DESIGN OF AN EVENT-BASED EARLY WARNING SYSTEM FOR PROCESS OPERATIONS

by

© Kosmapatabendige Pradeep Shiran Dalaptadu

A Thesis submitted to the

School of Graduate Studies

in partial fulfillment of the requirements for the degree of

Master of Engineering

Faculty of Engineering and Applied Science

Memorial University of Newfoundland

May 2014

St. John's Newfoundland

ABSTRACT

This thesis proposes a new methodology to design an event-based warning system as an alternative to the conventional variable-based alarm system.

This study initially explores the options for grouping process variables for alarm allocation. Several grouping methods are discussed and an event-based grouping procedure is detailed. Selection of the key variables for a group is performed considering the information that the variables contain to distinguish between an abnormal and a normal condition. The information theory is used to quantify the information content of a variable about an event to select the key variables. The cross-correlation analysis between pairs of key variables is used to identify the redundant variables. Simulation study using the model of a continuous stirred tank reactor (CSTR) is used to demonstrate the methodology.

The proposed event-based early warning system utilizing online measurements is detailed in the thesis. In this approach, warnings are assigned to plant abnormal events instead of individual variables. To assess the likelihoods of undesirable events, the Bayesian Network is used; the event likelihoods are estimated in real time utilizing online measurements. Diagnostic analysis is conducted to identify root-causes of events. By assigning warning to events, the methodology results in significantly lower number of warnings compared to traditional variable-based warning (alarms) system. It also enables early warning of a possible event along with an efficient diagnosis of the root-causes of the event. Experimental testing using a level control system is presented to demonstrate the efficacy of the proposed method. Simulation study using the model of a CSTR is also presented to demonstrate the performance of the algorithm. Both, experimental and simulation studies, have shown promising results.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deep gratitude to my supervisor Dr. Faisal Khan, who gave me the opportunity to conduct my studies in his research group and also for his valuable guidance, support and encouragement. Also, I greatly acknowledge my co-supervisor Dr. Salim Ahmed for his valuable comments, support, and guidance throughout my research work.

I would gratefully acknowledge the financial support provided by School of Graduate study, Vale Research Chair Grant, Research and Development Corporation (RDC), and Natural Sciences and Engineering Research Council (NSERC) of Canada.

My deep appreciation goes to all my friends and colleagues in the Safety and Risk Engineering Group (SREG) and everyone else who helped me with my research in different ways. I would like to thank especially Samith Rathnayake for his great support.

I would also like to say a heartfelt thank you to all my dearest friends, especially for Oscar and Nilanthi, for making my experience in St John’s memorable. Lastly, I would like to thank my parents and my three brothers for all their love and encouragement. Without their support it is impossible for achieve this far. And most of all for my loving wife Bhagya, who has always stood by me throughout this journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ASM	Abnormal Situation Management
BN	Bayesian Network
BP	British Petroleum
BPCS	Basic Process Control System
CPT	Conditional Probability Table
CSTR	Continuous Stirred Tank Reactor
DCS	Distributed Control System
EEMUA	The Engineering Equipment and Materials Users' Association
FMEA	Failure Mode and Effect Analysis
HAZOP	Hazard and Operability Study
HSE	Health and Safety Executive
ISA	International Society of Automation
LCL	Lower Control Limit
LOPA	Layer In Layer Of Protection Analysis
NSERC	Natural Sciences and Engineering Research Council
P&ID	Pipe and Instrumentation Diagram
PCA	Principle Component Analysis
RDC	Research and Development Corporation
SCADA	Supervisory Control And Data Acquisition

SREG	Safety and Risk Engineering Group
UCL	Upper Control Limit
USCSB	U.S. Chemical Safety and Hazard Investigation Board

LIST OF APPENDICES

1 INTRODUCTION

This chapter presents a brief introduction to process monitoring and warning systems in the process industry as well as the limitations of the standard variable-based alarm systems. The motivations and objectives of this research are also presented.

1.1 Early warning systems in the process industry

Process industries deal with hazardous materials, high intensive energy and complex equipment in day-to-day operations. It is important to monitor the state of a process in real time to identify any vulnerable condition before it leads to a more severe event, which may be harmful in safety, economical, social and environmental aspects. Due to the high level of complexity of modern plants, process industries use process control systems such as the supervisory control and data acquisition (SCADA) system and the distributed control system (DCS) to monitor a large number of process variables and store vast amount of data during plant operations. In a complex process plant, the number of observed variables may be in thousands (Venkatasubramanian, Rengaswamy, & Yin, 2003). Early warning systems are installed in the control rooms to monitor and detect deviation of variables from their normal operating range. The warning system communicates to the operator to take corrective actions required to maintain the operation under safe conditions.

Conventional warning systems are single variable-based where the warning provides information regarding the status of the respective variable; however, little or no information about the possible event or plant state is obtained. When the plant is in the initial phase of an abnormality, many secondary process variables may exceed their threshold limits. However, operators may not identify the possible key event until some primary variables exceed their threshold limit. The capability of early detection of abnormal situation is limited with conventional alarm systems. However, early detection of process abnormality is critical as some abnormality can quickly propagate to more severe and uncontrollable conditions. It is essential for operators to have enough time to analyze the situation to determine the root causes and to take corrective action to bring back the process state to normal operating conditions.

Due to the ability to monitor large numbers of variables with low cost, plant designers/engineers tend to assign alarms to as many variables as they can. It simply costs less to add an alarm than to discuss whether it is needed or not. Hence, the number of alarms in process plants has dramatically increased over the last two decades. As a result, even a minor disturbance may trigger many low information secondary alarms and give the operator a false impression about the plant state. On the other hand, a major disturbance, which can propagate to a severe event, can trigger many redundant alarms along with a primary alarm. As a result, alarm rates regularly exceed the operators' physical ability to handle alarms. This phenomenon is called `alarm flooding' that can reduce the motivation of the operators to check on alarms. It can also reduce the ability of the operator to detect the root-causes (Izadi, Shah, Shook, & Chen, 2009). According to the EEMUA, an operator should not handle more than six alarms per hour (EEMUA, 2007). In reality, according to different reviews, numbers of alarms exceed this value by a wide margin during both normal and abnormal conditions (Rothenberg, 2009; Y. Chang, Khan, & Ahmed, 2011). Minimizing the number of alarms without compromising the ability to identify failures is a crucial factor in process alarm system design.

Over the past few decades, numerous process related accidents took place during regular operations. One example is the BP Texas refinery explosion that killed 18 people and injured 180, and resulted in 1.5 billion dollars losses. One of the findings from the investigation of the accident is that the process monitoring system did not give adequate information and warning about the dangerous plant status timely (USCSB, 2007). The explosion at the Texaco Refinery Milford Haven is another example of process accident where alarm flooding contributed to the accident. With an upset situation, an alarm flooding condition occurred and 2 or 3 alarms per second were displayed in the control panel before the accident. Health and Safety Executive (HSE) reported that the accident could have been prevented if the operators were able to find the root-cause of the upset. Due to the accident 26 people were injured; damage and production losses amounted to £ 48 million (HSE, 1994).

1.2 Motivation and Objectives

Recent studies (Ahmed, Gabbar, Chang, & Khan, 2011; Bao, Khan, Iqbal, & Chang, 2011; Y. Chang et al., 2011) have proposed the concept of an event-based alarm system that groups process variables according to abnormal process events and uses the risk of an event as the annunciating agent of an alarm. In this approach, the probabilities and severities of the events are required to calculate the real time risk of the event. However, how to select the group of key variables to be monitored to predict an event, how to define and estimate a single indicator to represent the current state of the process with respect to an associated event, how to define the alarm annunciation philosophy, and finally how to analyze the root-causes of an event are not outlined in the published literature. This thesis attempts to address these issues and develop necessary techniques to design an event-based early warning system for process industries.

The possibilities of utilizing groups of variables to develop a warning system that can mitigate the limitation of the variable-based system are explored. Several grouping methods are discussed along with their advantages and disadvantages. Grouping of variables according to the possible abnormal event was found to be the most suitable for the purpose of warning system design. An event-based grouping procedure is proposed in this thesis.

The focus of the thesis is to design an event-based early warning system that is able to annunciate warnings based on the probability of abnormal events using real time process measurements. The specific objectives are:

- Explore the options for limiting the number of warning annunciation during an abnormal condition by allocating key process variables into event-based groups.
- Use Bayesian Network to define the relationship between events and associated factors to update the probabilities at real time utilizing sensor measurements.
- Design an event-based early warning system having significant early warning capability compared with the conventional variable-based alarm system.
- Develop a methodology to identify root-causes of abnormal events using real time sensor measurements.

1.3 Thesis Structure

This thesis is organized in 6 Chapters. The first chapter is a brief introduction to process monitoring and warning systems in the process industry and the limitations of the standard variable-based alarm systems. The motivations and objectives of this research are also presented. Chapter 2 presents a comprehensive overview of the literature. The review details the early warning methods that are currently used in the industry along with the advanced early warning methods proposed in the literature. Chapter 3 describes the variable grouping methods and details a methodology to allocate process variables to abnormal events using the information theory and cross correlation analysis. Chapter 4 presents the methodology for early warning system design. In this chapter a detail study of identification of events, related process variables, and root-causes, calculation of events probabilities, and root-cause analysis of events are presented. The methodology for development of a Bayesian Network model for the warning system is also detailed. In Chapter 5, an experimental application along with a simulation study, are used to demonstrate the methodology and evaluate its performance. Also, this chapter presents a detailed analysis of the proposed early warning system and discusses its unique features. Chapter 6 concludes the thesis by presenting the conclusions and recommendations and outlining the scopes for further improvement of this work.

2 LITERATURE REVIEW

A plant abnormal situation is defined as "a disturbance or a series of disturbances in a process that causes a plant operation to deviate from its normal operating state" ("Abnormal Situation Managment Consortium," 2008). Abnormal situations can range from minor process disturbances to major process upset, which require the operators to intervene and perform corrective actions. It is estimated that abnormal conditions cost at least $20 billion annually in the USA only (Cochran & Bullemer, 1996). Therefore, abnormal situation management (ASM) is a critical task in the process industry. ASM can be defined as early detection of an abnormal event, root-cause diagnosis of the abnormality and taking corrective action to bring back the process to normal and safe operating state (Venkatasubramanian, Rengaswamy, & Yin, 2003). For early detection and root-cause diagnosis of abnormal situations, process monitoring and warning system are extensively used in process industries. Warning system can be defined as "a system designed to direct the operator's attention towards significant aspects of the current plant status" (Bransby & Jenkinson, 1998). Some of the required characteristics of a warning system are relevant, unique, timely, prioritized, understandable, diagnostic, advisory, and focusing (EEMUA, 2007). The warning system is one of the most critical safety element in process industries and it is the 3rd layer of protection after the process design layer and the basic process control system (BPCS) layer (Crowl & Louvar, 2001).

Many types of alarms are used in the process industry such as variable based alarms, deviation alarms, rate of change alarms, and calculated alarms (EEMUA, 2007). However, common warning systems are based on monitoring of individual variables to check whether a variable value is exceeding its threshold limit. Input to the warning system is usually a measurement of a process variable from a sensor (Izadi, Shah, Shook, et al., 2009). However conventional variable-based warning systems have many limitations that arise mostly due to their single variable setting.

With the introduction of distributed control system, almost all the variables can be monitored. As a result, process-monitoring systems have large number of alarms, many of which are poorly designed. Due to the high number of warning variables, the number of nuisance alarms has increased significantly. Hence, even during a minor abnormality, alarm flooding may occur (Izadi, Shah, Shook, et al., 2009).

Many alarm management techniques have been proposed in the literature to prevent alarm flooding. EEMUA 191 is a key guide for design, management and procurement of alarm systems. Alarm management techniques such as grouping alarms, alarm suppression, and shelving have been discussed in the EEMUA guideline (EEMUA, 2007). Alarm management life cycle approach has been proposed in the ISA standard to manage alarms at the design stage as well as during operation. It consists of the following stages: philosophy, identification, rationalization, design, implementation, operation, monitoring, maintenance, and change activity. An effective tool that can be used to reduce number of redundant alarms is alarm rationalization (ISA, 2009).

Alarm rationalization methods, that can identify redundant or unwanted alarms, have been proposed by many researchers to reduce the number of alarms in plant operations. Kondaveeti et al. (Kondaveeti, Izadi, Shah, & Black, 2009) proposed visualization tools using the High Density Alarm Plot (HDAP) and the Alarm Similarity Color Map (ASCM) to identify the nuisance alarms and thus to improve the performance of alarm systems. Noda at al. (Noda, Higuchi, Takai, & Nishitani, 2011) proposed an event correlation analysis to detect the statistical similarities among alarms and operation alerts. In this approach, correlated alarms are grouped together according to the similarity to reduce alarms. Fuzzy clustering methodology has been proposed to identify similar alarms (Qunxiong & Zhiqiang, 2005).

Proper alarm designing techniques such as threshold designing, multivariate process monitoring, and data processing can be used to reduce false and missed alarms (Izadi, Shah, Shook, et al., 2009; Izadi, Shah, & Shook, 2009). However, these techniques are yet to find widespread industrial application.

As an alternate to the variable-based alarm system, Bao et al. (Bao et al., 2011) proposed a risk based alarm design method. Risk is a function of the probability of occurrence of an accident, and its consequences. According to the risk based alarm methodology, warnings are considered as the faulty state of the primary process variables. The probability and the severity of a fault are calculated using the deviations of process variables from corresponding normal operation. The risk-based alarms are triggered if the respective risk level exceeds the threshold level. If there are more than one high-risk process variable, alarms are prioritized according to the risk level. However, the proposed method (Bao et al., 2011) uses univariate analysis and its early warning capability is limited. Recent studies (Ahmed et al., 2011; Bao et al., 2011; Y. Chang et al., 2011) extended risk-based alarm design by proposing an event-based alarm system that groups process variables according to the events and use risk as the final indicator of the alarm. An event is defined as an undesirable condition that initiates as process deviates from its steady state. In the risk-based approach, the probabilities and severities of the events are required to calculate the real time risk of the event. However, the above-mentioned literature does not define how to calculate the probabilities and severity using the measurements of a group of variables.

Many advanced multivariate fault detection methods have been proposed in the literature to utilize the process measurements for the purpose of early fault detection and diagnosis. These methods are classified as qualitative model based, quantitative model based, and historical data based. Fault detection and diagnosis methods have been reviewed in the literature (Venkatasubramanian, Rengaswamy, & Kavuri, 2003a, 2003b;

Venkatasubramanian, Rengaswamy, & Yin, 2003). In the quantitative model based approach, the state of the actual process is observed and compared with an estimated state from a first principle model to determine abnormalities. The Kalman filter is frequently used to correct the measurement and process error to estimate the state of the process (Villez, Srinivasan, Rengaswamy, Narasimhan, & Venkatasubramanian, 2011). Use of dynamic models to predict process variables and their violation of emergency limit in future time steps by estimating unknown disturbances using the Kalman filter, has been discussed by Juricek et al. (Juricek, Seborg, & Larimore, 2001). Frameworks to integrate fault detection and diagnosis with risk based monitoring systems have also been suggested in the literature. Zadakbar et al. (O Zadakbar, Imtiaz, & Khan, 2013) proposed a model-based method to calculate the multivariate residual error between plant model and actual plant data using the Kalman filter to detect faulty condition in the process. Residuals generated form the Kalman filter is used to calculate the real time risk in the process. However, development of a comprehensive process model for a complex process plant is a challenging task. Qualitative model based methods for fault detection have been proposed in the literature using causal model such as the digraph, and the fault tree (Ram Maurya, Rengaswamy, & Venkatasubramanian, 2004). However, the qualitative methods have limited capability to detect faults in real time. Many researchers have suggested historical data based methods for fault detection and diagnosis applications as an alternative to the first principles approach. Data mining and knowledge discovery using unsupervised statistical multivariate techniques such as the Principle Component Analysis

(PCA) and the Partial Least Square (PLS) and supervised learning method such as the Neural Network have been detailed in Wang et al. (X.Z. Wang, 1999). Zadakbar et al. (Omid Zadakbar, Imtiaz, & Khan, 2012) proposed PCA to convert high dimensional monitoring variable sets to low dimensional variable sets according to their correlation to detect process abnormality and calculate the process risk in real time. Contribution plots from the PCA are used for fault diagnosis. However, a large set of historical data is required to develop a statistical model.

Recently, the Bayesian Network (BN), which is a probabilistic graphical method, has been used in many applications and it has shown promising results. BN is able to integrate expert subjective knowledge with plant data to do probabilistic prediction and diagnostic inference (Oniésko, Lucas, & Druzdzel, 2001). Khakzad et al. (Khakzad, Khan, & Amyotte, 2011) proposed a BN based methodology for safety analysis in process industry which was further extended to perform dynamic safety analysis (Khakzad, Khan, & Amyotte, 2012). The 'Pathfinder project' (Heckerman, Horvitz, & Nathwani, 1992) that used BN expert system to diagnose medical conditions of patients has reported successful use of BN in performing critical analysis in many complex situations. BN has also been used for many real time safety related accident prediction applications. Real time traffic accident prediction on urban expressway using traffic data has been proposed by (Hossain & Muromachi, 2012) and Argiolas et al. (Argiolas, Carbonari, Melis, & Quaquero, 2012) used BN to predict the accident in a construction site using site information. Early fault detection of a boiler using BN by utilizing real time process deviation data has also been reported (Widarsson & Dotzauer, 2008). Real time fire detection model using sensor measurements to predict fire is proposed by Jing et al. (Jing & Jingqi, 2012). Using uncertain real time sensor data, BN model was capable to detect the fire at the initial stage according to the fire symptoms. Many natural disaster early warning system that rely on uncertain data have used BN to model the situation and early warn the disaster (Blaser, Ohrnberger, Riggelsen, Babeyko, & Scherbaum, 2011; Zazzaro, Pisano, & Romano, 2012). BN is a powerful tool to do fault diagnosis due to its ability to do inference under uncertainty. Cause and effect relationship of the online process information is used to conduct root-cause analysis (Alaeddini & Dogan, 2011; Dey & Stori, 2005; Pradhan, Singh, Kachru, & Narasimhamurthy, 2007). The above literature demonstrates the potential use of BN as a tool to develop real time early warning systems.

3 VARIABLE ALLOCATION FOR EVENT-BASED GROUPS

3.1 Introduction

The ability to monitor large number of variables and the ability to assign alarms to each variable led to a substantial increase in the numbers of alarms in industrial plants. This, in turn, increased the numbers of false and redundant alarms. In plant operations, the numbers of annunciated alarms regularly exceed the acceptable rates that operators can handle. This chapter explores the options for grouping variables for alarm allocation. Several grouping methods are discussed and an event-based grouping procedure is detailed. Selection of the key variables for a group is performed using the information that the variables can have to distinguish between an abnormal and a normal condition. The concept of mutual information is used to quantify the information. Variables with high information gain are grouped together for each respective abnormal event. To identify the redundant variables within the groups to further reduce the number of variables to be monitored, the maximum cross-correlation between pairs of key variables are used. A case study using the example of a continuous stirred tank reactor is used to demonstrate the methodology.

3.2 Grouping methods

This section addresses the concept of grouping of variables to assign alarms. Allocation of an alarm to a group of variables will result in the annunciation of one alarm when one or more variables within the group deviate. Grouping can be performed considering various factors. Variables can be grouped according to their types, or the equipment that they are associated with, or according to their correlations. Variables can also be grouped according to the events they are associated with.

3.2.1 Grouping based on variable types

Different types of measurements such as temperatures, pressures and levels are available from industrial plants. By grouping variables according to their types and assigning alarms to groups may significantly reduce the number of alarms in a plant. For example, if there are number of thermo-couples along the length of a distillation column, instead of assigning alarms to each of the measurements, one alarm can be allocated to the set of temperature measurements. Annunciation of the alarm would indicate an abnormality related to the temperature in the column. Thus, in a particular system, which has a high number of monitored variables of the same type, the operator can efficiently identify a faulty situation without causing alarm flooding. However the operator will need more information to identify the root-causes of any failure.

3.2.2 Grouping by plant unit or equipment

In a complex process plant monitoring system, variables can be grouped unit- or equipment-wise. For example, measurements from the stripping section of a distillation column can be grouped together to assign an alarm whose annunciation would direct the operator to focus on that section and take actions. Thus the operator can effectively identify the failure location and further analyze the situation to find the root-cause without having many alarms from the same unit or system. But due to correlation of plant variables, one unit failure can be affected by other upstream variables and this can mislead the operators.

3.2.3 Grouping based on correlations

Strong correlations exist among plant variables due to their interactions and also due to plant connectivity. For example, the composition of the feed to a reactor may affect the conversion in the reactor leading to a changed product composition, product flow rate and/or the temperature in the reactor. If alarms are assigned to each of the variables, a change in the feed may cause a number of alarms to annunciate. Thus one failure may lead to many alarms. If variables are grouped according to their correlations, number of redundant alarms can be significantly reduced. However, the information from the alarm will be unclear. Also prioritizing of alarms can be difficult for the operator.

3.2.4 Grouping based on abnormal events

Variables related to an abnormal event may be grouped together to assign an alarm. For example, for a simple tank process, the flow rates of the inlet and the outlet streams along with the level of liquid in the tank may be related to an overflow condition of the tank. However, instead of assigning alarms to each of the variables, an overflow alarm can be defined based on the above measurements. Thus number of alarms can be reduced. In addition, the annunciation of the alarm would inform the operator about a defined event. Considering these advantages, the event-based grouping method was preferred over the other methods. Following section will demonstrate how to select process variables for event-based groups.

3.3 Variable selection methods

Identification of the key variables related to an abnormal event is a challenging task. The most important variables can be identified by various data-based techniques, or based on expert knowledge. Expert knowledge can be integrated with process risk assessment methods to identify variables that influence an abnormal event. However, if the process plant is large and complex or if there is not sufficient expert knowledge on the process, data based variable selection methods are more efficient. Z.Yang et al. (Z. Yang, Wang, & Chen, 2012) proposed a variable selection method based on the principle component analysis (PCA) and the resulting contribution plot to detect important variables to classify fault conditions. The concept of entropy from the information theory has been used to estimate the most informative variable related to a failure for the purpose of selecting sensor locations (Orantes, Kempowsky, Le Lann, & Aguilar-Martin, 2008). Mutual information concept for key variables selection by using information theory for Gaussian random variables is used for grouping variables to assign alarms by Pérez et al. (Pérez, Larrañaga, & Inza, 2006).

Based on variables information content, variables can be grouped together to represent different abnormal events. However, there can be highly correlated variables within a group. In order to identify the redundant variable within a group, correlation analysis can be performed. Various methods have been proposed in the literature to cluster process variables or alarms according to their correlation. Noda et al. and Yang et al. (Noda et al., 2011; Z. Yang et al., 2012) proposed methods to analyze correlated alarms by using binary alarm data. Geng et al. (Geng, Zhu, & Gu, 2005) proposed a method to cluster variables by fuzzy clustering method. Independent grouping analysis is proposed by Alhoniemi et al. (Alhoniemi et al., 2007) considering mutually dependent variables using a cost function. Information redundancy between variables using the concept of mutual information has been discussed in (Yu & Liu, 2003). The maximum cross-correlation among variables has also been used to identify the redundant variables (Swift, Tucker, Martin, & Liu, 2001).

In this study, selection of the key variables for a group is performed using the information that the variables can distinguish between an abnormal and a normal condition. The concept of mutual information is used to quantify the information. Variables with high information gain are grouped together for the respective abnormal event. To identify the redundant variables within a group to further reduce the number of variables to be monitored, the maximum cross-correlation between pairs of key variables are used. The following sections present the detailed methodology to allocate variables into event-based groups.

3.4 Methodology for allocating process variables to event-based groups

The proposed methodology for selecting process variables for abnormal event groups is presented in Figure 3-1. The initiating step is identification of abnormal event and corresponding failures (root causes) to generate process data or acquire historical data from the particular unit or process. Two subsequent steps to be followed: (i) to select the variable group with high information content about the event and (ii) identify redundant variables within the selected group. Mutual information between events and variables is used to calculate the information gain to identify the process variables that have high information gain for each abnormal event. Afterwards, cross correlation analysis between pairs of process variables for each abnormal event data set is used to identify redundant variables within each event group. Finally, using process knowledge and the grouping results, process variables are allocated for each abnormal event group.

