:1 for water level in the Not_Ok to Ok state, and the Not_Ok state is updated with hard evidence. Updated probabilities in Figure 3.6(e) reflect node status in Figure 3.6(c). Table 3.9 shows CPT for Z2.

Table 3.9: CPT for likelihood node Z2

Density		Yes	No
Z2	Yes	0.75	0.25
	No	0.25	0.75

3.4. Applications of Proposed Methodology

The proposed methodology has been applied to two benchmark processes, the continuous stirred tank heater (CSTH) in 2 fault scenarios and the Tennessee Eastman (TE) chemical process in 10 fault scenarios. Among the 10 fault scenarios of the TE chemical process, two fault scenarios will be discussed in detail. Sampling frequency is 1 second for both the process models.

3.4.1. Continuous Stirred Tank Heater (CSTH)

The continuous stirred tank heater (CSTH) is a common unit in the chemical processes. (Thornhill et al., 2008) developed a CSTH simulator based on the first principle model, which possesses real disturbances data from the pilot plant, at the University of Alberta. In a CSTH (shown in Figure 3.7), both cold and hot water are supplied to the continuously stirred tank and further heated by a steam heated coil. The output measurements are the water level in the tank, water flow rate and temperature. All measurements are presented as electric signals on a scale of 4-20 mA. Water level in the tank, water flow rate and temperature are also measured in cm, m^3/sec and °C respectively. This model is highly non-linear due to the presence of the heating coil at the bottom of the tank.

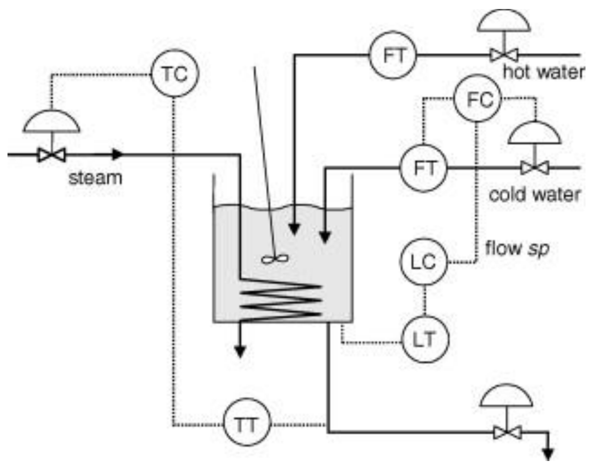


Figure 3.7: The continuous stirred tank heater

There are two standard linearized operating points. Hot water flow does not prevail at standard operating point 1. In both operating points, hot and cold water temperatures are kept constant at 50°C and 24°C respectively. Two fault scenarios have been generated in standard operating point 2 (Table 3.10) to test the proposed methodology.

Table 3.10: Fault descriptions in a CSTH

Fault scenario no	Fault description	Root cause
1	Leak in the tank	Level
2	Steam valve stiction	Steam valve

For determining the threshold value of T^2 statistics using PCA and construction of the BN, 500 fault free samples have been collected. 3 PCs can explain more than 90% variations.

Threshold for Hotelling's T^2 is calculated as 7.90 with 95% level of confidence. Figure 3.8 shows the BN in normal operating condition. Diagnosing true root cause in standard operating point 2 is challenging due to multiple parent nodes for level and temperature.

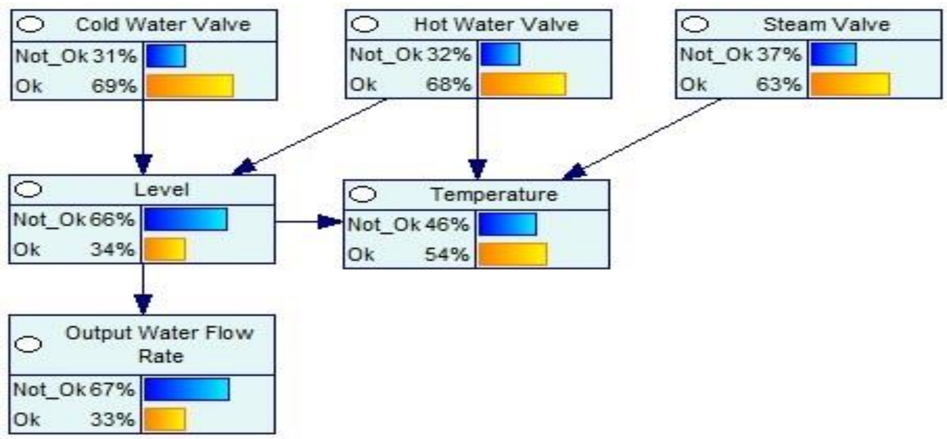


Figure 3.8: BN for CSTH in standard operating point 2

3.4.1.1. Fault scenario 1 (leak in the tank)

The tank can have a leak, especially if it is old. After operating for 500 samples in normal operating condition, a small leak occurs at the bottom of the tank, and a small amount of water goes out of the tank, resulting in level loss. Since the system is operated in a closed loop system, the controller can identify the level reduction, and increases the cold water valve demand to make up for the water loss to maintain the set point of level. This fault does not disturb the outputs (level, flow rate, and temperature) to a greater extent, but will increase operational cost. Therefore, this type of abrupt faults needs to be identified.

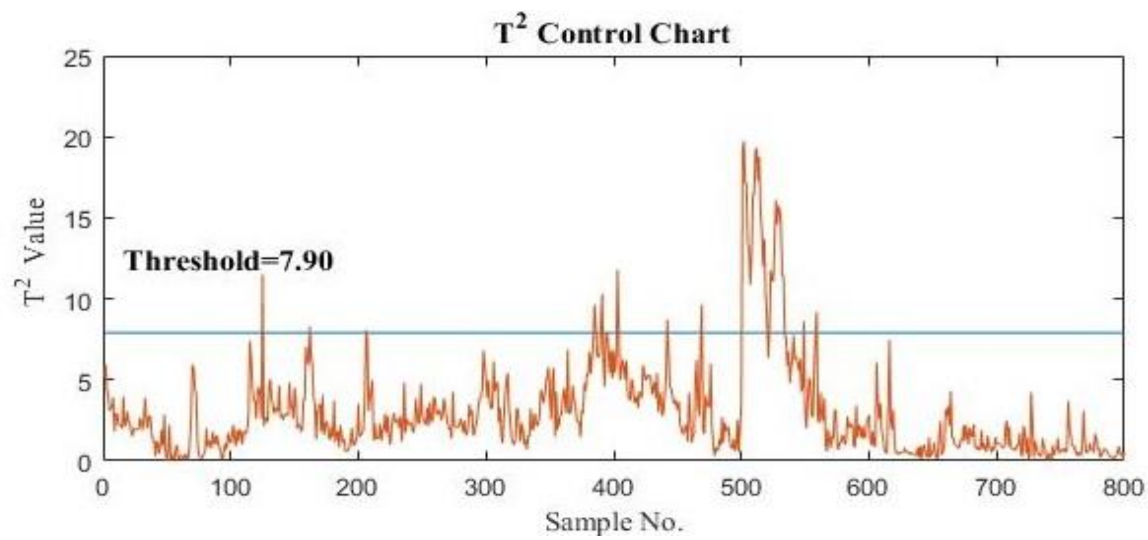


Figure 3.9: T^2 control chart for fault scenario 1

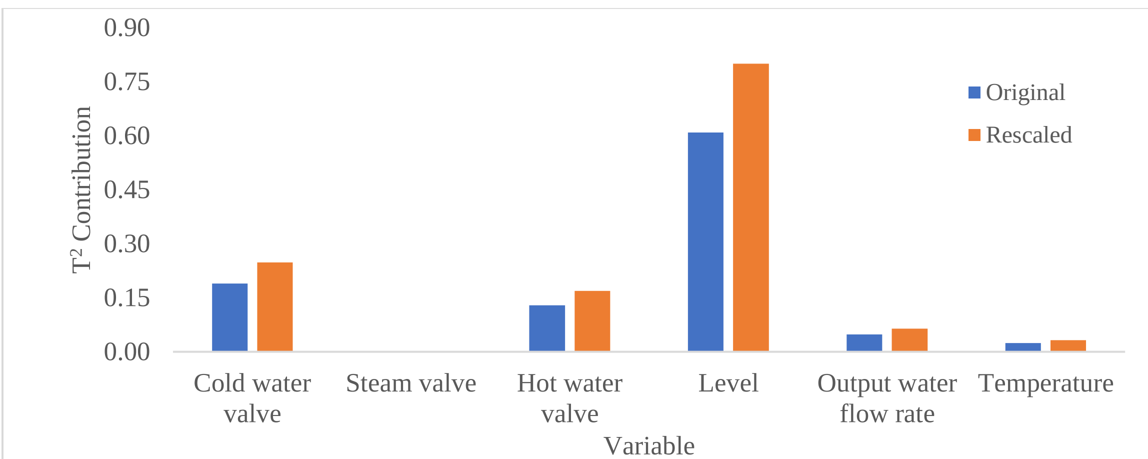
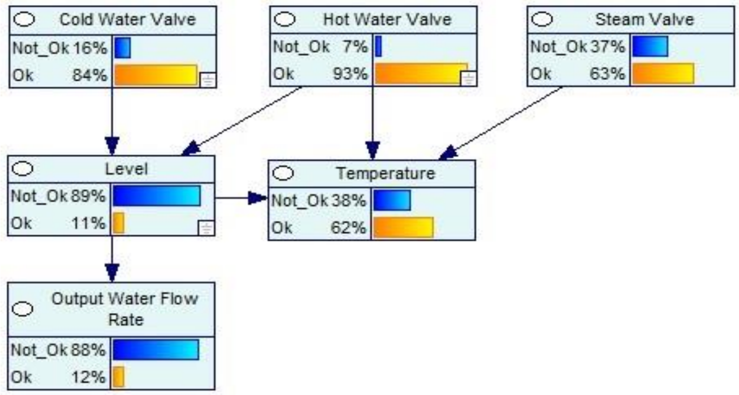


Figure 3.10: T^2 contribution plot for fault scenario 1

The T^2 control chart can identify this fault very quickly. Since it is a very small leak, the controller takes corrective actions and compensates for the leak. After a brief violation of the threshold, T^2 values return to being within the control limit. Figure 3.10 shows the original and rescaled T^2 contribution plot. The left and right-side bar for each variable represent the original and rescaled T^2 contributions respectively. Level is identified as being the variable contributing the most to the abnormality in Figure 3.10.

(a)



(b)

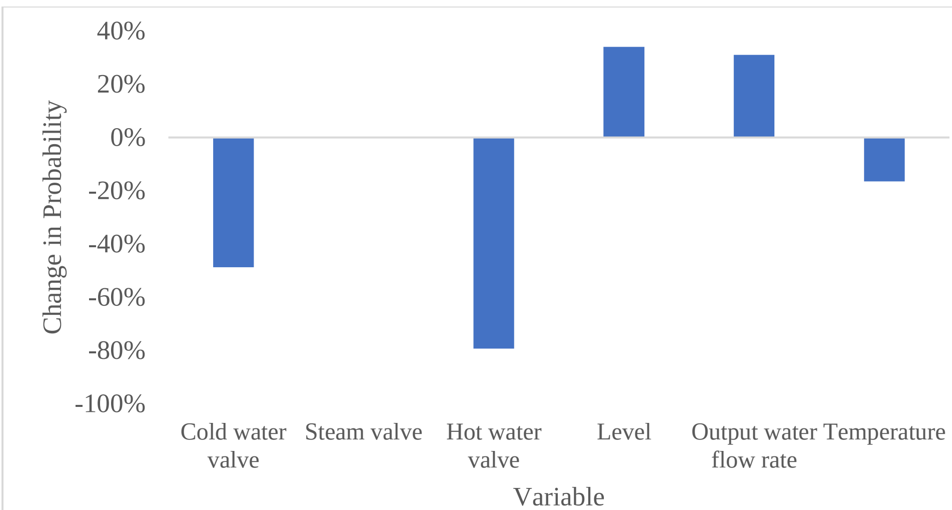
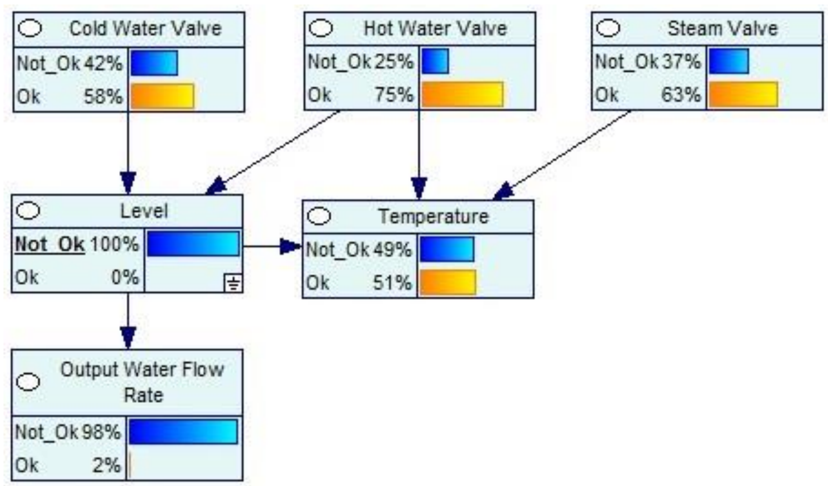


Figure 3.11: Root cause diagnosis by proposed methodology for fault scenario 1 (a) updated BN with rescaled multiple likelihood evidence (b) percentage change in probability

Level, cold water valve and hot water valve have more than a 10% rescaled contribution, and are selected to update the BN. The updated BN is shown in Figure 3.11(a). Figure 3.11(b) shows the percentage change in probability of all the variables. Level has the highest percentage increase (34.09%) in the faulty state in the updated BN. Level has two parent nodes, cold and hot water valve. Both these nodes have an increased proclivity in the normal state, which confirms level as the root cause of the abnormality. It should be noted that any increase towards a negative direction in the change in probability graph indicates the increasing tendency towards the normal state for a variable.

(a)



(b)

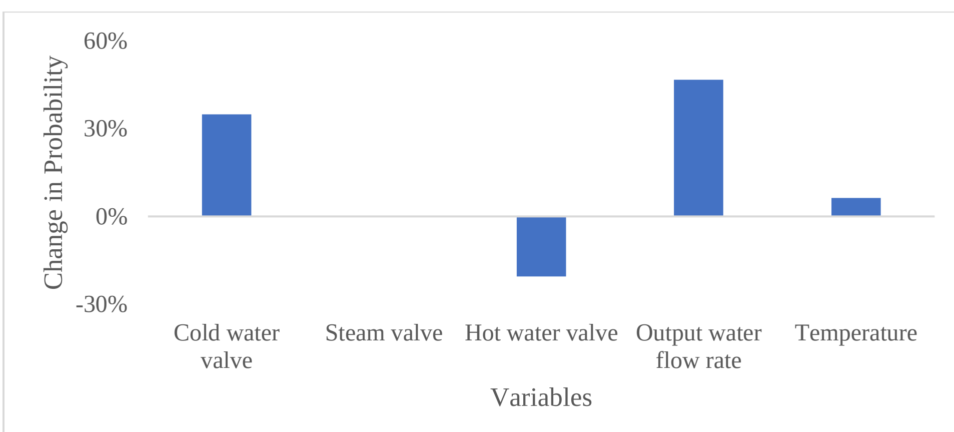


Figure 3.12: Root cause diagnosis by conventional approach for fault scenario 1 (a) updated BN with hard evidence (b) percentage change in probability

In the conventional approach, hard evidence will be given in the faulty state of level node. Figure 3.12(a) shows the updated BN with hard evidence. In the updated BN, cold water valve has the highest percentage increase (35.40%), and is one of the parent nodes of level (Figure 3.12(b)). Hence, cold water valve will be falsely diagnosed as the root cause.

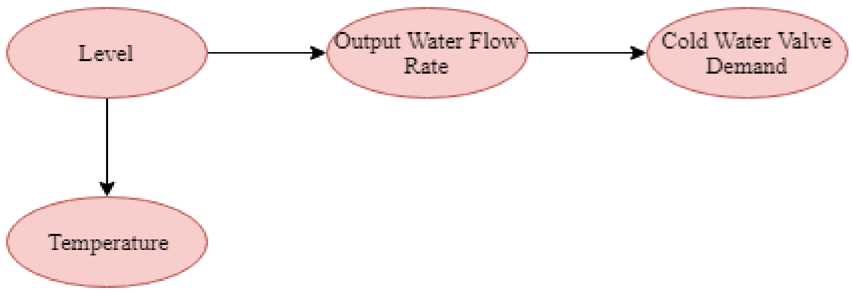


Figure 3.13: Fault propagation pathway for fault scenario 1

Next, hard evidence is given in the Not_Ok state of level node to check the fault propagation pathway. The variables which increase in the faulty state, lie inside the fault propagation pathway. Level will upset both temperature and output water flow rate. Temperature will have a minimal effect. However, cold water flow will change significantly because of the control action of the level loop to indemnify the leak. These phenomena are accurately captured by the fault propagation pathway shown in Figure 3.13.

3.4.1.2. Fault scenario 2 (steam valve stiction)

Valve stiction is a very common operational problem in process industries. In this fault scenario, steam valve gets stuck from 501 samples. As a result, temperature decreases due to reduced steam supply. The T^2 control chart quickly detects the fault. Controller tries to maintain the set point, and after a while the T^2 value goes below the threshold value. The reason may be that the linear correlation between variables remain, which does not significantly shift the process mean (Zeng, 2016).

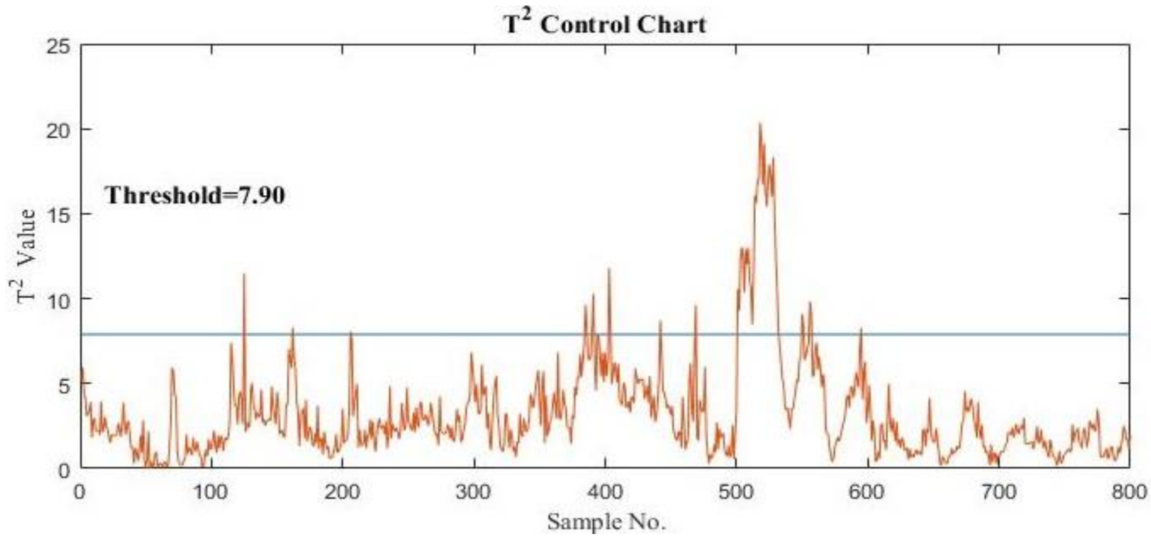


Figure 3.14: T^2 control chart for fault scenario 2

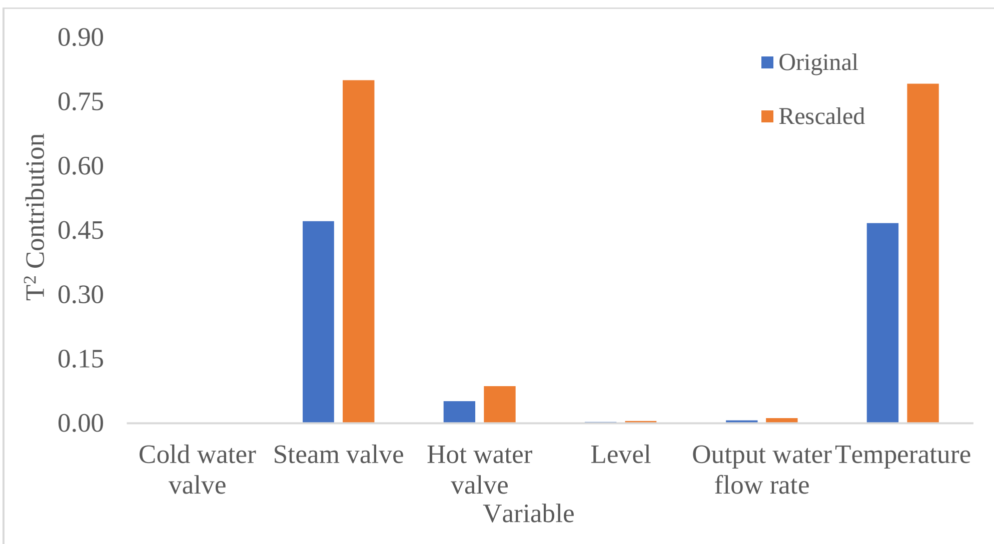
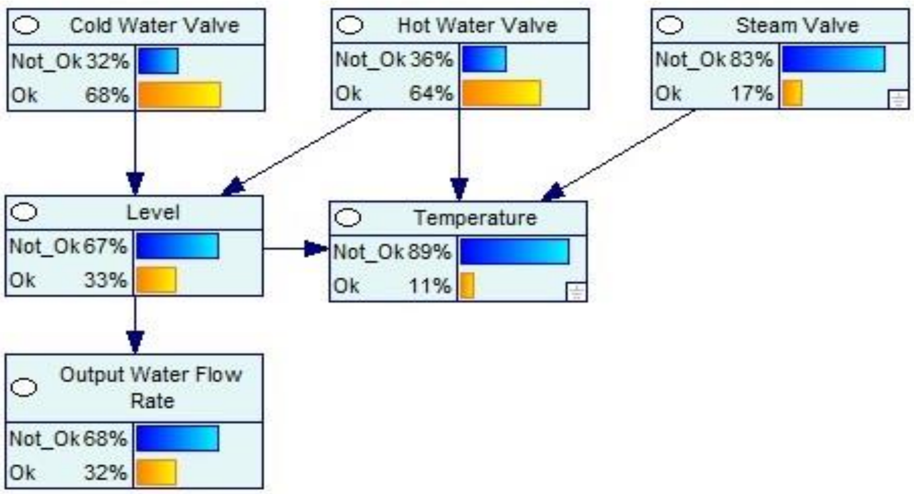


Figure 3.15: T^2 contribution plot for fault scenario 2

The next step is to generate the contribution plot and rescale the contributions (Figure 3.15). The rescaled likelihood evidence of steam valve and temperature are used to update the BN. The updated BN is shown in Figure 3.16(a). The increase in the faulty state for steam valve, temperature, hot water, cold water valve, level and output water flow rate are 126.22%, 93.43%, 13.54%, 2.83%, 1.43% and 1.30% respectively. All the root nodes (cold water valve, steam valve, hot water valve) in the BN have an increasing tendency in the faulty state. Since steam valve has the highest percentage increase in the faulty state, it can be easily identified as the true root cause of the abnormality.

(a)



(b)

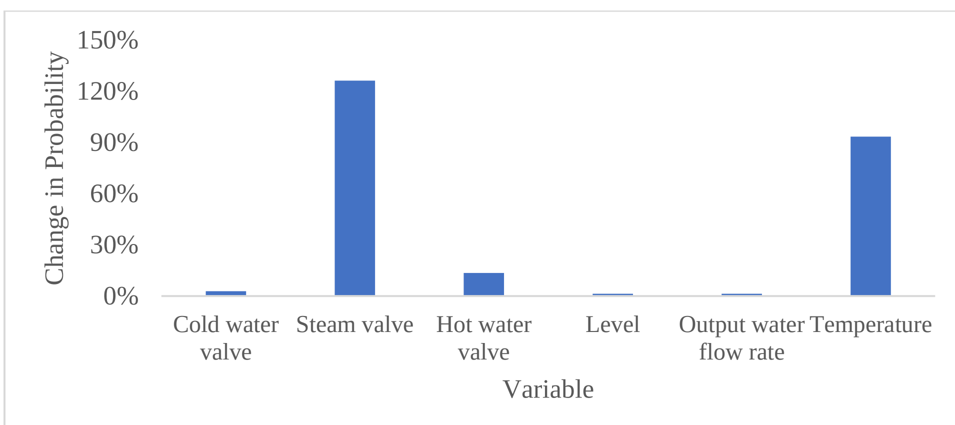
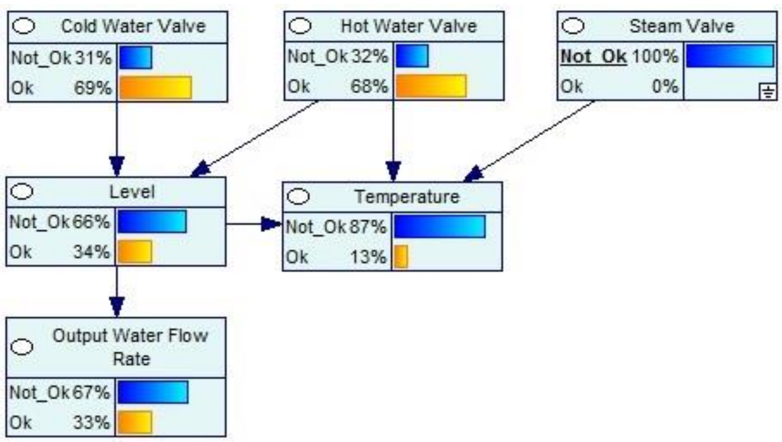


Figure 3.16: Root cause diagnosis by proposed methodology for fault scenario 2 (a) updated BN with rescaled multiple likelihood evidence (b) percentage change in probability

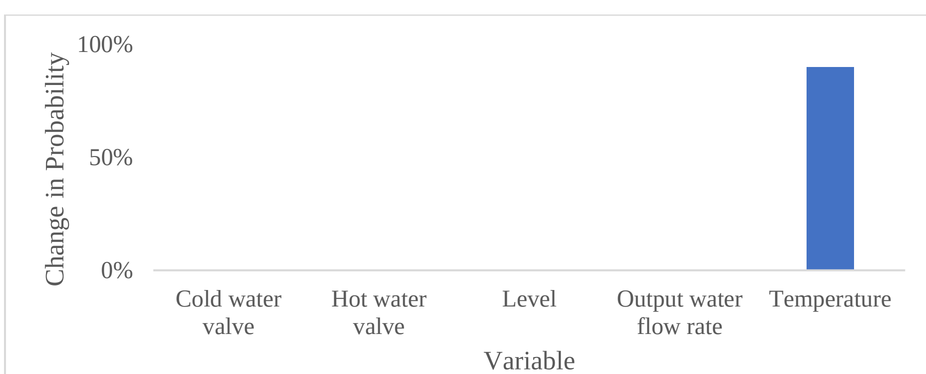
A BN is not required to diagnose the root cause in this fault scenario in conventional practice, as steam valve is a root node and it has the highest contribution in the T^2 contribution plot (Figure 3.15).

Hard evidence is given to the Not_Ok state of steam valve node in the BN to observe the inclination of other variables towards the faulty state. Only temperature has an increase of 89.87% towards the faulty state. Stuck steam valve will provide less steam, which will result in reduced water temperature. The identified fault propagation pathway can accurately reflect the scenario (Figure 3.17(c)).

(a)



(b)



(c)

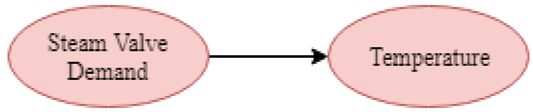


Figure 3.17: Fault propagation pathway identification for fault scenario 2 (a) hard evidence to steam valve node (b) percentage change in probability for other variables (c) identified fault propagation pathway

3.4.2. Tennessee Eastman (TE) Chemical Process

The Tennessee Eastman (TE) chemical process has five major units: a reactor, a product condenser, a vapor-liquid separator, a recycle compressor and a product stripper. Three gaseous reactants are fed to the reactor, where a catalyzed chemical reaction forms the liquid products. The product stream enters the condenser as vapor, and get condensed. Then product stream passes through the vapor-liquid separator, where the condensed and non-condensed products are separated. A centrifugal compressor recycles the non-condensed product back to the reactor, and the condensed product moves into the stripper to be stripped. The final product stream exits from the base of the stripper, and is pumped to the downstream for further refinement (Downs and Vogel, 1993). The PFD of the TE chemical process is shown in Figure 3.18. The TE chemical process consists of 41 measured variables and 12 manipulated variables. Among the measured variables, 22 variables are continuous process variables and 19 variables are related to composition measurements. These 22 continuous process variables have been considered in this work, and their description is shown in Table 3.11. There are 15 known and 5 unknown types of faults in the TE chemical process (Downs and Vogel, 1993; Yu et al., 2015). 10 fault scenarios have been tested. The tested fault IDs and true root cause for each fault type are summarized in Table 3.12.

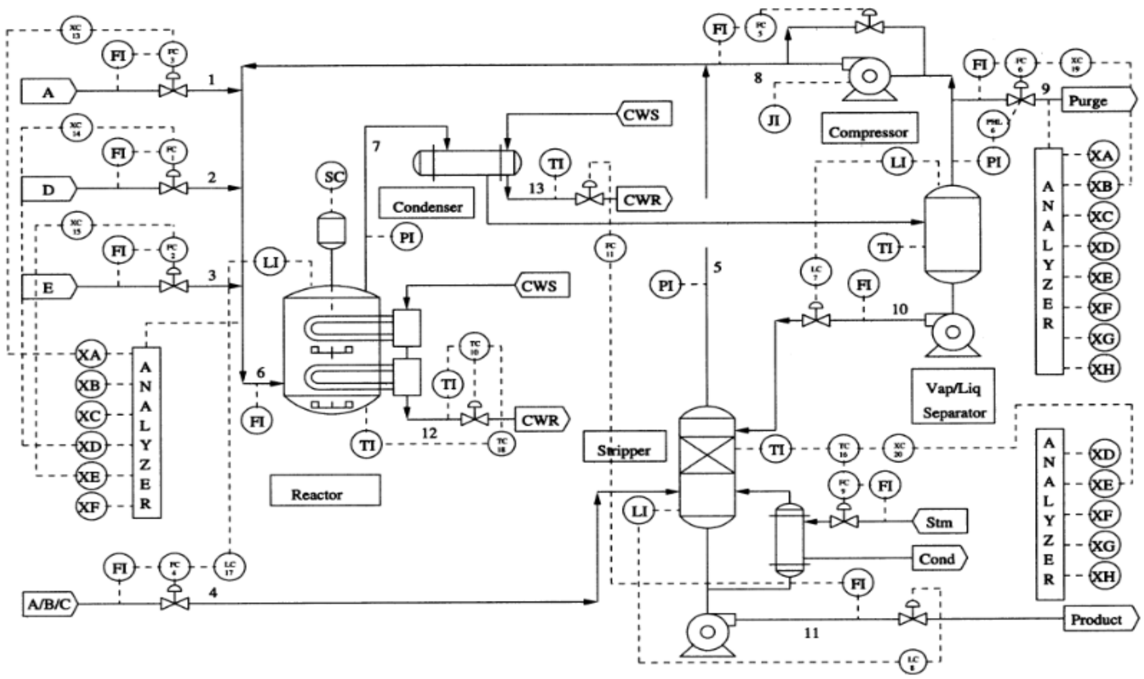


Figure 3.18: Process flow diagram of Tennessee Eastman chemical process

Table 3.11: Description of continuous process variables of TE chemical process

Variable No.	Description	Unit
XMEAS (1)	A Feed (Stream 1)	kscmh
XMEAS (2)	D Feed (Stream 2)	kg/hr
XMEAS (3)	E Feed (Stream 3)	kg/hr
XMEAS (4)	A and C Feed (Stream 4)	kscmh
XMEAS (5)	Recycle Flow (Stream 8)	kscmh
XMEAS (6)	Reactor Feed Rate (Stream 6)	kscmh
XMEAS (7)	Reactor Pressure	kPa gauge
XMEAS (8)	Reactor Level	%
XMEAS (9)	Reactor Temperature	ºC
XMEAS (10)	Purge Rate (Stream 9)	kscmh
XMEAS (11)	Product Separator Temperature	ºC
XMEAS (12)	Product Separator Level	%
XMEAS (13)	Product Separator Pressure	kPa gauge

XMEAS (14)	Product Separator Underflow (Stream 10)	m3/hr
XMEAS (15)	Stripper Level	%
XMEAS (16)	Stripper Pressure	kPa gauge
XMEAS (17)	Stripper Underflow (Stream 11)	m3/hr
XMEAS (18)	Stripper Temperature	ºC
XMEAS (19)	Stripper Steam Flow	kg/hr
XMEAS (20)	Compressor Work	kW
XMEAS (21)	Reactor Cooling Water Outlet Temperature	ºC
XMEAS (22)	Separator Cooling Water Outlet Temperature	ºC

Table 3.12: True root causes for tested fault conditions

Fault ID	**True Root Variable**
IDV 1	XMEAS (4)
IDV 4	XMEAS (9)
IDV 5	XMEAS (11)
IDV 6	XMEAS (1)
IDV 11	XMEAS (9)
IDV 12	XMEAS (11)
IDV 14	XMEAS (9)
IDV 15	XMEAS (11)
Stripper steam valve stiction	XMEAS (19)
E Feed Loss	XMEAS (3)

IDV 4 and IDV 15 are presented as case studies, and the results of other tested fault scenarios are shown in the results and discussion Section. IDV 4 is difficult to classify because of data overlapping (Zhang and Dong, 2014), and IDV 15 shows how this methodology improves diagnostic capacity of PCA-T^2. In both cases, the fault has been introduced after 1000 samples of normal operation. Auto-standardized (zero median, unit variance) data from 1000 samples in NOC, have been used to calculate the threshold value of T^2 statistics. The threshold value of T^2 statistics is computed as 29.56 with a 95% confidence level. 18 PCs can explain more than 85% of the total variation.

The TE chemical process has a recycle variable, XMEAS (5). A BN is acyclic in nature. To capture this cyclic nature in a BN, a duplicate dummy node of the recycle flow, XMEAS (5), has been created. Simply, XMEAS (5) has been used twice in the BN, one as a parent node, and another one as a child node, depending on the PFD. Figure 3.19 shows the BN in normal operating state.

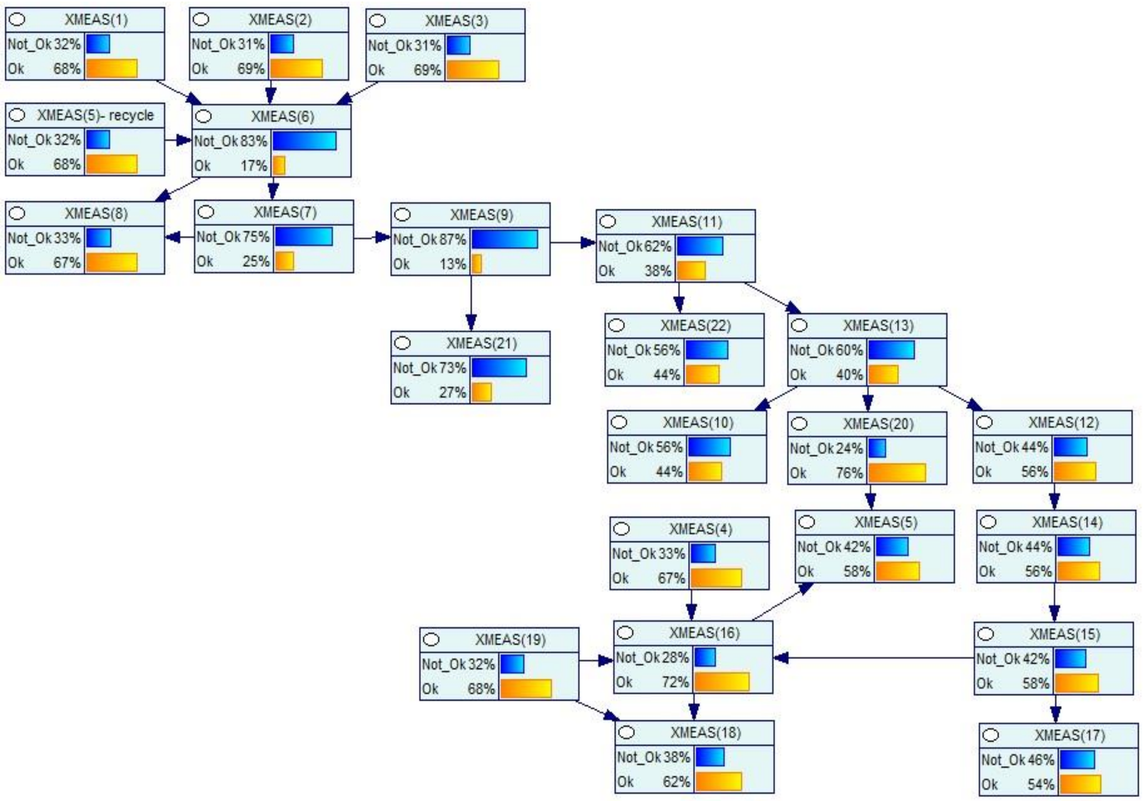


Figure 3.19: BN for the TE chemical process

3.4.2.1. IDV 4 (step disturbance in reactor cooling water inlet temperature)

This scenario has been generated by reducing the reactor cooling water flow. The T^2 control chart detects the fault at the 1339th sample. The reason for the delayed detection is the small magnitude of reduction in the cooling water flow. It results in an average increase of 0.014% for the 500 test samples from the 1000 NOC samples of reactor temperature. Hence, it takes longer to upset the reactor temperature significantly. The contribution plot is generated after detecting the fault. Figure 3.21 shows that six continuous variables, XMEAS (3), XMEAS (7), XMEAS (8), XMEAS (9), XMEAS (16) and XMEAS (21) have more than 10% rescaled contribution. Corresponding likelihood evidence of these six variables are used to update the BN. The updated BN with multiple likelihood evidence is shown in Figure 3.22(a).

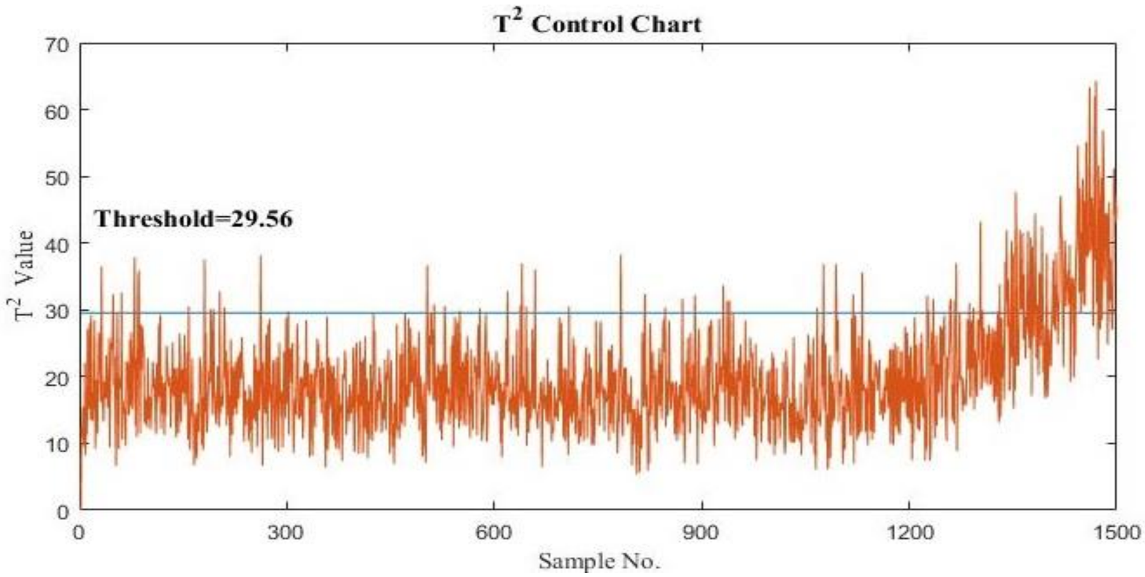


Figure 3.20: T^2 control chart for IDV 4

Figure 3.22(b) shows that XMEAS (9) has the highest increase in the faulty state (6.59%) in the updated BN. XMEAS (5) has an increase of 4.12% in the faulty state. XMEAS (7) is the only parent node of XMEAS (9). XMEAS (7) has 12.07% increase in the normal state, which implies that the fault has not been initiated form XMEAS (7), and it confirms XMEAS (9) as the true root cause of the abnormality.

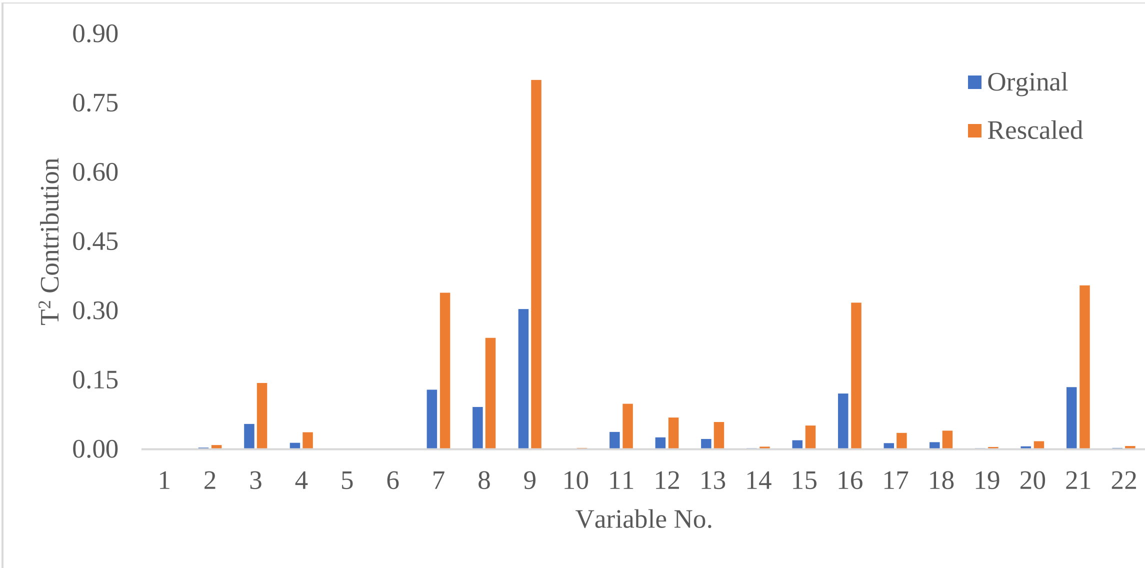
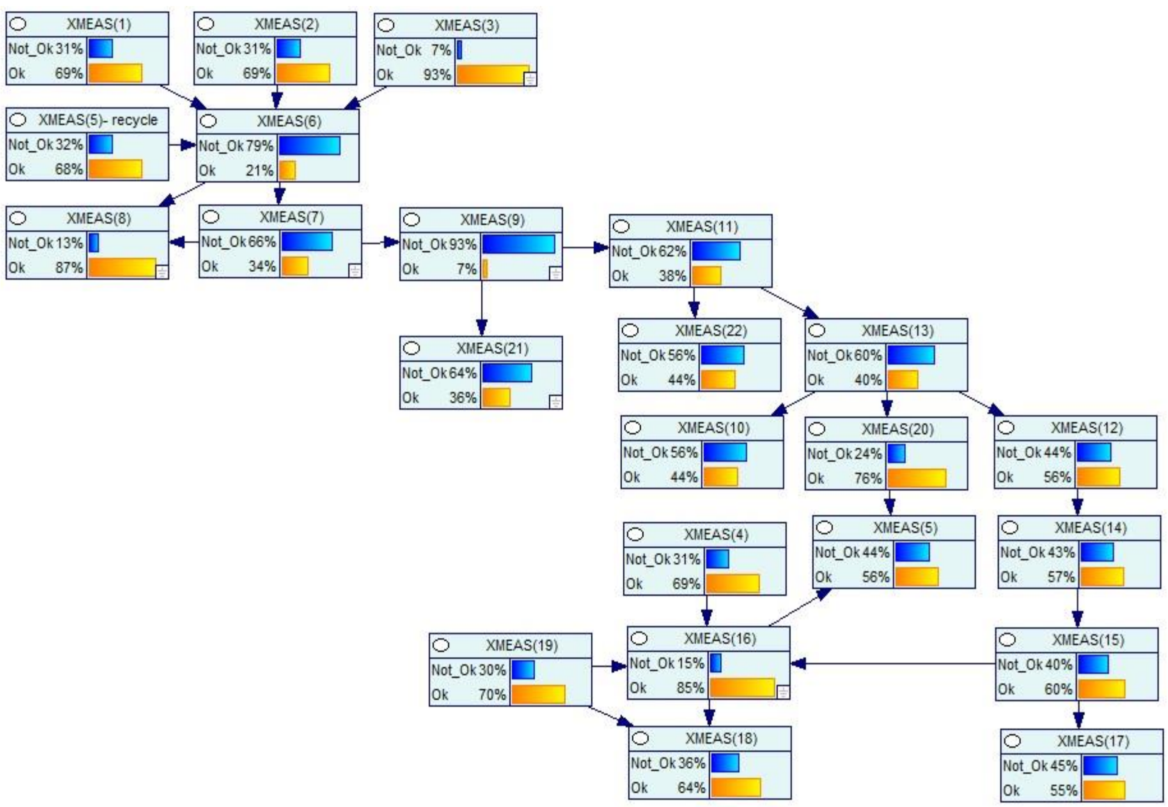


Figure 3.21: T^2 contribution plot for IDV 4

(a)



(b)

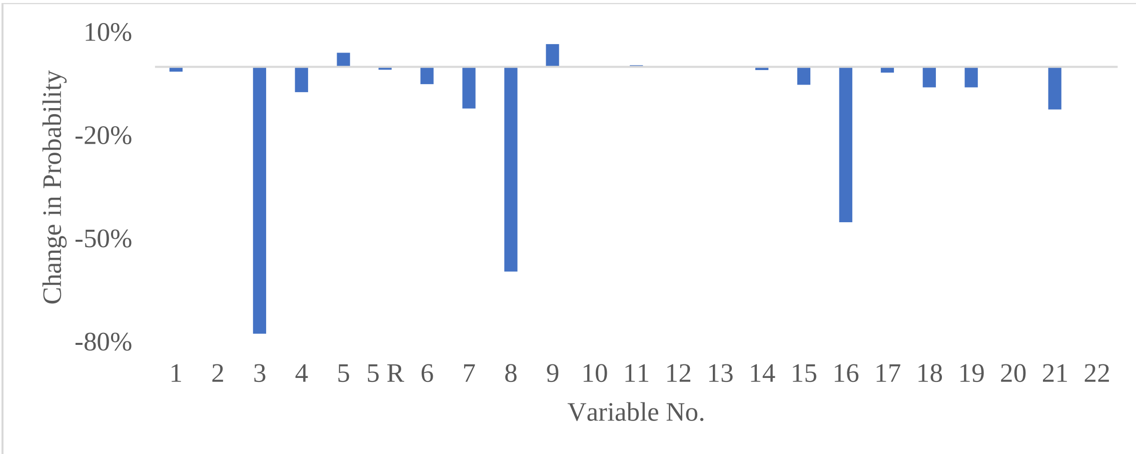
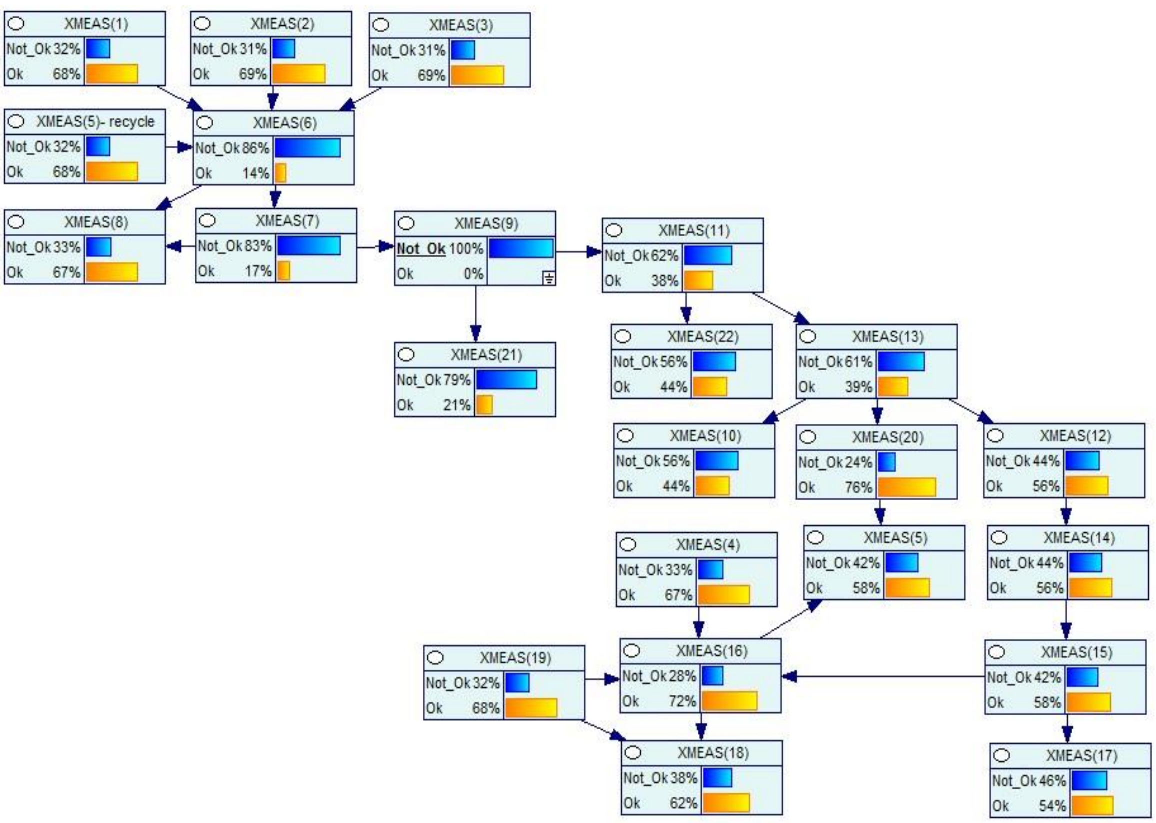


Figure 3.22: Root cause diagnosis by proposed methodology for IDV 4 (a) updated BN with rescaled multiple likelihood evidence (b) percentage change in probability

(a)



(b)

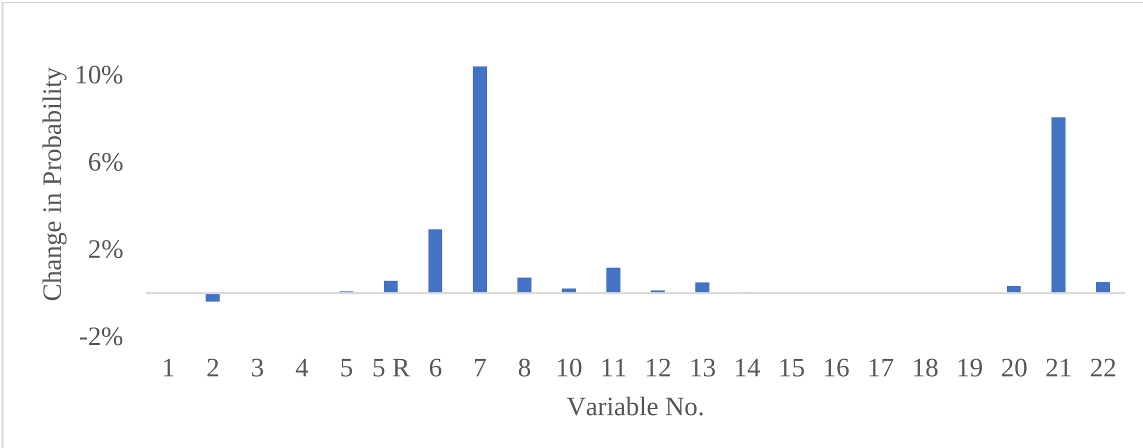


Figure 3.23: Root cause diagnosis by conventional approach for IDV 4 (a) updated BN with hard evidence (b) percentage change in probability

Conventional approaches will provide hard evidence to the Not_Ok state of XMEAS (9) and look for the root cause from other variables. The updated BN with hard evidence is shown in Figure 3.23(a). Figure 3.23(b) shows that XMEAS (7) and XMEAS (21) have a larger increase than other variables after updating the BN. XMEAS (7) is the only parent node of XMEAS (9), while XMEAS (21) is a child node of XMEAS (9). Hence, XMEAS (7) will be falsely diagnosed as the root cause.