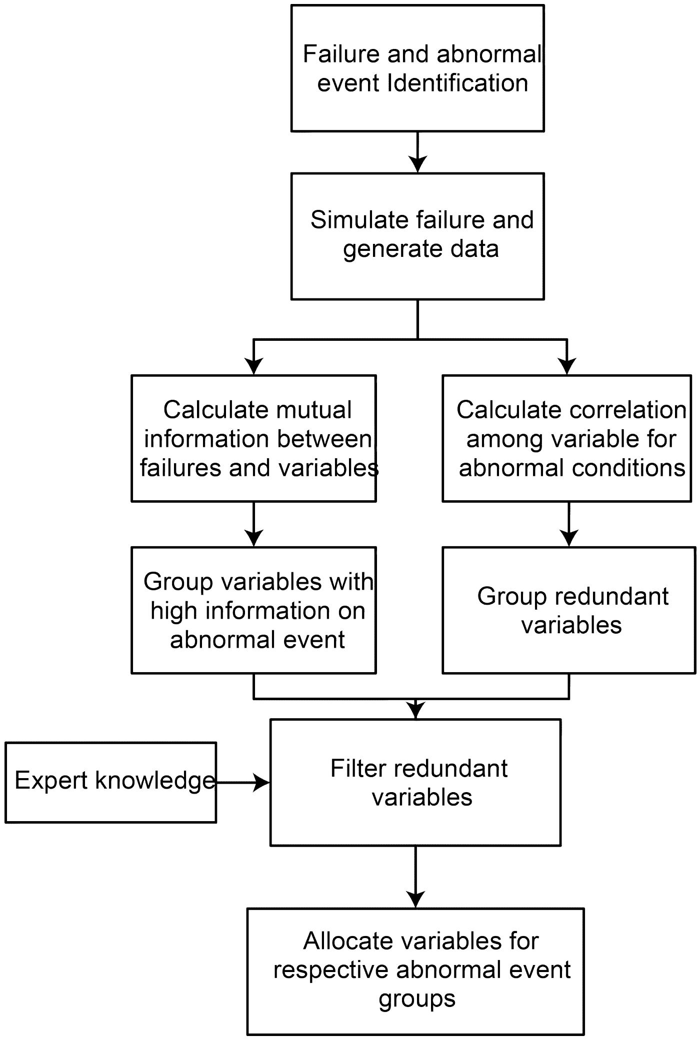


Figure 3-1: Methodology of allocation of variable for abnormal event groups

3.4.1 Process data generation

In order to select variables to form a group, the first step is to identify the abnormal events and root causes (failures) that can occur in the unit, equipment or a system. This is done by Hazard and Operability Study (HAZOP) (Crowl & Louvar, 2001). Information of the HAZOP study can also be used to select variables for each abnormal event.

However, if the plant is complex or processes are integrated, data based methodology can be used to get efficient results. Once the failures and the abnormal events are identified, process data are required to group variables. If the plant is at the design stage, simulation can be carried out to generate data for the abnormal events. For an operational plant, historical data can be used along with simulations to meet data requirements. Using data for both normal operations and for abnormal events, the information theory is used to select the key variables associated with an event.

3.4.2 Grouping variables according to information gain

3.4.2.1 The Information Theory

The information theory, proposed by Shannon (Shannon, 1948), which is routinely used in communication systems, measures the information content of a random variable in a quantitative manner. According to the theory, uncertainty associated with a random discrete variable X can be measured by its entropy $H(X)$ using equation 3-1

$$H(X) = -\sum_{x} P(x)\, log_2(P(x)) \qquad (3\text{-}1)$$

Here, X is assumed to be a discrete random variable. $P(x)$ is the probability mass density function of $X = x$ occurrence. Entropy is measured in unit `bits', therefore $\log_2$ is considered in the calculations.

For example, a fair coin toss, with $P(Head = True) = 0.5$ and $P(Tails = True) = 0.5$ have 1 bit of entropy. If the coin is not fair and the probability of getting a tails is p, then the probability of getting a head is $1 - p$. Following equation 3-1 the entropy of tossing a coin can be obtained as

$$H(p) = -p\, log_2\, p - (1 - p)\, log_2(1 - p)$$

The function $H(p)$ is plotted in Figure 3-2 to illustrate the basic characteristic of entropy. Entropy will maximize if the uncertainty of the random variable is maximum. In the above case, a fair coin represents maximum uncertainty $(P(Head = T) = 0.5)$ and results in the maximum entropy. Entropy will be minimum when the uncertainty is minimum. A biased coin that always get either head or tail $(P(Head = T) = 0 \text{ or } 1)$ has zero entropy.

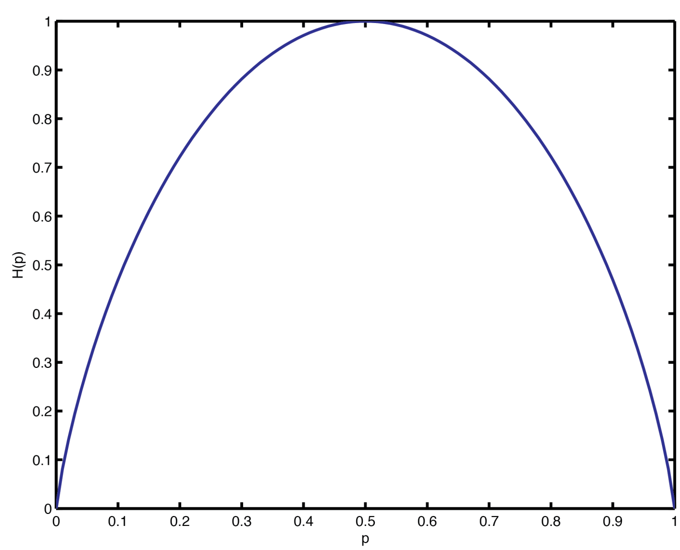


Figure 3-2: Entropy of tossing a coin.

For 2 discrete random variables X and Y, the joint entropy $H(X;\ Y)$ can be defined as,

$$H(X;\ Y\ ) = \sum_{x,y} P(x,y)\, log_2(P(x,y)) \tag{3-2}$$

$P(x,y)$ is the discrete joint probability distribution of $X = x$ and $Y = y$.

Conditional entropy between discrete random variables X and Y with joint probability distribution $P(x,y)$ and conditional distribution $P(x|y)$ is defined by,

$$H(X|\ Y\ ) = \sum_{x,y} P(x,y)\, log_2(P(x|y)) \tag{3-3}$$

3.4.2.2 Mutual Information

The mutual information which one random variable contains about another random variable can be derived as outlined in (Cover & Thomas, 1991).

$$I(X;\ Y\ ) = \sum_{x,y} P(x,y)\, \log_2 \frac{p(x,y)}{P(x)P(y)} \tag{3-4}$$

$I(X;\ Y\ )$ is the mutual information between random variables X and Y. Equation 3-4 can be simplified as follows

$$I(X;\ Y\ ) = H(X) - H(X|Y\ ) \tag{3-5}$$

$H(X|Y\ )$ is the entropy of the random variable X given Y. If $H(X|Y)$ is same as $H(X)$, then it is considered that the variable Y does not have any information about the variable X. However, if $H(X|Y)$ is lower than $H(X)$, it is considered that the variable Y does have some information about the variable X. That is uncertainty about X has reduced if information about variable Y is known. Hence, that reduction of uncertainty is defined as the mutual information gain of the variables X and Y.

3.4.2.3 Selection of variables using mutual information

The mutual information between two random variables can be used to select the key variables that contain significant information regarding an event. In order to do that, variable Y needs to be defined as a random variable that indicates the failures that can propagate to a specific abnormal event. For an example, if there is only one failure that can propagate to an abnormal condition then Y can be defined by two random numbers, $Y = 0$ (normal) and $Y = 1$ (failure). On the other hand if there are $k-1$ number of failures that can propagate to a specific abnormal event, then Y can have k random numbers (k being the number of failures plus the normal condition). If the variable does not have the ability to distinguish an abnormal event from the normal condition, then the amount of uncertainty do not change. If a variable can distinguish between conditions, then the amount of uncertainty will be reduced. Reduction of the uncertainty or entropy is the information gain that a variable contains. The following equation presents the information gain between a continuous random variable X and a discrete random variable Y that have k values.

$$I(X;\, Y\,) = \sum_{y_1}^{y_k} . \int_x P(y,x)\, log_2 \frac{p(y,x)}{P(y)p(x)} \mathrm{d}x \qquad (3\text{-}6)$$

Entropy of a continuous random variable having a normal distribution has been defined in (Cover & Thomas, 1991) as,

$$H(X) \ = \ \int_x p(x) \, log_2(p(x)) \, dx \ = \frac{1}{2} log_2(2\pi e \sigma^2) \qquad (3\text{-}7)$$

$p(x)$ is the probability distribution of a continuous random variable X and σ is the standard deviation of X. Pérez et al (Pérez et al., 2006) proved that if the variable X has a normal distribution and if C is a multinomial random variable having 1 to k finite outcome with a probability distribution of $P(C \ = \ c)$, and $p(c, x)$ is the joint probability distribution of $C \ = \ c$ and $X \ = \ x$, then the information that the variable can have for all the C values is given by,

$$I(X; C) \ = \ \sum_{c_1}^{c_k} \int_x P(c, x) \, log_2 \frac{p(c, y)}{P(c)P(y)} dx \qquad (3\text{-}8)$$

$I(X; C)$ is the mutual information between X and C. During normal conditions, variations of data occur only due to measurement noise, which is typically white noise with small magnitude. But if the variable contains high information, then for each failure condition the variation in data will be significant. Therefore entropy between failures and the variable will decrease. Information gain for a continuous random variable X having a Gaussian distribution and a multinomial variable C is derived in Pérez et al (Pérez et al., 2006) as follows,

$$I(X; \ C \ ) = \sum_{c_1}^{c_k} . \int_x P(c, x) \log_2 \frac{p(c, x)}{P(c)p(x)} dx = \sum_{c_1}^{c_k} . \int_x P(c) p(x|c) \, log_2 \frac{p(x|c)}{p(x)} dx$$

$$I(X;\ C\ ) = \sum_{c_1}^{c_k} P(c) \int_x p(x|c)\, log_2\, p(x|c)\, dx - \sum_{c_1}^{c_k} . \int_x P(c)p(x|c)\, \log_2 p(x)\, dx$$

The second integral can be expressed as,

$$\sum_{c_1}^{c_k} . \int_x P(c)p(x|c)\, log_2\, p(x)\, dx = \int_x \sum_{c_1}^{c_k} p(x,c)\, log_2\, p(x)\, dx$$

$$\sum_{c_1}^{c_k} . \int_x P(c)p(x|c)\, log_2\, p(x)\, dx = \int_x p(x)\, log_2\, p(x)\, dx = -\frac{1}{2} log_2(2\pi e \sigma^2)$$

and then,

$$I(X;\ C\ ) = \sum_{c_1}^{c_k} P(c) \left(-\frac{1}{2} log_2(2\pi e \sigma^2)\right) + \frac{1}{2} log_2(2\pi e \sigma^2)$$

$$I(X;\ C\ ) = -\frac{1}{2} log_2(2\pi e) - \frac{1}{2} \sum_{c_1}^{c_k} P(c)\, log_2(\sigma_c^2) + \frac{1}{2} log_2(2\pi e) + \frac{1}{2} log_2(\sigma^2)$$

$$I(X;C)\ = \frac{1}{2}\left[log_2(\sigma^2) - \sum_{c_1}^{c_k} P(c)\, log_2(\sigma_c^2)\right] \qquad (3\text{-}9)$$

Equation 3-9 is used to calculate the information gain. Where σ is the standard deviation of the random variable X and σ_c is the standard deviation of the random variable X given $C = c$. $P(C)$ is assumed to have uniform distribution, implying that the information about the normal and the failure conditions are unknown and their probability of occurrence are the same. It is also assumed that the data acquired for each variable in different conditions are normally distributed and C is considered as multinomial random variable. Finally variables having high I values are selected as the most suitable variables to monitor the respective event.

3.4.3 Identification of the redundant variables

Variables that are selected for each event group may contain the same information and thus they can be considered to be redundant. The redundant variables are identified using the correlation analysis. The purpose is to identify redundant variables within a group and thus to exclude all but one from a redundant set for monitoring. To perform the correlation analysis, data are standardized to have zero means. The cross-correlation between pairs of variables are estimated. There can be time lags between variables. Hence, to calculate the maximum correlation, time lag is varied and the correlations are calculated to get the maximum positive or maximum negative value. Maximum time lag can be decided using process knowledge (Swift et al., 2001). Pearson correlation coefficient is used to calculate the similarity between x and y continuous process variables as follows (F. Yang, Sirish, & Xiao, 2010)

$$\phi_{x,y}(lag) = \frac{E\left[(x_i - \mu_x)\left(y_{i+lag} - \mu_y\right)\right]}{\sigma_x \sigma_y} \tag{3-10}$$

$$\phi_{x,y}(lag) = \frac{1}{n - lag} \sum_{i=1}^{n-lag} \frac{(x_i - \mu_x)\left(y_{i+lag} - \mu_y\right)}{\sigma_x \sigma_y} \quad lag \geq 0 \tag{3-11}$$

$$\phi_{x,y}(lag) = \frac{1}{n - lag} \sum_{i=1-lag}^{n} \frac{(x_i - \mu_x)\left(y_{i+lag} - \mu_y\right)}{\sigma_x \sigma_y} \quad lag < 0 \tag{3-12}$$

σ_x *and* σ_y are the standard deviations of x and y, and μ_x *and* μ_y are their mean values. lag is the time lag of y with respect to x. At the maximum positive correlation coefficient ϕ_{max}, time lag is lag_{max} and at maximum negative correlation coefficient ϕ_{min}, time lag is lag_{min}. Then the maximum absolute correlation coefficient is calculated as follows (F. Yang et al., 2010),

$$\phi_{max}\ is\ taken\ at\ lag_{max}\ if\ \phi_{max} \geq\ -\phi_{min}$$

$$-\phi_{min}\ is\ taken\ at\ lag_{min}\ if -\phi_{min}\ \geq\ \phi_{max}$$

The correlation matrix is thus developed. Variables that are highly correlated with each other are grouped together. For the purpose of better visual representation, grouping is done by calculating the similarity distance between each pair of variables. After getting the distance between variables in the data, variables close to each other can be linked and presented in clusters in a hierarchical tree dendrogram (Martinez, Martinez, & Solka, 2004).

3.4.4 Allocating variables for abnormal event group

Finally after calculating the information that process variables have regarding the event and the redundant variable within the group, variable allocation is carried out. Expert knowledge is also used to justify the choice of variables within a redundant group.

3.5 Application of the variable allocating methodology

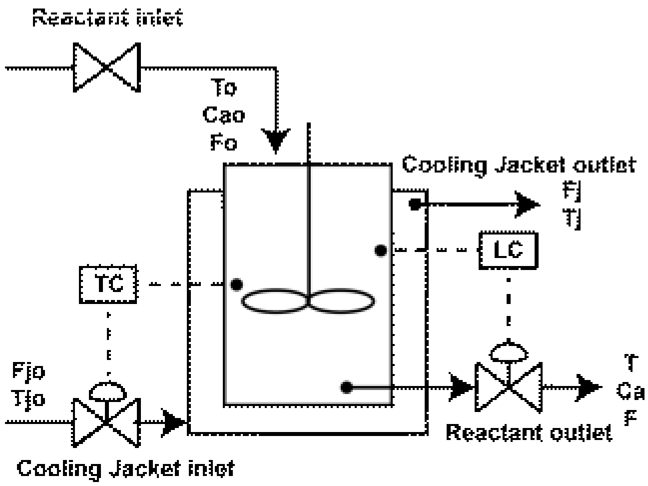


Figure 3-3: CSTR process diagram

To demonstrate the methodology and its applicability, a jacketed continuous stirred tank reactor (CSTR) is considered. An irreversible exothermic reaction $A \rightarrow B$ is assumed to take place in the reactor with a first order kinetics. A temperature controller is used to control the reactor temperature by manipulating the coolant flow rate. The level of liquid in the reactor is also maintained by manipulating the reactor outlet flow. Heat losses are considered negligible and a perfect mixing condition is assumed. All the parameters for the model are taken from literature (Luyben, 1996) and the controller parameters are taken from Chang et al. (C. Chang & Yu, 1990). Simulation model used due to the unavailability of real process operation data. This study is further extended to develop an early waning system and the same CSTR operation is used through the study due to the convenient. To demonstrate the methodology, Simulink is used to build a plant model. Detail model is presented in the appendix. Different failure conditions are simulated with

Simulink to generate data. For the CSTR, 11 variables are identified as measurement variables as presented in Table 3-1.

Table 3-1: List of variables that are measured in the CSTR process

No	***Measured variable***
1	*Reactor liquid level*
2	*Coolant utility outlet temperature*
3	*Reactant concentration*
4	*Reactor vessel temperature*
5	*Reactor output flow rate*
6	*Coolant utility flow rate*
7	*Reactant feed temperature*
8	*Reactant feed flow rate*
9	*Coolant inlet temperature*
10	*Level controller output*
11	*Temperature controller output*

Variables 5 and 6 are manipulated to control variables 1 and 4, respectively. Other variables are uncontrolled variables. Ten failures are considered for this study. Using Simulink, data for all the failures and the normal condition are generated. Table 3-2 presents the failures.

Table 3-2: Possible failure conditions for the CSTR process

No Failure	*Failures (Root-causes)*
F1	*Reactant feed flow disturbance - High flow*
F2	*Reactant feed flow disturbance- Low Flow*
F3	*Coolant system failure - High coolant temperature*
F4	*Coolant system failure - Low coolant temperature*
F5	*Reactor out flow valve failure- High flow*
F6	*Reactor out flow valve failure- Low flow*
F7	*Coolant flow valve failure - High flow*
F8	*Coolant flow valve failure - Low flow*
F9	*Reactant feed quality failure - High concentration*
F10	*Reactant feed quality failure - Low concentration*

Some of these failures can propagate to more severe abnormal events. Table 3-3 presents the abnormal events and the respective failures (root-causes) that can lead to the events.

Table 3-3: Possible abnormal events for the CSTR process

Abnormal Event	*Failure (Root-causes)*
Runaway	*F3, F8*
Flooding	*F1, F6*
Low quality products	*F9, F4, F7*

3.5.1 Grouping variables

Data were gathered from simulating the failure conditions. For each variable, information gain was calculated by using equation 3-9. Information gains for failures are calculated considering failure and normal condition data variation. For abnormal event all the related failures and normal condition data are used to calculate the information gain according to the equation 3-9.

Then variables are grouped according to their information gain. First a pairwise comparison for the normal and a failure is carried out to identify variables that can reduce the uncertainty of the failure under consideration.

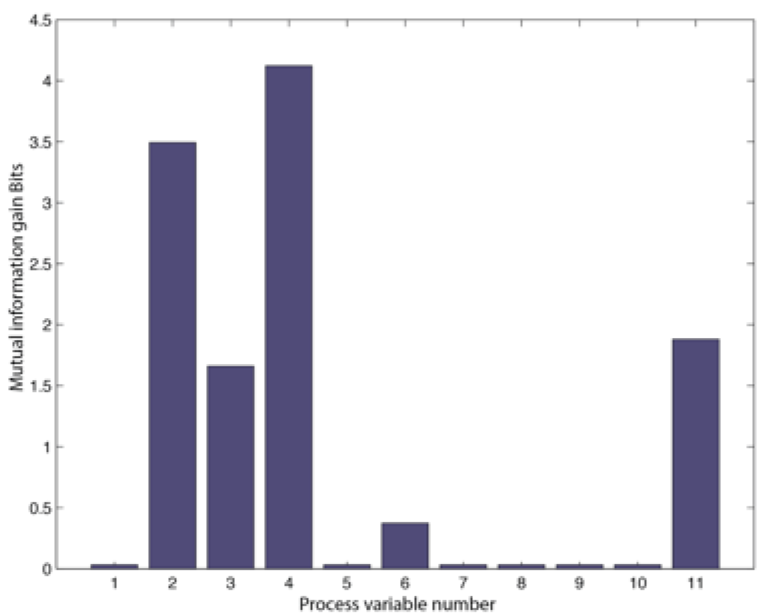


Figure 3-4: Information gain of different variable corresponding to failure F8: coolant valve failure - low coolant flow

As observed from Figure 3-4, there are 5 main variables 4, 2, 11, 3 and 6 having significant information about the coolant valve failure. Accordingly, these sets of five variables are considered as the key variables for the failure F8. Following the same procedure, key variables are identified for all the failures.

In order to group variables according to the abnormal events, information gains are calculated by considering all the failures that can propagate to the corresponding abnormal event. Figure 3-5 presents the mutual information gain for all of the listed variables for the abnormal event, runaway reaction.

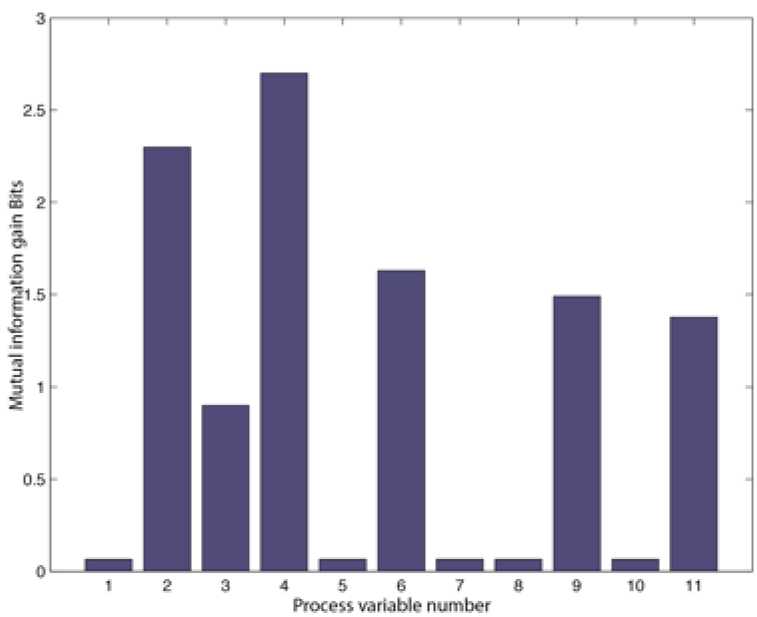


Figure 3-5: Information gain of different variable for the event- runaway reaction

As shown in Figure 3-5, the variables 4, 2, 11, 6, 3 and 9 can give significant information about the runaway event. Accordingly, these set of six variables are considered as the key variables for the event runaway reaction. Following the same procedure, key variables are identified for all the events.

3.5.2 Redundant variable selection

Cross correlation analyses are carried out to identify the redundant variables within a group of key variables, which are selected for each event. The maximum cross correlation matrix is generated by varying the time lag between each pair of variables for each abnormal condition data. For the purpose of visualization, a hierarchical cluster tree is developed in the form of a dendrogram as presented in Figure 3-6 that shows the correlations among variables related to the runaway event. Finally it is required to choose one variable from each redundant group for the purpose of minimizing the monitored variables.

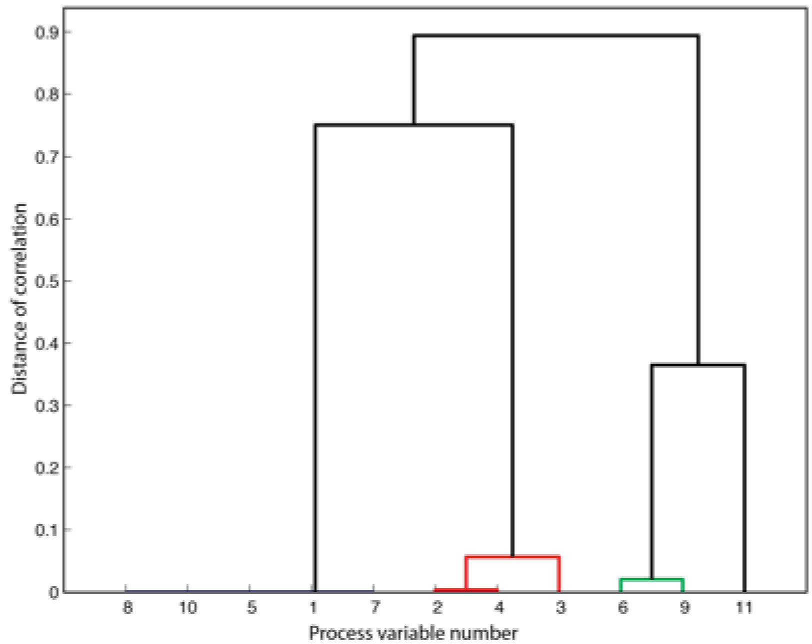


Figure 3-6: Correlation among variables for the runaway event

3.5.3 Allocating variables to groups

Choosing the most suitable variable from a redundant set may become a challenging task. Process knowledge as well as sensor characteristics may be required to consider for this purpose. In this case the variable with the highest information gain among the redundant variables is chosen. Table 3-4 presents the list of variables that can be allocated to the groups corresponding to the individual events and the highly correlated variables within different groups. It also shows the final group selection for each of the events.

Table 3-4: Selected variable groups for different abnormal events

Event	*Key variables*	*Redundant variable*	*Chosen groups*
Runaway	*(2,3,4,6,9,11)*	*(2,3,4) & (6,9)*	*(2,6,11)*
Flooding	*(1,3,5,8,10)*	*(3,5,8)*	*(1,8,10)*
Low Quality	*(2,3,4,11)*	*(2,4)*	*(3,4,11)*

3.6 Discussion

From the results, reactor vessel temperature (variable 2), coolant utility flow rate (variable 6) and temperature controller output (variable 11) are the main variables that have most information regarding runaway reaction. It is obvious that the reactor temperature is the main variables that can be used to detect a runaway. Main root cause for the runaway reaction is the failure of the coolant system. Variable 6 and 11 are directly related to the coolant system failures.