To find the variables that fall inside the fault propagation pathway, hard evidence is given in the Not_Ok state of XMEAS (9). Reactor temperature affects separator temperature and reactor cooling water inlet temperature. Separator temperature will upset separator pressure and level. Separator pressure will affect both the purge rate and compressor work. As a result, recycle flow will be affected, and total feed will change, which will affect both reactor pressure and level. Separator level will hamper separator underflow. The stripper unit will be affected in similar fashion. Figure 3.24 shows the identified fault propagation pathway using the proposed methodology.

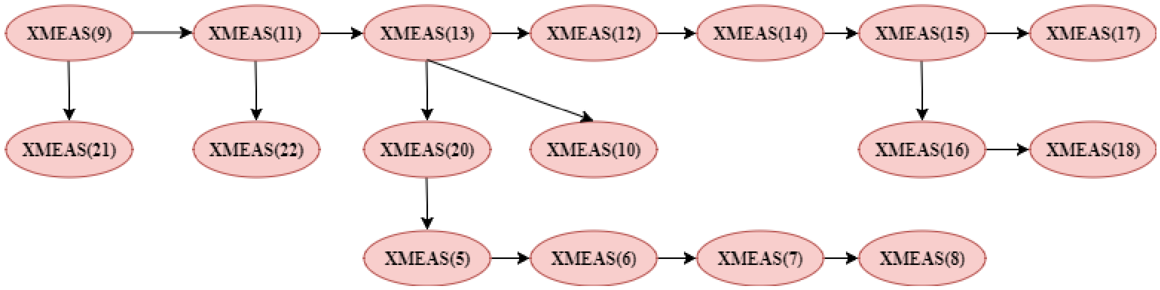


Figure 3.24: Fault propagation pathway for IDV 4

3.4.2.2. IDV 15 (condenser cooling water valve stiction)

This scenario has been created by fixing the manipulated variable XMV (11). The T^2 control chart can detect the fault as soon as it affects the process significantly. The T^2 contribution plot diagnoses XMEAS (22) as the root cause of the abnormality (Figure 3.26). Since it is a child node, a BN must be used to diagnose the true root cause.

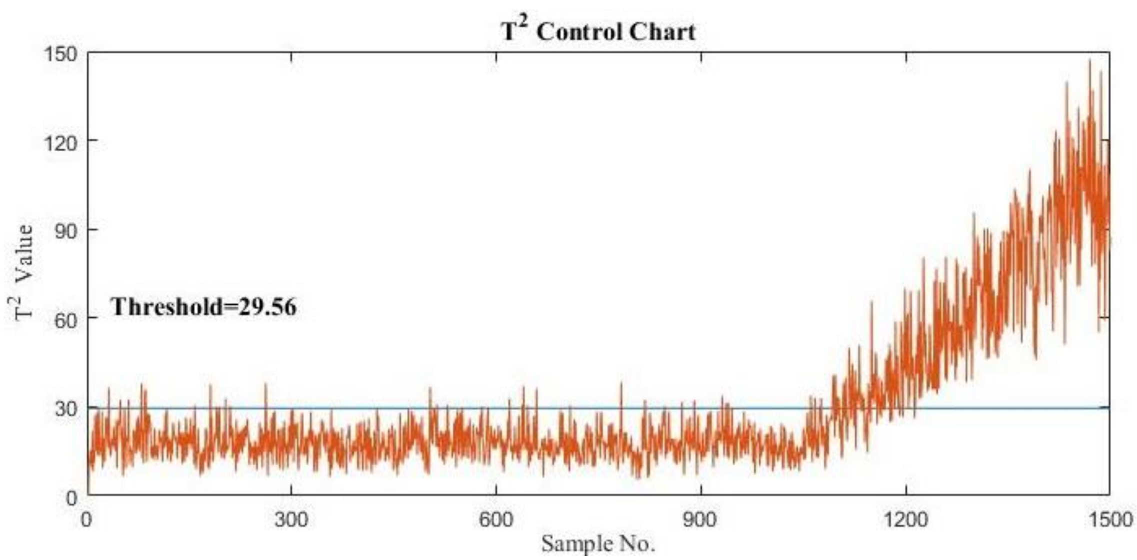


Figure 3.25: T^2 control chart for IDV 15

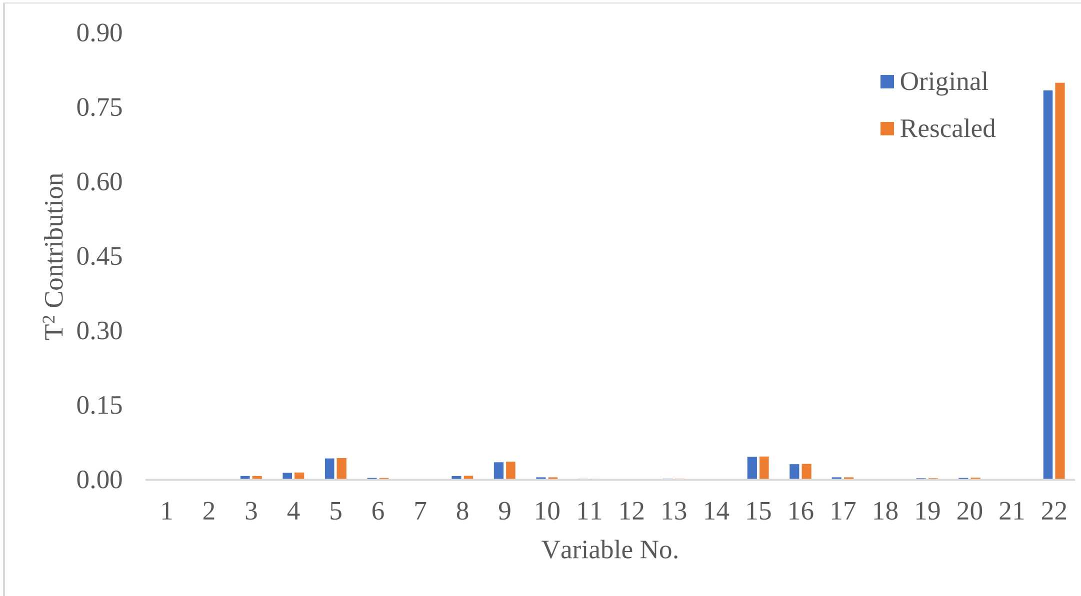
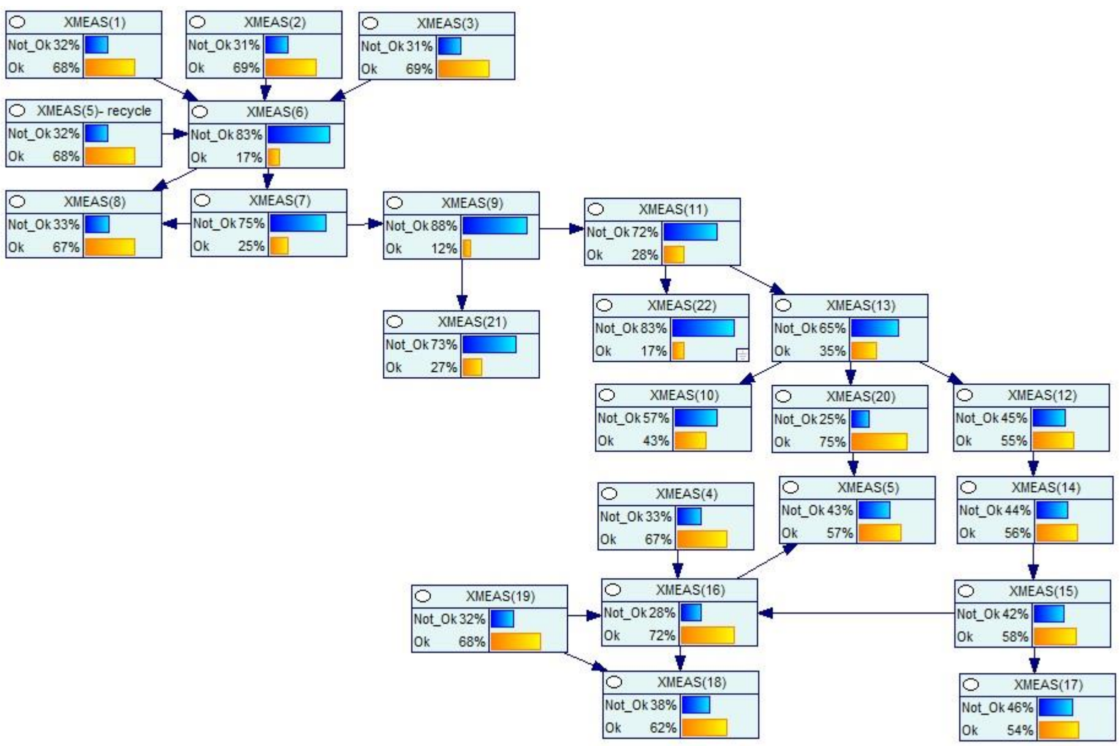


Figure 3.26: T^2 contribution plot for IDV 15

(a)



(b)

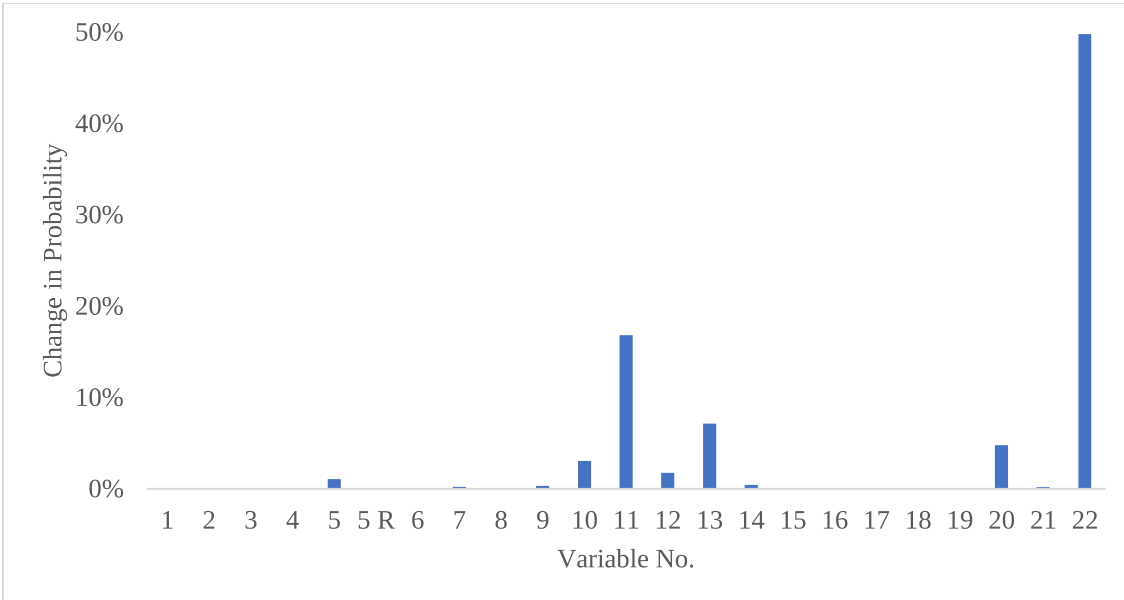
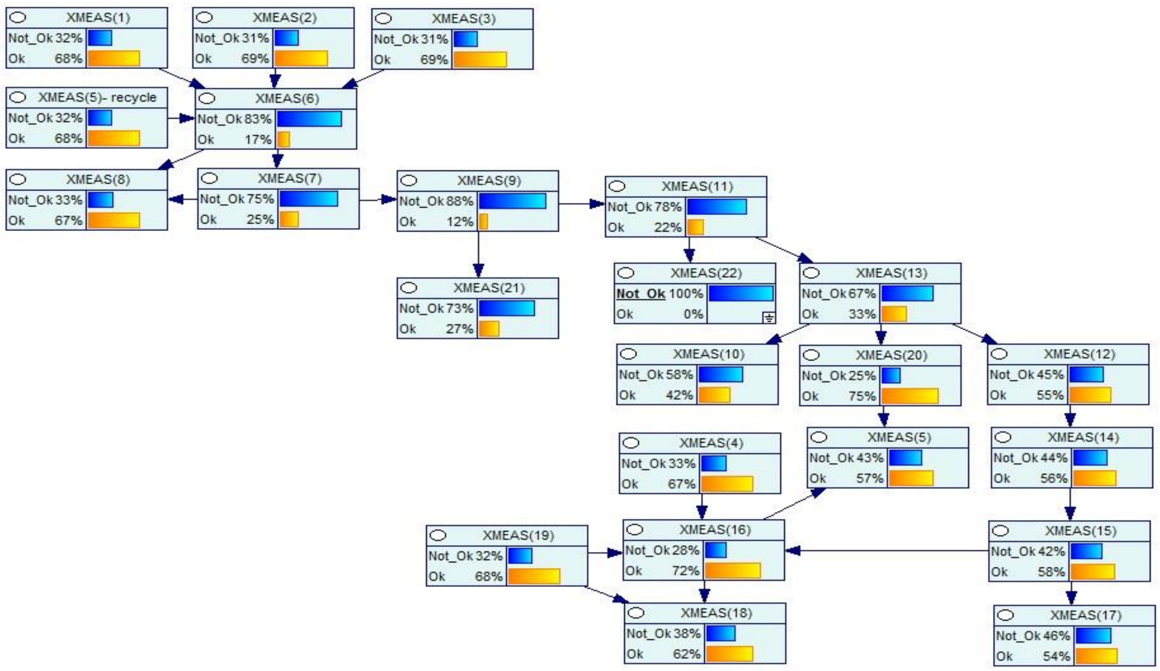


Figure 3.27: Root cause diagnosis by proposed methodology for IDV 15 (a) updated BN with rescaled likelihood evidence (b) percentage change in probability

To do so, the contributions are rescaled. Only XMEAS (22) has more than a 10% rescaled contribution and is used to update the BN. The updated BN in Figure 3.27 finds XMEAS (22) as most likely to be in the faulty state, as it has an increase of 49.81% in the faulty state. XMEAS (22) is a child node. Therefore, search for root cause is performed among its successive parent nodes. Its only parent node, XMEAS (11) increases 16.84% in the faulty state after updating the BN. Other variables increase less than 8% in the faulty state. This confirms XMEAS (11) as the true root cause.

In the conventional approach, hard evidence will be given to the Not_Ok state of XMEAS (22). Figure 3.28(a) shows the updated BN. XMEAS (10), XMEAS (11), XMEAS (13) and XMEAS (20) have greater probability to be in faulty state (Figure 3.28(b)). However, XMEAS (11) has much larger increase in the faulty state and can be accurately diagnosed as the true root cause.

(a)



(b)

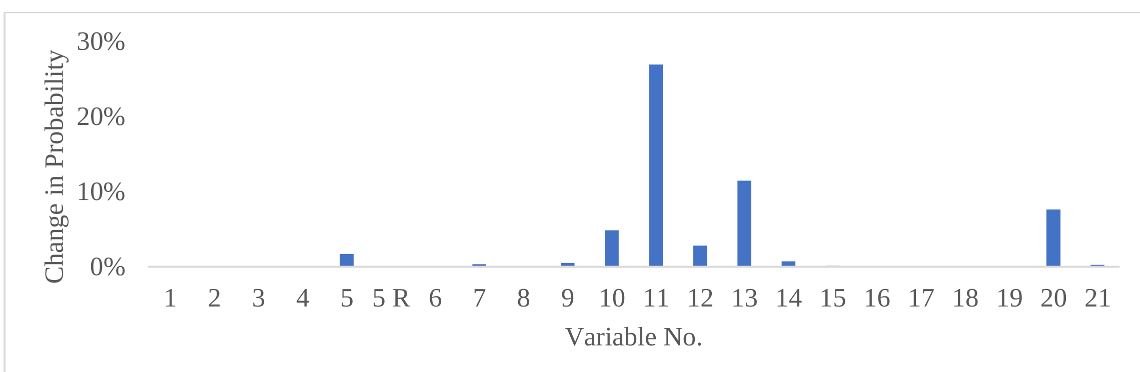
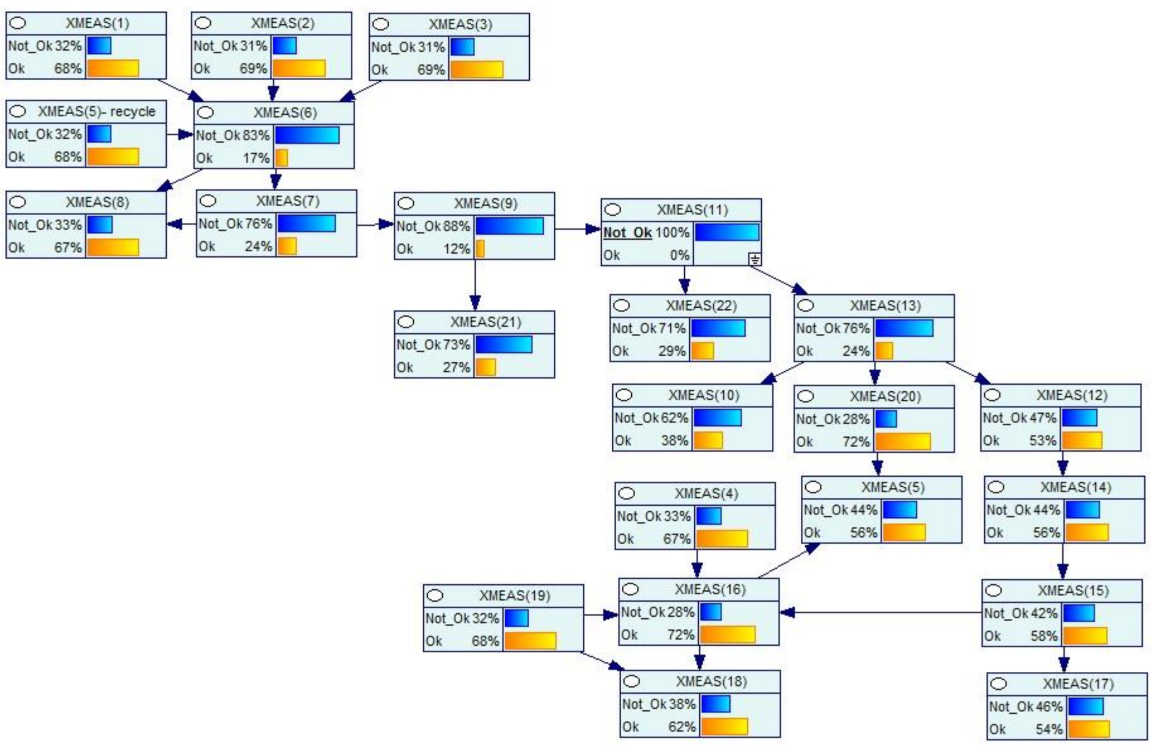


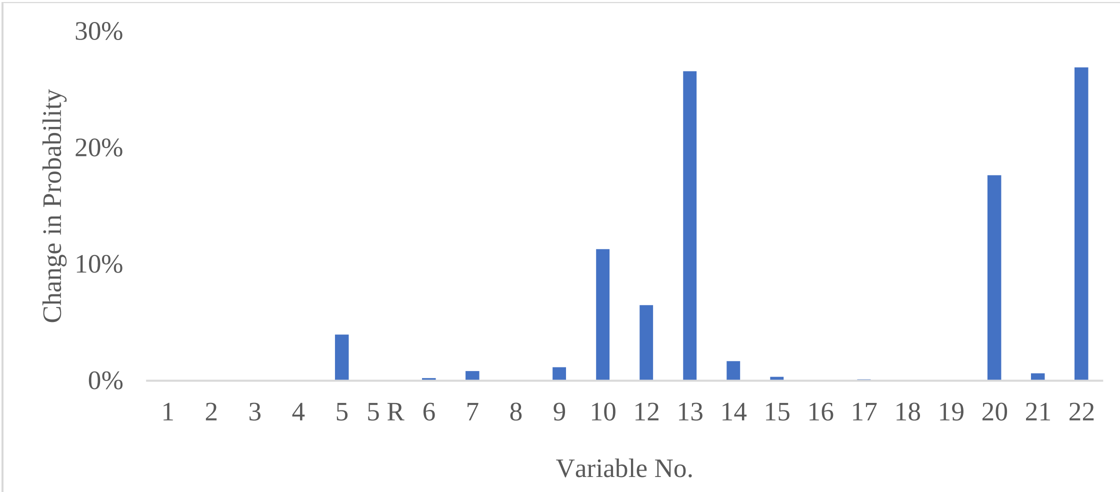
Figure 3.28: Root cause diagnosis by conventional approach for IDV 15 (a) updated BN with hard evidence (b) percentage change in probability

Next, the fault propagation pathway is checked by giving hard evidence (P(Not_Ok=100%)) to XMEAS (11). The aim is to observe the behavior of other variables towards the faulty state. Any upset in reactor temperature will upset the separator temperature and reactor cooling water outlet temperature. Separator temperature will create abnormality in separator pressure, separator pressure will cause abnormal behavior in the separator level and there will be underflow in the separator. Stripper unit will be affected in an analogous way. The compressor work will be changed, which subsequently affects the recycle flow, and the recycle flow will affect the total feed rate. Since feeds are gaseous, change in a feed will change reactor pressure, and reactor pressure will further upset reactor temperature and level. Figure 3.29(c) captures the phenomena accurately.

(a)



(b)



(c)

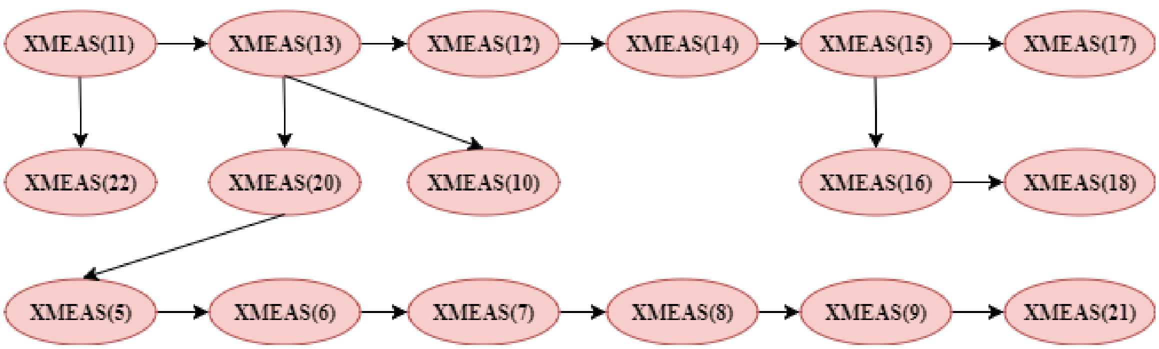


Figure 3.29: Fault propagation pathway identification for IDV 15 (a) hard evidence to XMEAS (11), (b) percentage change in probability for other variables (c) identified fault propagation pathway

3.5. Results and Discussion

This paper demonstrates the better suitability of PCA-T^2-BN with multiple likelihood evidence over conventional PCA and PCA-BN using hard evidence based techniques in dynamic process monitoring. It appears to be very effective, while conventional hard evidence based techniques fail to diagnose the root cause of process abnormality. Table

3.13 shows the comparative diagnostic performance among the proposed methodology, the conventional PCA and PCA-BN with hard evidence based techniques for all twelve tested fault scenarios studied in CSTH and TE chemical process.

Table 3.13: Diagnostic performance comparison among proposed methodology and different PCA-BN based techniques

Process model	**Fault description**	**PCA-T^2**	**PCA-T^2-BN**	
			Hard evidence	**Multiple likelihood evidence**
CSTH	Leak in the tank	Yes	No	Yes
	Steam valve stiction	Yes	Yes	Yes
TE Chemical Process	IDV 1	No	Yes	Yes
	IDV 4	Yes	No	Yes
	IDV 5	No	Yes	Yes
	IDV 6	Yes	Yes	Yes
	IDV 11	No	Yes	Yes
	IDV 12	No	Yes	Yes
	IDV 14	Yes	No	Yes
	IDV 15	No	Yes	Yes
	Stripper steam valve stiction	Yes	Yes	Yes
	E feed loss	Yes	Yes	Yes

PCA-T^2 performs well in diagnosing the root cause in CSTH. PCA-T^2-BN with a conventional hard evidence based approach, fails to diagnose the true root cause in the case of a leak in the tank. Diagnostic performance of PCA-T^2 is not consistent for several cases in the TE chemical process. It fails to identify the true root cause for IDV 1, IDV 5, IDV 11, IDV 12 and IDV 15. It gives an indication that PCA-T^2 may not fulfil the diagnostic requirement for large scale processes. A BN with conventional hard evidence can relax the limitation of PCA-T^2 by diagnosing the true root cause accurately in these five cases.

The performance of these conventional hybrid approaches becomes unsuitable, while PCA-T^2 already accurately diagnoses an intermediate node as the true root cause. By convention, hard evidence will be given to that intermediate node, and a search for the root cause will be carried out among other variables. Because of this inherent characteristic, PCA-T^2-BN with a hard evidence based approach fails to diagnose the true root cause for a leak in the tank in CSTH and IDV 4, IDV 14 in the TE chemical process. The updated BN with multiple likelihood evidence overcomes this limitation for all these cases, as it also takes the highest contributing variable of the multivariate contribution plots into consideration, while searching for the root cause.

3.6. Conclusion

A new hybrid methodology (PCA-T^2-BN) has been proposed in process FDD. PCA has been used at the first stage to detect and diagnose the fault. A BN has been used to reach a certain conclusion from uncertain observed information provided by PCA. Higher diagnostic accuracy has been achieved by applying a multiple likelihood evidence based updating technique in the BN which curtails the possibility of false diagnosis. Although PCA takes a little longer to detect the fault in many cases, its combination with BN and multiple likelihood evidence make it a comprehensive tool for accurate diagnosis. However, detection delay is normal when the fault magnitude is very low, or when it takes longer to affect the process significantly. The methodology has been examined in twelve fault scenarios in two benchmark processes, the CSTH and TE chemical process. The results suggest that it can diagnose the root cause accurately, where convention PCA and PCA-BN based approaches fail. It also preserves the strength of conventional techniques.

The main contributions of this research work are: (1) it improves the limited diagnostic capacity of the conventional PCA, (2) it provides superior performance over conventional PCA-BN with hard evidence based techniques, (3) it proposes a unique multivariate contribution rescaling technique, which helps the BN to represent the variable states approximately accurately and (4) it provides a guideline to improve the diagnostic capacity of other BN based hybrid methodologies. The main advantages of the proposed methodology are: (1) it utilizes the advantage of data-driven and knowledge based methods, (2) it can diagnose the root cause of abnormality in any node of a BN, (3) it is very easy to apply and computationally inexpensive, (4) it can identify the fault propagation pathway and (5) it enhances the uncertainty handling capacity of a BN by enabling it to get updated with real time multivariate contributions.

Future work may include the application of the proposed methodology to industrial systems and combining the BN with non-linear MSPM tools to capture the non-Gaussian features also to have better detection capability. In industries, it is not expedient to monitor all the variables due to cost optimization. The methodology needs to be applied to diagnose true root cause in an unmonitored variable.

Chapter 4

Fault Detection and Pathway Analysis using Dynamic Bayesian Network

Abstract: A dynamic Bayesian network (DBN) based fault detection, root cause diagnosis, and fault propagation pathway identification scheme is proposed. The proposed methodology generates evidence from monitored process data and uses the information to update DBN that captures the process knowledge. A new dynamic Bayesian anomaly index (DBAI) based control chart is proposed for detection purpose. Following the detection of the fault, root cause(s) is diagnosed using the smoothing inference of a DBN, and fault propagation pathway is identified from the cause-effect relationships among the process variables. The proposed methodology is applied to a binary distillation column and a continuous stirred tank heater (CSTH). The result shows that it can detect the fault and diagnose the root cause of the fault precisely. The result has been compared to the performance of the Shewhart control chart, principal component analysis (PCA) and static BN. The comparative study confirms that the proposed methodology is a more efficient fault detection and diagnosis (FDD) tool.

Keywords: Fault detection, fault propagation pathway, root cause diagnosis, dynamic Bayesian network, cause-effect relationship

4.1. Introduction

Process industries suffer significant economic losses due to abnormal situations emerged from the faults during operation (Nimo, 1995). Fault detection and diagnosis (FDD) are the first steps of abnormal situation management (ASM) (Kresta et al., 1991). Timely detection and diagnosis of the root cause of the fault are important to assure process safety, reliable operation, product quality and optimum operational cost. Data driven methods rely on the process data collected from normal operating condition (NOC). These fault free data are used to define the control limit (CL), and later fault is detected when on-line monitored samples violate the threshold of the CL (Venkatasubramanian et al., 2003b). These tools are easier to apply and provide very quick detection performance. Hence, data based statistical process monitoring (SPM) tools are very popular in the process industries (Qin, 2012).

Data based SPM tools are mainly classified into two categories: univariate and multivariate tools. Univariate tools monitor the individual signal of process variables to detect and diagnose the fault. The Shewhart chart is one of the earliest univariate control charts applied in process monitoring (Shewhart, 1930). This chart defines the CLs (e.g. upper control limit (UCL), lower control limit (LCL)) using mean and standard deviation of individual variable. UCL and LCL are determined based on the target standard deviation(s) from the mean depending on the product quality requirement. In spite of ease of application and reliable performance, the Shewhart chart is vulnerable to process noise since it only considers current measurement from the sensors without filtering. Exponentially weighted moving average (EWMA) and cumulative sum (CUSUM) control charts are two other univariate control charts, which possess memory effect of data and suitable for small mean shifts. However, these techniques require expert opinion for desired performance and are computationally more expensive than the Shewhart control chart (Montgomery and Runger, 2010). The major disadvantage of the univariate control charts is that these make the monitoring complex since a dedicated control chart is required to monitor each variable. Moreover, these charts ignore the change in operating condition.

To overcome the aforementioned limitations of the univariate monitoring techniques, multivariate statistical process monitoring (MSPM) tools are widely used in the process industries. These tools monitor the process by a combined index as the CL using different multivariate statistics such as T^2, SPE and I^2 statistics. Thus, monitoring of numerous control charts is avoided. MSPM tools detect the fault by observing the breakdown in correlation among process variables and diagnose the root cause by generating the multivariate contribution plots. Principal component analysis (PCA) and partial least square (PLS) are the most widely used MSPM tools (Bakshi, 1998; Nomikos and MacGregor, 1995). However, these tools are most suitable when process data follow a multivariate Gaussian distribution. Nonlinear PCA (NPCA) and kernel PCA (KPCA) have been proposed to handle non-linearity (Choi and Lee, 2004; Kramer, 1991). These tools can detect the fault effectively. Although these MSPM tools provide robust detection performance, diagnosis of root cause is often misleading and incomplete. Support vector machine (SVM) and artificial neural network (ANN) has also been applied (Hoskins et al., 1991; Kulkarni et al., 2005). However, SVM and ANN are suitable when in-depth process knowledge is available, since they need both normal and faulty data for training purpose.

The knowledge based tools (e.g. fault tree, signed digraph, possible cause-effect graph, Bayesian network etc.) are popularly used for fault diagnosis. These tools are mainly built on the expert opinion and using if-else-then logics (Venkatasubramanian et al., 2003a). While the knowledge based tools provide excellent performance in fault diagnosis, they lack in robust detection capacity. Hybrid methods have also been proposed to eliminate the limitation of an individual method. In many hybrid methods, a data based MSPM tool is used to detect the fault in the first stage, and a knowledge based tool is used in the second stage to diagnose the root cause of the fault utilizing the diagnostic information provided by the first stage detection tool (Gharahbagheri et al., 2017; Mallick and Imtiaz, 2013; Vedam and Venkatasubramanian, 1999; Y. Wang et al., 2017; Yu et al., 2015).

Bayesian network (BN) is an emerging tool in process FDD. It has ample application in the fields of risk analysis, dependability, and maintainability (Weber et al., 2012). BNs are mostly used for root cause diagnosis in process monitoring. It is constructed combining process historical data and expert knowledge. However, it is possible to build a BN using expert opinion only when process data are unavailable. (Mehranbod et al., 2005, 2003) used BN in sensor fault detection and isolation in both steady and transient operating conditions. (Dey and Stori, 2005) proposed a BN based root cause identification technique for several process variations. (Azhdari and Mehranbod, 2010) demonstrated the application of a BN in detecting and diagnosing the faults in the Tennessee Eastman (TE) chemical process. An application of a BN in dimensionality reduction is available in literature (Gonzalez et al., 2015). The BN used in existing literature is mostly static in nature. (Yu and Rashid, 2013) used a dynamic BN (DBN) for process fault detection, root cause diagnosis and fault propagation pathway identification. Two indicators abnormality likelihood index (ALI) and dynamic Bayesian probability index (DBPI) were proposed to detect the fault, and dynamic Bayesian contribution index (DBCI) was used to diagnose the root cause of the fault. (Zhang and Dong, 2014) incorporated the output of a Gaussian mixture model (GMM) with a three time-slice DBN to detect and isolate the fault. Hence, it is evident that a DBN has the potential to perform as an efficient solitary process FDD tool.

The main objective of this research is to develop a technique which can detect the fault and diagnose the root cause of the fault accurately by utilizing the power of DBN. A DBN based methodology has been proposed and tested on two common industrial sub-systems: a binary distillation column and a continuous stirred tank heater (CSTH). The results show that the proposed DBN based methodology can detect the fault as well as precisely diagnose the root cause. The performance of the proposed methodology has been also compared with static BN and two data driven methods: a univariate tool (the Shewhart control chart) and a multivariate tool (PCA). The comparative performance evaluation suggests that the proposed approach is a more efficient process monitoring tool for fault detection and root cause diagnosis.

The rest of the paper is organized as follows: Section 4.2 will provide a brief description on the BN and DBN. The methodology will be explained in Section 4.3. Application of the proposed methodology will be demonstrated in Section 4.4. Comparative results will be shown in Section 4.5. The contribution and merits will be discussed in Section 4.6.

4.2. Preliminaries

4.2.1. Bayesian Network (BN)

A BN is a causal network which belongs to the family of probabilistic graphical models (Pearl, 1988). It represents the knowledge in the graphical form. It is a directed acyclic graph, which links up the uncertain observations and helps to reach a certain conclusion (Neapolitan, 2004). It contains node and directed arcs. Each node represents the probability distribution of a random variable, while an arc determines the probabilistic relationship between the two connected variables. The node from which the arc is created is called the parent node, and the node to which the arc is directed is called the child node. An arc from a child node can never come back to its parent nodes. The node which does not have any child node is called a leaf node. On the other hand, any parentless node is called a root node. An intermediate node serves as both a parent and a child node in the network (Yu et al., 2015). Bayes' theorem (shown in Equation 4.1) is the governing equation of a BN.

$$P(\theta/X) = \frac{P(X/\theta) \times P(\theta)}{P(X)} \tag{4.1}$$

where P(Θ) is the prior belief, and P(X) is the probability of an observation or evidence. P(X/Θ) is the conditional probability of X given Θ. It is also called the likelihood probability. P(Θ/X) is the conditional probability of Θ given X. It can also be referred to as the posterior probability, since it contains the degree of belief relying on an observation of X. For a certain evidence of X, the updating equation can be written as:

$$P(\theta/X) = P(X/\theta) \times P(\theta) \tag{4.2}$$

A BN works by propagating the belief in the entire network. So, it is often termed as a Bayesian Belief Network (BBN) (Mallick and Imtiaz, 2013). The joint probability distribution (JPD) of the network can be obtained from the product of all likelihood probabilities (Y. Wang et al., 2017). Using chain rule, the JPD can be expressed as:

$$\mathrm{P}(\mathrm{X}_1, \mathrm{X}_2, \dots, \mathrm{X}_\mathrm{n}) = \prod_{i=1}^{n} \mathrm{P}(\mathrm{X}_\mathrm{i}/\mathrm{Pa}(\mathrm{X}_\mathrm{i})) \quad (4.3)$$

where Pa (X_i) is the parent set of any node X_i.

A BN is constructed in three main steps: (1) determining the causal dependency among the variables, (2) estimating the prior probability distribution, and (3) estimating the conditional probability distribution. The main advantage of using a BN is that any state of a node can be updated and the renewed probabilities obtained after belief propagation can be utilized in decision making. It also provides a pictorial view of the entire process operation.

4.2.2. Dynamic Bayesian Network (DBN)

The conventional BN is discrete and static. It cannot model dynamic systems. As chemical processes are dynamic, static BN fails to capture this dynamic nature. A dynamic Bayesian network (DBN) is an extension of a static BN. It can represent the temporal relationships. A static BN can be extended to a DBN by the following ways (Mihajlovic and Petkovic, 2001):

1. Adding the state of a node to describe the temporal relationship with time slice.
2. Modifying the structure of the BN based on process dynamics.

3. Repeating the static BN with time if all the variables exert influence on the process and updating the belief of current time-step.

Figure 4.1 shows a simple DBN of three nodes for N time-steps. The static BN has been converted into a DBN by applying the first order Markov chain. A and B are the parent nodes of C. The probability of node C depends on its parent nodes in a time slice, as well as on its own value of previous time slice.

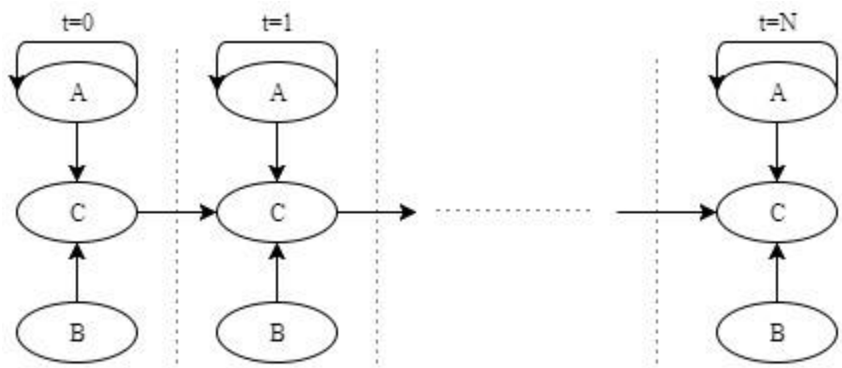


Figure 4.1: A simple DBN of three nodes

The arcs can be connected to another node, as well as return to itself. The later type of arc is called a self-rolling arc. For example, if a self-rolling arc is used at node A in Figure 4.1, it implies that one more conditional probability table (CPT), $P(A_t/A_{t-1})$ is required to model the random variable A. The arcs are always forward moving among time slices. No arc can return to its previous time slices.

If Z represents a family of random variables $X_1, X_2, \ldots.., X_n$ (i.e. $X \in Z$); the transition model from the previous time slice to the current time slice for a DBN can be expressed as (Murphy, 2002):

$$P(Z_t/Z_{t-1}) = \prod_{i=1}^{n} P(Z_{i,t}/Pa(Z_{i,t})) \tag{4.4}$$

where $Z_{i,t}$ is i^{th} node at time t and $Pa(Z_{i,t})$ is the parent nodes of $Z_{i,t}$ from same and previous time slice. n is the number of nodes in the network.

The joint probability density function of a DBN for time t=1 to N can be expressed as:

$$P(Z_{1:N}) = \prod_{t=1}^{N}\prod_{i=1}^{n} P(Z_{i,t}/Pa(Z_{i,t})) \tag{4.5}$$

Unlike static BN, time dependency is required to be included in the CPTs to model a DBN, since variables are also dependent on the previous time slice. It introduces one more step in estimating the CPTs. Table 4.1 shows an illustrative example of a CPT of node C in Figure 4.1.

Table 4.1: CPT of node C

A_t		Not_Ok				Ok			
B_t		Not_Ok		Ok		Not_Ok		Ok	
C_{t-1}		Not_Ok	Ok	Not_Ok	Ok	Not_Ok	Ok	Not_Ok	Ok
C_t	Not_Ok	0.998	0.954	0.923	0.871	0.534	0.641	0.462	0.119
	Ok	0.002	0.046	0.077	0.129	0.466	0.359	0.538	0.881

The CPTs can be estimated either from historical process data or from expert opinion. Maximum likelihood estimation (MLE) and Bayesian estimation (BE) techniques can be employed to the data to estimate the CPTs. In MLE, CPTs are calculated counting the relative frequencies of variables that maximizes the likelihood of data. In the BE technique, the prior distributions are usually estimated and consolidated with respect to Dirichlet distribution or Wishart distribution (Spiegelhalter and Lauritzen, 1990). If there is any time lag in the process, it needs to be adjusted before applying the MLE to have better estimation.

A DBN has two robust features: smoothing and prediction inferences. Prediction is forecasting the future of a state in a node based on the current evidence, while smoothing refers to the estimation of the probability of a node in the past based on collected evidence up to current time slice. Consider a variable is continuously observed, an evidence of a fault is received at 50th time-step, it is possible to estimate the probability of a fault of this variable at 45th time-step (i.e.P(Fault$_{t=45}$/evidence$_{t=50}$)) and 55th time-step (i.e. P(Fault$_{t=55}$/evidence$_{t=50}$)). Smoothing inference enables to estimate the probability of a fault at t=45, and prediction inference helps to predict what will be the probability of a fault at t=55 given that a fault has occurred at t=50. When evidence is given, probability of fault increases in the adjacent time-slices due to the belief propagation. Smoothing provides vigorous conclusion, since sufficient information are available about the events that already have occurred (Łupińska-Dubicka, 2012).

4.3. DBN based FDD Methodology

Figure 4.2 shows the flow chart of the proposed methodology. A DBN is used to detect the fault, diagnose the root cause of the fault, and identify the fault propagation pathway. First, prior knowledge and process flow diagrams (PFDs) are utilized to build the DBN model. CPTs are estimated using MLE. Fault can initiate from any variable in a process. Hence, a first order self-rolling arc has been considered in this study for all the monitored variables so that the corresponding node can be updated, and fault is detected earlier. Time invariant network structure and likelihood probabilities have been used to model the DBN.

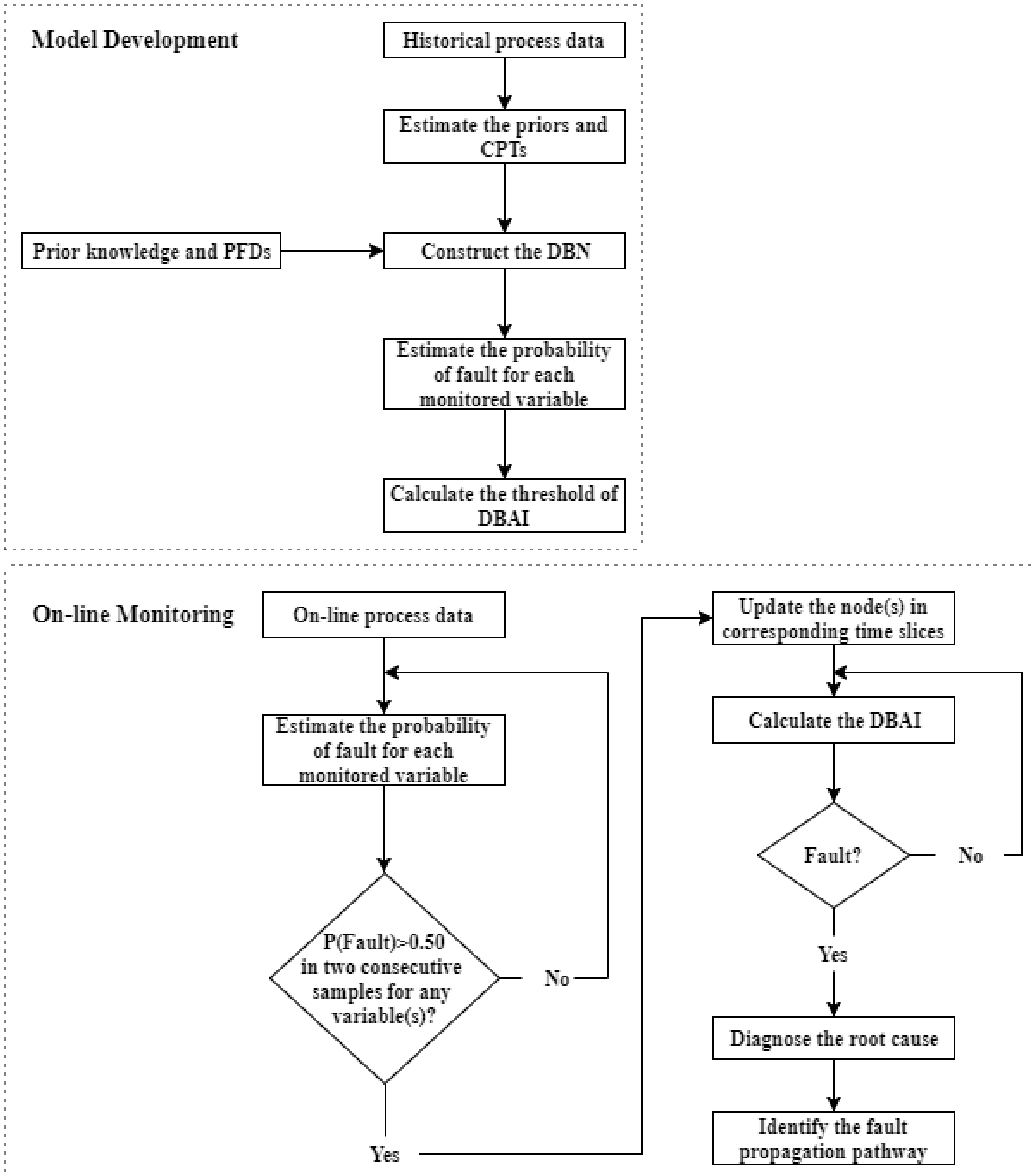


Figure 4.2: DBN based FDD methodology

The network needs evidence to get updated and provide conclusion. The new information considered as evidence, is modelled as a Gaussian distribution. A fault is classified when a variable starts to move from an expected value (mean, UCL, LCL). The probability of fault for a variable can be estimated using Equation 4.6.

$$\mathrm{P(Fault)} = \varphi\left(\frac{X - \mu}{\sigma}\right) \tag{4.6}$$

where X is any arbitrary value in the domain of Z, μ is the mean and σ is the standard deviation for a variable.