Primary variable for flooding condition monitoring is level of the reactor (Variable 1). Therefore it should be a key variable. The main root causes are the level controller failure and feed flow valve failure. Variables which are directly related to both failures and identified as the key variables are the reactant feed flow rate (variable 8) and the level controller output (variable 10).

Low quality production can be quantified by reactant concentration (variable 3), which is a key variable according to the methodology. Incomplete reaction due to the low temperature is the main reason for low quality production. The Proposed methodology has identified reactor vessel temperature (variable 4) and temperature controller output (variable 11) as other key variables to detect low quality production. The case study demonstrates that process knowledge justifies the selection of the key variables by the proposed methodology.

As mentioned in section 3.2, there are other options to group variables. Table 3-5 presents the selected group of variables according to different grouping methods. As shown in the table, different methods may result in significantly different results.

Table 3-5: Results on group formation using different methodologies

Groups	***G1***	***G2***	***G3***	***G4***	***G5***
Variable type	*1*	*2,4,7,9*	*3*	*5,6,8*	*10,11*
Plant/Unit	*4,2,6,9,11*	*1,4*	*3,7,8*	*1,5,10*	
Correlation	*2,3,4*	*6,9*	*7,11*	*5,8,10*	*1*
Event based	*2,6,11*	*1,8,10*	*3,4,11*		

3.7 Conclusion

A procedure for selection of variables to form groups for an event-based alarm system is detailed. The method uses the information theory and the concept of mutual information to select the key variables to allocate to a group. Correlation analyses are then carried out to select the redundant variables within a group. A case study using the example of a CSTR is used to elaborate the proposed methodology. Following the same procedure, variable selection to design an event-based alarm system can be carried out for an entire plant. Once variables are selected to form groups, one alarm will be assigned to each group. Finally, an event-based approach will be used to estimate the probability of event occurrence. The warning will be annunciated if the overall probability is higher than a pre-chosen threshold. This chapter outlines the grouping methods; event probability estimation and root-cause diagnosis of event warning have been addressed in the next chapter.

4 EARLY WARNING SYSTEM DESIGNING METHODOLOGY

The proposed early warning system design methodology includes two main elements. First, a Bayesian Network (BN) is constructed to define the relationship of each identified event with the factors associated with it. Second, the BN is used to calculate the real time probability of the event occurrence and to diagnose the root-causes of the event utilizing sensor measurements.

Figure 4-1 presents the complete methodology for the early warning system. The proposed procedure starts with identifying (a) the significant abnormal events, (b) scenarios associated with each event, (c) process measurements (symptoms) related to each event and (d) the root-causes of the events. Subsequently, Bayesian networks are developed to define the following relationships (i) between process event and scenarios (ii) between scenarios and symptoms, and (iii) between symptoms and root-causes. The first two relations are used to issue warnings and the third relation is used to identify the root causes of a warning. Real time measurements of process variables are used as inputs to the warning system. The measured variables are used as evidences to update the probabilities of symptom nodes in the Bayesian Network. Next, the probabilities of scenarios and event occurrences are assessed using BN forward inference. If the probability of an event occurrence exceeds a predefined threshold value, a warning is annunciated. When a warning is annunciated, the root-cause analysis is carried out using the evidences. The methodology includes four main steps described in the following sections.

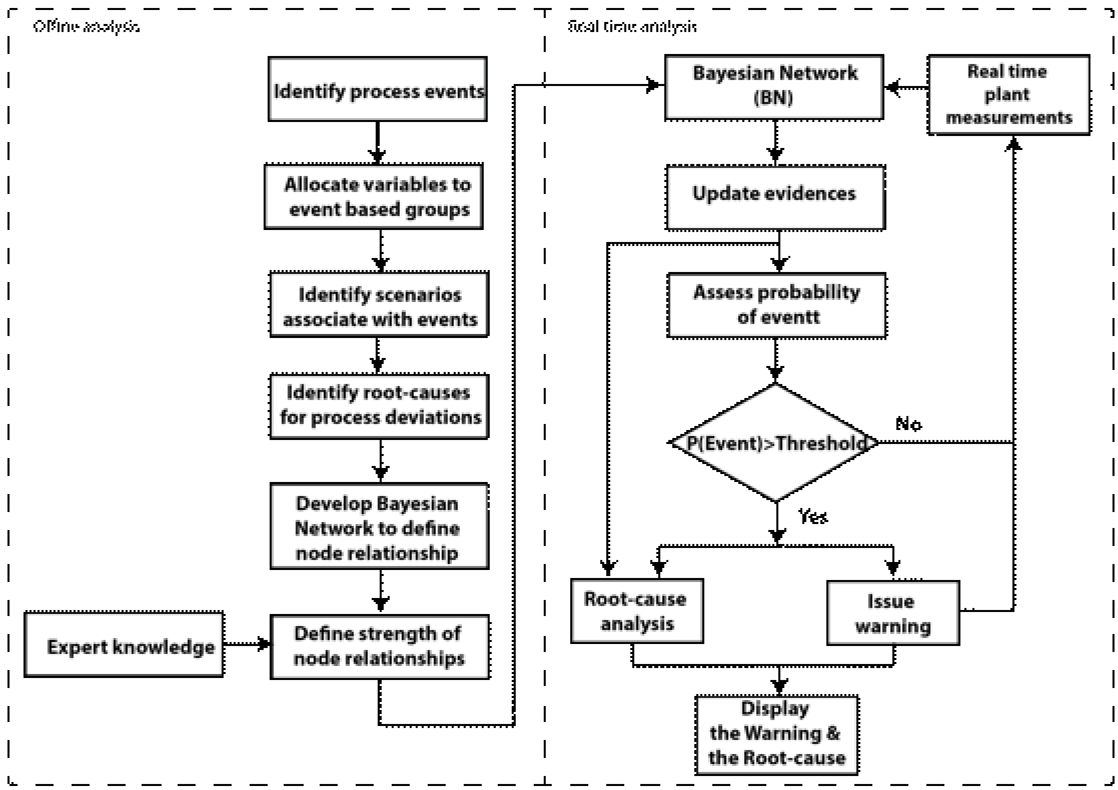


Figure 4-1: Event-based early warning system design methodology

4.1 Step 1: Identification of events and corresponding scenarios and root-causes

4.1.1 Identification of events

In the proposed methodology, warnings are assigned to undesirable events instead of individual variables. Events, in the context of the proposed warning system, are defined as undesirable abnormal conditions such as runaway reaction in a reactor, flooding of a tank or operational problems e.g. plant shutdown or product degradation. Figure 4-2 presents the propagation of an abnormal event. One or several initial causes such as failures can cause deviation of the process variables away from their normal operating conditions. The process can further deviate from normal operation range due to failure of process safety barriers. Subsequently, an undesirable event can occur.

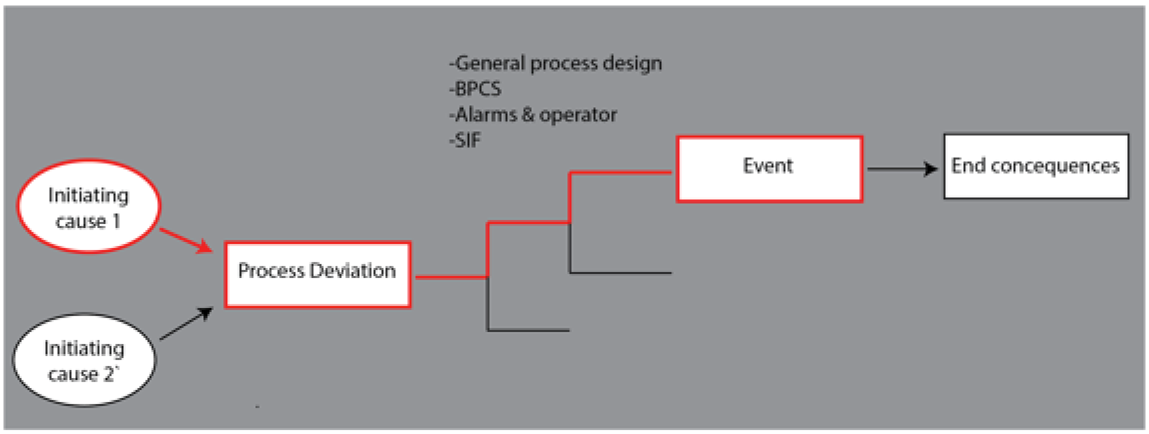


Figure 4-2: Abnormal event propagation

The first step in the methodology is to identify the undesired events, which can take place during the operation of the particular process. There are many risk assessment tools that can be used to identify events. In the proposed procedure, the Hazard and Operability Study (HAZOP) is used to identify the potential events. HAZOP study is a qualitative risk assessment method to identify process hazards that can occur due to deviations of the process variables. Piping and instrumentation diagrams are used to identify the propagation of major process variable deviations. The process variables are systematically analyzed using the so-called guidewords to identify potential consequences. The root-causes of the corresponding deviations are also identified. Table 4-1 presents the list of HAZOP guidewords.

Table 4-1: HAZOP guideword list

Guidewords	*Meaning*
NO, NOT, NONE	*The complete negation of the intention*
MORE, HIGHER, GREATER	*Quantitative increase*
LESS, LOWER	*Quantitative decrease*
AS WELL AS	*Quantitative increase*
PART OF	*Quantitative decrease*
REVERSE	*The logical opposite of*
OTHER THAN	*Complete substitution*
SOONER THAN	*Too early or in the wrong order*
WHERE ELSE	*In additional location*
LATER THAN	*Too late or in the wrong order*

After carrying out the study, the significant abnormal conditions that need to be warned in the operation are identified as events. For example, in a simple tank process, a deviation in the flow rate of the inlet or the outlet stream may lead to over-flow of the tank with significant consequences. Hence, the overflow is considered as an event for the tank process. Table 4-2 presents the HAZOP table for a buffer tank operation.

Table 4-2: Example HAZOP study for tank operation

	Process variables	***Deviation***	***Consequences***	***Causes***
1	*Level*	*High*	*1. Overflow of tank*	*1. High inlet flow*
			2. Environment hazard	*2. Failure of the level control system*
				3. Blockage in outlet pipe
				4. Secondary inlet valve failure
		Low	*1. Possible damage to Pump*	*1. High flow*
				2. Failure of the level control system
				3. Tank leakage
2	*Outlet flow*	*High*	*1. None identified*	*1. Failure of the level control system*
				2. Secondary inlet valve failure
		Low/No	*1. Overflow of tank*	*1. Outlet valve blockage*
			2. Environment hazard	
3	*Inlet flow*	*High*	*1. Overflow of tank*	*1. Failure of the level control system*
			2. Environment hazard	*2. Pump high speed*
		Reverse	*1. None identified*	*1. Pump mechanical problem*
	. . .	. . .	. . .	. . .

4.1.2 Variable allocation for event-based groups

After identifying the significant events, monitored process variables are selected according to their abilities to distinguish between an abnormal event and a normal condition. The deviations of these process variables are defined as the symptoms of the event. If the process plant is large and complex or if there is not sufficient expert knowledge about the process, data-based variable selection methods are more efficient to allocate variables to abnormal event-based groups. A detailed methodology on allocation of variables to event-based groups using mutual information and cross correlation analysis has been described in chapter 3. In the proposed early warning design methodology, mutual information between an event and a process variable is used to select the key variables if the plant is complex. Process knowledge is used for allocation of variables for simple processes.

4.1.3 Identification of corresponding scenarios

Scenarios are defined as the process operating conditions that influence an event. Deviations of process variables and their correlations during an event are considered when determining scenarios. Deviation of one variable or combination of different variables causes a scenario. For example, saturation of a flow control valve along with an increase in the level of liquid in a tank can be considered as one scenario. Both of these conditions are needed to occur for this scenario to happen. If the controller valve reaches to the saturation limit without level deviation then this condition is controllable and it will not influence the overflow event.

4.1.4 Identification of corresponding root-causes

Possible root-causes are identified that can influence the process variables to deviate from their respective normal limits and thus the occurrence of an event. These root-causes can be external disturbances, equipment malfunctions, control system failures or human errors. In this methodology, all the root-causes that are associated with the events are identified using the information from the HAZOP study. An alternative to the HAZOP study is the Failure Mode and Effect Analysis (FMEA). FMEA identifies the failures in process equipment as well as the effect of these failures. In this method, process components are considered and the failure modes that can occur during the operation are examined. Also, the undesirable effects of each failure are evaluated. Failures, which influence the process variable deviation and finally propagate to an event, are considered as root-causes of the corresponding event.

4.2 Step 2: Development of Bayesian Network (BN)

After identifying the event and the corresponding scenarios, symptoms and root-causes, Bayesian Network is used to develop the event-based early warning system.

4.2.1 Bayesian Network

Bayesian Network (BN) is a graphical method that is used to model causal relationship among random variables. BN contains two parts, a graphical structure that defines the qualitative representation, and the conditional probabilities that define the quantitative relations. The network structure is a directed acyclic graph containing nodes and arcs: nodes represent uncertain variables and arcs represent the direct casual relationship or the influence between linked nodes (Pearl, 1988; Korb & Nicholson, 2003).

BN represents the joint probability distribution of discrete random variable nodes $X = (X_1, X_2, \ldots X_n)$. According to the chain rule of probability theory, joint probability $P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = P(x_1, x_2, \ldots\ldots, x_n)$ is factorized as,

$$P(x_1, x_2, \ldots\ldots, x_n) = P(x_1) * P(x_2|x_1) * \ldots.* P(x_n|x_1, x_2, \ldots\ldots x_{n-1}) \qquad (4\text{-}1)$$

According to the d-separation property of BN, root nodes are conditionally independent and the rest of the nodes are conditionally dependent with their direct parents. Hence, the joint probability distribution is compacted as,

$$P(x_1, x_2, \ldots\ldots, x_n) = \prod_{i}^{n} P(x_i | Parents(X_i)) \quad (4\text{-}2)$$

$Parents(X_i)$ are the parent nodes that are directly connected with the node X_i (Korb & Nicholson, 2003).

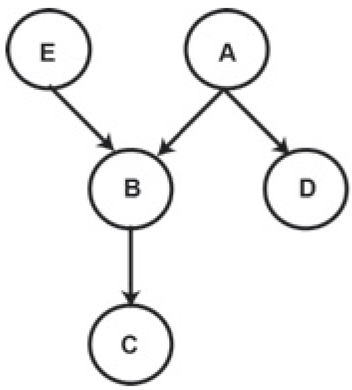


Figure 4-3: Bayesian Network example 1 network topology

Example 1: Figure 4-3 presents a simple Bayesian Network having random variable A, B, C, D, and E. Joint probability distribution of the BN can be presented as,

$$P(A, B, C, D, E) \; = \; P(E|A, B, C, D) \; P(D|A, B, C) \; P(C|B, A) \; P(B|A) \; P(A)$$

However due to the variable independency that represented in the network, above joint probability can compress as,

$$P(A, B, C, D, E) \; = \; P(E) \; P(D|A) \; P(C|B) \; P(B|A, E) \; P(A)$$

In order to quantify the network, conditional probability tables of $P(D|A)$, $P(C|A)$, and $P(B|A, E)$ are required.

The most important property of the BN is its ability to perform probabilistic inference. Any evidence (E) can be entered into any node and according to the evidence, the belief of the other nodes are updated (Khakzad et al., 2011; Korb & Nicholson, 2003).

$$P(\mathrm{X}|evidence) = \frac{P(X, evidence)}{P(evidence)} = \frac{P(X, evidence)}{\sum_n P(X_i, evidence)} \tag{4-3}$$

In this methodology, probabilities of the symptoms nodes are updated using real time process evidences. According to the symptoms, beliefs about the other nodes probabilities are updated. There are two cases of belief updating, first is the forward inference which is the prediction of a child node probability using the evidences of parent nodes. The forward inference is used to calculate the $P(Event|Symptops)$. Backward inference, which diagnoses a parent node probability, using evidences of the child nodes is used to calculate the $P(RootCauses_i|Symptoms)$ using the Bayes theorem (Korb & Nicholson, 2003).

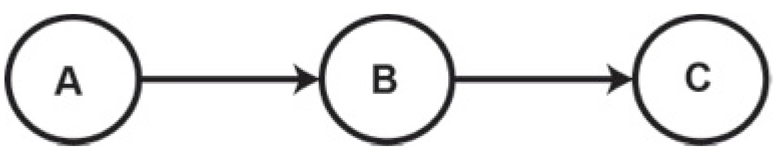


Figure 4-4: Bayesian Network example 2 network topology

Example 2: Figure 4-4 present a simple Bayesian network about a student getting a scholarship.

A: Study hard for exam, B: Get high score for exam and C: Get a scholarship.

Conditional probabilities of the BN:

$$P(B = T|A = T) = 0.8, \qquad P(B = T|A = F) = 0.3$$

$$P(C = T|B = T) = 0.75, \qquad P(C = T|B = F) = 0.1$$

Case 1: If there is an evidence that probability of student studying hard is $P(A = T) = 0.7$, then from BN forward inference, the probability of getting a high score is estimated as follows,

$$P(B = T, A) = P(B = T|A = T) * P(A = T) + P(B = T|A = F) * P(A = F)$$

$$= 0.8 * 0.7 + 0.3 * 0.3 = 0.65$$

Subsequently, probability of student getting a scholarship is estimated as follows,

$$P(C = T, B) = P(C = T|B = T) * P(B = T) + P(C = T|B = F) * P(B = F)$$

$$= 0.75 * 0.65 + 0.1 * 0.35 = 0.5225$$

Case 2: If there is uncertainty regarding whether the student has studied hard or not, then from the history of the student, probability of student studying hard is believed as $P(A = T) = 0.4$. However, if there is evidence that the student got 80% score, $P(B = T)$ can be updated to 0.8. From the evidence of the exam score, belief about the student has studied hard can be updated by BN backward inference as follows,

$$P(A = T|B = T) = \frac{P(A = T, B = T)}{\sum_{A=T,F} P(A, B = T)} = \lambda P(B = T|A = T)P(A = T)$$

$$= \lambda * 0.8 * 0.4 \;=\; \lambda 0.32$$

$$P(A = F|B = T) = \frac{P(A = F, B = T)}{\sum_{A=T,F} P(\text{A}, B = T)} \;=\; \lambda P(B = T|A = F)P(A = F)$$

$$= \lambda * 0.3 * 0.6 \;=\; \lambda 0.18$$

$$1 = P(A = T|B = T) + P(A = F|B = T) = \lambda 0.32 + \lambda 0.18 = \lambda 0.5$$

$$\lambda = 2$$

$$P(A = T|B = T) = 0.64$$

Similarly $P(A = T|B = F)$ is calculated $P(A = T|B = F) = 0.16$

But with the evidence, $P(B = T) = 0.8$, we can estimate,

$$P(A = T, B) = P(A = T|B = T) * P(B = T) + P(A = T|B = F) * P(B = F)$$

$$= 0.64 * 0.8 + 0.16 * 0.2 = 0.544$$

BN forward inference is used to calculate the probability of getting a scholarship as follows,

$$P(C = T, B) = P(C = T|B = T) * P(B = T) + P(C = T|B = F) * P(B = F)$$

$$= 0.75 * 0.8 + 0.1 * 0.2 = 0.62$$

According to the evidence, the belief about whether the student has studied hard has increased from 0.4 to 0.544 and the likelihood of student getting a scholarship is estimated as 0.62.

4.2.2 Bayesian Network topology development

Construction of the BN topology depends on the qualitative relationships among variable nodes. The proposed network consists of four different layers of nodes. Root nodes are root-causes that influence the deviation of the process variables. First intermediate nodes that are influenced by the root-cause nodes are the symptom nodes. Other intermediate nodes, which are influenced by symptoms, are the scenario nodes. Finally leaf nodes, which do not have any child, are considered as event node. Figure 4-5 presents the BN structure for the proposed early warning system.

According to the network in Figure 4-5, joint probability of the event (E) and the corresponding scenarios (SC), symptoms (SY) and root-causes (RC) are expressed as follows,

$$P(\mathrm{E}, \mathrm{SC}, \mathrm{SY}, \mathrm{RC}) = \mathrm{P}(\mathrm{E}|\mathrm{SC}) \prod_{i=1}^{l} P(\mathrm{SC_i}|\mathrm{SY}) * \prod_{j=1}^{m} P(\mathrm{SY_j}|\mathrm{RC}) * \prod_{k=1}^{n} P(\mathrm{RC_k}) \qquad (4\text{-}4)$$

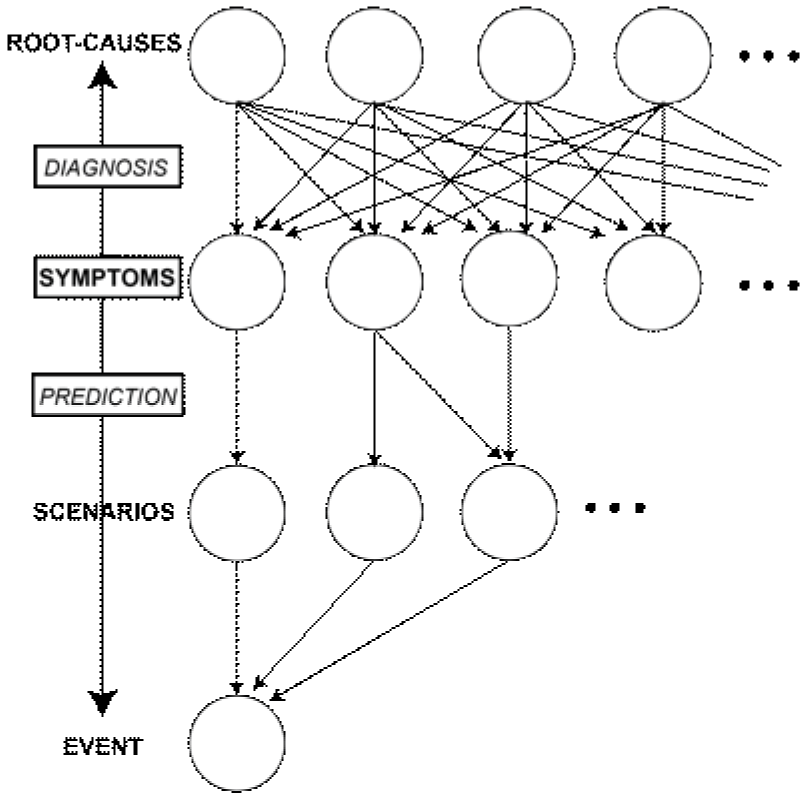


Figure 4-5: Bayesian Network topology representing early warning system

4.2.3 Defining strength of the variables

The conditional probabilities define the strength of the causal relationships among parent nodes and child nodes. The probabilities can be defined based on expert knowledge or from historical data using Bayesian learning methods. In this study, expert knowledge is used to define the strength of the relationship among nodes due to the unavailability of the large number of historical process data on different abnormal events.

The main obstacle in determining the conditional probabilities arises when the number of parent variable nodes is large. With the increase in parent nodes, the number of conditional probability requirement increase exponentially. Variable nodes in the proposed methodology consider the occurrence and nonoccurrence of respective variable node as their states. Hence, the proposed BN has only binary state variable nodes. Therefore, if there are n number of parents for a single child, then 2^n number of parameters are required to define the network strength (Pearl, 1988; Bobbio, Portinale, Minichino, & Ciancamerla, 2001; Flairs, Barr, Markov, Zagorecki, & Druzdzel, 2004). There can be a large number of parent nodes (root-cause or scenario nodes) that influence the child node (symptom or event node). Hence, a large number of conditional parameters are needed to define the proposed network. This situation would not be desirable or practical for a large industrial plant. In order to reduce the number of parameters, Noisy OR is used that simplifies the Bayesian Network (Bobbio et al., 2001; Flairs et al., 2004; Khakzad, Khan, Amyotte, & Cozzani, 2012). If there is n number of parent nodes directly connected to a child node, then the conditional probability parameters that are needed to define are reduced from 2^n to n when the Noisy OR logic algorithm is used. Noisy OR can be used if a variable Y has binary parent variables $parents(Y) = (X_1, X_2 \dots X_n)$ and also if the following two assumptions are true,

1. Parent node X_i influences child node Y independently from other parent nodes. In the proposed network, each root-cause nodes and scenario nodes can independently influence the symptom nodes and event nodes without occurrence of other respective parent nodes.