A normal operating zone needs to be defined to reduce false alarms and efficient fault detection. The upper and lower threshold values for the normal operating zone are selected as $\mu+3\sigma$ and $\mu-3\sigma$ respectively. The probability of fault is considered as 0.50 at $\mu\pm3\sigma$ and 0 at μ. The individual control limit for any variable is 0.50. From this point the variable can move towards a faulty state or come back to a normal state (Zadakbar et al., 2012).

for $x_{ij}>\mu_j$,

$$\mathrm{P(Fault)} = \varphi\left(\frac{x_{ij} - (\mu_j + 3\sigma_j)}{\sigma_j}\right)$$

$$= \int_{-\infty}^{x_{ij}} \frac{1}{\sigma_j\sqrt{2\pi}}\, e^{\left[\left\{x_{ij}-\left(\mu_j+3\sigma_j\right)\right\}^2/2\sigma_j^2\right]} dx \tag{4.7}$$

for $x_{ij}<\mu_j$,

$$\mathrm{P(Fault)} = 1 - \varphi\left(\frac{x_{ij} - (\mu_j - 3\sigma_j)}{\sigma_j}\right)$$

$$= 1 - \int_{-\infty}^{x_{ij}} \frac{1}{\sigma_j\sqrt{2\pi}}\, e^{\left[\left\{x_{ij}-\left(\mu_j-3\sigma_j\right)\right\}^2/2\sigma_j^2\right]} dx \tag{4.8}$$

where i=1,2, …., n and j=1, 2, …., m

n and m represent the number of samples and variables respectively. Figure 4.3 illustrates how probability of fault is estimated.

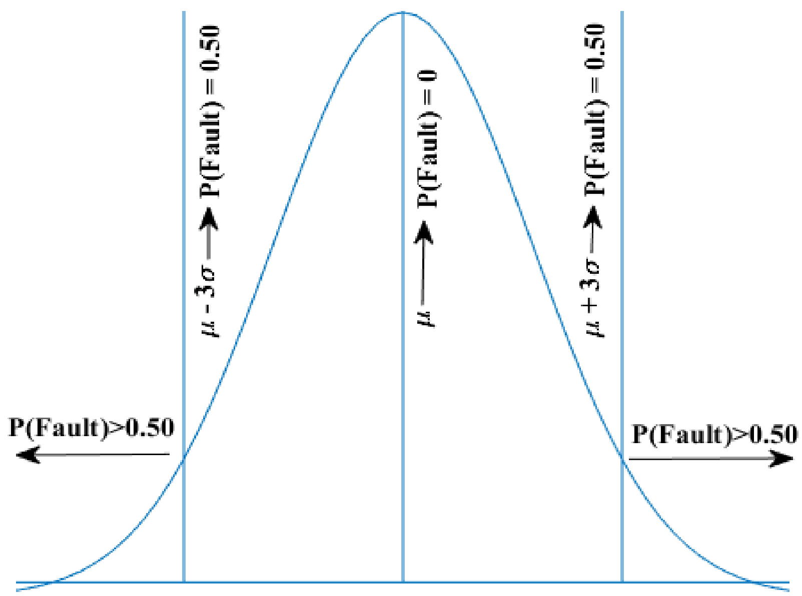


Figure 4.3: Visual depiction of the probability of fault estimation

To eliminate the complexity of monitoring individual control chart for each variable, a single consolidated control chart is required. We propose a new dynamic Bayesian anomaly index (DBAI) based control chart for fault detection. Consider, each node consists of two states in a DBN: one state represents the probability of being in a faulty state and another state implies to be in a normal state. The DBAI at any time can be calculated by summing up the probability of abnormal state for all the variables and then dividing by the number of variables. So, the upper and lower limit of DBAI will be 1 and 0 respectively. When the process fails, the DBAI value may reach up to 1.

The mathematical formulation of DBAI at any time slice, i can be expressed as:

$$\mathrm{DBAI_i} = \frac{1}{m}\sum_{j=1}^{m} \mathrm{P(abnormality)} \tag{4.9}$$

The threshold value for DBAI can be calculated as:

$$\mathrm{DBAI} = \frac{1}{mn}\sum_{i=1}^{n}\sum_{j=1}^{m} \mathrm{P(abnormality)} \tag{4.10}$$

In essence, the proposed methodology works in nine steps. The monitoring model development part consists of five steps:

Step 1: Prior knowledge and process flow diagrams (PFDs) are used to determine the qualitative network structure.

Step 2: Collection of historical process data in normal operation condition.

Step 3: Priors and CPTs are defined from this data set and integrated with the network for quantitative analysis.

Step 4: P(Fault) is calculated using Equations 4.7 and 4.8. It should be noted that the DBN is updated only when the probability of fault for two consecutive samples exceeds 0.50 for any variable to avoid false alarms caused by noise. Both hard and likelihood evidence are used to update the DBN to preserve the nature of the evidence. Updating any state of a node in the DBN with 100% true value is called hard evidence based updating, while likelihood evidence implies updating the DBN with probabilistic evidence (i.e. P(Fault)=0.75). Details of different types of evidence used to update a BN can be found in literature (Mrad et al., 2015; Pearl, 1988).

Step 5: Threshold of DBAI is calculated using Equation 4.10. The probability of fault for all the variables get steady after a brief period. These values are used for root cause diagnosis.

The online part of the methodology is comprised of four steps:

Step 6: P(Fault) of on-line samples is calculated using Equations 4.7 and 4.8. If P(Fault)>0.50 for two consecutive samples for any variable, the corresponding node is updated with hard or likelihood evidence in these specific time slices.

Step 7: Fault is detected when the threshold of DBAI is violated.

Step 8: Smoothing capacity of a DBN is utilized in root cause diagnosis. The DBAI value in previous time-slices may exceed the threshold after updating the DBN and detecting the fault due to smoothing inference. Average probability of fault for all the variables from the time-step when the DBAI value exceeded the threshold for the first time up to the detection time-step is calculated and compared with the steady state failure probability. Percentage change in the probability in the faulty state for all the variables are plotted. If a root node has the highest percentage increase in the probability in the faulty state, it is diagnosed as the root cause. When a child node gets the highest increase in the probability in the faulty state, root cause is diagnosed among its parent nodes.

Step 9: Fault propagation pathway is identified using the cause-effect relationships among the process variables. It includes the root, intermediate and leaf nodes. Fault propagation pathway starts with the root cause variable (cause) and ends up with the affected variables (effects). All the intermediate variables that lie between the cause and effect variables are also included in the fault propagation pathway.

4.4. Application of the Proposed Methodology

The application of the proposed DBN based FDD methodology is demonstrated on two process systems: a binary distillation column and a continuous stirred tank heater (CSTH) for four fault scenarios. Each test case consists of 500 samples. First 400 samples are fault free and used to calculate the threshold of DBAI. Fault is initiated at 401 samples in all four cases. Samples have been collected at 1 second interval.

4.4.1. Binary Distillation Column

The binary distillation column considered in this study has 40 stages that separates a mixture of relative volatility of 1.5 into products of 96% purity. Figure 4.4 shows the schematic diagram of a binary distillation column. Equilibrium condition in all stages, linearized liquid dynamics, constant pressure and relative volatility, no vapor holdup and total condenser have been considered to model the distillation unit (Skogestad, 1997).

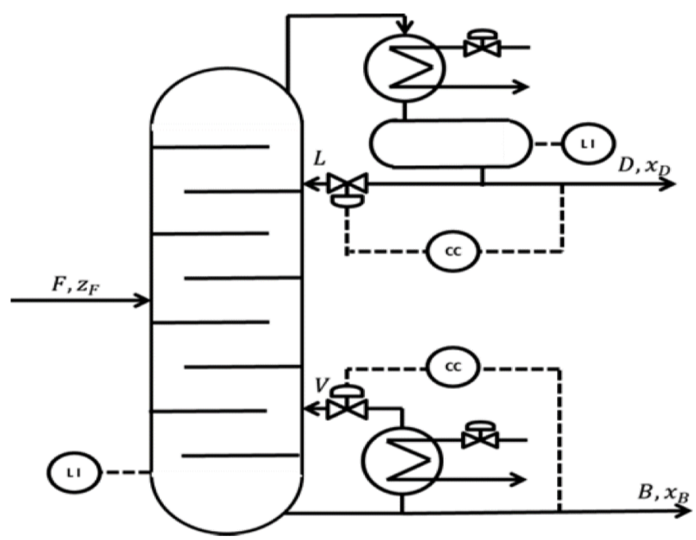


Figure 4.4: Schematic diagram of a binary distillation column

There are six inputs and four output variables in the distillation unit. The input variables are feed rate (F), feed composition (z_F), reflux flow rate (L), boil up flow rate (V), top product flow (D) and bottom product flow (B). The output variables are top composition (x_D), bottom composition (x_B), condenser holdup (M_D) and reboiler holdup (M_B). In this study, five process variables: feed rate (F), feed composition (z_F), reflux flow rate (L), top composition (x_D) and bottom composition (x_B) are monitored.

First, 400 samples are generated in normal operating condition. Prior knowledge and PFDs are used to construct the qualitative part of the network. CPTs are defined from MLE. The probability of fault for all the variables get steady from 16th to the rest of the samples since no evidence of fault is received for two consecutive samples for any variable in between these time-steps. These steady state failure probabilities for all five monitored variables are recorded. The threshold of DBAI is calculated as 0.2942 from these 400 samples. The developed DBN model is shown in Figure 4.5. GeNIe 2.0 is used for modeling the DBN. Two fault scenarios have been generated. The fault descriptions are shown in Table 4.2.

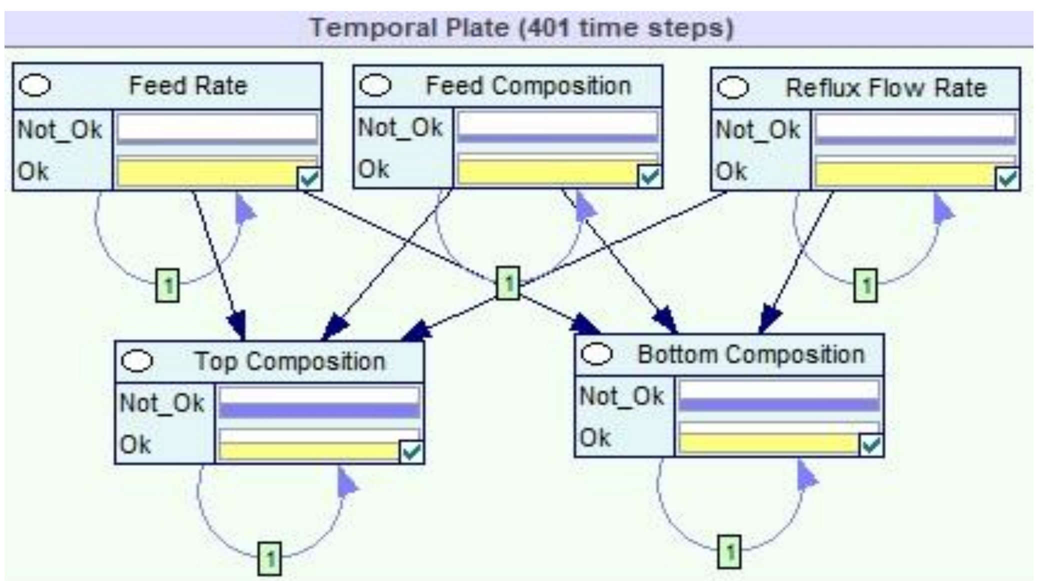


Figure 4.5: DBN model for a binary distillation column

Table 4.2: Fault descriptions in a binary distillation column

Fault Scenario	Description	Root cause
A1	5% feed loss	Feed rate
A2	Random variation in feed composition	Feed composition

4.4.1.1. 5% Feed Loss (Fault Scenario A1)

A step type signal is used to model this fault scenario. Figure 4.6 shows the probability of fault for all the variables. It can be seen that the feed rate gets more perturbation than other two input variables from 401 samples. Although it exceeds the CL twice at 431^{st} and 451^{st} samples, it does not exceed the limiting value of 0.50 for two consecutive samples. As a result of continuous feed loss, both top and bottom composition get reduced. Among the five monitored variables, top composition first exceeds the individual CL at 424^{th} sample. Bottom composition also crosses the CL at 445^{th} sample. Since no corrective measure was taken, these output variables do not come back inside the normal operating zone.

All these evidence are provided in the developed DBN model (Figure 4.7). Then, DBAI value for all time slices are calculated and compared with the threshold value. The fault is detected at 425^{th} sample with a sharp jump in the DBAI value (Figure 4.8). The DBAI value never returns under the threshold for the rest of the test samples.

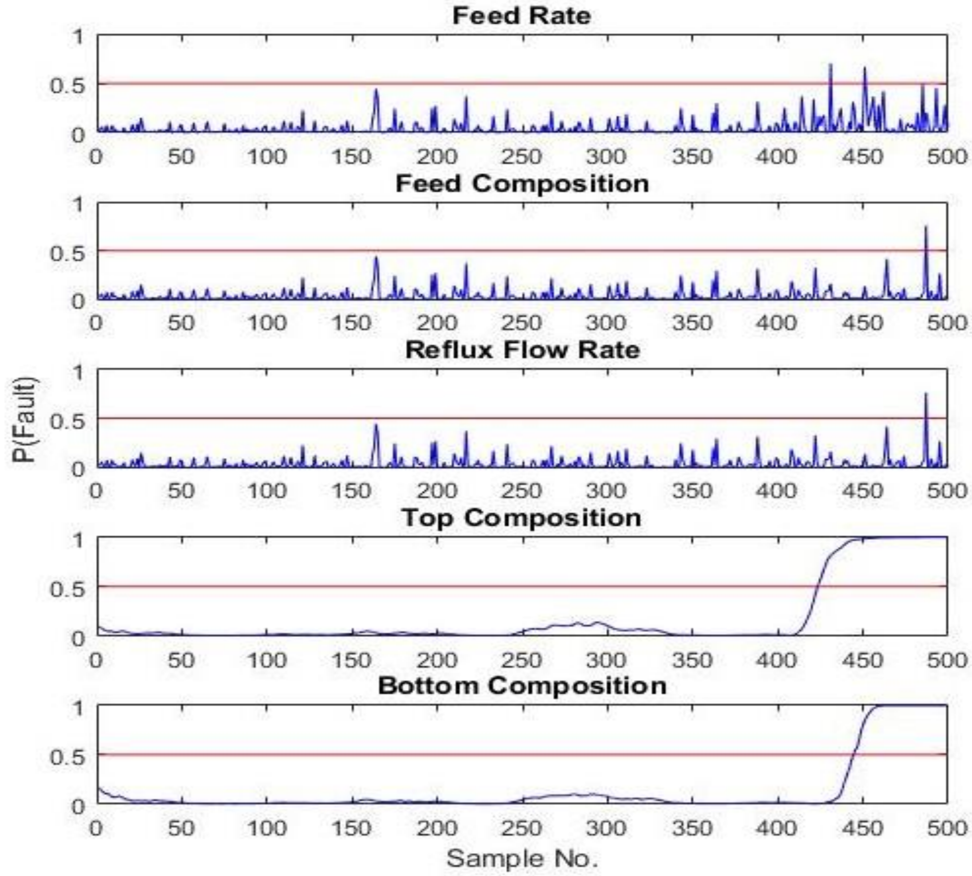


Figure 4.6: Probability of fault for fault scenario A1 in a binary distillation column

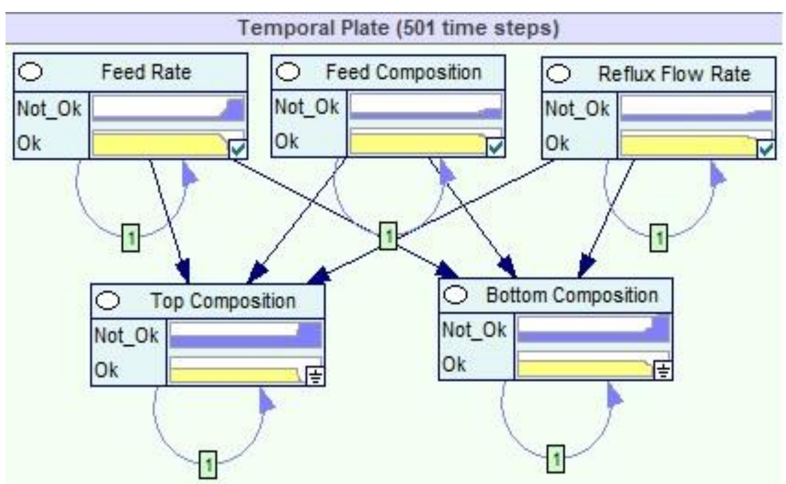


Figure 4.7: DBN for fault scenario A1 in a binary distillation column

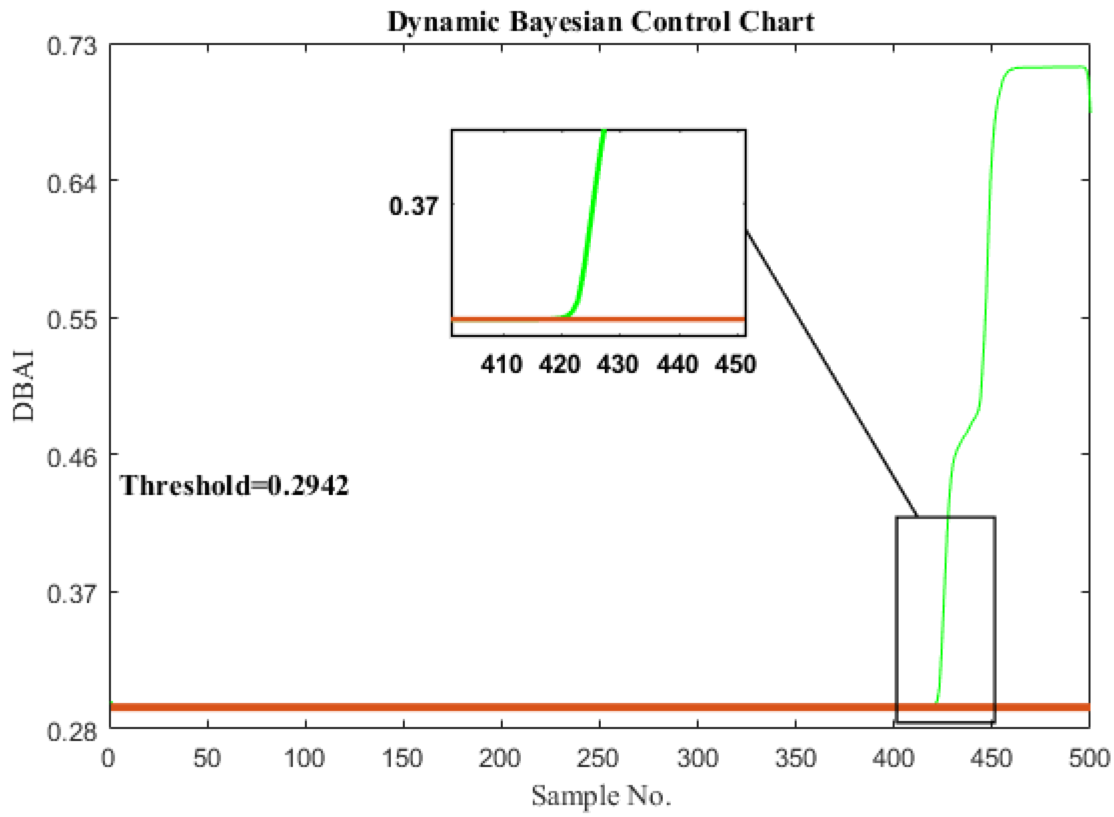


Figure 4.8: Dynamic Bayesian control chart for fault scenario A1 in a binary distillation column

The next step is to diagnose the root cause of the fault. Due to smoothing inference, the DBAI value first exceeds the threshold at 417^{th} sample after updating the BN with the fault information received for top composition at 424^{th} and 425^{th} samples. The average probability of fault from 417^{th} to 425^{th} time-steps for the monitored variables are computed and compared with the steady state values. Figure 4.9 shows the percentage change in the probability in the faulty state. Top composition has the highest increase (3.97%) in the faulty state. However, it is a child node. It has three parent nodes: feed rate, feed composition and reflux flow rate. Among these three input variables, feed rate has the highest increase (3.5%) in the faulty state and can be accurately diagnosed as the root cause.

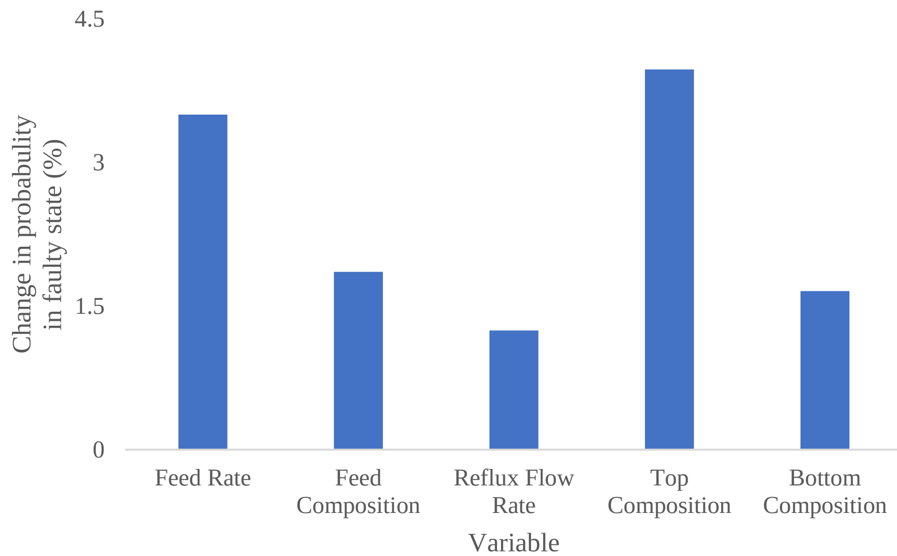


Figure 4.9: Root cause diagnosis for fault scenario A1 in a binary distillation column

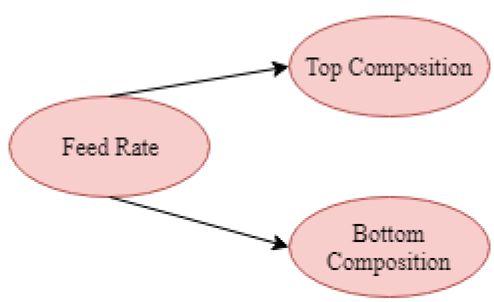


Figure 4.10: Fault propagation pathway for fault scenario A1 in a binary distillation column

After diagnosing feed rate as the root cause of the fault, we check for the fault propagation pathway. Feed rate has two child nodes: top composition and bottom composition. A fault in the feed rate will affect both, and these variables are selected as the terminating nodes in the fault propagation pathway (Figure 4.10).

4.4.1.2. Random Variation in Feed Composition (Fault Scenario A2)

A random variation starts in the feed composition after 400 samples of normal operation. A lot of fluctuations in the probability of fault is observed in Figure 4.11 for feed composition. Feed composition exceeds the individual CL from 407th sample for three consecutive samples. Top composition crosses the CL at 422nd sample and returns within the limit at 498th sample. Bottom composition shows higher fluctuation than top composition.

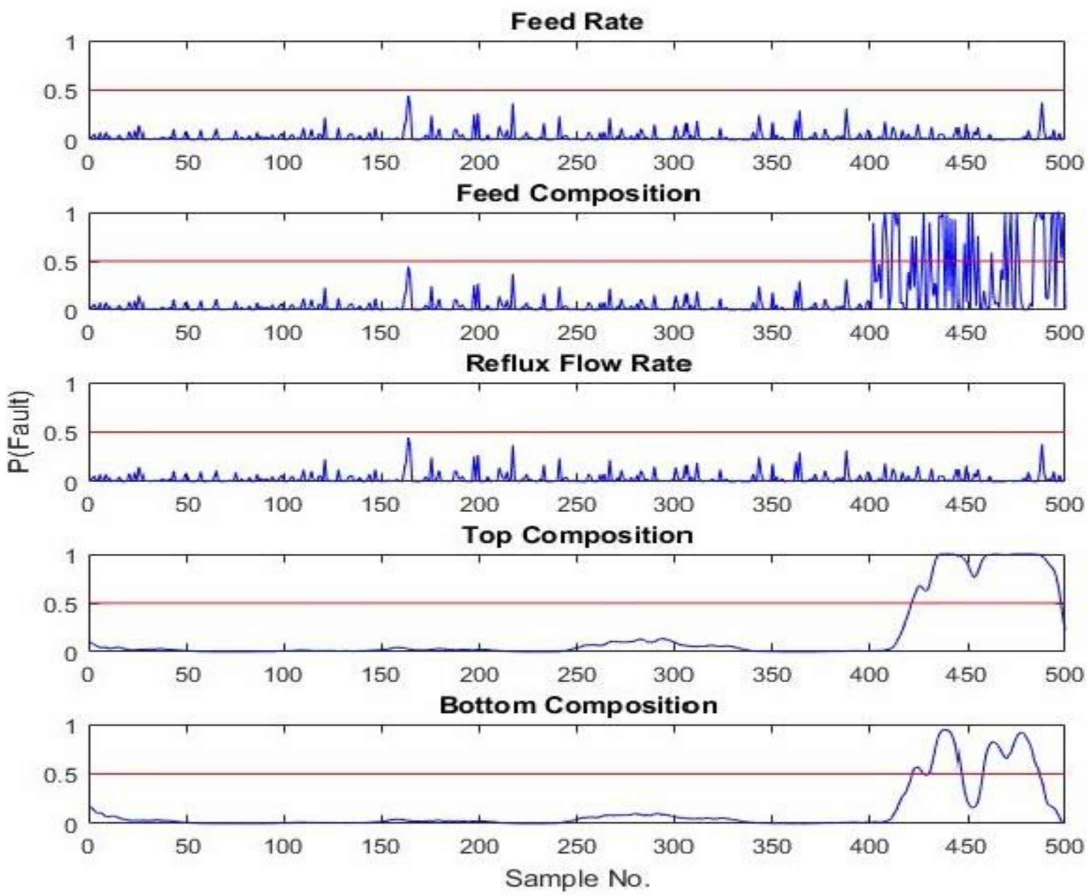


Figure 4.11: Probability of fault for fault scenario A2 in a binary distillation column

This new fault information is used to update the DBN. Figure 4.12 shows the DBN for a random variation in feed composition. The DBAI values for all the time slices are calculated and plotted in the dynamic Bayesian control chart (Figure 4.13). It exceeds the threshold immediately after updating the DBN with the evidence of abnormal behavior received from the feed composition. If the DBN structure was determined only using the self-rolling arcs on the state variables (top and bottom composition), a delayed detection would happen in this fault scenario. Fluctuating DBAI values are observed due to the random nature of the fault. However, it never returns below the threshold limit.

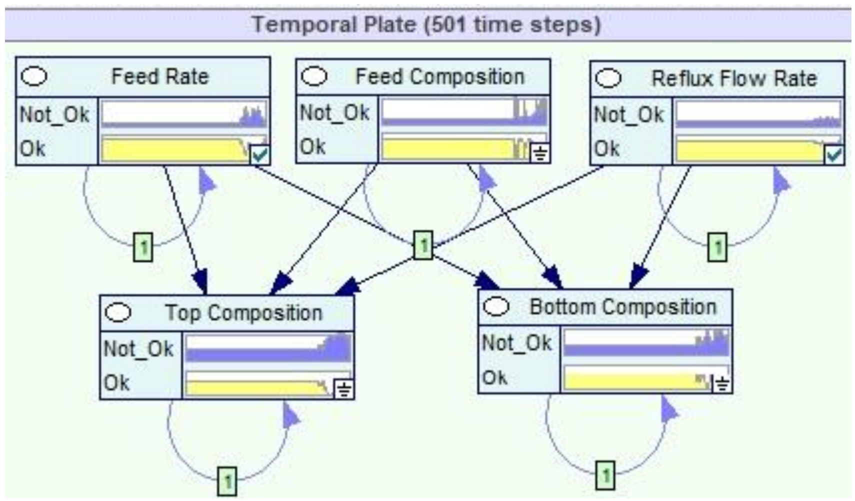


Figure 4.12: DBN for fault scenario A2 in a binary distillation column

To find out the root cause of the abnormality, we look for the time slice when smoothing inference first reported anomalous activity. It is observed that the DBAI value first exceeds the threshold at 401st sample. The average probability of fault for all the variables from 401st to 408th samples are calculated and compared with the steady state values. Figure 4.14 shows the percentage change in the probability to be in the faulty state. It can be seen that the feed rate and reflux flow rate have no significant increase in the probability to be in a faulty state, while feed composition, top composition and bottom composition show increased probability. Feed composition has the highest percentage increase and can be successfully diagnosed as the root cause of the process abnormality.

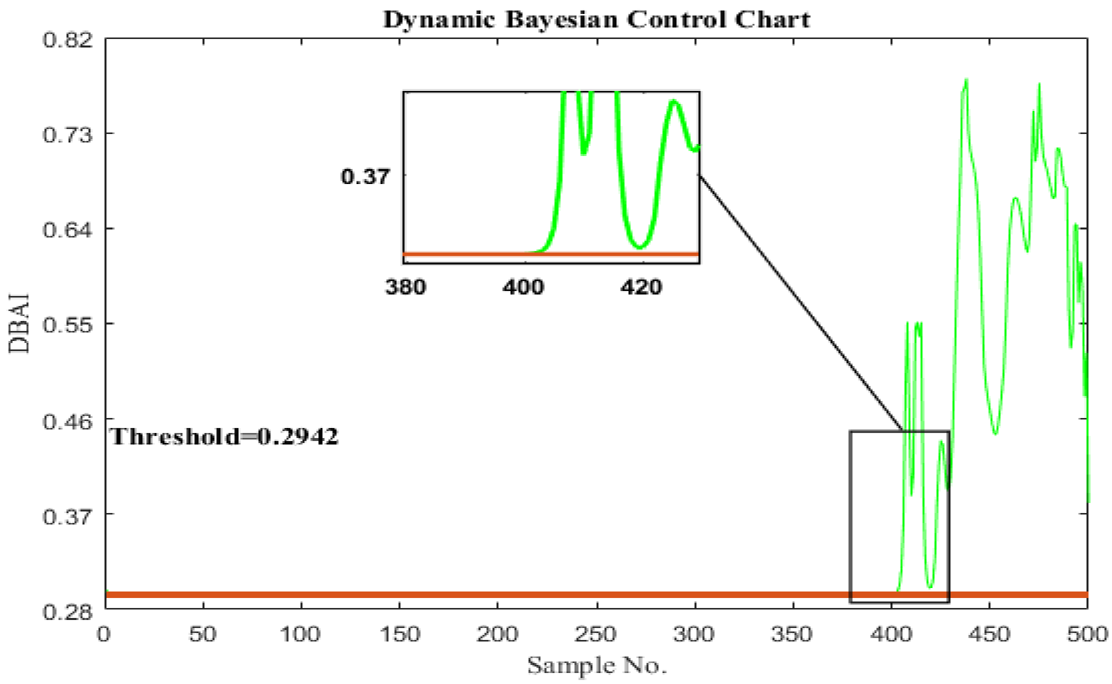


Figure 4.13: Dynamic Bayesian control chart for fault scenario A2 in a binary distillation column

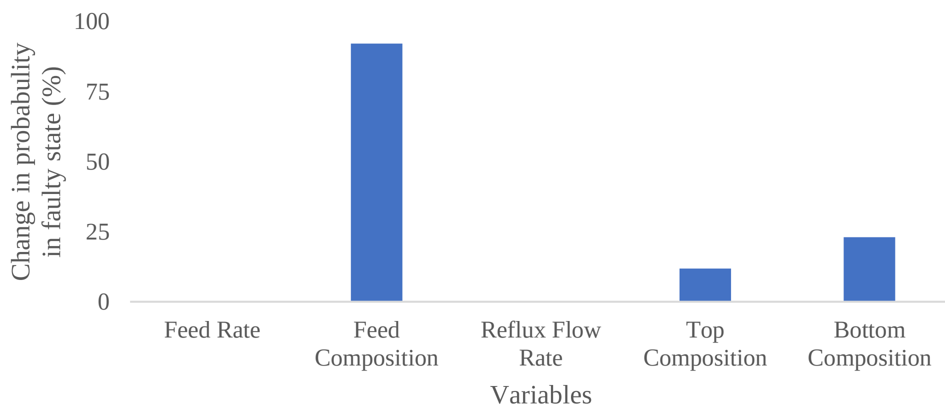


Figure 4.14: Root cause diagnosis for fault scenario A2 in a binary distillation column

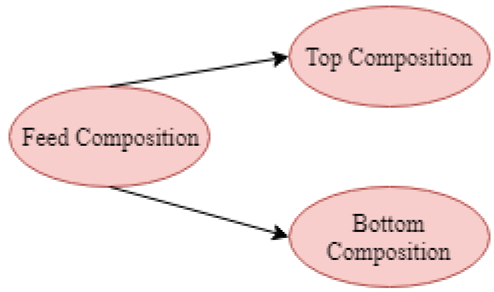


Figure 4.15: Fault propagation pathway for fault scenario A2 in a binary distillation column

The fault propagation pathway can be obtained from Figure 4.14, as it can identify the affected variables due to a fault in the feed composition. This fault affects both top and bottom composition. The fault propagation pathway is shown in Figure 4.15.

4.4.2. Continuous Stirred Tank Heater (CSTH)

The continuous stirred tank heater (CSTH) is a common unit in the chemical process industry. The CSTH model considered in this study was developed by (Thornhill et al., 2008). It is built using the first principal models and possess the real disturbance data from a pilot plant located in the University of Alberta. The schematic diagram of a CSTH is shown in Figure 4.16. There are six variables: cold water valve demand, steam valve demand, hot water valve demand, level, output water flow rate and temperature. The first three variables are the inputs in the system, while the rest of the three variables are the outputs. All measurements are converted to electrical signal and presented on a scale of 4-20 mA. However, level, output water flow rate and temperature can also be measured in cm, m^3/sec and ºC respectively. In a CSTH, cold water comes to the tank to maintain the level. Steam valve mainly controls the temperature of the tank by providing required quantity of steam to maintain the desired temperature. Cold water flow has minimal effect on the temperature. Hot water supply can also influence both level and temperature. Two PID controllers are controlling the cold water and steam valve demand depending on the set points of level and temperature. In this experiment, hot water flow is kept constant at 5.5 mA, while set points for level and temperature are selected as 12 and 10.50 mA respectively. There is a heating coil at the bottom of the tank which makes the model highly non-linear.

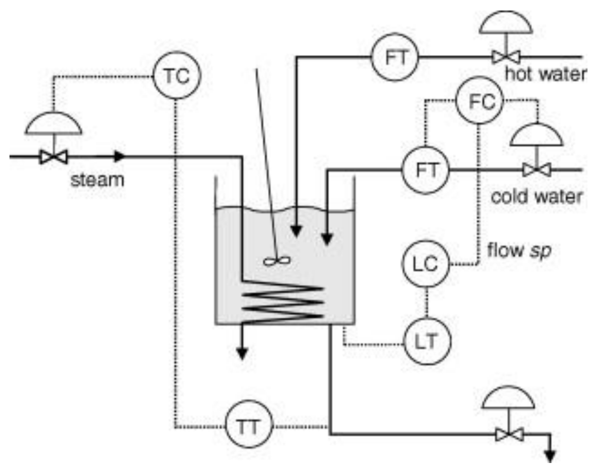


Figure 4.16: Schematic diagram of a continuous stirred tank heater

Prior knowledge and PFDs are utilized to construct the qualitative part of the DBN. 400 samples are generated to define the priors and CPTs. The probability of fault for all the variables become steady after 22nd sample. These steady state failure probabilities for the monitored variables are recorded. The threshold of DBAI is calculated as 0.2546. The developed DBN model of a CSTH is shown in Figure 4.17. Two fault scenarios have been considered in this paper. The tested fault scenarios in a CSTH are shown in Table 4.3.

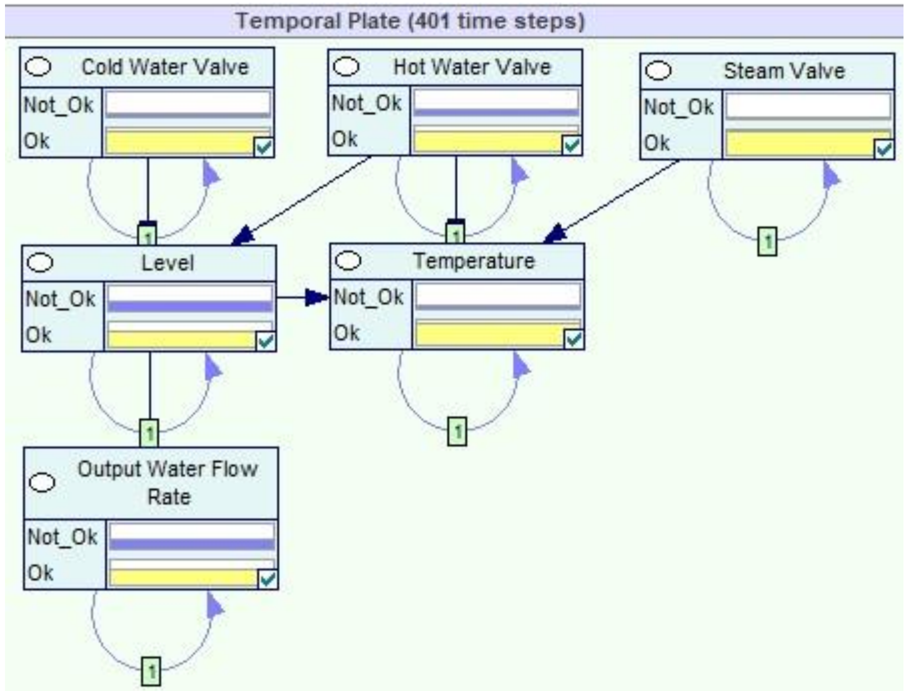


Figure 4.17: DBN model for a CSTH

Table 4.3: Fault descriptions in a CSTH

Fault Scenario	Description	Root cause
B1	Gradual increase in cold water valve demand	Cold water valve
B2	Steam valve stiction	Steam valve

4.4.2.1. Gradual Increase in Cold Water Valve Demand (Fault Scenario B1)

A ramp type signal is used to generate this fault scenario. First probability of fault for all the variables is calculated (Figure 4.18). Cold water valve, level and output water flow rate are highly affected by this fault and keep fluctuating around the individual CL. Increase in cold water valve demand results in increasing cold water supply. Hence, both level and output water flow rate will be increased. To maintain the set point, controller reduces the valve opening. Once level comes to 10.50 mA position, this fault again starts affecting the valve opening.

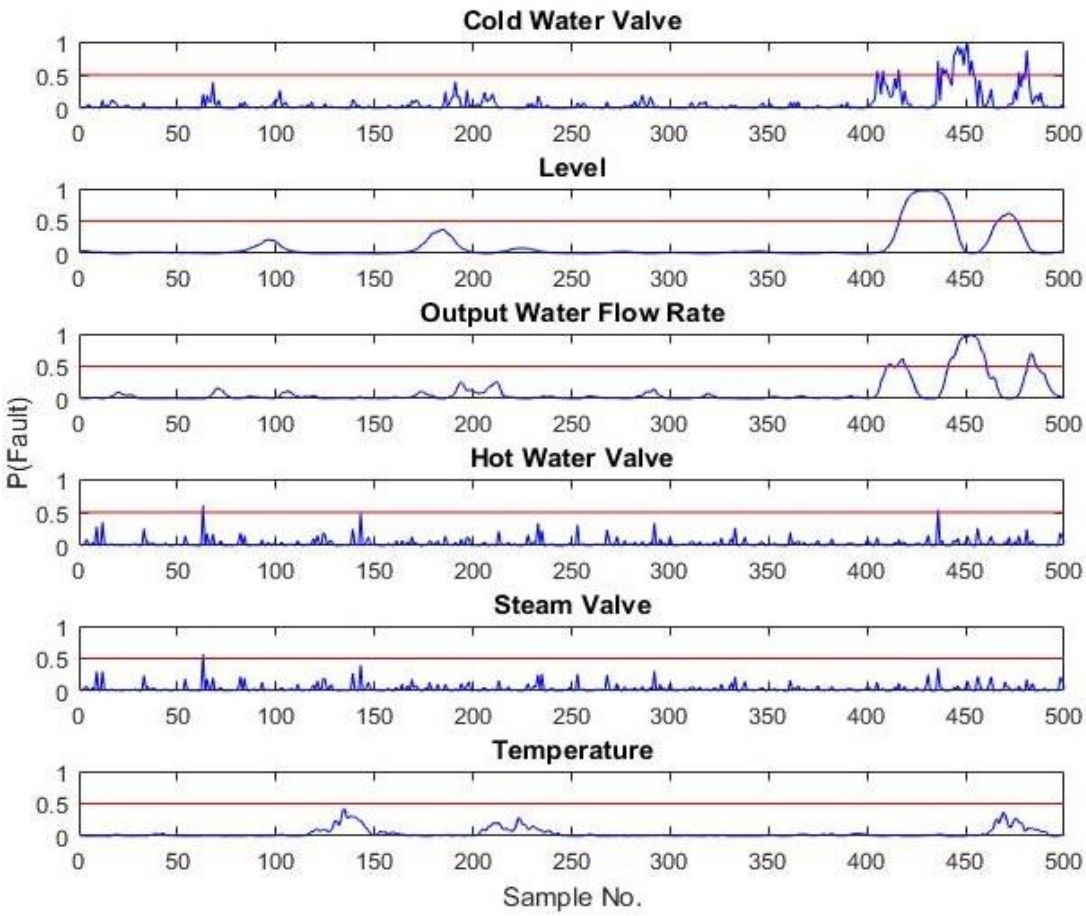


Figure 4.18: Probability of fault for fault scenario B1 in a CSTH

Output water flow rate exceeds the individual CL at 411th, 412th and 413th samples and come back within the normal operating zone again. Probability of fault for level first exceeds 0.50 at 417th sample. Although cold water valve demand exceeds the CL at 405th sample, due to the controller action it immediately comes back to normal operating regime. It violates the CL at 438th sample for four consecutive samples. Steam valve, temperature and hot water valve remain inside the fault free zone for the test samples.

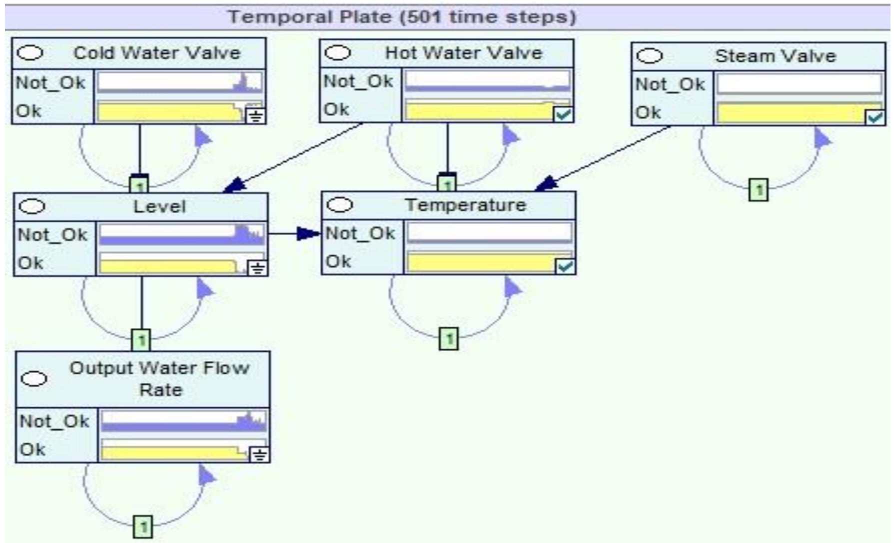


Figure 4.19: DBN for fault scenario B1 in a CSTH

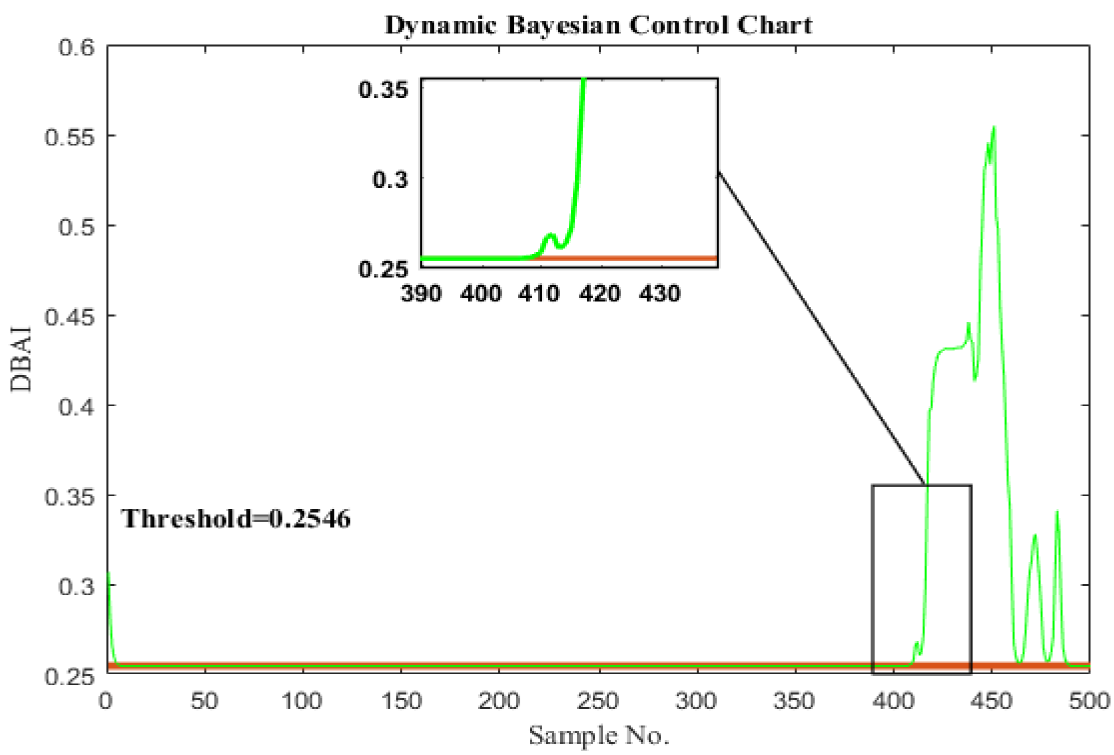


Figure 4.20: Dynamic Bayesian control chart for fault scenario B1 in a CSTH

Whenever a variable exceeds the individual CL (P(Fault)=0.50) for two consecutive samples, evidence are provided in the DBN. In this case, collected evidence of fault for cold water valve, level and output water flow rate are used to update the DBN. Figure 4.19 shows the DBN for a gradual increase in cold water valve demand in a CSTH.

Figure 4.20 shows the dynamic Bayesian control chart for this fault scenario. The fault is detected at 412th sample as soon as it receives the evidence of a fault from output water flow rate. Although the probability of fault for the monitored variables keep fluctuating, the DBAI value continuously stay above the threshold till 489th sample. It returns under the threshold at 490th sample and never exceeds again for the rest of the test samples.

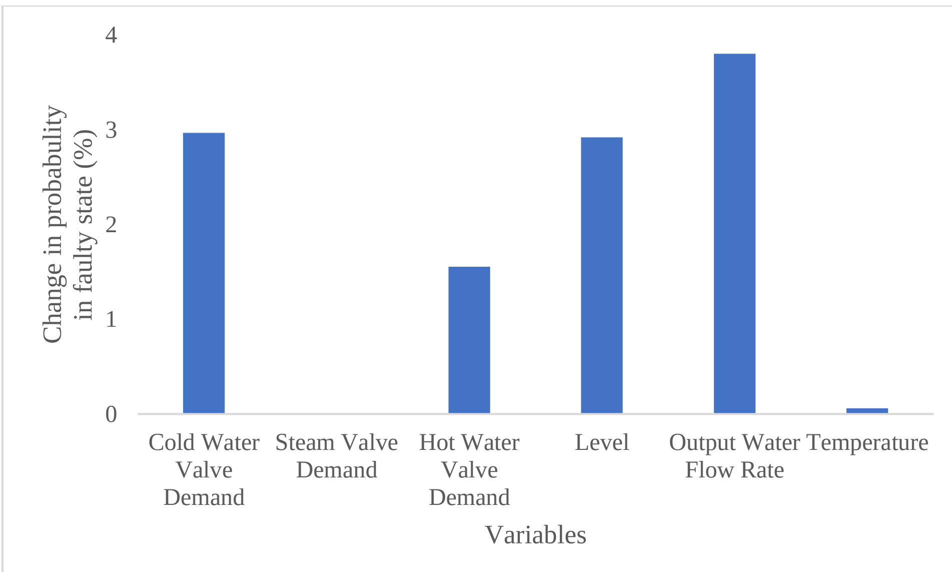


Figure 4.21: Root cause diagnosis for fault scenario B1 in a CSTH

After providing the evidence at 411th and 412th samples, it is observed that the DBAI violates the threshold at 407th sample. The average probability of fault for the variables from 407th to 412th samples are calculated and compared with that of steady state probability of fault. Figure 4.21 shows the percentage change in the probability of fault for all the variables. It shows that this fault does not affect steam valve demand and temperature. Cold water valve, level and output water flow rate have the most prominent increase in the faulty state. Output water flow rate has the highest increase (3.80%) in the faulty state. Level is the only parent node of output water flow rate, which has an increase of 2.92%. Level has two parent nodes: cold water valve demand and hot water valve demand. Among these, cold water valve has the highest increase (2.97%) in the faulty state and can be diagnosed as the root cause of the fault.