2. Y is false if none of the parent is true, $P(E|\overline{X_1} \dots \overline{X_n}) = 0$. It is assumed that when identifying the root-causes and scenarios, all the possible conditions are identified. Hence,

if all the root-causes do not occur, the probability of symptom occurrence is zero and if all the scenarios do not occur, the probability of event occurrence is zero. The conditional probability of Y can be calculated as follows,

$$P(Y|X) = 1 - \prod_{X_i=\pi}(1 - p_i) \tag{4-5}$$

Where, π is the set of true variables in X and $p_i = P(Y|\overline{X_1}, \overline{X_2}, \dots, \overline{X_{i-1}}, X_i, \overline{X_{i+1}}, \dots \overline{X_n})$ is the probability of Y given X_i is true and all other parents are false (for all *i=1* to *n*) (Onisko, Druzdzel, & Wasyluk, 2000; Bobbio et al., 2001). Therefore, Noisy OR is used between root-causes and symptoms and between scenarios and event nodes. Only the individual influence of parent to its child needs to be defined and the other conditional probabilities are calculated using the Noisy OR algorithm.

Example 3: Consider the simple Bayesian Network with binary variables present in Figure 4-6 (A: Flue, B: Cold C: Malaria and D: Fever). Each parent nodes can influence the child node. In order to complete the conditional probability table for this particular BN, 8 conditional probabilities are needed. However, with Noisy OR technique the conditional probability requirement can reduce to 3. However, the following assumptions should be satisfied, (1) each parent node have to independently influence the child node and (2) if all the parent node are false then the child node should be false.

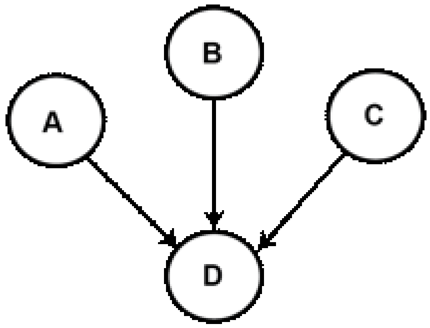


Figure 4-6: Bayesian Network example 03 Network topology

Then only the 3 probabilities that are needed to complete the CPT, $P(D|A,\bar{B},\bar{C}) = 0.6$; $P(D|\bar{A},B,\bar{C}) = 0.9$; $P(D|\bar{A},\bar{B},C) = 0.98$. Table 4-3 presents the calculation of conditional probabilities.

Table 4-3: Bayesian Network conditional probability table for example 3

Cold	***Flue***	***Malaria***	$\boldsymbol{P(Fever \mid Cold, Flue, Malaria)}$	
T	*F*	*F*		*0.3*
T	*F*	*T*	*1-(1-0.3)*(1-0.98)*	*0.986*
T	*T*	*F*	*1-(1-0.3)*(1-0.9)*	*0.93*
T	*T*	*T*	*1-(1-0.3)*(1-0.9)*(1-0.98)*	*0.9986*
F	*F*	*F*		*0*
F	*F*	*T*		*0.98*
F	*T*	*F*		*0.9*
F	*T*	*T*	*1-(1-0.9)*(1-0.98)*	*0.998*

4.3 Step 3: Real time probability estimation and alarm annunciation

4.3.1 Evidence updating from real time sensor measurements

Evidences can be updated to any node in the Bayesian Network. In the proposed warning system model, evidences are the real time sensor measurements. From these evidences, probabilities of the symptom nodes are updated. Symptoms are defined as the occurrence of the variable deviation to a faulty state. Normal range of a variable is defined by statistical process control 3-sigma method or according to the desirable process limits of the respective variables. In this study, process variable values between upper control limit (UCL) and lower control limit (LCL) are considered as the normal range. Values, which fall outside that range, are considered as abnormal. If a variable is at its threshold limit, then it can either move back to normal condition or propagate to an abnormal condition; hence, the probability of the fault in the process variable is considered as 0.5. Using this concept, following equations are proposed by Bao et al. (Bao et al., 2011) to calculate the probability of variable deviation to a faulty state.

$If\ v_i > \mu_i$

$$P(v_i > v_{UF}) = \varphi\left[\frac{v_i - (\mu_i + 3\sigma_i)}{\sigma_i}\right] = \int_{-\infty}^{v_i} \frac{1}{\sqrt{2\pi}\sigma_i} e^{\frac{(v_i-(\mu_i+3\sigma_i))^2}{2\sigma_i^2}}\, dv \tag{4-6}$$

$If\ v_i < \mu_i$

$$P(v_i < v_{LF}) = \varphi\left[\frac{v_i - (\mu_i - 3\sigma_i)}{\sigma_i}\right] = \int_{-\infty}^{v_i} \frac{1}{\sqrt{2\pi}\sigma_i} e^{\frac{(v_i-(\mu_i-3\sigma_i))^2}{2\sigma_i^2}}\, dv \tag{4-7}$$

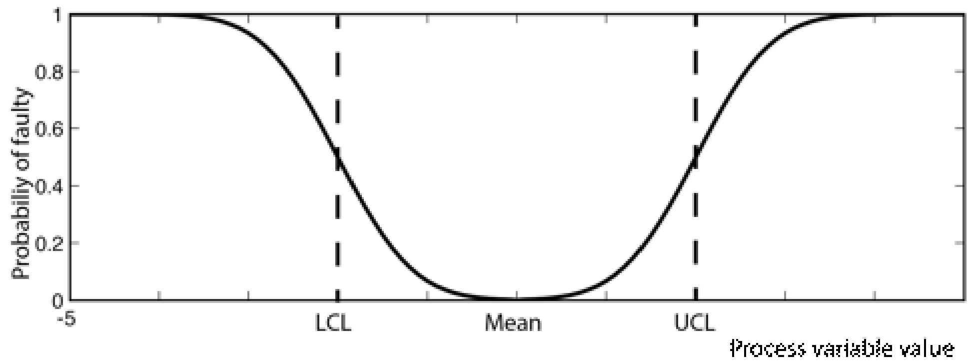


Figure 4-7: Probability of process variable being faulty

Where, v_i is the value of the V_i process variable. v_{UF} and v_{LF} are the fault thresholds. UCL and LCL are considered as $\mu_i - 3\sigma_i$ and $\mu_i + 3\sigma_i$ respectively. μ_i is the steady state value and σ_i is selected according to the faulty limit. Consequently, with the real time measured data, probability of each process variable moving to a faulty condition is calculated to update the probability of all symptom nodes.

4.3.2 Real time prediction of event occurrence

After updating the symptom nodes probabilities using the evidences at each time step, the developed Bayesian Network is used to predict the probability of the scenario nodes and event nodes using BN forward inference. In order to minimize the false alarm condition, noise filtering is carried out using moving average filter for all sensor measurements (Izadi, Shah, & Shook, 2009; Smith, 2009).

$$\widehat{x_k} = \frac{1}{n}(x_{k-n+1} + \cdots + x_{k-1} + x_k); \qquad k = n, n+1, \ldots \tag{4-8}$$

Where, n is the number of data points that are considered and x_k is the value of the variable x at kth time step. $\widehat{x_k}$ is the moving average value of the data set at kth time step.

Further, a forecasting step is introduced at the warning annunciation to increase the warning system robustness. This step is carried out when the probability of an event occurrence exceeds the threshold level. A moving window linear regression is carried out to calculate the probabilities of event occurrence for the next n time steps as follows,

$$P(Event)_{t_{(i+n)}} = b * t_{i+n} + a \tag{4-9}$$

n is the number of steps that need to forecast and it is decided considering the maximum safety time required for the operator to take corrective action or according to the acceptable false warning rate. i is the current time step. Following equations are used to calculate the a and b parameters of the equation (4-9).

$$a = \frac{\sum_{i-w}^{i} P(Event)_{t_j} - b \sum_{i-w}^{i} t_j}{w} \tag{4-10}$$

$$b = \frac{w \sum_{i-w}^{i} t_j * P(Event)_{t_j} - \left(\sum_{i-w}^{i} t_j\right) * \left(\sum_{i-w}^{i} P(Event)_j\right)}{w \sum_{i-w}^{i} {t_j}^2 - \left(\sum_{i-w}^{i} t_j\right)^2} \tag{4-11}$$

$P(Event)_{t_j}$ is the probability of event at time t_j. w is the window width of the data set. The window width can be decided according to the time steps as in Figure 4-8. Once the real time probability of the event occurrence exceeds its threshold limit, forecast probability is estimated to annunciate the event warning. If the forecasted probability of event occurrence for next n time step $\left(P(Event)_{t_{i+n}}\right)$ approach to one, warning will trigger to inform the operator about the unsafe event condition and if not an alert will be issued to inform the vulnerable plant state. In both cases, root-cause diagnosis is carried out.

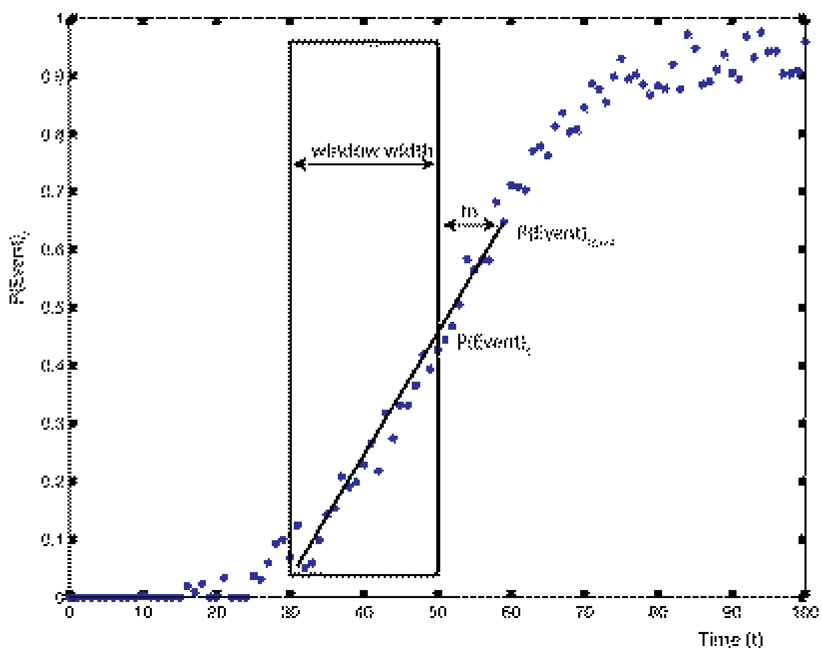


Figure 4-8: Moving window linear regression for forecasting

4.4 Step 4: Root-cause analysis

If the event probability exceeds the threshold limit, root-cause diagnosis is carried out using the BN backward inference. This is also done using the symptom node probabilities that are updated by real time measurements. The following equation is used to update the parent node (root-causes) probabilities, given the child nodes (symptoms) probabilities using Bayes Theorem. It is assumed that each root-cause is independent.

$$P(RootCause_i|Symptoms) = \frac{P(Symptoms|RootCause_i) * P(RootCause_i)}{\sum_{RC=T,F}(Symptoms|RootCause_i) * P(RootCause_i)} \tag{4-12}$$

Example 4: Consider the BN presented in Figure 4-9 that represent the root-causes and variable deviations.

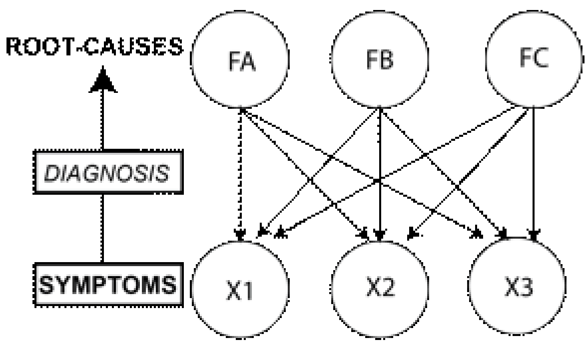


Figure 4-9: Bayesian Network example 4 network topology

It assumes that all the root-causes can influence the symptom variables and conditional probabilities are known. Then if the probabilities of each symptom are known from the process measurements, the belief about the failures can be update from BN backward inference as follows:

Updating the FA root-cause given process deviations

$$P(FA|X_1, X_2, X_3,) = \frac{P(X_1, X_2, X_3, |FA) * P(FA)}{\sum_{FA=T,F} P(X_1, X_2, X_3, |FA) * P(FA)}$$

$$= \propto P(X_1, X_2, X_2, |FA) * P(FA)\,, \qquad \propto is\ a\ constant$$

$$= \propto P(FA) \sum_{b=T,F} \sum_{c=T,F} \left(P(X_1, X_2, X_3, |FA, FB_b, FC_c) P(FB_b) P(FC_c)\right)$$

$$=\propto P(FA) \sum_{b=T,F} \sum_{c=T,F} \left[\left(\prod_{i=3} (X_i|FA, F_b, FC_c\,)\right) *\ P(FB_b) P(FC_c)\right]$$

Likewise, $P(\overline{FA}|X_1, X_2, X_3)$ is also calculated. Then, $\propto = \frac{1}{P(FA|X_1,X_2,X_3)+P(\overline{FA}|X_1,X_2,X_3)}$

Similarly, $P(FA|X_1, X_2, X_3)$ for all True and False of X_i (i=2,3) combination are calculated. Next, diagnostic inference probability of faults FB and FC are calculated for all T and F. However, at real time, probability of the variable deviation, X_i(i=1,2,3), are uncertain. To calculate the real time root-causes probabilities, the total probability is needed to calculate. Following equation present the calculation of FA,

$$P(FA = T, X_1, X_2, X_3) = \sum_{X_1=T,F} . \sum_{X_2=T,F} . \sum_{X_3=T,F} P(FA|X_1, X_2, X_3,)\ P(X_1)P(X_2)P(X_3)$$

Likewise for faults FB and FC probabilities are calculated.

The proposed methodology is summarized as follows; the abnormal events and the associated scenarios, symptoms and root-causes, which can take place during the process operation, are identified at the initial step. BN is constructed to represent the relationship between the event and associated factors corresponding to the BN topology presented in Figure 4-5. The conditional probabilities, that are needed to define the quantitative relationship of the BN, are determined by expert knowledge. In order to minimize the number of conditional probabilities, Noisy OR condition is adopted between root-causes and symptoms and between scenarios and events using equation 4-5. Real time process measurements are used to update the probabilities of the symptoms using 4-6 and 4-7. According to the symptoms probabilities, child nodes of the symptom nodes are updated by forward inference using equation 4-4 to calculate the probability of event occurrence. Consequently, parent nodes of the symptom nodes are updated by backward inference using equation 4-12 to calculate the probability of the root-causes. According to the threshold level, warning is annunciated and possible root-causes are displayed in the warning system. In the next chapter, an experimental application along with a simulation study, are used to demonstrate the methodology and evaluate its performance.

5 METHODOLOGY TESTING AND APPLICATIONS

Application of the proposed methodology is demonstrated using two case studies. The first study is based on a level control tank process. This experimental setup is used to test and validate the concept. The second case study is based on the simulation of a complex jacketed Continuous Stirred Tank Reactor (CSTR) system. This case study demonstrates the performance of proposed method.

5.1 Application 1: Level control tank experiment

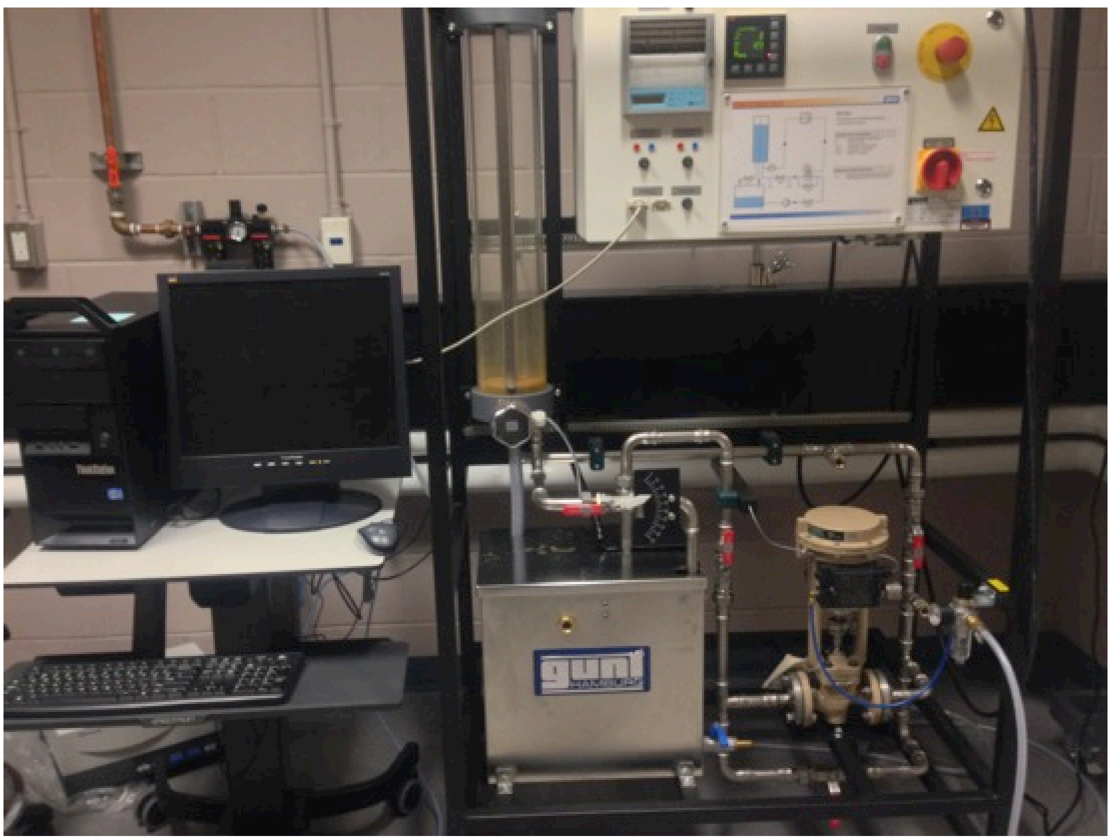


Figure 5-1: Photograph of the tank level controller set-up

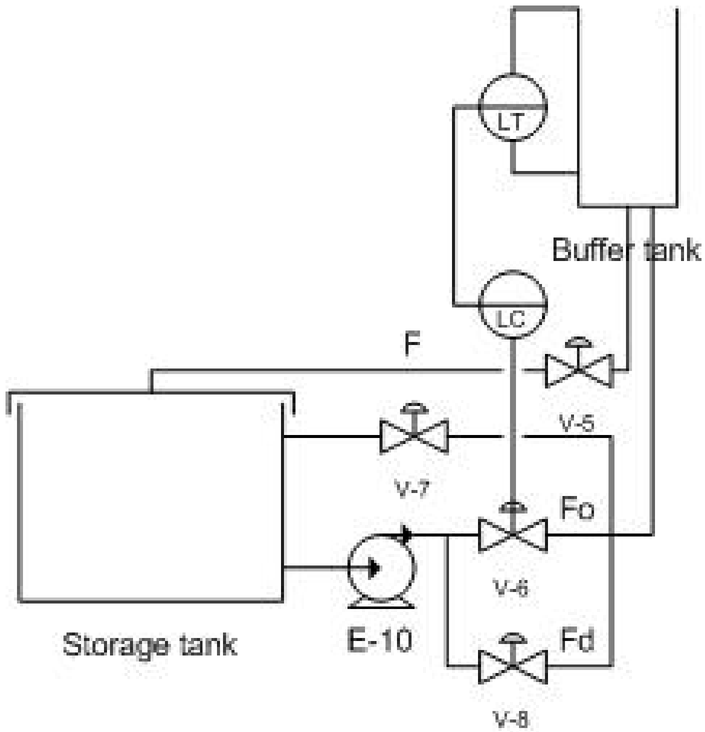


Figure 5-2: Diagram of the tank level controller set-up

The level control set-up in the Process Dynamics and Control Laboratory at the Memorial University is used to experimentally demonstrate the methodology. A photograph of the tank set-up is presented in Figure 5-1. The instrumentation diagram of the process is presented in Figure 5-2. There are two flow inlets, one is to supply water (Fo) and the other is to create disturbance (Fd). The bottom outlet (F) is used to remove water from the tank. The inlet flow rate is manipulated using a level controller (LC) (proportional-integral controller) by adjusting the control valve (V-6) to maintain the required level in the tank. Two process variables, the level of liquid in the tank and the percentage opening of the flow control valve, are measured in real time with a sampling interval of 0.5 second. Data were captured and store from a computer connected to the control system. The tank height is 0.6m. At steady state, water level is at 0.3m with the flow control valve opening at 50%.

5.1.1 Identification of event and corresponding scenarios and root-causes

Table 5-1: HAZOP study for level control tank

Variable		Deviation	Consequences	Causes
1	*Level*	*High*	*Overflow of tank*	*High inlet flow (Disturbance 1)*
				Failure of the level control system
				Outlet valve blockage (Disturbance 2)
				Secondary inlet valve failure
		Low	*Pump Damage*	*High outlet flow*
				Failure of the level control system
				Tank leakage
2	*Outlet flow*	*High*	*None*	*Failure of the level control system*
				Secondary inlet valve failure
		Low/No	*Overflow of tank*	*Outlet valve blockage (Disturbance 2)*
				Tank leakage
3	*Inlet flow*	*High*	*Overflow of tank*	*Failure of the level control system*
				Pump problem - high speed
		Reverse	*None*	*Pump mechanical problem*
		Low/No	*None*	*Pump failure*

HAZOP study is carried out to identify potential events that can occur during operation with the tank. According to the results of the HAZOP study, presented in Table 5-1, a significant abnormal event that can occur in the experiment setup is the flooding of the tank. Two process variables that are monitored in real-time in this experiment, namely, the tank level and the control valve position, are selected to the flooding event group. Hence, the valve position deviating to the minimum limit and the level deviating to the faulty limit are considered as the symptoms. Two scenarios, which influence the flooding event, are high level condition (S1) and controller valve saturation before controlling the level increment (S2). Identified root causes are presented in Table 5-2.

Table 5-2: Tank level controller root-causes for flooding event

Node	*Root-causes*
FA	*Disturbance inflow/outflow*
FB	*Flow controller failure*

5.1.2 Development of Bayesian Network

Figure 5-3 presents the Bayesian Network for the flooding event, which is constructed according to the proposed methodology. HHL and HL symptom nodes represent the level reaching to the High-High limit and the High limit, respectively. LV represents the valve reaching to its minimum opening position. Conditional probabilities are assigned to define the causal relationships among parent nodes and their child nodes based on process knowledge. Noisy OR is used to define relationships between a root-cause and a symptom and between a scenario and an event.

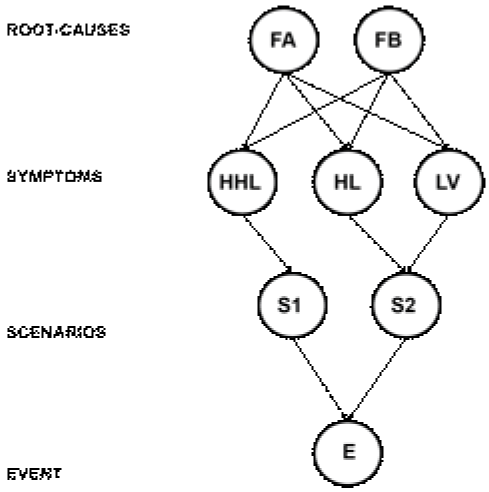


Figure 5-3: Tank level controller Bayesian Network

5.1.3 Real time probability calculation and warning annunciation

Using equations 4-6 and equation 4-7, the symptom node probabilities are updated using the process measurements. Then, according to the symptom nodes probabilities, two scenarios are updated by BN forward inference. Finally the flooding event (E) probability is updated to check the process condition. Warning is triggered when the flooding probability exceeds 0.9 and the forecast flooding probability reaches to 1. Finally, root-cause nodes probabilities are updated by BN backward inference.

5.1.4 Results

The level is the primary variable to monitor the overflow condition. The secondary variable, level controller opening percentage also gives information about the overflow condition. However, in conventional alarm system, most secondary process variables are considered as nuisance alarms when detecting abnormal events and only deviations of primary variables are considered. Thus, the variable-based High-High (HH) level alarm, which triggers to inform that the water level has exceeded the overflow limit, is defined as the conventional alarm. To demonstrate the effectiveness of the proposed method, its performance is compared with the HH level alarm. Figure 5-4 presents the performance of the conventional level alarm and the event-based early warning system for disturbance in the inlet flow. With this disturbance, the level increases steadily and the control valve moves toward its minimum position to bring the level back to the set point. Unlike the conventional method, the proposed method considers both the level and the valve position to estimate the probability of occurrence of flooding. When the valve opening reached to its minimum position, the certainty of the flooding event occurrence increases. In this situation, there is no more control action available to prevent the level increase. The proposed warning system triggered before the conventional level alarm. In this case, conventional level alarm triggers at 20.5s and the proposed warning system triggers at 18s. Flooding occurs at 24s, therefore, correction time has increase by 71.4%. . Time saving % represents the percentage increase in operator's correction time with the proposed method with reference to conventional alarm. Time saving % is calculated as follows,

$$\frac{(Time\ of\ proposed\ warning - Time\ of\ conventional\ alarm)}{(Time\ of\ event\ occurance - Time\ of\ conventianal\ alarm)} * 100$$

Numerator of the equation considered the time difference between the proposed and conventional system annunciation. Denominator of the equation considers the correction time that the operator has if the conventional alarm system is used. If the numerator increases, the saving from the proposed warring will also increase, as the denominator is a constant.