Level is the only intermediate variable between the root cause variable: cold water valve demand and terminating variable: output water flow rate, and the fault propagation pathway can be drawn as shown in Figure 4.22.

Figure 4.22: Fault propagation pathway for fault scenario B1 in a CSTH

4.4.2.2. Steam Valve Stiction (Fault Scenario B2)

Valve stiction is a frequently encountered problem in the process industries. In this fault scenario, steam valve gets stuck at 6.75 mA position from 401st sample till the rest of the test samples. As a result, temperature get increased and crosses the threshold limit at 417th sample and never returns below the CL. It is because the controller is unable to handle this type of operational problem. Human intervention is required to solve it. The probability of fault never rises over 0.50 for steam valve demand, since the valve got stuck within the normal operating zone. Fluctuation in cold water valve demand is observed 477th to 481st sample, which affects the output water flow rate.

Figure 4.24 shows the DBN for this fault scenario. It can show which sub-system of the process system is highly affected by this fault. The dynamic Bayesian control chart detects the fault at 418th sample, when the likelihood evidence received from temperature, are used to update the DBN for 417th and 418th samples (Figure 4.25). The DBAI stays well above the threshold for the remaining test samples.

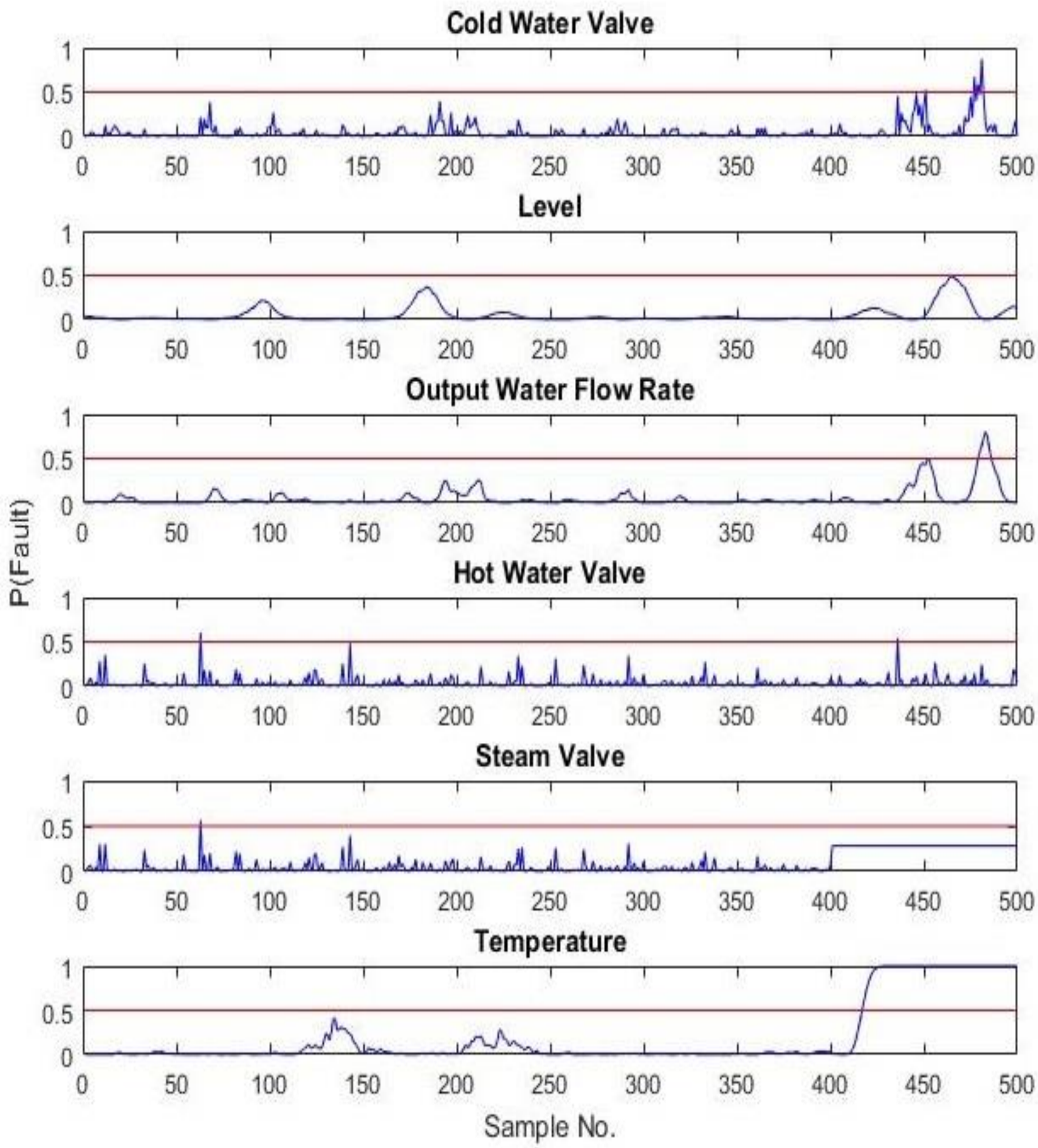


Figure 4.23: Probability of fault for fault scenario B2 in a CSTH

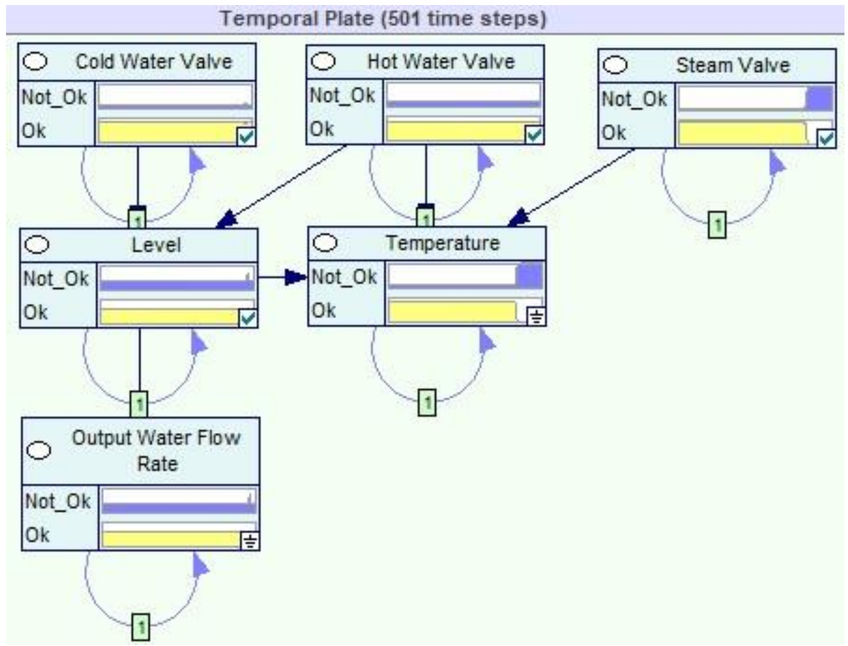


Figure 4.24: DBN for fault scenario B2 in a CSTH

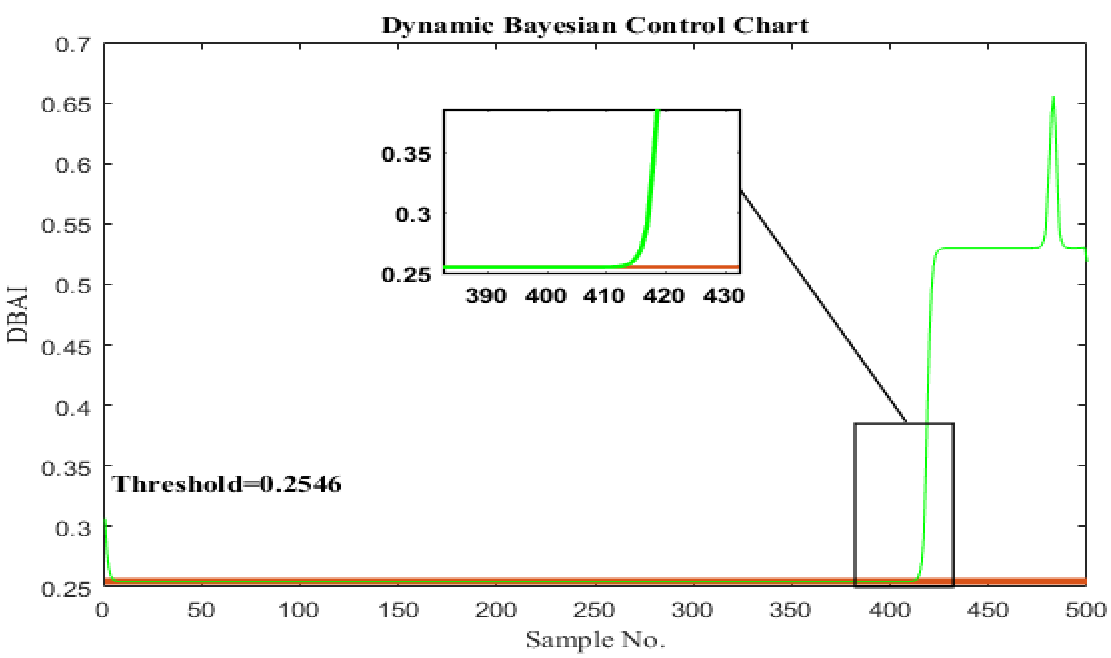


Figure 4.25: Dynamic Bayesian control chart for fault scenario B2 in a CSTH

To diagnose the root cause, we look for the time-step when DBAI first exceeds the threshold. It is observed that it crosses the threshold at 413th sample. Then, average probability of fault for all the variables are calculated from 413th to 418th samples and compared with the steady state values. Figure 4.26 shows the percentage change in the faulty state for all the monitored variables. It is observed that steam valve has the highest percentage increase (10.26%) in the faulty state and can be swiftly diagnosed as the root cause. Its only child node: temperature has 9.41% increase in the faulty state and can be identified as the terminating variable in the fault propagation pathway (Figure 4.27).

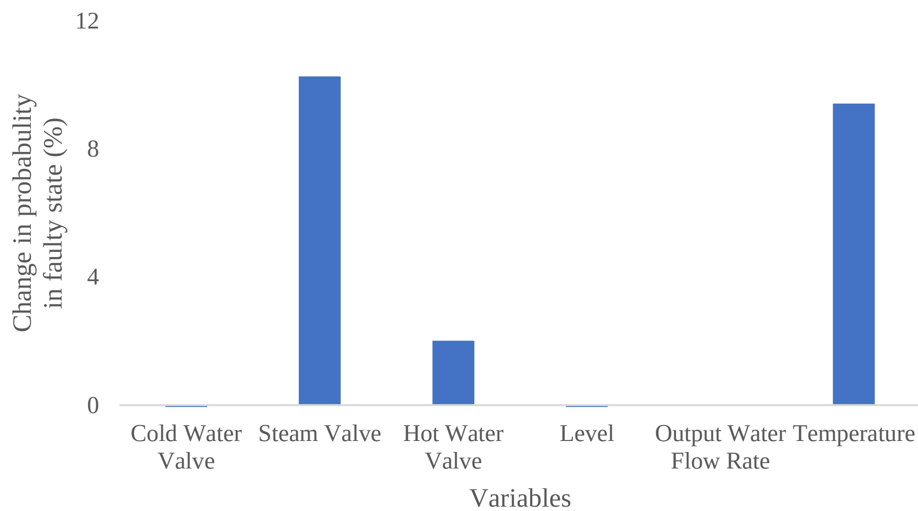


Figure 4.26: Root cause diagnosis for fault scenario B2 in a CSTH

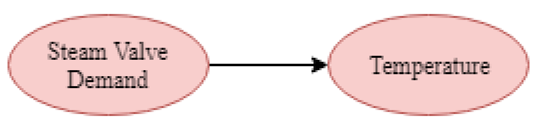


Figure 4.27: Fault propagation pathway for fault scenario B2 in a CSTH

4.5. Results and Discussion

The proposed methodology is compared with a univariate monitoring technique (the Shewhart chart), and multivariate FDD tools (PCA and a static BN) for various fault scenarios. The performance measurement criteria are false alarm rate (FAR), detection rate

(DR) and accurate diagnosis capacity. FAR is the number of trained samples exceeding the threshold divided by the total number of trained samples, while DR is the number of faulty samples successfully detected divided by the total number of faulty samples. The comparative study on these four test cases is shown in Table 4.4.

Mean (μ) and standard deviation (σ) for all the variables are calculated from 400 trained samples. To have a fair comparison, UCL and LCL of the Shewhart chart for all the variables have been selected as $\mu+3\sigma$ and $\mu-3\sigma$ respectively, and when two consecutive samples of any variable exceed the UCL or LCL, fault is detected, and this variable is considered as the root cause of the fault since the Shewhart chart cannot provide any causal information.

Two statistics are commonly used in PCA based monitoring: Hotelling's T^2 and squared prediction error (SPE) or Q statistics. T^2 indicates how far a sample lies from the center of the feature space, while SPE is the measurement of lack of fitness of a sample from the residual space. Whenever any test sample exceeds the threshold of T^2 or SPE value fault is detected, and multivariate contribution plots are generated at the detection point. Required number of principal components (PCs) to monitor the process is selected from the cumulative percent variance (CPV) approach. Two PCs can explain 99.85% of total process variation for a binary distillation column while three PCs can explain 86.78% process variation in a CSTH. These PCs are used to build the monitoring model.

The threshold of Bayesian anomaly index (BAI) for fault detection by the static BN is calculated for time, t=0 using Equation 4.9. This index has been also calculated from the arithmetic mean of abnormal states of all the variables, and it is computed as 0.3092 and

0.3899 for the binary distillation column and CSTH respectively. Evidence of fault is generated using Equations 4.7 and 4.8 to update the BN. For an equitable comparison, the BN is only updated when two consecutive samples exceed the individual CL for any variable. Root cause diagnosis technique slightly varies from DBN, since static BN has no smoothing inference. In this case, percentage change in the probability to be in the faulty state for all the monitored variables between the steady state and updated BN is used to diagnose the root cause of the fault. Cause-effect relationship is utilized if required.

Table 4.4: Comparative performance of the Shewhart chart, PCA, static BN and the proposed DBN based methodology (best performance is marked bold for each condition)

<table>
<tr><th>Process Model</th><th>Fault Description</th><th>Performance Criterion</th><th>Shewhart Chart</th><th>PCA-T^2</th><th>PCA-SPE</th><th>Static BN</th><th>DBN</th></tr>
<tr><td rowspan="6">Distillation Column</td><td rowspan="3">5% feed loss (Fault Scenario A1)</td><td>FAR (%)</td><td>0</td><td>4</td><td>5</td><td>0</td><td>1</td></tr>
<tr><td>DR (%)</td><td>76</td><td>72</td><td>99</td><td>76</td><td>76</td></tr>
<tr><td>Diagnosis</td><td>No</td><td>No</td><td>Yes</td><td>Yes</td><td>Yes</td></tr>
<tr><td rowspan="3">Random variation in feed composition (Fault Scenario A2)</td><td>FAR (%)</td><td>0</td><td>4</td><td>5</td><td>0</td><td>1</td></tr>
<tr><td>DR (%)</td><td>75</td><td>85</td><td>98</td><td>75</td><td>93</td></tr>
<tr><td>Diagnosis</td><td>Yes</td><td>No</td><td>Yes</td><td>Yes</td><td>Yes</td></tr>
<tr><td rowspan="6">CSTH</td><td rowspan="3">Gradual increase in cold water valve demand (Fault Scenario B1)</td><td>FAR (%)</td><td>0.25</td><td>4.25</td><td>12.75</td><td>0.25</td><td>1.75</td></tr>
<tr><td>DR (%)</td><td>36</td><td>70</td><td>61</td><td>36</td><td>78</td></tr>
<tr><td>Diagnosis</td><td>No</td><td>No</td><td>No</td><td>Yes</td><td>Yes</td></tr>
<tr><td rowspan="3">Steam valve stiction (Fault Scenario B2)</td><td>FAR (%)</td><td>0.25</td><td>4.25</td><td>12.75</td><td>0.25</td><td>1.75</td></tr>
<tr><td>DR (%)</td><td>83</td><td>76</td><td>99</td><td>83</td><td>83</td></tr>
<tr><td>Diagnosis</td><td>No</td><td>No</td><td>Yes</td><td>Yes</td><td>Yes</td></tr>
</table>

The Shewhart control chart and the static BN has the lowest FAR for all four faulty scenarios. PCA-SPE provides the maximum FAR in both process models. Although the DBN gets updated taking evidence from the univariate monitoring, it provides higher FAR than the Shewhart chart because of the initial time required to get stable in the probabilistic domain. PCA-SPE provides the best detection rate in the fault scenario A1, A2 and B2. Static BN has the same DR as the Shewhart control chart. DBN gives better DR than the static BN in the fault scenario A2 and B1. It provides the best detection rate in case of a gradual increase in the cold water valve demand in a CSTH because of the prediction inference.

PCA-T^2 provides the worst diagnostic performance. PCA-SPE can identity the root cause in the fault scenario A1, A2 and B2 while the univariate Shewhart chart can diagnose the root cause for a random variation in feed composition in the binary distillation column. Both static and dynamic BN can consistently diagnose the root cause.

4.6. Conclusion

In this paper, a DBN based process fault detection, root cause diagnosis, and fault propagation pathway identification methodology is proposed. Fault information from the monitored variables has been modelled as a Gaussian distribution, and evidence are generated from the output of these distributions using the cumulative Gaussian distribution to update the DBN. A new dynamic Bayesian anomaly index based control chart is developed to detect the fault. Smoothing inference of a DBN and cause-effect relationship among the process variables are utilized to diagnose the root cause of the fault. Furthermore, fault propagation pathway is identified taking the advantage of networked process monitoring. The methodology has been tested for four fault scenarios in two process models. The result suggests that the methodology can provide desired FDD performance. Although there is a delay in detecting the fault compared to PCA, it persistently diagnoses the root cause of the fault which is not obtained by the conventional univariate or multivariate monitoring methods. The main contributions of this work are:

- Converting the continuous process variable into a probability index appropriate for using as evidence to the DBN.
- A new dynamic Bayesian anomaly index (DBAI) for process fault detection.
- Use of the smoothing capacity of DBN for robust fault diagnosis.
- Fault propagation pathway showing how fault propagated in the system from the root variable, thus helping in the recovery of the system.

Chapter 5

Summary Conclusions and Future Work Scopes

A hybrid methodology combining PCA, BN with multiple likelihood evidence is presented in the first part of this research which provides a comprehensive solution for real time fault detection and diagnosis. This methodology updates the BN with more fault information to ensure precise diagnosis that is not obtainable by conventional PCA and PCA-BN with hard evidence based approach. Furthermore, a DBN based scheme is proposed which enables FDD in a single tool which is still a challenge in the existing literatures. It discusses how continuous process data can be converted to generate evidence to update the DBN. A new DBAI based control chart is proposed for fault detection. Both these methods can provide in-depth information about the fault propagation pathway which may help the operators to restore the process quickly from the abnormal state. Several process models are used to demonstrate the efficiency of the proposed FDD algorithms by comparing with some other conventional approaches.

5.1. Conclusions

- PCA is sensitive to the process variations. Although it takes longer time to detect the lower magnitude faults, it can detect the faults as soon as the process is significantly affected.
- PCA cannot guarantee accurate diagnosis. However, it provides significant information about a fault which can be utilized to diagnose the root cause of the fault (Section 3.4.2.2).
- Updating the BN from PCA contributions using heuristic rule based method can lead to false diagnosis. A BN needs to be updated with multiple likelihood evidence to ensure accurate diagnosis. However, PCA contributions are required to be rescaled in this approach (Sections 3.4.1.1 and 3.4.2.1).
- DBN can provide a comprehensive solution in FDD (Chapter 4).
- Smoothing inference of a DBN secures robust root cause diagnosis (Sections 4.4.1.1, 4.4.1.2, 4.4.2.1 and 4.4.2.2).
- Fault propagation pathway can be identified using the predictive feature of a BN as well as cause-effect relationships among the process variables.

5.2. Future Work Scopes

- PCA is optimal when process data follow a multivariate Gaussian distribution. Other MSPM tools such as semi-parametric PCA, KPCA can be used as the first stage detection tool which will ensure robust detection in case of non-Gaussian and non-linear process data.

- Sensor faults are often encountered in the process industries. Including sensor fault module in detection stage will strengthen the proposed methodologies. A bank of Kalman filters can be used to identify the sensor faults.
- Prior knowledge and PFDs have been used in qualitative network construction. Different algorithm for determining the structure of the Bayesian network from data mentioned in Section 2.3 can be applied to build the network.
- Kernel distribution can be used to generate the evidence to update the DBN. This will allow to make assumptions about any specific data distribution.
- Detection by a DBN can be made earlier by reducing the span of the normal operating zone.
- Loss function can be integrated with these methods to develop real time dynamic risk management frameworks.

References

Aamodt, A., Plaza, E., 1994. Case-based reasoning: Foundational issues, methodological variations, and system approaches. AI Commun. 7, 39–59.

Abimbola, M., Khan, F., Khakzad, N., Butt, S., 2015. Safety and risk analysis of managed pressure drilling operation using Bayesian network. Saf. Sci. 76, 133–144.

Adedigba, S.A., Khan, F., Yang, M., 2017. Dynamic Failure Analysis of Process Systems Using Principal Component Analysis and Bayesian Network. Ind. Eng. Chem. Res. 56, 2094–2106.

Ali, J.M., Hoang, N.H., Hussain, M.A., Dochain, D., 2015. Review and classification of recent observers applied in chemical process systems. Comput. Chem. Eng. 76, 27–41.

Ashley, K.D., 2003. Case-Based Reasoning Research and Development: 5th International Conference on Case-Based Reasoning, ICCBR 2003, Trondheim, Norway, June 23-26, 2003, Proceedings. Springer Science & Business Media.

Atoui, M.A., Verron, S., Kobi, A., 2016. A Bayesian network dealing with measurements and residuals for system monitoring. Trans. Inst. Meas. Control 38, 373–384.

Azhdari, M., Mehranbod, N., 2010. Application of Bayesian belief networks to fault detection and diagnosis of industrial processes, in: Chemistry and Chemical Engineering (ICCCE), 2010 International Conference on. IEEE, pp. 92–96.

Bakshi, B.R., 1998. Multiscale PCA with application to multivariate statistical process monitoring. AIChE J. 44, 1596–1610.

Bea, R., 2011. Final Report on the Investigation of the Macondo Well Blowout. Technical report, Deepwater Horizon Study Group, Department of Civil and Environmental Engineering, University of California Berkeley, Berkeley, CA, USA.

Board, M., 2012. Macondo Well Deepwater Horizon Blowout: Lessons for Improving Offshore Drilling Safety. National Academies Press.

Bobbio, A., Portinale, L., Minichino, M., Ciancamerla, E., 2001. Improving the analysis of dependable systems by mapping fault trees into Bayesian networks. Reliab. Eng. Syst. Saf. 71, 249–260.

Che Mid, E., Dua, V., 2017. Model-Based Parameter Estimation for Fault Detection Using Multiparametric Programming. Ind. Eng. Chem. Res. 56, 8000–8015.

Chen, L.W., Modarres, M., 1992. Hierarchical decision process for fault administration. Comput. Chem. Eng. 16, 425–448.

Cheng, J., Bell, D.A., Liu, W., 1997. Learning belief networks from data: An information theory based approach, in: Proceedings of the Sixth International Conference on Information and Knowledge Management. ACM, pp. 325–331.

Chiang, L., Russell, E., Braatz, R., 2000. Fault diagnosis in chemical processes using Fisher discriminant analysis, discriminant partial least squares, and principal component analysis. Chemom. Intell. 50, 243–252.

Chiang, L.H., Kotanchek, M.E., Kordon, A.K., 2004. Fault diagnosis based on Fisher discriminant analysis and support vector machines. Comput. Chem. Eng. 28, 1389–1401.

Chiang, L.H., Russell, E.L., Braatz, R.D., 2001. Fault detection and diagnosis in industrial systems. Springer-Verlag London, UK.

Chickering, D.M., Heckerman, D., Meek, C., 2004. Large-sample learning of Bayesian networks is NP-hard. J. Mach. Learn. Res. 5, 1287–1330.

Cho, J.-H., Lee, J.-M., Wook Choi, S., Lee, D., Lee, I.-B., 2005. Fault identification for process monitoring using kernel principal component analysis. Chem. Eng. Sci. 60, 279–288.

Choi, S.W., Lee, C., Lee, J.-M., Park, J.H., Lee, I.-B., 2005. Fault detection and identification of nonlinear processes based on kernel PCA. Chemom. Intell. Lab. Syst. 75, 55–67.

Choi, S.W., Lee, I.-B., 2004. Nonlinear dynamic process monitoring based on dynamic kernel PCA. Chem. Eng. Sci. 59, 5897–5908.

Chow, E., Willsky, A.S., 1984. Analytical redundancy and the design of robust failure detection systems. IEEE Trans. Automat. Contr. 29, 603–614.

Cooper, G.F., Herskovits, E., 1992. A Bayesian method for the induction of probabilistic networks from data. Mach. Learn. 9, 309–347.

Das, A., Maiti, J., Banerjee, R.N., 2012. Process monitoring and fault detection strategies: a review. Int. J. Qual. Reliab. Manag. 29, 720–752.

Dey, S., Stori, J.A., 2005. A Bayesian network approach to root cause diagnosis of process variations. Int. J. Mach. Tools Manuf. 45, 75–91.

Ding, S.X., 2014. Data-driven design of monitoring and diagnosis systems for dynamic processes: A review of subspace technique based schemes and some recent results. J. Process Control 24, 431–449.

Dong, J., Zhang, K., Huang, Y., Li, G., Peng, K., 2015. Adaptive total PLS based quality-relevant process monitoring with application to the Tennessee Eastman process. Neurocomputing 154, 77–85.

Downs, J.J., Vogel, E.F., 1993. A plant-wide industrial process control problem. Comput. Chem. Eng. 17, 245–255.

Dunia, R., Qin, S.J., Edgar, T.F., McAvoy, T.J., 1996. Identification of faulty sensors using principal component analysis. AIChE J. 42, 2797–2812.

Frank, P.M., 1996. Analytical and qualitative model-based fault diagnosis–a survey and some new results. Eur. J. Control 2, 6–28.

Frank, P.M., 1990. Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy: A survey and some new results. Automatica 26, 459–474.

Frank, P.M., Ding, X., 1997. Survey of robust residual generation and evaluation methods in observer-based fault detection systems. J. Process Control 7, 403–424.

Friedman, N., Linial, M., Nachman, I., Pe'er, D., 2000. Using Bayesian networks to analyze expression data. J. Comput. Biol. 7, 601–620.

Gao, Z., Cecati, C., Ding, S.X., 2015. A survey of fault diagnosis and fault-tolerant techniques—Part I: Fault diagnosis with model-based and signal-based approaches. IEEE Trans. Ind. Electron. 62, 3757–3767.

Ge, Z., Song, Z., Gao, F., 2013. Review of recent research on data-based process monitoring. Ind. Eng. Chem. Res. 52, 3543–3562.

Gertler, J., 2015. Fault Detection and Diagnosis. Springer.

Gertler, J., 1997. Fault detection and isolation using parity relations. Control Eng. Pract. 5, 653–661.

Gertler, J., 1991. Analytical redundancy methods in fault detection and isolation, in: Preprints of IFAC/IMACS Symposium on Fault Detection, Supervision and Safety for Technical Processes SAFEPROCESS'91. pp. 9–21.

Gharahbagheri, H., Imtiaz, S.A., Khan, F., 2017. Root Cause Diagnosis of Process Fault Using KPCA and Bayesian Network. Ind. Eng. Chem. Res. 56, 2054–2070.

Gonzalez, R., Huang, B., Lau, E., 2015. Process monitoring using kernel density estimation and Bayesian networking with an industrial case study. ISA Trans. 58, 330–347.

Grant, P.W., Harris, P.M., Moseley, L.G., 1996. Fault diagnosis for industrial printers using case-based reasoning. Eng. Appl. Artif. Intell. 9, 163–173.

Grossman, D., Domingos, P., 2004. Learning Bayesian network classifiers by maximizing conditional likelihood, in: Proceedings of the Twenty-First International Conference on Machine Learning. ACM, p. 46.

Guo, L., Kang, J., 2015. A hybrid process monitoring and fault diagnosis approach for chemical plants. Int. J. Chem. Eng. 2015.

Heckerman, D., 1998. A tutorial on learning with Bayesian networks. Nato Asi Ser. D Behav. Soc. Sci. 89, 301–354.

Heckerman, D., Geiger, D., Chickering, D.M., 1995. Learning Bayesian networks: The combination of knowledge and statistical data. Mach. Learn. 20, 197–243.

Himmelblau, D.M., 1978. Fault detection and diagnosis in chemical and petrochemical processes. Elsevier Science Ltd.

Höfling, T., Pfeufer, T., 1994. Detection of additive and multiplicative faults-parity space vs. parameter estimation. IFAC Proc. Vol. 27, 515–520.

Hoo, K.A., Piovoso, M.J., Schnelle, P.D., Rowan, D.A., 2003. Process and controller performance monitoring: overview with industrial applications. Int. J. Adapt. Control Signal Process. 17, 635–662.

Hoskins, J.C., Kaliyur, K.M., Himmelblau, D.M., 1991. Fault diagnosis in complex chemical plants using artificial neural networks. AIChE J. 37, 137–141.

Iri, M., Aoki, K., O'Shima, E., Matsuyama, H., 1979. An algorithm for diagnosis of system failures in the chemical process. Comput. Chem. Eng. 3, 489–493.

Isermann, R., 2006. Fault-diagnosis systems: an introduction from fault detection to fault tolerance. Springer Science & Business Media.

Isermann, R., 2005. Model-based fault-detection and diagnosis - Status and applications. Annu. Rev. Control.

Isermann, R., 1997. Supervision, fault-detection and fault-diagnosis methods—an introduction. Control Eng. Pract. 5, 639–652.

Isermann, R., 1993. Fault diagnosis of machines via parameter estimation and knowledge processing—tutorial paper. Automatica 29, 815–835.

Isermann, R., 1985. Process fault diagnosis with parameter estimation methods. IFAC Proc. Vol. 18, 51–60.

Isermann, R., 1982. Process fault detection based on modeling and estimation methods. IFAC Proc. Vol. 15, 7–30.

Isermann, R., Balle, P., 1997. Trends in the application of model-based fault detection and diagnosis of technical processes. Control Eng. Pract. 5, 709–719.

Jackson, J.E., 2005. A user's guide to principal components. John Wiley & Sons, Toronto, Canada.

Jackson, J.E., Mudholkar, G.S., 1979. Control Procedures for Residuals Associated with Principal Component Analysis. Technometrics 21, 341–349.

Jeng, J.-C., 2010. Adaptive process monitoring using efficient recursive PCA and moving window PCA algorithms. J. Taiwan Inst. Chem. Eng. 41, 475–481.

Joe Qin, S., 2003. Statistical process monitoring: basics and beyond. J. Chemom. 17, 480–502.

Jung, D., Ng, K.Y., Frisk, E., Krysander, M., 2016. A combined diagnosis system design using model-based and data-driven methods, in: Control and Fault-Tolerant Systems (SysTol), 2016 3rd Conference on. IEEE, pp. 177–182.

Kano, M., Hasebe, S., Hashimoto, I., Ohno, H., 2001. A new multivariate statistical process monitoring method using principal component analysis. Comput. Chem. Eng. 25, 1103–1113.

Kano, M., Tanaka, S., Hasebe, S., Hashimoto, I., Ohno, H., 2003. Monitoring independent components for fault detection. AIChE J. 49, 969–976.

Katipamula, S., Brambley, M.R., 2005. Methods for fault detection, diagnostics, and prognostics for building systems—a review, part I. Hvac&R Res. 11, 3–25.

Kolodner, J.L., 1992. An introduction to case-based reasoning. Artif. Intell. Rev. 6, 3–34.

Kourti, T., MacGregor, J.F., 1995. Process analysis, monitoring and diagnosis, using multivariate projection methods. Chemom. Intell. Lab. Syst. 28, 3–21.

Kramer, M.A., 1991. Nonlinear principal component analysis using autoassociative neural networks. AIChE J. 37, 233–243.

Kresta, J. V., Macgregor, J.F., Marlin, T.E., 1991. Multivariate statistical monitoring of process operating performance. Can. J. Chem. Eng. 69, 35–47.

Ku, W., Storer, R.H., Georgakis, C., 1995. Disturbance detection and isolation by dynamic principal component analysis. Chemom. Intell. Lab. Syst. 30, 179–196.

Kuhner, M.K., 2006. LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. Bioinformatics 22, 768–770.

Kulkarni, A., Jayaraman, V.K., Kulkarni, B.D., 2005. Knowledge incorporated support vector machines to detect faults in Tennessee Eastman Process. Comput. Chem. Eng. 29, 2128–2133.

Lauritzen, S.L., Spiegelhalter, D.J., 1988. Local computations with probabilities on graphical structures and their application to expert systems. J. R. Stat. Soc. Ser. B 157–224.

LaViola, J.J., 2003. A comparison of unscented and extended Kalman filtering for estimating quaternion motion, in: American Control Conference, 2003. Proceedings of the 2003. IEEE, pp. 2435–2440.

Lee, C.K.H., 2017. A knowledge-based product development system in the chemical industry. J. Intell. Manuf. 1–16.

Lee, J.-M., Qin, S.J., Lee, I.-B., 2006. Fault Detection and Diagnosis Based on Modified Independent Component Analysis. AIChE J. 52, 3501–3514.

Lee, J.-M., Yoo, C., Choi, S.W., Vanrolleghem, P.A., Lee, I.-B., 2004a. Nonlinear process monitoring using kernel principal component analysis. Chem. Eng. Sci. 59, 223–234.

Lee, J.-M., Yoo, C., Lee, I.-B., 2004b. Statistical process monitoring with independent component analysis. J. Process Control 14, 467–485.

Lee, W.-S., Grosh, D.L., Tillman, F.A., Lie, C.H., 1985. Fault Tree Analysis, Methods, and Applications ⑀ A Review. IEEE Trans. Reliab. 34, 194–203.

Lemmer, J.F., Kanal, L.N., 2014. Propagating uncertainty in Bayesian networks by probabilistic logic sampling, in: Uncertainty in Artificial Intelligence. 149-163.

Li, W., Yue, H.H., Valle-Cervantes, S., Qin, S.J., 2000. Recursive PCA for adaptive process monitoring. J. Process Control 10, 471–486.

Liu, J., 2012. Fault diagnosis using contribution plots without smearing effect on non-faulty variables. J. Process Control 22, 1609–1623.

Liu, J., Chen, D.-S., 2009. Fault Detection and Identification Using Modified Bayesian Classification on PCA Subspace. Ind. Eng. Chem. Res. 48, 3059–3077.

Łupińska-Dubicka, A., 2012. Modeling dynamical systems by means of dynamic Bayesian networks. Sci. Bull. Bialystok Univ. Technol. Informatics 77–92.

Madsen, A.L., 2008. Belief update in CLG Bayesian networks with lazy propagation. Int. J. Approx. Reason. 49, 503–521.

Magni, J.-F., Mouyon, P., 1994. On residual generation by observer and parity space approaches. IEEE Trans. Automat. Contr. 39, 441–447.

Mahadevan, S., Shah, S.L., 2009. Fault detection and diagnosis in process data using one-class support vector machines. J. Process Control 19, 1627–1639.

Mallick, M.R., Imtiaz, S.A., 2013. A hybrid method for process fault detection and diagnosis. IFAC Proc. Vol. 46, 827–832.

Martin, T.G., Burgman, M.A., Fidler, F., Kuhnert, P.M., LOW-CHOY, S., McBride, M., Mengersen, K., 2012. Eliciting expert knowledge in conservation science. Conserv. Biol. 26, 29–38.

Maurya, M.R., Rengaswamy, R., Venkatasubramanian, V., 2003a. A systematic framework for the development and analysis of signed digraphs for chemical processes. 1. Algorithms and analysis. Ind. Eng. Chem. Res. 42, 4789–4810.

Maurya, M.R., Rengaswamy, R., Venkatasubramanian, V., 2003b. A systematic framework for the development and analysis of signed digraphs for chemical processes. 2. Control loops and flowsheet analysis. Ind. Eng. Chem. Res. 42, 4811–4827.

Mehranbod, N., Soroush, M., Panjapornpon, C., 2005. A method of sensor fault detection and identification. J. Process Control 15, 321–339.

Mehranbod, N., Soroush, M., Piovoso, M., Ogunnaike, B.A., 2003. Probabilistic model for sensor fault detection and identification. AIChE J. 49, 1787–1802.

Mihajlovic, V., Petkovic, M., 2001. Dynamic bayesian networks: A state of the art. Technical report, Computer Science Department, University of Twente, The Netherlands, 2001.

Misra, M., Yue, H.H., Qin, S.J., Ling, C., 2002. Multivariate process monitoring and fault diagnosis by multi-scale PCA. Comput. Chem. Eng. 26, 1281–1293.

Montgomery, D.C., Runger, G.C., 2010. Applied statistics and probability for engineers. John Wiley & Sons.

Mrad, A. Ben, Delcroix, V., Piechowiak, S., Leicester, P., Abid, M., 2015. An explication of uncertain evidence in Bayesian networks: likelihood evidence and probabilistic evidence. Appl. Intell. 43, 802.

Murphy, K.P., 2002. Dynamic Bayesian networks: representation, inference and learning. PhD Thesis 1–281.

Musharraf, M., Hassan, J., Khan, F., Veitch, B., MacKinnon, S., Imtiaz, S., 2013. Human reliability assessment during offshore emergency conditions. Saf. Sci. 59, 19–27.

Mylaraswamy, D., Venkatasubramanian, V., 1997. A hybrid framework for large scale process fault diagnosis. Comput. Chem. Eng. 21, S935–S940.

Nan, C., Khan, F., Iqbal, M.T., 2008. Real-time fault diagnosis using knowledge-based expert system. Process Saf. Environ. Prot. 86, 55–71.

Neapolitan, R.E., 2004. Learning bayesian networks. Pearson Prentice Hall Upper Saddle River, NJ.

Nielsen, T.D., Jensen, F.V., 2009. Bayesian networks and decision graphs. Springer Science & Business Media.

Nimmo, I., 1995. Adequately address abnormal operations. Chem. Eng. Prog. 91, 36–45.

Nomikos, P., MacGregor, J.F., 1995. Multi-way partial least squares in monitoring batch processes. Chemom. Intell. Lab. Syst. 30, 97–108.

Nomikos, P., MacGregor, J.F., 1994. Monitoring batch processes using multiway principal component analysis. AIChE J. 40, 1361–1375.

Odendaal, H.M., Jones, T., 2014. Actuator fault detection and isolation: An optimised parity space approach. Control Eng. Pract. 26, 222–232.

Olivier-Maget, N., Negny, S., Hétreux, G., Le Lann, J.-M., 2009. Fault diagnosis and process monitoring through model-based and case based reasoning. Comput. Aided

Chem. Eng. 26, 345–350.

Oyeleye, O.O., Kramer, M.A., 1988. Qualitative simulation of chemical process systems: Steady-state analysis. AIChE J. 34, 1441–1454.

Patton, R.J., Chen, J., 1991. A review of parity space approaches to fault diagnosis, in: IFAC/IMACS Safeprocess Conference. pp. 65–81.

Pearl, J., 1988. Probabilistic reasoning in intelligent systems: Networks of plausible inference. San Fransisco, CA: Morgan Kaufman.

Peng, Y., Zhang, S., Pan, R., 2010. Bayesian network reasoning with uncertain evidences. Int. J. Uncertainty, Fuzziness Knowledge-Based Syst. 18, 539–564.

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., Reich, D., 2006. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38, 904-909.

Qin, S.J., 2012. Survey on data-driven industrial process monitoring and diagnosis. Annu. Rev. Control 36, 220–234.

Qin, S.J., Valle, S., Piovoso, M.J., 2001. On unifying multiblock analysis with application to decentralized process monitoring. J. Chemom. 15, 715–742.

Quantrille, T.E., Liu, Y.A., 2012. Artificial intelligence in chemical engineering. Elsevier.

Rhoads, T.R., Montgomery, D.C., 1996. Process monitoring with principal components and partial least squares, in: Proceedings of the 1996 5th Industrial Engineering Research Conference. IIE.

Russell, E.L., Chiang, L.H., Braatz, R.D., 2012. Data-driven methods for fault detection and diagnosis in chemical processes. Springer Science & Business Media.

Severson, K., Chaiwatanodom, P., Braatz, R.D., 2016. Perspectives on process monitoring of industrial systems. Annu. Rev. Control.

Shewhart, W.A., 1930. Economic quality control of manufactured product. Bell Labs Tech. J. 9, 364–389.

Shiozaki, J., Matsuyama, H., O'shima, E., Iri, M., 1985. An improved algorithm for diagnosis of system failures in the chemical process. Comput. Chem. Eng. 9, 285–293.

Simani, S., Fantuzzi, C., Patton, R.J., 2003. Model-Based Fault Diagnosis Techniques, in: Model-Based Fault Diagnosis in Dynamic Systems Using Identification Techniques. Springer, pp. 19–60.

Skogestad, S., 1997. Dynamics and control of distillation columns: A tutorial introduction. Chem. Eng. Res. Des. 75, 539–562.

Sorsa, T., Koivo, H.N., 1993. Application of artificial neural networks in process fault diagnosis. Automatica 29, 843–849.

Sotomayor, O.A.Z., Odloak, D., 2005. Observer-based fault diagnosis in chemical plants. Chem. Eng. J. 112, 93–108.

Spiegelhalter, D.J., Lauritzen, S.L., 1990. Sequential updating of conditional probabilities on directed graphical structures. Networks 20, 579–605.

Staroswiecki, M., Cassar, J.P., Cocquempot, V., 1993. Generation of optimal structured residuals in the parity space. IFAC Proc. Vol. 26, 535–542.

Thornhill, N.F., Patwardhan, S.C., Shah, S.L., 2008. A continuous stirred tank heater simulation model with applications. J. Process Control 18, 347–360.

Van Der Gaag, L.C., 1996. Bayesian belief networks: odds and ends. Comput. J. 39, 97–113.

Vedam, H., Venkatasubramanian, V., 1999. PCA-SDG based process monitoring and fault diagnosis. Control Eng. Pract. 7, 903–917.

Venkatasubramanian, V., Chan, K., 1989. A neural network methodology for process fault diagnosis. AIChE J. 35, 1993–2002.

Venkatasubramanian, V., Rengaswamy, R., Kavuri, S.N., 2003a. A review of process fault detection and diagnosis part II: Qualitative models and search strategies. Comput. Chem. Eng.

Venkatasubramanian, V., Rengaswamy, R., Yin, K., Kavuri, S.N., 2003b. A review of process fault detection and diagnosis Part III: Process history based methods. Comput. Chem. Eng.

Venkatasubramanian, V., Rengaswamy, R., Yin, K., Kavuri, S.N., 2003c. A review of process fault detection and diagnosis, Part I Quantitative model-based methods. Comput. Chem. Eng. 27, 293–311.

Venkatasubramanian, V., Rich, S.H., 1988. An object-oriented two-tier architecture for integrating compiled and deep-level knowledge for process diagnosis. Comput. Chem. Eng. 12, 903–921.

Verron, S., Tiplica, T., Kobi, A., 2008. Distance rejection in a bayesian network for fault diagnosis of industrial systems, in: Control and Automation, 2008 16th Mediterranean Conference on. IEEE, pp. 615–620.

Vesely, W.E., Goldberg, F.F., Roberts, N.H., Haasl, D.F., 1981. Fault tree handbook.

Nuclear Regulatory Commission Washington dc.

Wan, E.A., Van Der Merwe, R., 2000. The unscented Kalman filter for nonlinear estimation, in: Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC. The IEEE 2000. Ieee, pp. 153–158.

Wang, J., Ge, W., Zhou, J., Wu, H., Jin, Q., 2017. Fault isolation based on residual evaluation and contribution analysis. J. Franklin Inst. 354, 2591–2612.

Wang, Y., Liu, Y., Khan, F., Imtiaz, S., 2017. Semiparametric PCA and bayesian network based process fault diagnosis technique. Can. J. Chem. Eng.

Wang, Z., Zhao, J., Shang, H., 2012. A hybrid fault diagnosis strategy for chemical process startups. J. Process Control 22, 1287–1297.

Watson, I., Marir, F., 1994. Case-based reasoning: A review. Knowl. Eng. Rev. 9, 327–354.

Weber, P., Medina-Oliva, G., Simon, C., Iung, B., 2012. Overview on Bayesian networks applications for dependability, risk analysis and maintenance areas. Eng. Appl. Artif. Intell. 25, 671–682.

Weerasinghe, M., Gomm, J.B., Williams, D., 1998. Neural networks for fault diagnosis of a nuclear fuel processing plant at different operating points. Control Eng. Pract. 6, 281–289.

Weidl, G., Madsen, A.L., Israelson, S., 2005. Applications of object-oriented Bayesian networks for condition monitoring, root cause analysis and decision support on operation of complex continuous processes. Comput. Chem. Eng. 29, 1996–2009.

Westerhuis, J.A., Kourti, T., MacGregor, J.F., 1998. Analysis of multiblock and hierarchical PCA and PLS models. J. Chemom. 12, 301–321.

Wilcox, N.A., Himmelblau, D.M., 1994a. The possible cause and effect graphs (PCEG) model for fault diagnosis—I. Methodology. Comput. Chem. Eng. 18, 103–116.

Wilcox, N.A., Himmelblau, D.M., 1994b. The possible cause and effect graphs (PCEG) model for fault diagnosis-II. applications. Comput. Chem. Eng. 18, 117–127.

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

Wise, B.M., Veltkamp, D.J., Davis, B., Ricker, N.L., Kowalski, B.R., 1988. Principal component analysis for monitoring the West Valley liquid fed ceramic melter. Waste Manag. 88, 811–818.

Yang, F., Shah, S., Xiao, D., 2012. Signed directed graph based modeling and its validation from process knowledge and process data. Int. J. Appl. Math. Comput. Sci. 22, 41–53.

Yin, S., Ding, S.X., Haghani, A., Hao, H., Zhang, P., 2012. A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process. J. Process Control 22, 1567–1581.

Yin, S., Ding, S.X., Xie, X., Luo, H., 2014. A review on basic data-driven approaches for industrial process monitoring. IEEE Trans. Ind. Electron. 61, 6418–6428.

Yoon, S., MacGregor, J.F., 2001. Fault diagnosis with multivariate statistical models part I: using steady state fault signatures. J. Process Control 11, 387–400.

Young, P., 1981. Parameter estimation for continuous-time models—a survey. Automatica 17, 23–39.

Yu, D., Shields, D.N., Daley, S., 1996. A hybrid fault diagnosis approach using neural networks. Neural Comput. Appl. 4, 21–26.

Yu, H., Khan, F., Garaniya, V., 2015. Modified Independent Component Analysis and Bayesian Network-Based Two-Stage Fault Diagnosis of Process Operations. Ind. Eng. Chem. Res. 54, 2724–2742.

Yu, J., Rashid, M.M., 2013. A novel dynamic bayesian network-based networked process monitoring approach for fault detection, propagation identification, and root cause diagnosis. AIChE J. 59, 2348–2365.

Zadakbar, O., Imtiaz, S., Khan, F., 2012. Dynamic Risk Assessment and Fault Detection Using Principal Component Analysis. Ind. Eng. Chem. Res. 52, 809–816.

Zarei, J., Poshtan, J., 2010. Design of nonlinear unknown input observer for process fault detection. Ind. Eng. Chem. Res. 49, 11443–11452.

Zeng, Q., 2016. Data-driven Process Monitoring and Fault Detection with Convex Geometry. M.Sc. thesis, University of Alberta, Edmonton.

Zerrouki, H., Smadi, H., 2017. Bayesian Belief Network Used in the Chemical and Process Industry: A Review and Application. J. Fail. Anal. Prev. 17, 159–165.

Zhang, H., Tangirala, A.K., Shah, S.I., 1999. Dynamic process monitoring using multiscale PCA, in: Electrical and Computer Engineering, 1999 IEEE Canadian Conference on. IEEE, pp. 1579–1584.

Zhang, Y., Zhang, Y., 2010. Fault detection of non-Gaussian processes based on modified independent component analysis. Chem. Eng. Sci. 65, 4630–4639.

Zhang, Z., Dong, F., 2014. Fault detection and diagnosis for missing data systems with a three time-slice dynamic Bayesian network approach. Chemom. Intell. Lab. Syst. 138, 30–40.

Zhao, H., Liu, J., Dong, W., Sun, X., Ji, Y., 2017. An improved case-based reasoning method and its application on fault diagnosis of Tennessee Eastman process. Neurocomputing 249, 266–276.

Zhong, M., Song, Y., Ding, S.X., 2015. Parity space-based fault detection for linear discrete time-varying systems with unknown input. Automatica 59, 120–126.