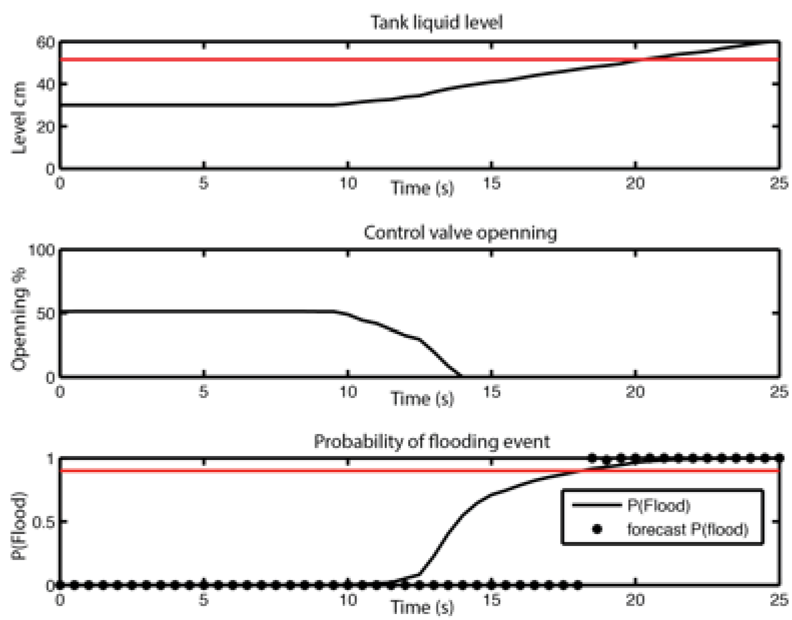


Figure 5-4: Real time alarm annunciation for tank Inlet flow disturbance

Disturbances and failures are introduced to the experimental setup to analyze the early warning capability and root-cause diagnosis ability of the proposed method. Table 5-3 presents a comparison between the conventional variable-based level alarm and the proposed warning system.

For failure of the level controller, there was no time saving as the proposed warning system acts similar to the conventional alarm. The reason for this is that both alarms consider only the level variable to trigger the alarm. With disturbances in the inflow and the outflow, both level and controller action measurements have deviated from the normal range. Hence, the proposed method was able to detect the event condition earlier than the conventional alarm as it utilized both sensor measurements to calculate the probability of the flooding occurrence.

Table 5-3: Tank level controller time-savings

Root-causes	***Time of level alarm (s)***	***Time of proposed warning (s)***	***Time of event occurrence (s)***	***Time saving %***
Level controller failure	*16.5*	*16.5*	*19*	*-*
Disturbance inflow	*20.5*	*18*	*24*	*71.4*
Disturbance outflow	*9*	*7*	*12*	*66.6*

After the warning was annunciated, proposed method was able to diagnose the root-cause correctly. Figure 5-5 presents the result from the root-cause analysis in the case of disturbance (FA).

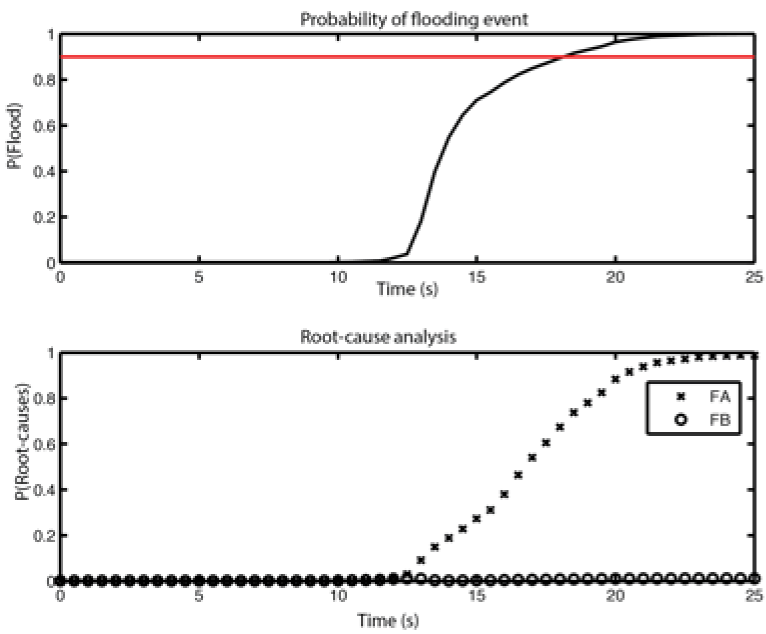


Figure 5-5: Tank level controller root-cause analysis for case of disturbance

5.2 Application 2: CSTR model simulation

In the next case study, a simulation based jacketed Continuous Stirred Tank Reactor (CSTR) is considered. The instrumentation diagram for the CSTR operation is presented in Figure 5-6.

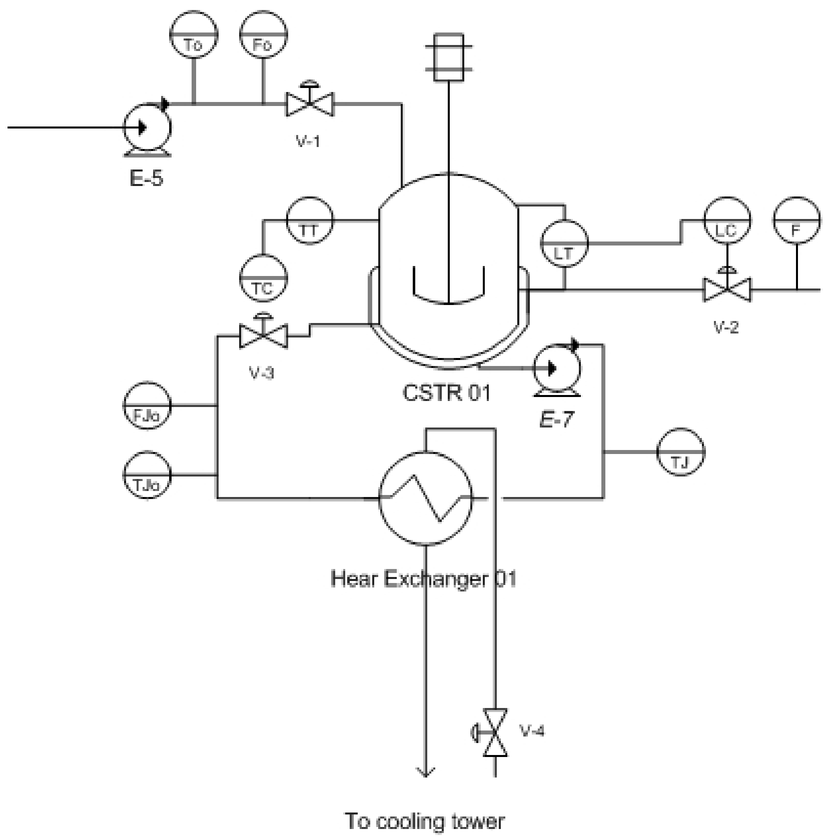


Figure 5-6: CSTR plant and controllers

An irreversible exothermic reaction A→ B is assumed to take place in the reactor with a first order kinetics. A temperature controller (TC) is used to control the reactor temperature by manipulating the coolant flow valve (V-3). Level controller (LC) maintained the level of the reactor. Reactor outlet valve (V-2) is manipulated to vary the reactor outlet flow rate (F). Heat losses are considered negligible and a perfect mixing condition is assumed. All parameters for the model are taken from literature (Luyben, 1996). Following are the ordinary differential equations and the respective parameters for the CSTR plant,

$$\frac{dV}{dt} = Fo - F$$

$$\frac{d(VCa)}{dt} = FoCao - FCa - VkCa$$

$$\frac{d(VT)}{dt} = FoTo - FT - \frac{\lambda VkCa}{\rho C_p} - \frac{UA_H}{\rho C_p}(T - TJ)$$

$$\frac{d(TJ)}{dt} = \frac{FJ(TJo - TJ)}{V_j} - \frac{UA_H}{\rho_j V_j C_j}(T - TJ)$$

$$k = \alpha e^{-E/RT}$$

Table 5-4: CSTR model parameters

Steady state values					
F	40	ft^3/h	V	48	ft^3
Cao	0.5	$lb.mol\ /ft^3$	Ca	0.245	$lb.mol\ /ft^3$
T	600	R	TJ	549.9	R
FJ	49.9	ft^3/h	To	530	R
Parameter values					
V_J	3.85	ft^3	α	$7.08*10^{10}$	h^{-1}
E	30000	$Btu/lb.mol$	R	1.99	$Btu/lb.mol.R$
U	150	$Btu/h.ft^2R$	A_H	250	Ft^2

TJo	530	R	λ	-30000	$Btu/lb.mol$
C_P	0.75	Btu/lb_mR	C_J	1	$Btu/lb.R$
ρ	50	lb_m/ft^3	ρ_J	62.3	FJ
K	4	$(Ft^3/h)/R$			

To demonstrate the methodology, the Matlab software is used to build a plant model. Simulink model is presented in the appendix. Different root-causes are simulated to generate data. Ten variables that are listed in Table 5-5 are measured in real time with a sampling time of 1 second. Concentrations of the reactants are not measured. At steady state, temperature of the reactor is at 60 ^{0}C and the level is at 3.81ft. The corresponding temperature control valve position is at 25% and level control valve position is at 50%.

Table 5-5: Measuring variables in CSTR

Variable	*Description*
T	*Temperature of the reactor*
L	*Level of the reactor*
To	*Temperature of reactant inflow*
Fo	*Flow rate of reactant inflow*
TJo	*Temperature of coolant inflow*
FJo	*Coolant flow rate*
F	*Flow rate of reactant outflow*
TJ	*Temperature of coolant outflow*
TV	*Temperature control valve position*
LV	*Level control valve position*

5.2.1 Identification of event and corresponding scenarios and root-causes

According to the HAZOP study, the runaway reaction and the overflow (flooding) condition are selected as significant abnormal events that are needed to monitor in real time. So the warning system consists of (i) runaway reaction warning and (ii) flooding warning. The mutual information between the event and the process variables are calculated to identify the key variables that have high information about each event. Details about the variable allocation methodology and the results on grouping process variable for both the runaway event and the flooding event for the CSTR simulation have been discussed in Chapter 3 and can also be found in Dalpatadu et al. (Dalpatadu, Ahmed, & Khan, 2013). Five key variables (T, FJo, TJo, TV and TJ) are selected for the runaway event group and 4 variables (L, F, Fo, LV) are selected for the flooding event group. From the grouped process variables, 5 scenarios are identified that influence the runaway event and 5 scenarios are identified for the flooding event. Scenarios corresponding to the runaway event are presented in Table 5-6.

Table 5-6: Scenarios corresponding to runaway event of CSTR

Node	*Scenario*
S1	*Reactor temperature reaching to the runaway temperature limit*
S2	*Temperature controller saturation at reactor high temperature*
S3	*Coolant flow rate decreasing at reactor high temperature*
S4	*Coolant temperature rise up at reactor high temperature*
S5	*Coolant outflow temperature rise up without reactor temperature increment*

Table 5-7 presents the identified root-causes that can influence the process variables to deviate from the respective normal operating range to cause the runaway event.

Table 5-7: Root-causes for CSTR runaway condition

Node	***Root-causes***
FA	*Temperature controller failure*
FB	*Temperature sensor failure*
FC	*Coolant pump failure*
FD	*Cooling towers failure*
FE	*Feed mixture failure*

5.2.2 Development of Bayesian Network

According to the methodology presented in section 4.2.2, the BN topology is developed for both of the events. Figure 5-7 presents the BN structure for the runaway event. Symptom nodes, HHT and HT, represent the temperature reaching to the High-High limit and the High limit, respectively. Other symptom nodes represent the corresponding measured values reaching to respective limits. Event node, E1, is the runaway occurrence node. All the conditional probabilities are assigned based on process knowledge. Noisy OR is used to define the conditional probabilities between root-cause and symptom nodes and between scenario and event nodes. The use of the Noisy OR reduces the number of conditional probabilities from 192 to 30 between root-causes and symptoms, and from 32 to 5 between scenarios and the runaway event.

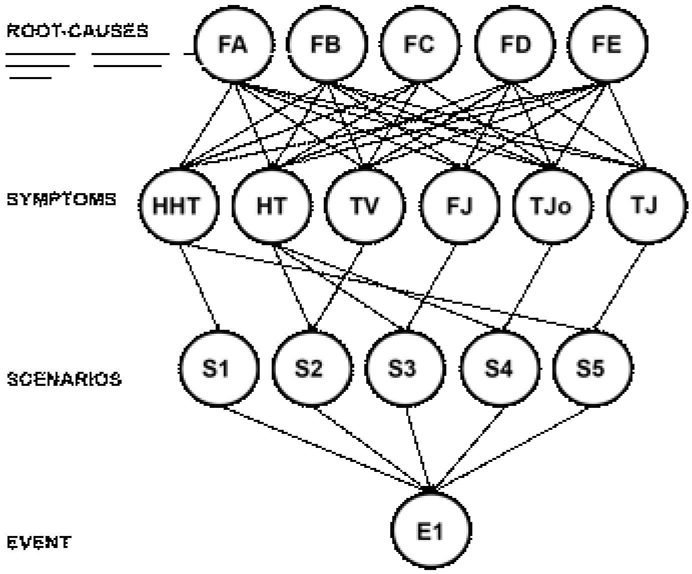


Figure 5-7: Bayesian Network topology for CSTR runaway event

5.2.3 Real time probability calculation and warning annunciation

Noisy process data from the CSTR model are send through a moving average filter to reduce the effect of measurement noise. Then, the filtered process data are used to calculate the probability of deviation of the process variables from their corresponding steady state to the faulty limit. These probabilities are used as the evidences in the Bayesian Network to update the symptom nodes. Scenario nodes and event nodes probabilities are calculated using BN forward inference. Root-cause nodes are updated using BN backward inference at each time step. If the probability of an event exceeds 0.9 and the forecasted probability reaches 1, then, the proposed warning system will annunciate a warning.

5.2.4 Results

Performances of the conventional alarm and the proposed method are compared to demonstrate the advantages of the proposed algorithm. Temperature is the primary process variable to monitor the runaway reaction. Hence, conventional variable-based alarm refers to the High-High temperature alarm that informs the operator about a high temperature state. Conventional alarm triggers when reactor temperature exceeds 74.5 ^{0}C. Different root-causes can cause the plant state to move from a steady state to an undesirable event state. Different root-causes are introduced to the simulated CSTR model to analyze the proposed warning system's capability of early warning and root-cause diagnosis. In Figure 5-8, the first plot presents the performance of variable-based temperature alarm and the second plot presents the event-based early warning for the runaway event. In this case, failure in the cooling tower is considered. With the cooling tower failure, temperature of the inlet coolant increases and it reduces the heat removal ability of the cooling jacket. Therefore, reactor temperature rises from its steady state value and subsequently leads to a runaway reaction. Conventional temperature alarm was able to detect High-High temperature at 38min. The proposed method was able to detect the runaway event at 34 min, which saved 4 min. This is a significant saving as the reactor reached to runaway condition at 44 min. Operator's correction time has increased by 66.6% with the proposed method. Also, the variable-based alarm system issued four-alarms that are T, Tjo, TJ and VT, in which T is the primary variable and others are secondary variables. But proposed warning system annunciates only one event alarm. Root-cause is diagnosed correctly by the proposed system.

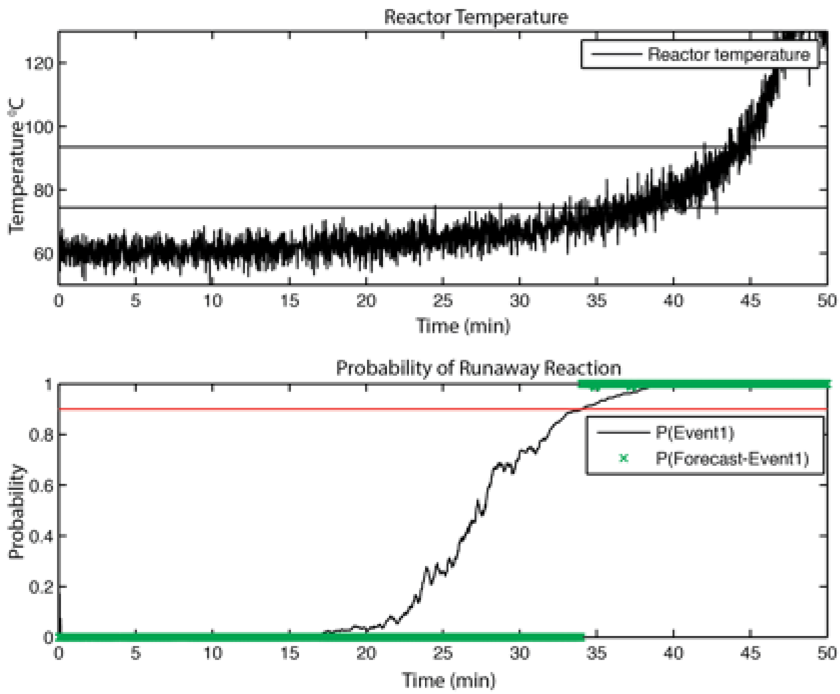


Figure 5-8: Probability of event occurrence prediction for cooling tower failure

Table 5-8 presents the time saving for five different root-causes. In the case of controller failure, there is no saving, as in this case only the temperature is used as a variable for annunciation of warning. In the case of sensor failure, temperature alarm cannot detect the runaway condition. However, the proposed method detects the situation by utilizing other variables. Considerable time saving was achieved.

Table 5-8: CSTR runaway event time-savings for different root-causes

Root-causes	*Time of temperature alarm (min)*	*Time of proposed warning (min)*	*Time of event occurrence (min)*	*Time saving %*
Temp controller failure	*10*	*10*	*15*	*0*
Temp sensor failure	*n/a*	*36*	*44*	*n/a*
Coolant pump failure	*31*	*28*	*35*	*75*
Cooling towers failure	*38*	*34*	*44*	*66.6*
Feeding operate failure	*40*	*37.5*	*44*	*62.5*

The proposed warning system is capable of diagnosing each root-cause accurately. Figure 5-9 presents the diagnosis of feeding system failure, which led to an increase in the reactant inflow concentration. The first plot presents the probability of the runaway event occurrence. The second plot presents root-cause diagnosis probability. From BN backward inference, proposed method detects that the failure has occurred in the feed mixture (FE).

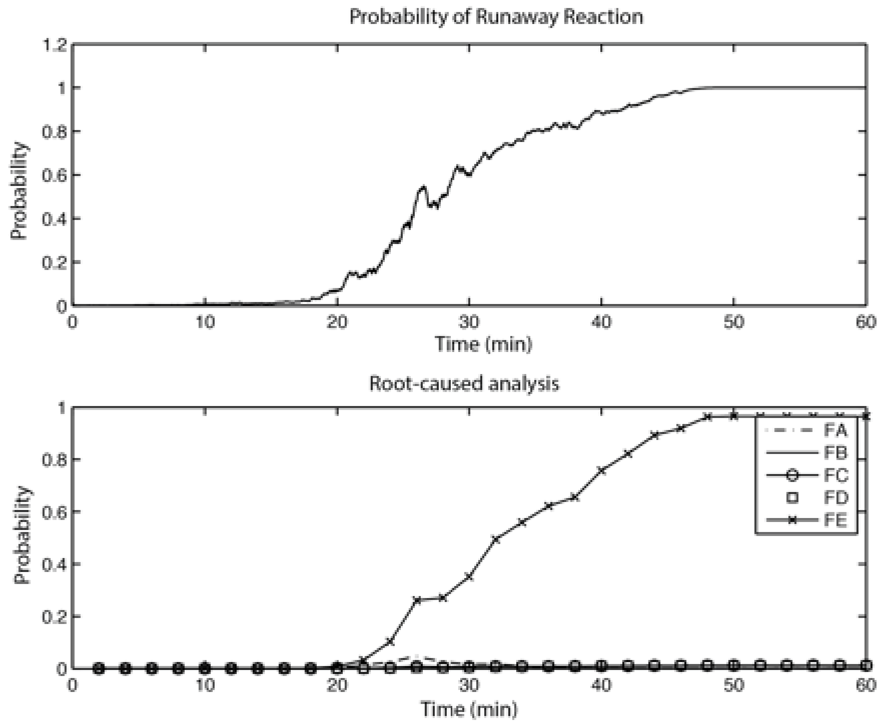


Figure 5-9: Root-cause diagnosis-feed controller failure

5.3 Discussion

The unique feature of the proposed method is to provide early warning of an abnormal event in real time. Conventional variable-based alarm system depends on the deviation of the primary variables to detect major abnormal conditions. Operators do not try to assess the relationship of the secondary variables with the primary variables. Therefore, the early warning capability of the variable-based alarm system is limited. The proposed warning system utilizes multiple process measurements to detect an abnormal event. Hence, this method showed significant early warning capability compared with the conventional variable-based alarm system in both experimental and simulation studies for different root-causes.

Since Bayesian Network is used to develop a model to define the relationship of abnormal events and the corresponding factors, development of the network topology and defining conditional probabilities are critical steps. Methodology has been proposed to define a generic structure for Bayesian Network topology. Using the methodology development of the structure for both the CSTR and the tank process for different abnormal events was very efficient. Conditional probability requirement for the tank process is very low as it is a very simple process with two process measurements. However, CSTR is a complex process and the BN contain many child nodes, which are influenced by many parent nodes. Hence the number of conditional probability requirement is very high. With the Noisy OR algorithm, requirement of conditional probabilities reduced significantly. Therefore the Noisy OR is very useful when defining conditional probabilities for the CSTR process.

Other unique feature of the proposed warning system is its ability to detect abnormal events in case of a failure of the primary variable sensor. Plant operators tend to give more attention for primary process variables than the secondary process variables. Also if the primary variables are not deviating beyond their threshold limits, operators may ignore the other variables. However, if the primary variable sensor fails, then the operator will not be able to diagnose the real situation and may perform wrong corrective action. However, the proposed method utilizes the process information of both the primary and the secondary variables. Hence, in the CSRT simulation case study, for both runway and flooding event, the proposed warning system was able to detect the abnormal event even when the primary process variable sensor failed.

Another advantage of the proposed system is its ability to perform root-cause analysis using real time sensor data. For both case studies, the root caused analysis algorithm correctly identified the main causes of the deviations. This feature is very useful if the process plant is very complex or the operators have limited knowledge about the process and the abnormal conditions.

In the conventional variable-based alarm system, several alarms are triggered to inform the same plant state. With the proposed warning system, only one warning is triggered for an event. For example, cooling tower failure in the CSTR case study triggered four process alarms in the conventional alarm system. The proposed method reduced four variable-based alarms to one event alarm without compromising the ability to perform a root-cause analysis. Thus this warning system can be used as an alarm reduction method to reduce alarm flooding. Also a minor disturbance, with less influence to lead the plant to an abnormal event, may trigger many secondary process variable alarms. However, these nuisance alarms can be reduced with the proposed methodology.

6 CONCLUSION AND FUTURE WORK

This thesis proposes a methodology to design an event-based early warning system as an alternative to the current alarm systems used in plant operations. Warnings are issued based on the estimated probability of an event. Bayesian Network is used to define the relationship between an abnormal event and corresponding symptoms, scenarios, and root-causes to calculate the probability of an event occurrence. The same network is also used to diagnose the root-causes of the event. An experimental case study using a tank level process was carried out to demonstrate the efficacy of the proposed method. Simulation study using the model of a continuous stirred tank reactor (CSTR) was carried out to demonstrate the performance of the algorithm.

- The developed event-based early warning system shows significant early warning capabilities compared with the conventional variable-based alarm system when detecting abnormal events.
- This method has the capability to identify the root-causes of an event using real time sensor measurements.
- The presented study has shown that Bayesian Network can be easily adopted to design the early warning system using the proposed generic network structure and using the Noisy OR algorithm.
- By assigning warnings to events, the methodology will result in significantly lower number of alarms compared to the variable-based alarm system during an

abnormal condition. Thus the proposed warning system can be used as an alarm reduction method to reduce alarm flooding.

6.1 Future work

The proposed method could be extended to a risk-based warning system by calculating the real time risk involved in the process operation. If there are two or more event conditions propagating in a process, an appropriate way to prioritize the event warnings will be by calculating the individual risk of each event in real time conditions. Risk-based approach is more effective in process alarm management.

In order to calculate the risk, the methodology to estimate the probability of events should be integrated with algorithms to evaluate the consequences of events.. Each abnormal event has different magnitude of severity. Hence a methodology is needed to calculate the severity in real time. Deviation of process variable related to each event can be integrated to develop such a methodology.

The main concern of this study was the false alarm reduction. In order to minimize false alarms, this research has proposed noise filtration and moving window linear regression. However it is required to investigate other methods that can be integrated with this methodology to reduce false alarms.

In the proposed methodology, expert knowledge is used to define the conditional probabilities of the Bayesian Network. However, if it is possible to acquire enough historical data, which are classified according to different abnormal events, Bayesian network learning algorithm can be used to learn the conditional probabilities of the network.

Finally, it is important to test this methodology with a real complex process system to identify further limitation.

7 REFERENCES

Abnormal Situation Managment Consortium. (2008). Retrieved from http://www.asmconsortium.net

Ahmed, S., Gabbar, H. A., Chang, Y., & Khan, F. I. (2011). Risk based alarm design: A systems approach. In *Advanced Control of Industrial Processes* (pp. 42–47). Thousand Islands Lake, Hangzhou, China.

Alaeddini, A., & Dogan, I. (2011). Using Bayesian networks for root cause analysis in statistical process control. *Expert Systems with Applications*, *38*(9), 11230–11243.

Alhoniemi, E., Honkela, A., Lagus, K., Sepp, J., Wagner, P., & Valpola, H. (2007). Compact Modeling of Data Using Independent Variable Group Analysis. In *IEEE Transactions on Neural Networks* (pp. 1–14).

Argiolas, C., Carbonari, A., Melis, F., & Quaquero, E. (2012). A Bayesian model for real-time safety management in construction sites. *Gerontechnology*, *11*(2).

Bao, H., Khan, F., Iqbal, T., & Chang, Y. (2011). Risk-based fault diagnosis and safety management for process systems. *Process Safety Progress*, *30*(1), 6–17.

Blaser, L., Ohrnberger, M., Riggelsen, C., Babeyko, A., & Scherbaum, F. (2011). Bayesian networks for tsunami early warning. *Geophysical Journal International*, *185*(3), 1431–1443.

Bobbio, A., Portinale, L., Minichino, M., & Ciancamerla, E. (2001). Improving the analysis of dependable systems by mapping fault trees into Bayesian networks. *Reliability Engineering & System Safety*, *71*(3), 249–260.

Bransby, M. L., & Jenkinson, J. (1998). *The management of alarm systems*. Health & Safety Executive.

Chang, C., & Yu, C. (1990). On-Line Fault Diagnosis Using the Signed Directed Graph. *Industrial & Engineering Chemistry Research*, 1290–1299.

Chang, Y., Khan, F., & Ahmed, S. (2011). A risk-based approach to design warning system for processing facilities. *Process Safety and Environmental Protection*, *89*(5), 310–316.

Cochran, E., & Bullemer, P. (1996). Abnormal Situation management: Not by New Technology Alone. In *Honeywell Technology Center* (pp. 1–6).

Cover, T., & Thomas, J. (1991). *Elements of Information Theory*. Wiley-Interscience.

Crowl, D., & Louvar, J. (2001). *Chemical process safety: fundamentals with applications*. Prentice Hall.

Dalpatadu, P., Ahmed, S., & Khan, F. I. (2013). Alarm allocation for event-based process alarm systems. In *IFAC International Symposium on Dynamics and Control of Process Systems* (pp. 815–820). Mumbai, India.

Dey, S., & Stori, J. a. (2005). A Bayesian network approach to root cause diagnosis of process variations. *International Journal of Machine Tools and Manufacture*, *45*(1), 75–91.

EEMUA. (2007). *Alarm Systems, a guide to design, management and procurement, Publication No 191* (2nd ed.). The Engineering Equipment and Materials Users Association publication.

Flairs, C., Barr, V., Markov, Z., Zagorecki, A., & Druzdzel, M. (2004). An Empirical Study of Probability Elicitation under Noisy-OR Assumption. In *Artificial Intelligence Research Society Conference* (pp. 880–885). Menlo Park, CA.

Geng, Z., Zhu, Q., & Gu, X. (2005). A fuzzy clustering-ranking algorithm and its application for alarm operating optimization in chemical processing. *Process Safety Progress*, *24*(1), 66–75.

Heckerman, D. E., Horvitz, E. J., & Nathwani, B. N. (1992). Toward normative expert systems: Part I. The Pathfinder project. *Methods of Information in Medicine*, *31*(2), 90–105.

Hossain, M., & Muromachi, Y. (2012). A Bayesian network based framework for real-time crash prediction on the basic freeway segments of urban expressways. *Accident; Analysis and Prevention*, *45*, 373–381.

HSE. (1994). Better alarm handling. *Health and Safety Executive*, (6), 1–4.

ISA. (2009). Management of Alarm Systems for the Process Industries. *International Society of Automation*.

Izadi, I., Shah, S. L., Shook, D., & Chen, T. (2009). An introduction to alarm analysis and design. In *Fault Detection, Supervision and Safety of Technical Processes* (pp. 645–650). Barcelona, Spain.

Izadi, I., Shah, S. L., & Shook, D. S. (2009). A framework for optimal design of alarm systems. In *Fault Detection, Supervision and Safety of Technical Processes* (pp. 651–656). Barcelona, Spain.

Jing, C., & Jingqi, F. (2012). Fire Alarm System Based on Multi-Sensor Bayes Network. *Procedia Engineering*, *29*, 2551–2555.

Juricek, B. C., Seborg, D. E., & Larimore, W. E. (2001). Predictive monitoring for abnormal situation management. *Journal of Process Control*, *11*(2), 111–128.

Khakzad, N., Khan, F., & Amyotte, P. (2011). Safety analysis in process facilities: Comparison of fault tree and Bayesian network approaches. *Reliability Engineering and System Safety*, *96*(8), 925–932.

Khakzad, N., Khan, F., & Amyotte, P. (2012). Dynamic safety analysis of process systems by mapping bow-tie into Bayesian network. *Process Safety and Environmental Protection*, 46–53.

Khakzad, N., Khan, F., Amyotte, P., & Cozzani, V. (2012). Domino Effect Analysis Using Bayesian Networks. *Risk Analysis*, *33*(2), 292–306.

Kondaveeti, S. R., Izadi, I., Shah, S. L., & Black, T. (2009). Graphical representation of industrial alarm data. In *Analysis, design, and evaluation of human-machine system* (Vol. 1, pp. 181–186).

Korb, K., & Nicholson, A. (2003). *Bayesian artificial intelligence* (2nd ed.). CRC Press.

Luyben, I. (1996). *Process Modelling, Simulation and Control for Chemical Engineers*. McGraw-Hill Higher Education.

Martinez, W., Martinez, A., & Solka, J. (2004). *Exploratory Data Analysis with MATLAB*. CRC Press.

Noda, M., Higuchi, F., Takai, T., & Nishitani, H. (2011). Event correlation analysis for alarm system rationalization. *Asia-Pacific Journal of Chemical Engineering*, 497–502.

Oniésko, A., Lucas, P., & Druzdzel, M. (2001). Comparison of rule-based and Bayesian network approaches in medical diagnostic systems. *Artificial Intelligence in Medicine*, 283–292.

Onisko, A., Druzdzel, M., & Wasyluk, H. (2000). Learning Bayesian Network Parameters from Small Data Sets : Application of Noisy-OR Gates. *International Journal of Approximate Reasoning*, 165–182.

Orantes, A., Kempowsky, T., Le Lann, M., & Aguilar-Martin, J. (2008). A new support methodology for the placement of sensors used for fault detection and diagnosis. *Chemical Engineering and Processing: Process Intensification*, *47*(3), 330–348.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann.

Pérez, A., Larrañaga, P., & Inza, I. (2006). Supervised classification with conditional Gaussian networks: Increasing the structure complexity from naive Bayes. *International Journal of Approximate Reasoning*, 1–25.

Pradhan, S., Singh, R., Kachru, K., & Narasimhamurthy, S. (2007). A Bayesian Network Based Approach for Root-Cause-Analysis in Manufacturing Process. *2007 International Conference on Computational Intelligence and Security (CIS 2007)*, 10–14.

Qunxiong, Z., & Zhiqiang, G. (2005). A New Fuzzy Clustring-Ranking Algorithm and Its Application in Process alarm Management. *Chinese Journal of Chemical Engineering*, 477–483.

Ram Maurya, M., Rengaswamy, R., & Venkatasubramanian, V. (2004). Application of signed digraphs-based analysis for fault diagnosis of chemical process flowsheets. *Engineering Applications of Artificial Intelligence*, *17*(5), 501–518.

Rothenberg, D. H. (2009). *Alarm Management for Process Control: A Best-practice Guide for Design, Implementation, and Use of Industrial Alarm Systems*. Momentum Press.

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Mobile Computing and Communications Review*, *5*(I), 3–55.

Smith, C. (2009). *Basic Process Measurements*. John Wiley & Sons.

Swift, S., Tucker, A., Martin, N., & Liu, X. (2001). Grouping multivariate time series variables : applications to chemical process and visual field data. *Knowledge Based System*, *14*, 147–154.

USCSB. BP America refinery explosion Texas City, TX, March 23, 2005. Final report finding. , U.S. Chemical Safety And Hazard Investigation Board (2007). Retrieved from http://www.csb.gov

Venkatasubramanian, V., Rengaswamy, R., & Kavuri, S. N. (2003a). A review of process fault detection and diagnosis Part II : Qualitative models and search strategies. *Computers and Chemical Engineering*, *27*, 313–326.

Venkatasubramanian, V., Rengaswamy, R., & Kavuri, S. N. (2003b). A review of process fault detection and diagnosis Part III : Process history based methods. *Computers and Chemical Engineering*, *27*, 327–346.

Venkatasubramanian, V., Rengaswamy, R., & Yin, K. (2003). A review of process fault detection and diagnosis Part I : Quantitative model-based methods. *Computers and Chemical Engineering*, *27*, 293–311.

Villez, K., Srinivasan, B., Rengaswamy, R., Narasimhan, S., & Venkatasubramanian, V. (2011). Kalman-based strategies for Fault Detection and Identification (FDI): Extensions and critical evaluation for a buffer tank system. *Computers & Chemical Engineering*, 806–816.

Widarsson, B., & Dotzauer, E. (2008). Bayesian network-based early-warning for leakage in recovery boilers. *Applied Thermal Engineering*, *28*(7), 754–760.

X.Z. Wang. (1999). *Data mining and knowledge discovery for process monitoring and control*. *Advances in Industrial Control, Springer*. Advances in Industrial Control, Springer.

Yang, F., Sirish, L. S., & Xiao, D. (2010). Signed Directed Graph modeling of industrial processes and their validation by data-based methods. In *Control and Fault-Tolerant Systems* (pp. 387–392).

Yang, Z., Wang, J., & Chen, T. (2012). On correlation analysis of bivariate alarm signals. In *International Conference on Information and Automation* (pp. 530–535).

Yu, L., & Liu, H. (2003). Feature Selection for High-Dimensional Data : A Fast Correlation-Based Filter Solution. In *International Conference on Machine Leaning* (pp. 856–863).

Zadakbar, O., Imtiaz, S., & Khan, F. (2012). Dynamic Risk Assessment and Fault Detection Using Principal Component Analysis. *Industrial & Engineering Chemistry Research*, 809–816.

Zadakbar, O., Imtiaz, S., & Khan, F. (2013). Dynamic risk assessment and fault detection using a multivariate technique. *Process Safety Progress*, *32*(4), 365–375.

Zazzaro, G., Pisano, F., & Romano, G. (2012). Bayesian Networks for Earthquake Magnitude Classification in a Early Warning System. *World Academy of Science, Engineering and Technology*, 85–95.

8 APPENDIX

APPENDIX A: Variable allocation for event-base group

1. Mathlab/Simulink model of the CSTR plant

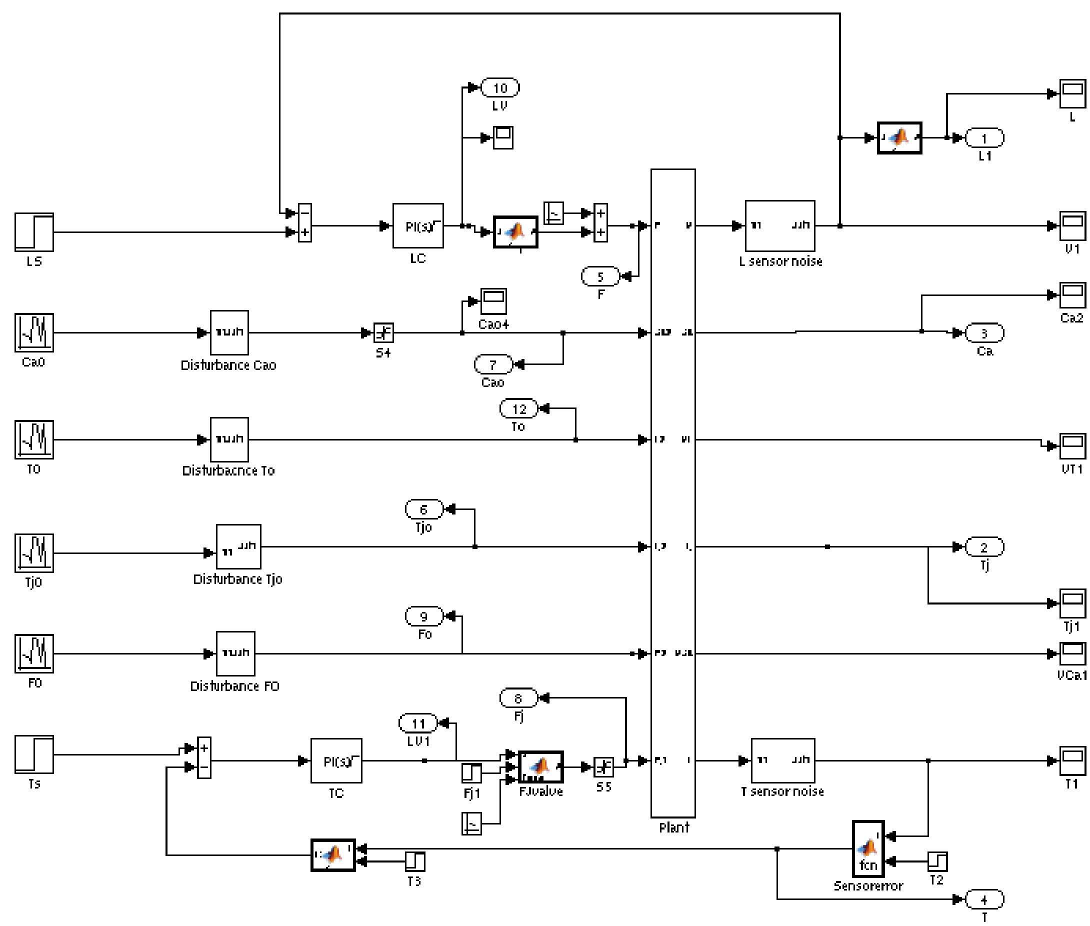


Figure 10: CSTR Matllab/Simulink model

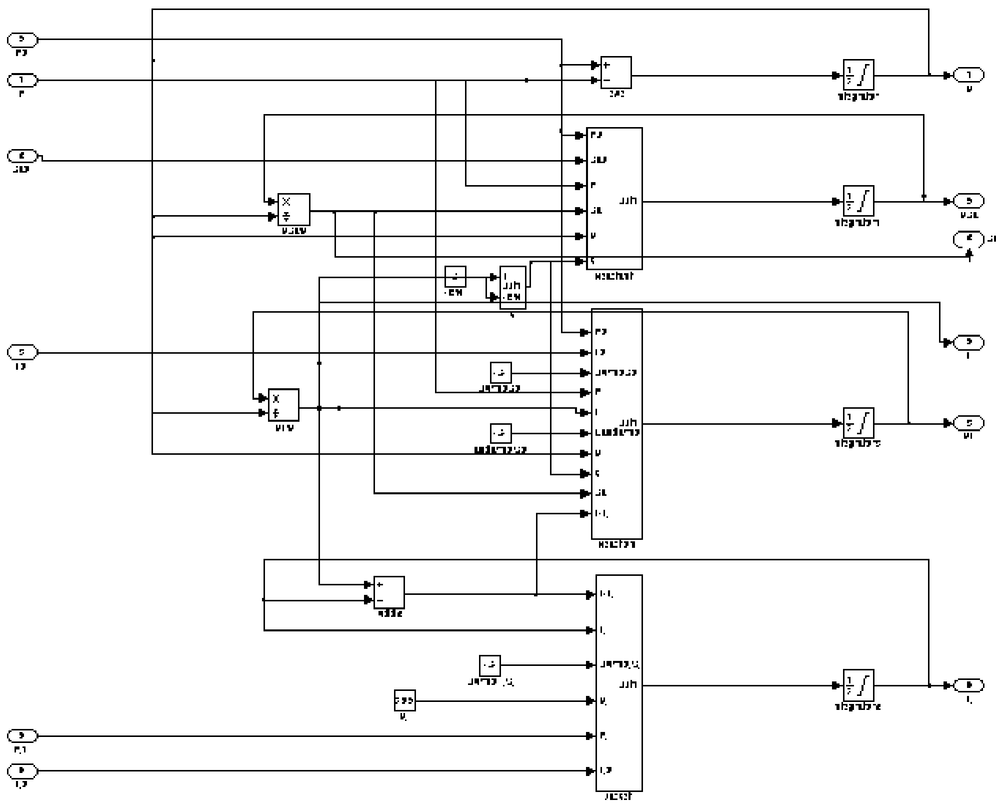


Figure 11: CSTR simulation model: subsystem - reactor

2. Mutual information gain results of each abnormla event

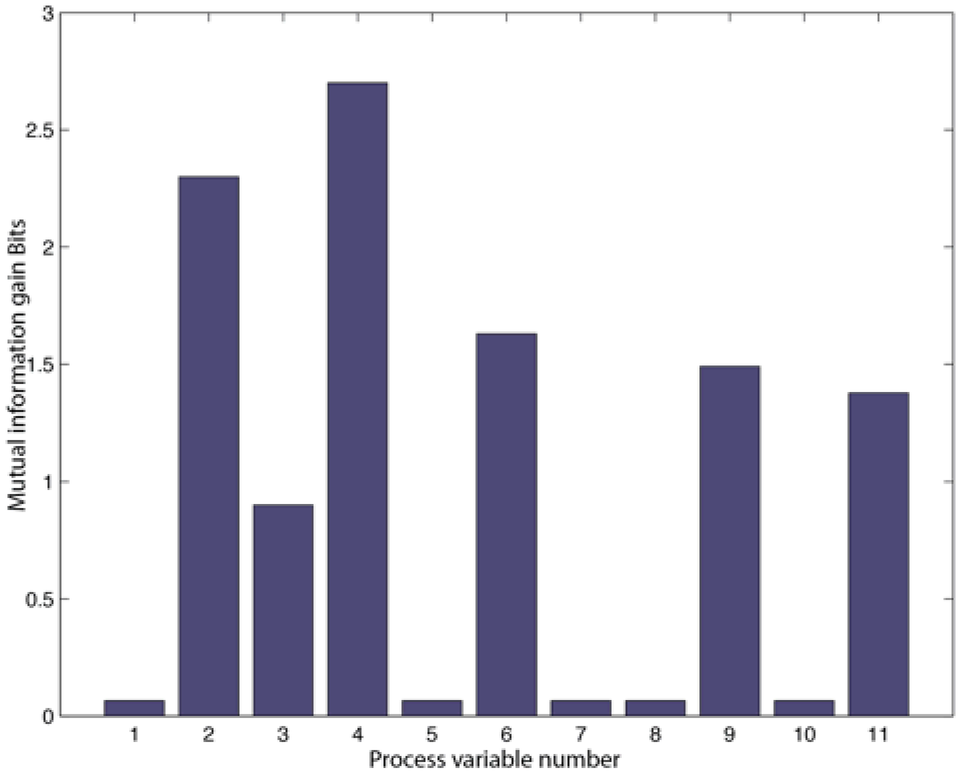


Figure 8-12: Mutual information gain between runaway event and each process variable

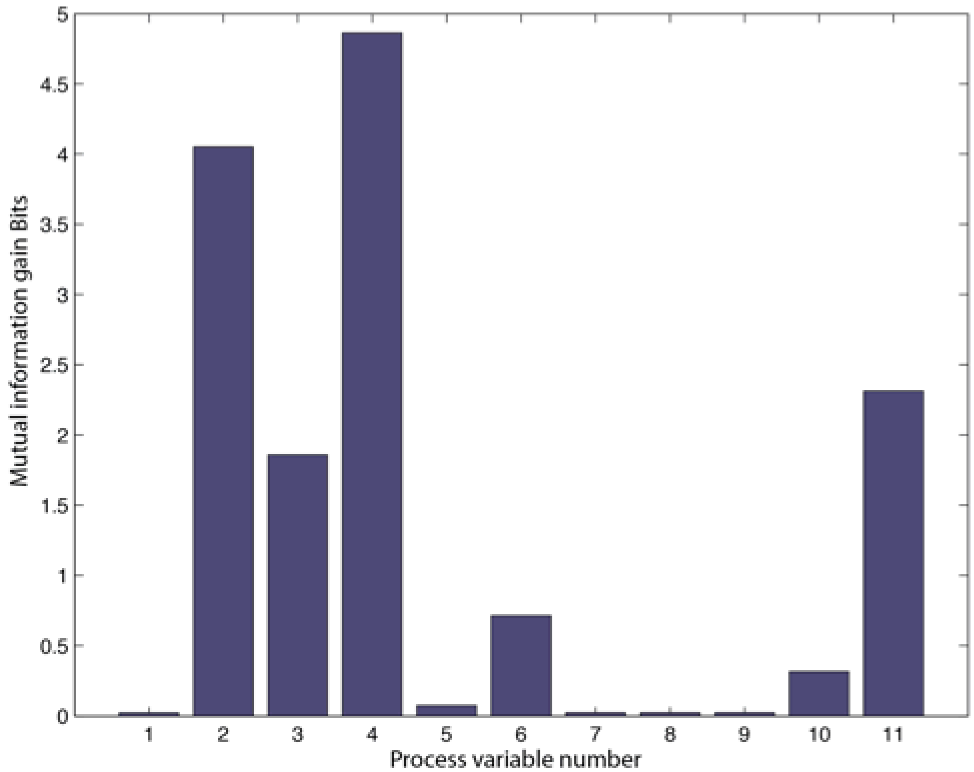


Figure 8-13: Mutual information gain between low quality operation event and each process variable

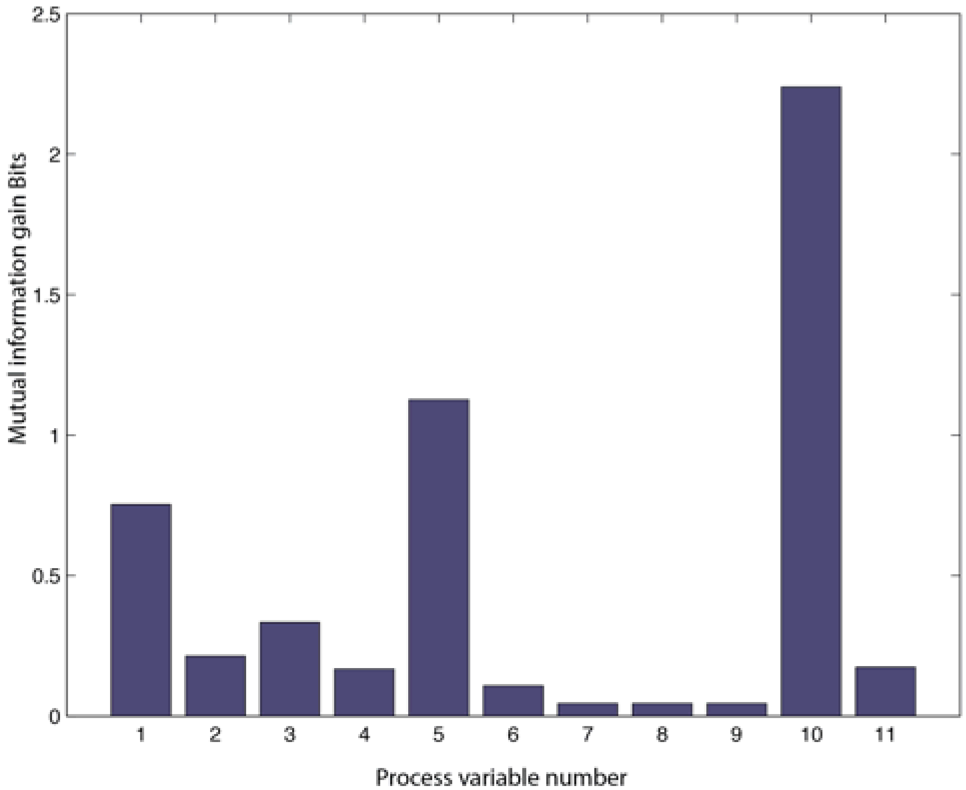


Figure 8-14: Mutual information gain between flooding event and each process variable

3. Cross correlation analysis results for each abnormal event

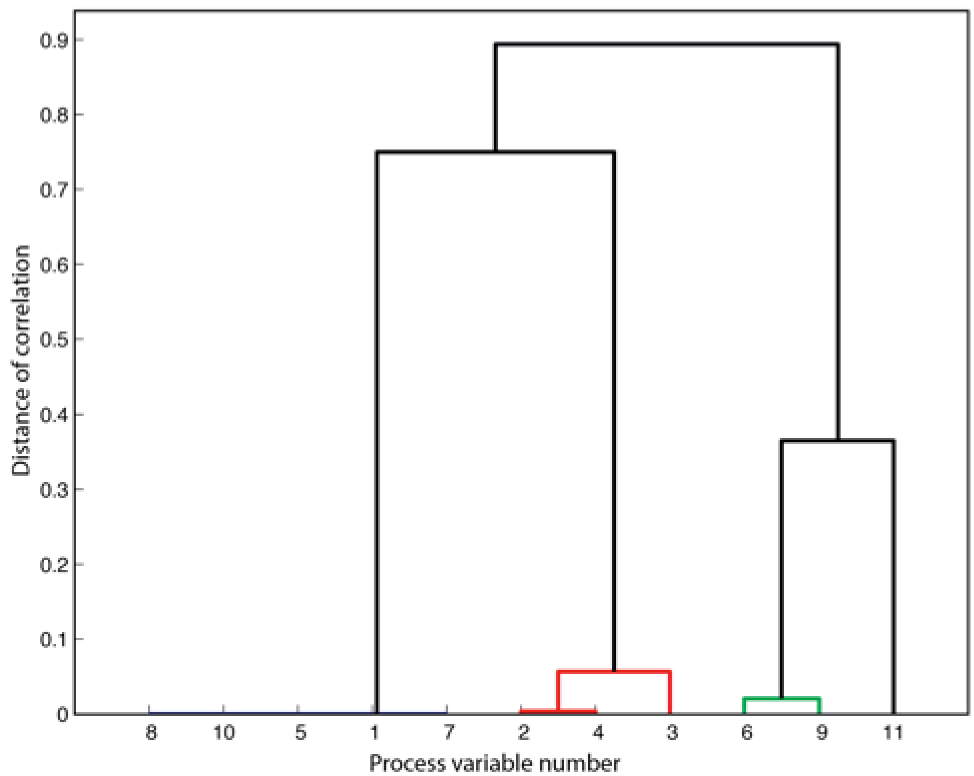


Figure 8-15: Correlation distance between variable at runaway event

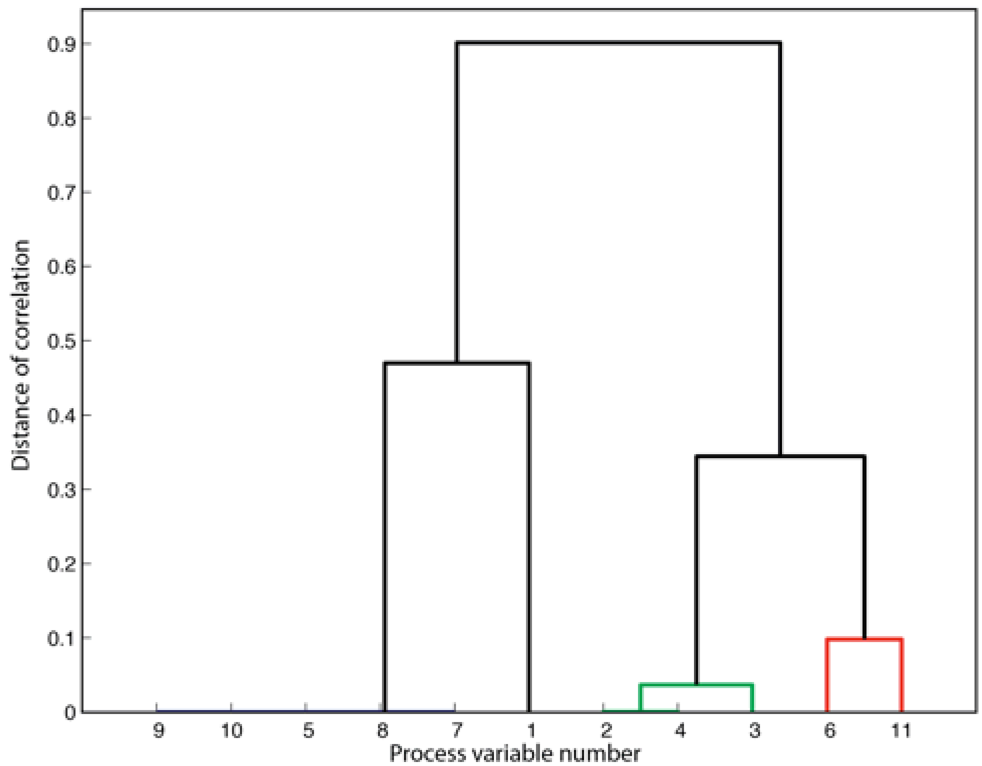


Figure 8-16: Correlation distance between variable at low quality operation event

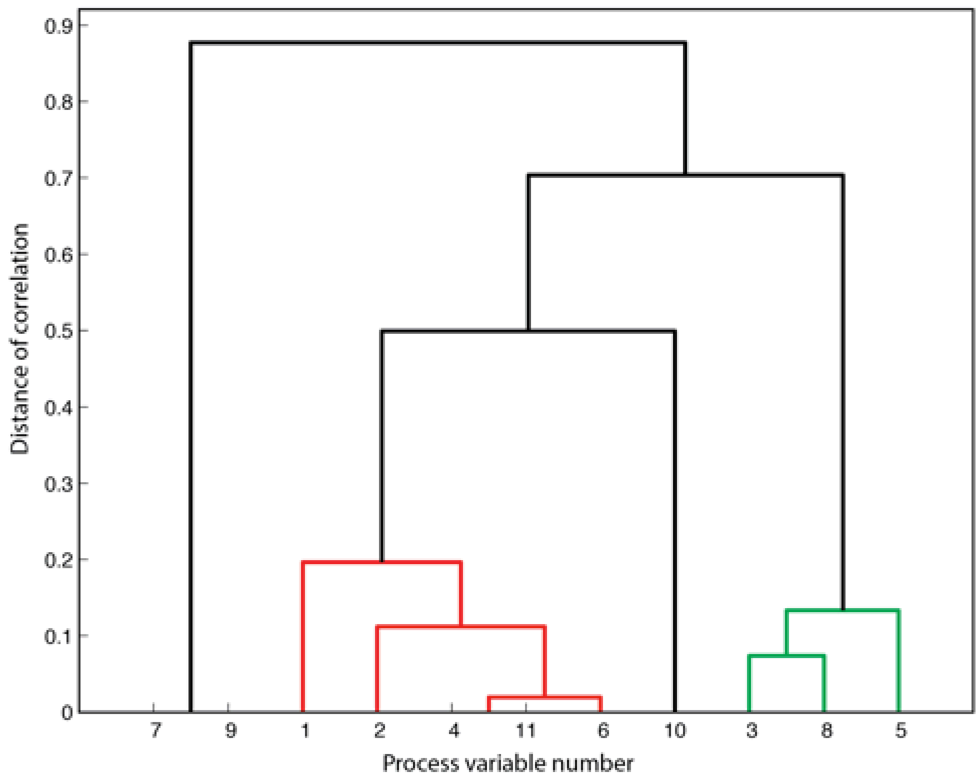
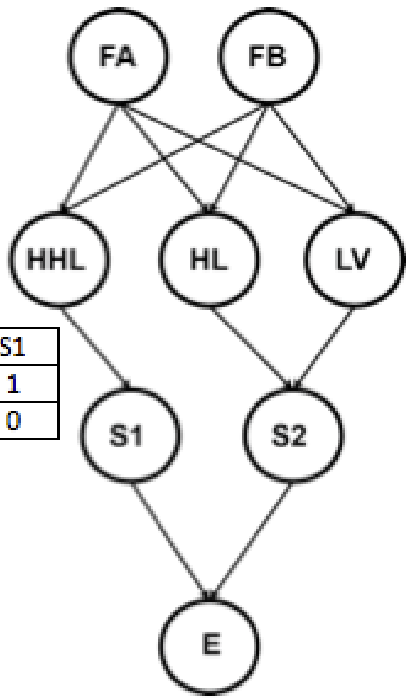


Figure 8-17: Correlation distance between variable at low flooding event

APPENDIX B: Early warning system designing

4. Tank level control process: Bayesian Network conditional probabiilty tables

FA	FB	HHL	HL	LV
1	1	1	1	0.01
1	0	0.99	1	0.9
0	1	0.99	1	0.01
0	0	0	0	0



HHL	S1
1	1
0	0

HL	lv	S2
1	1	1
1	0	0
0	1	0
0	0	0

S1	S2	E
1	1	1
1	0	1
0	1	0.7
0	0	0

Figure 8-18: Conditional probability tables of tank level control operation

5. Tank level control process: Alarm annunciation result for different root causes

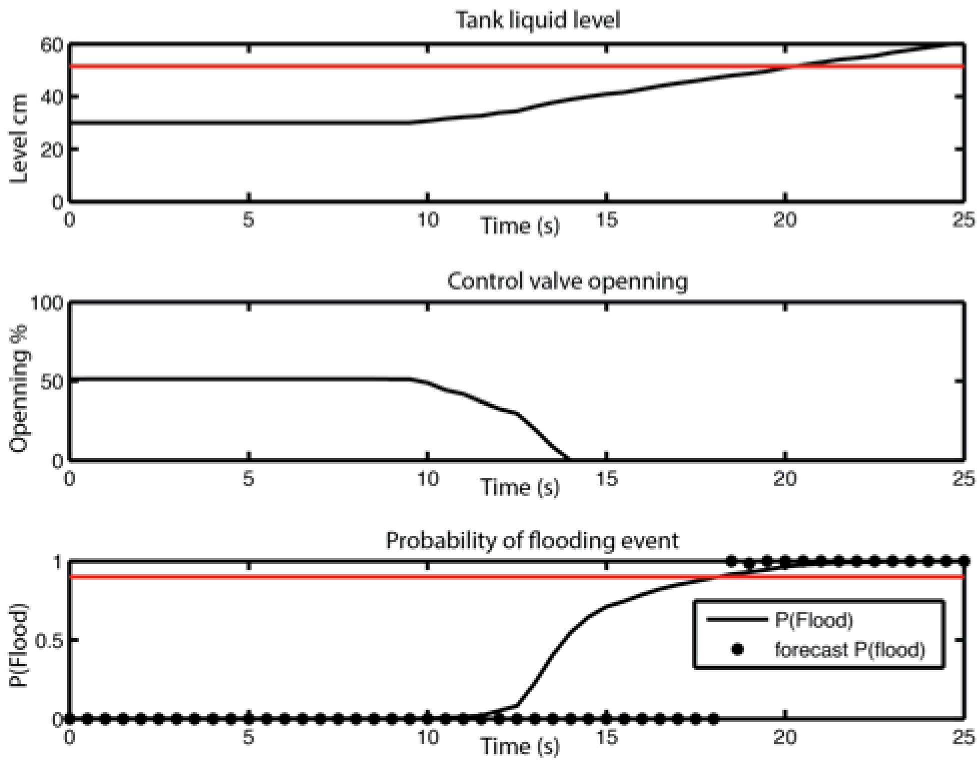


Figure 8-19: Warning annunciation of the inlet flow valve failure (High inlet flow rate)

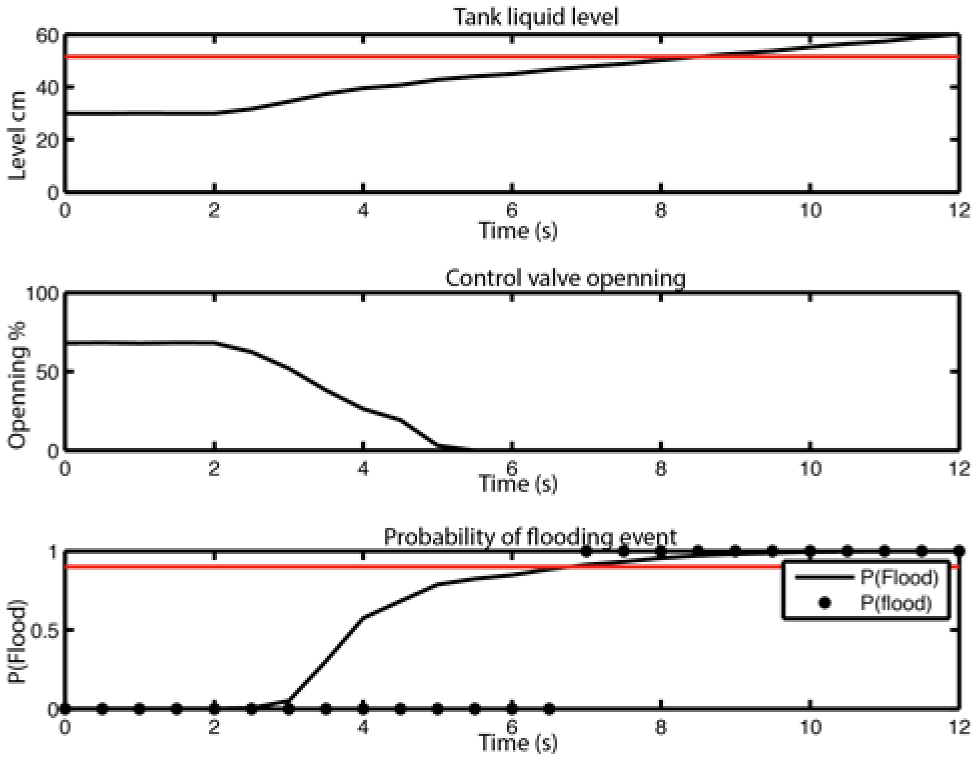


Figure 8-20: Warning annunciation of outlet flow valve failure (Low outlet flow rate)

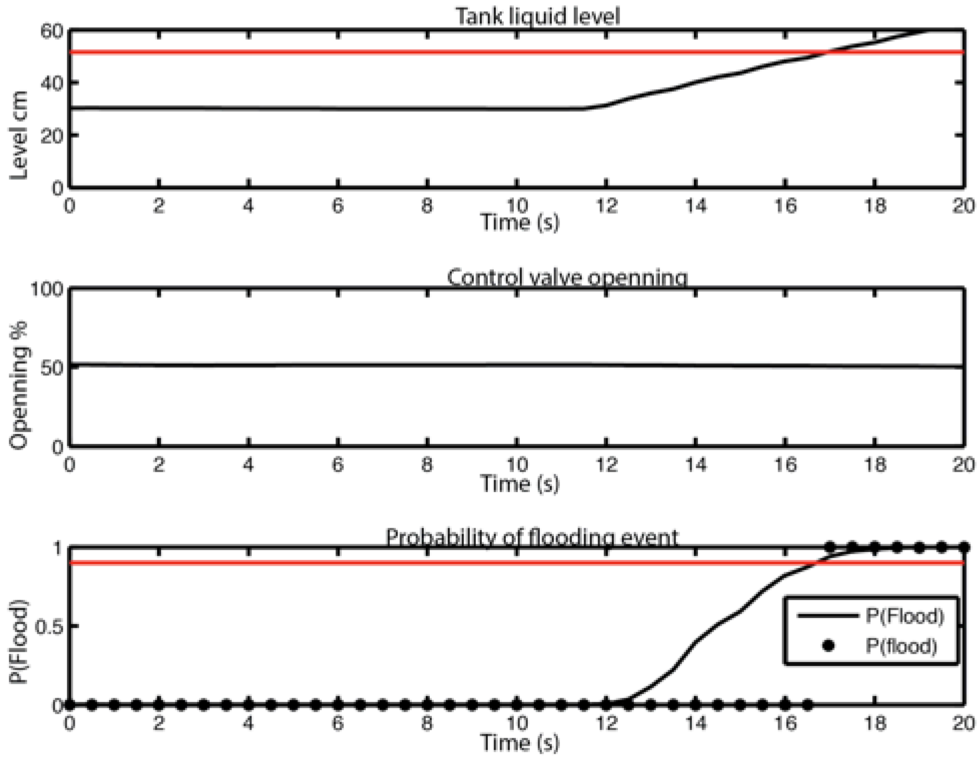


Figure 8-21: Warning annunciation of the level controller failure

6. Tank level control process: Root cause analysis

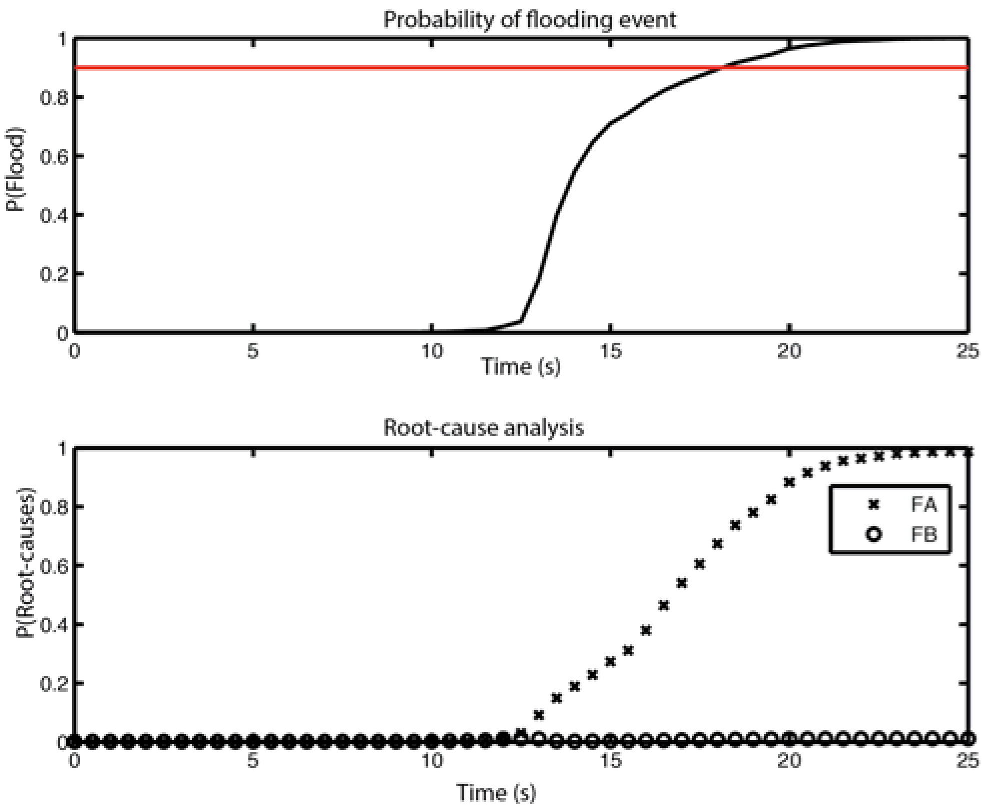


Figure 8-22: Root-cause analysis for disturbance

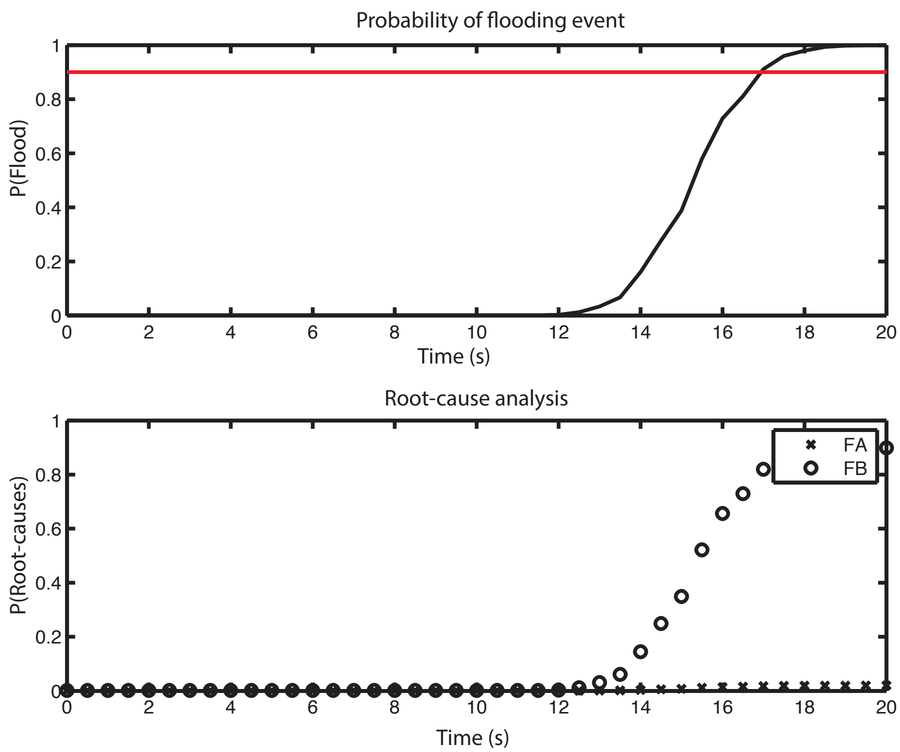


Figure 8-23: Root-cause analysis for level controller failure

7. CSTR HAZOP study table

Table 8-1: HAZOP study for CSTR

Variable		Deviation	Consequences	Causes
1	*FJ*	*No*	*Runaway reaction*	*Temperature controller failure*
			Off quality product	*Temperature sensor failure*
				Plug pipe line
		Reverse	*Runaway reaction*	*Backflow due to high back pressure*
			Off quality product	*Coolant pump failure*
		More	*Off quality product*	*Temperature controller failure*
				Coolant pump failure
		As well as	*None*	*Contamination of water supply*
		Low	*Runaway reaction*	*Temperature controller failure*
				Partially plug line
				Temperature sensor failure
				Cooling pump failure
2	*TJo*	*High*	*Runway reaction*	*Cooling tower failure*
3	*Cao*	*High*	*Runaway reaction*	*Feeder mixture system failure*
			Off quality product	*Feed flow controller failure*
		Low	*Off quality product*	*Feeder mixture system failure*
4	*To*	*High*	*High reactor temperature*	*Feed heater failure*
5	*T*	*High*	*Runaway reaction*	*Temperature controller failure*

		Low	*Off quality product*	*Temperature sensor failure*
6	*L*	*High*	*Overflow of reactor*	*Feeding pump failure*
				Failure of the level control system
				Outlet valve blockage
				Level sensor failure
		Low	*None identified*	*High outlet flow*
				Failure of the level control system
				Tank leakage
7	*F*	*High*	*None identified*	*Level controller failure*
				Level sensor failure
		Low/No	*Overflow of reactor*	*Outlet valve blockage*
				Level controller failure
				Level sensor failure
				Tank leakage
8	*Fo*	*High*	*Overflow of reactor*	*Feed flow controller failure*
		Reverse	*None identified*	*Pump mechanical problem*
		Low/No	*None identified*	*Pump failure*

8. Complete event-based warning system for the CSTR operation

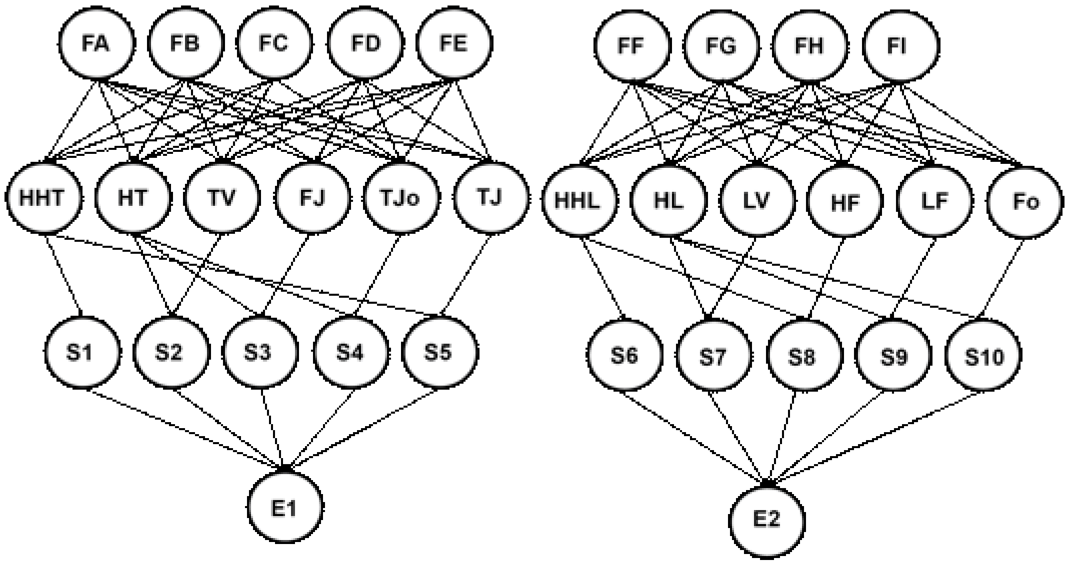


Figure 8-24: Runaway and flooding event network for CSTR operation

9. CSTR runaway event conditional tables

Table 8-2 : Runaway Event - Individual inference of root-causes to symptoms

	FA	*FB*	*FC*	*FD*	*FE*
HHT	*0.95*	*0.005*	*0.95*	*0.95*	*0.95*
HT	*0.99*	*0.01*	*0.99*	*0.99*	*0.99*
TV	*0.01*	*0.01*	*0.99*	*0.99*	*0.99*
FJ	*0.01*	*0.01*	*0.99*	*0.01*	*0.01*
TJo	*0.01*	*0.01*	*0.01*	*0.99*	*0.01*
TJ	*0.99*	*0.99*	*0.99*	*0.99*	*0.01*

Table 8-3: Runaway event - CPT between symptoms and scenarios

HHT	*S1*
1	*1*
0	*0*

HT	*TV*	*S2*
1	*1*	*1*
1	*0*	*0*
0	*1*	*0*
0	*0*	*0*

HT	*FJ*	*S3*
1	*1*	*1*
1	*0*	*0*
0	*1*	*0*
0	*0*	*0*

HT	*TJo*	*S4*
1	*1*	*1*
1	*0*	*0*
0	*1*	*0*
0	*0*	*0*

HHT	*TJ*	*S5*
1	*1*	*0*
1	*0*	*0*
0	*1*	*1*
0	*0*	*0*

Table 8-4: Runaway Event- Individual inference of scenario nodes and event node

	S1	*S2*	*S3*	*S4*	*S5*
E	*1*	*0.7*	*0.5*	*0.5*	*1*

Table 8-5: Conditional probability table of scenarios and event for runaway event

P(E=1 | S5, S4, S3, S2,S1) = P(g,h,I,j,k,1)

P(:,:,1,1,1,1) = *0.9993 0.9987* *0.9992 0.9985*	*P(:,:,1,2,2,1) =* *0.7750 0.5500* *0.7500 0.5000*
P(:,:,2,1,1,1) = *0.9987 0.9973* *0.9985 0.9970*	*P(:,:,2,2,2,1) =* *0.5500 0.1000* *0.5000 0*
P(:,:,1,2,1,1) = *0.9978 0.9955* *0.9975 0.9950*	*P(:,:,1,1,1,2) =* *0.0007 0.0014* *0.0008 0.0015*
P(:,:,2,2,1,1) = *0.9955 0.9910* *0.9950 0.9900*	*P(:,:,2,1,1,2) =* *0.0014 0.0027* *0.0015 0.0030*
P(:,:,1,1,2,1) = *0.9325 0.8650* *0.9250 0.8500*	*P(:,:,1,2,1,2) =* *0.0023 0.0045* *0.0025 0.0050*
P(:,:,2,1,2,1) = *0.8650 0.7300* *0.8500 0.7000*	*P(:,:,2,2,1,2) =* *0.0045 0.0090* *0.0050 0.0100*
P(:,:,1,1,2,2) = *0.0675 0.1350* *0.0750 0.1500*	*P(:,:,1,2,2,2) =* *0.2250 0.4500* *0.2500 0.5000*
P(:,:,2,1,2,2) = *0.1350 0.2700* *0.1500 0.3000*	*P(:,:,2,2,2,2) =* *0.4500 0.9000* *0.5000 1.0000*

10. CSTR runaway event probability calculation and warning annunciation

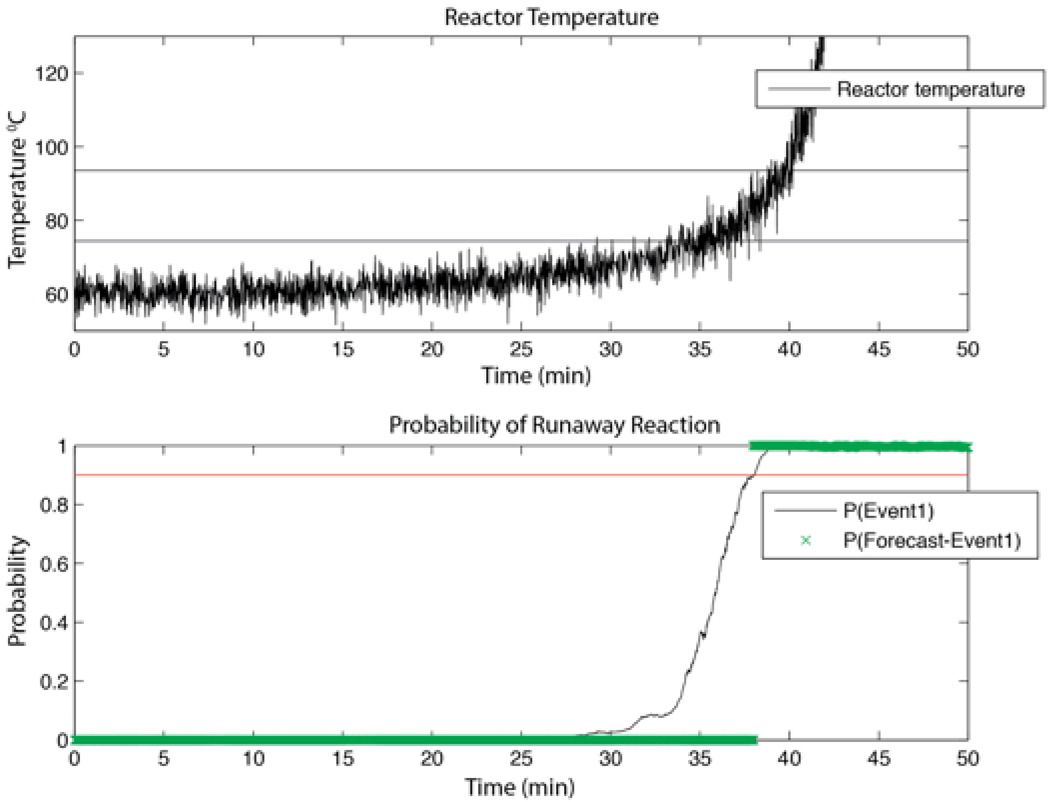


Figure 8-25: Warning annunciation for temperature sensor failure

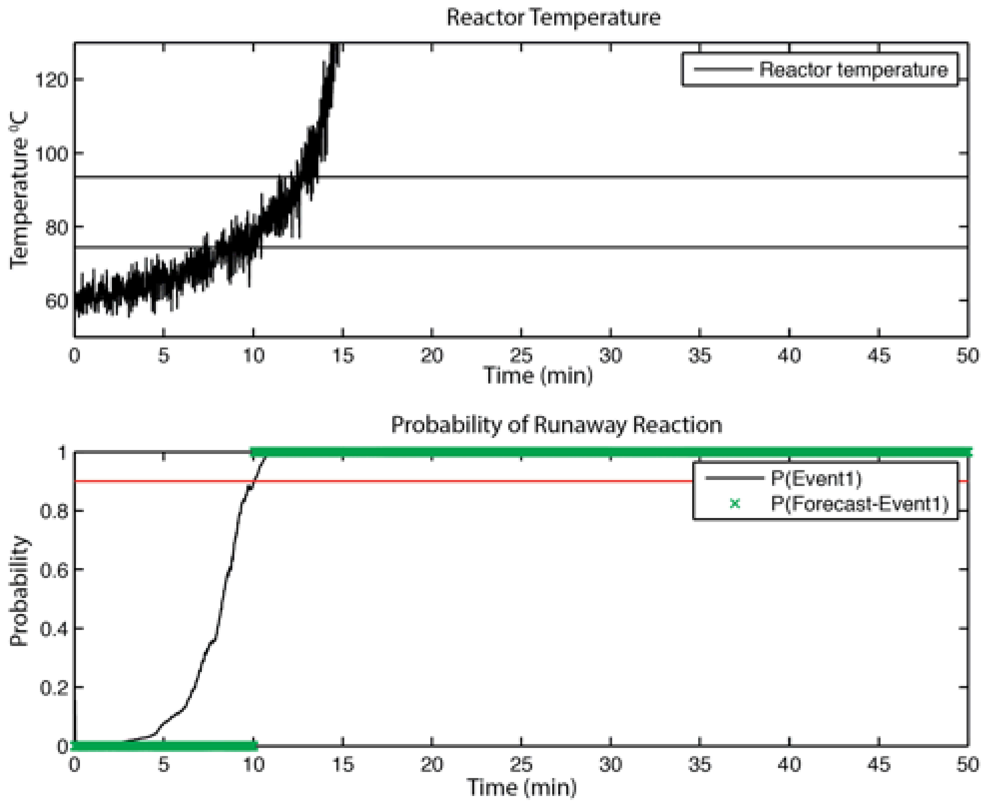


Figure 8-26: Warning annunciation for temperature controller failure

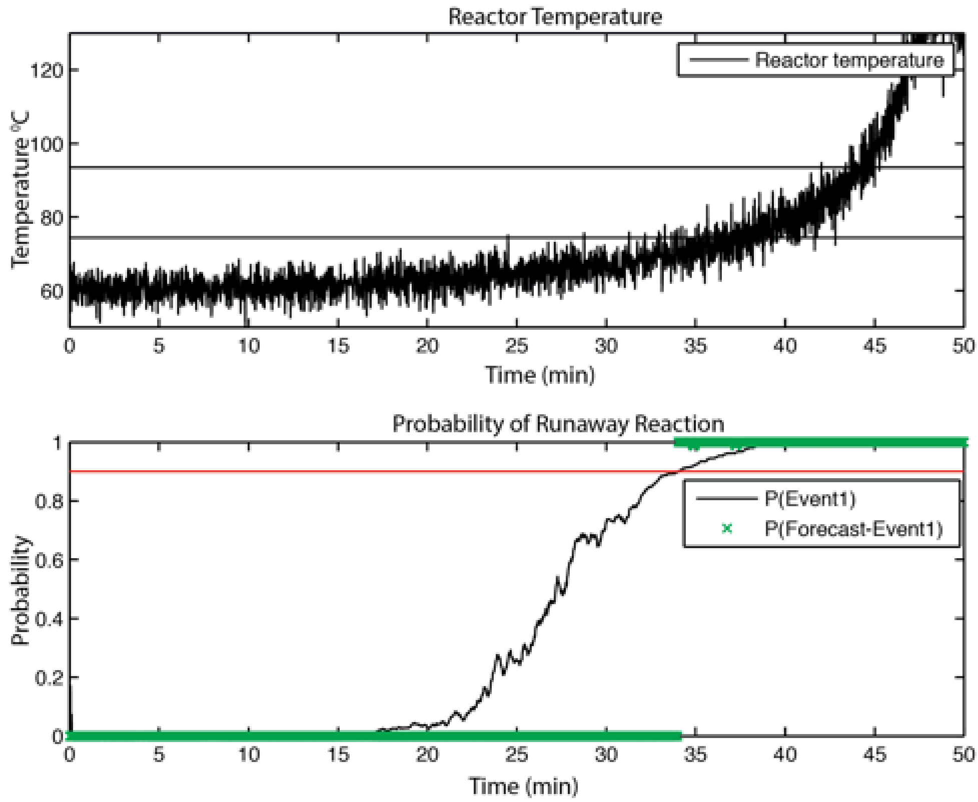


Figure 8-27: Warning annunciation for cooling tower failure

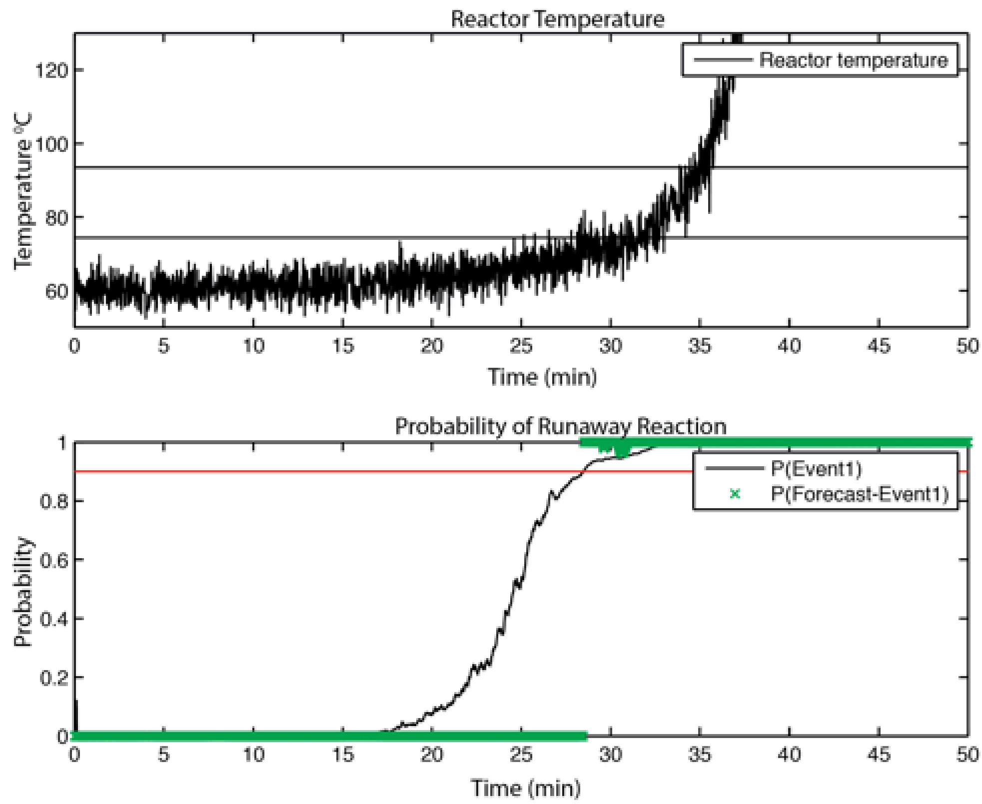


Figure 8-28: Warning annunciation for coolant pump failure

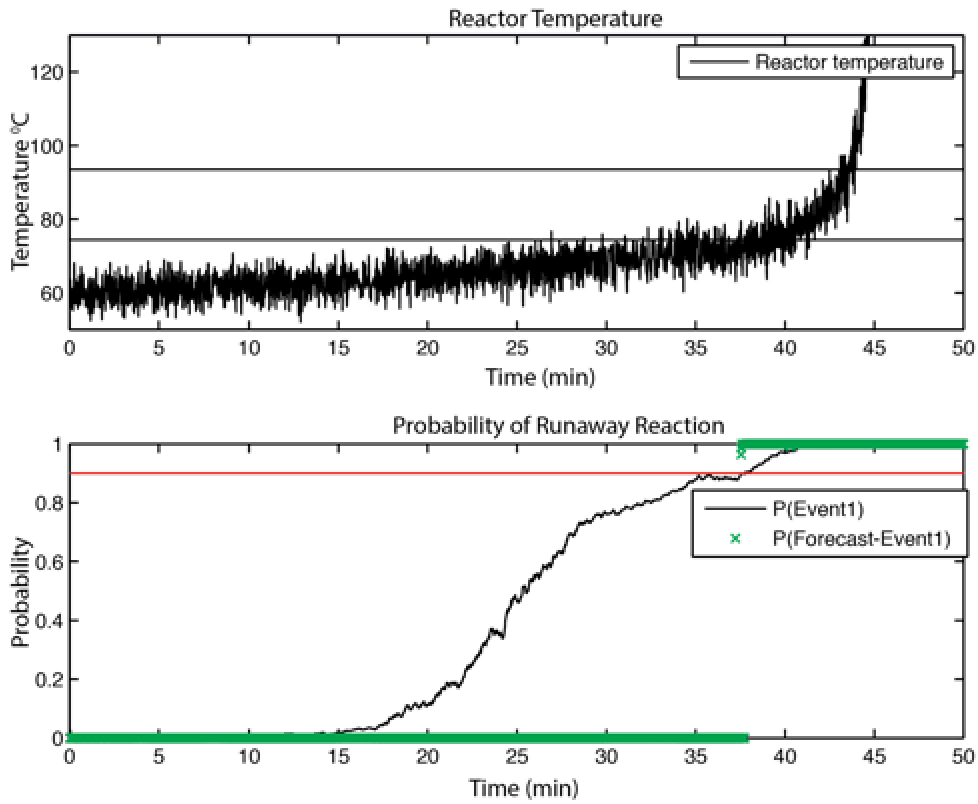


Figure 8-29: Warning annunciation reactant feeding mixture failure

11. CSTR runaway event root-cause analysis

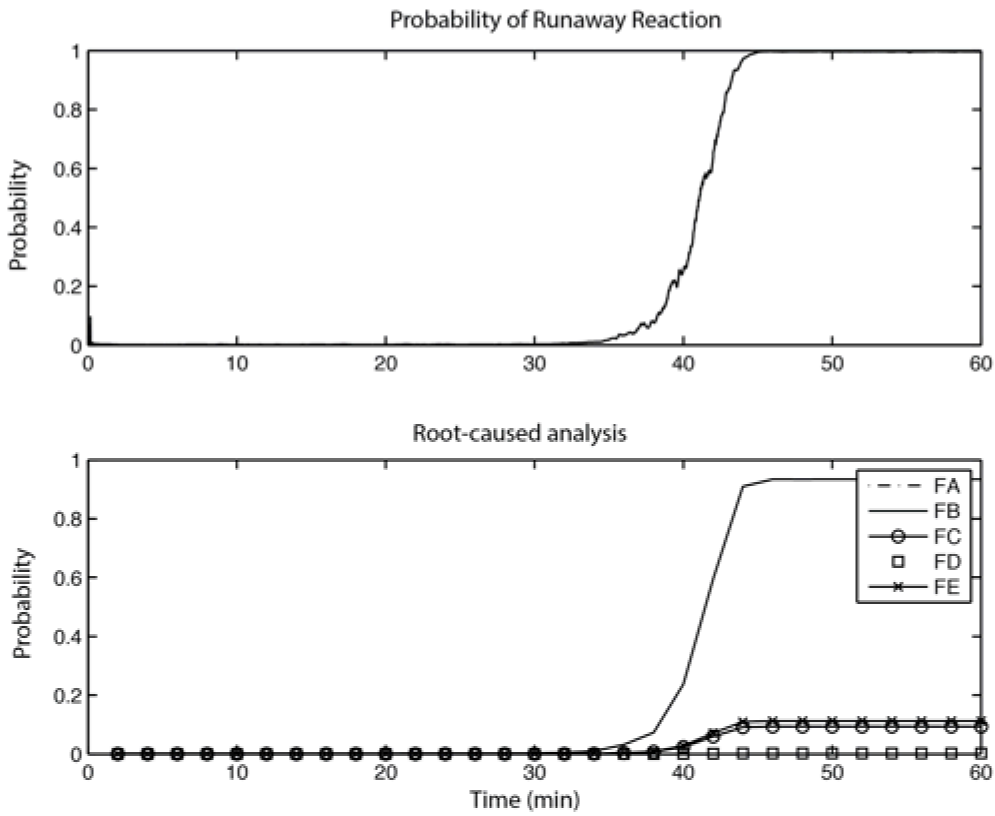


Figure 8-30: Temperature sensor failure root-cause analysis

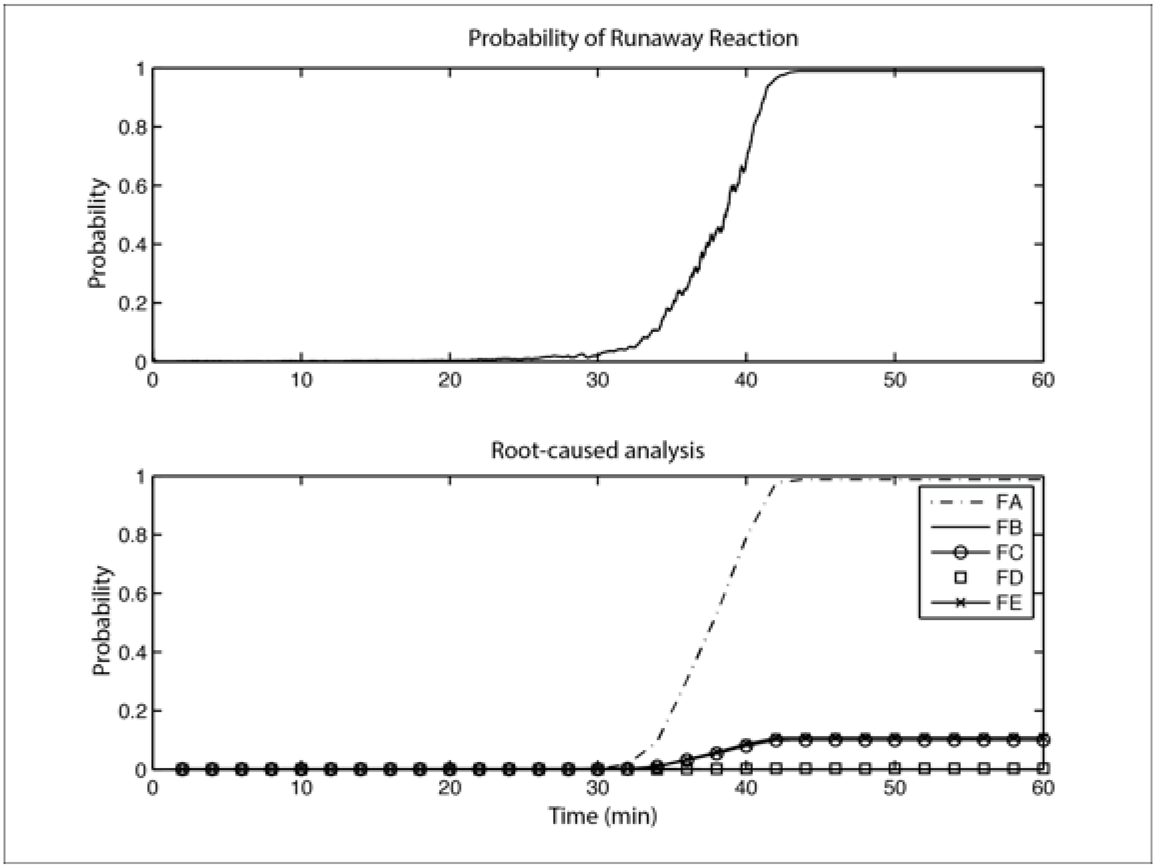


Figure 8-31: Temperature controller failure root-cause analysis

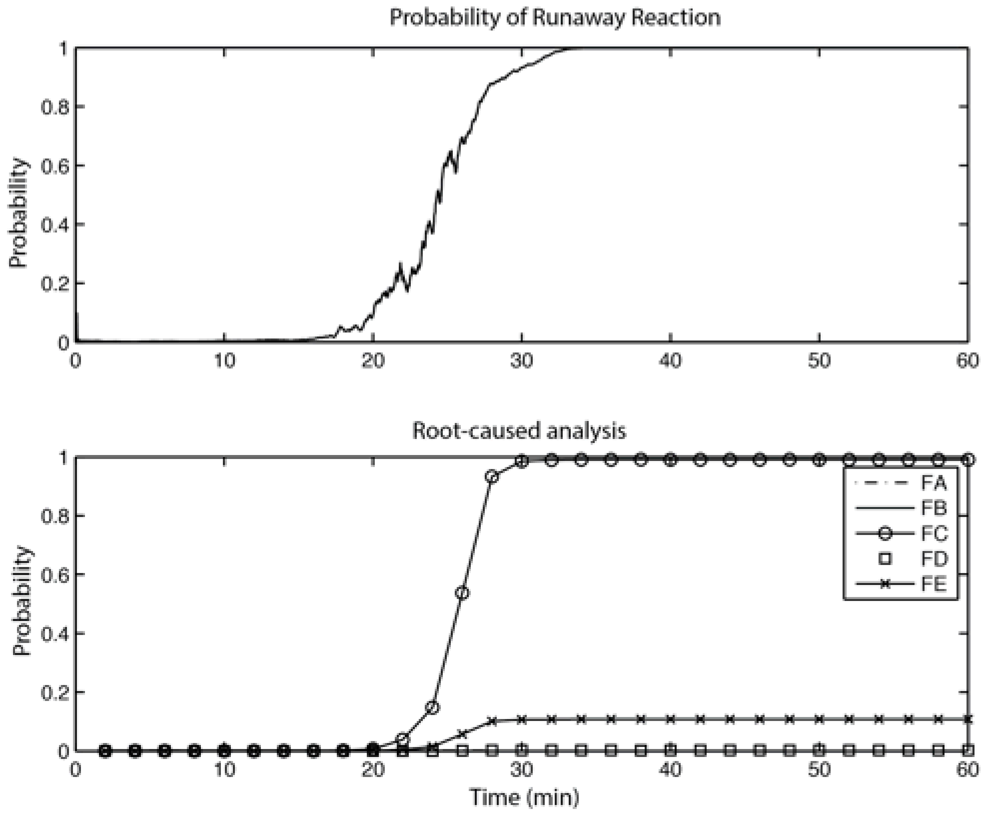


Figure 8-32: Cooling tower failure root-cause analysis

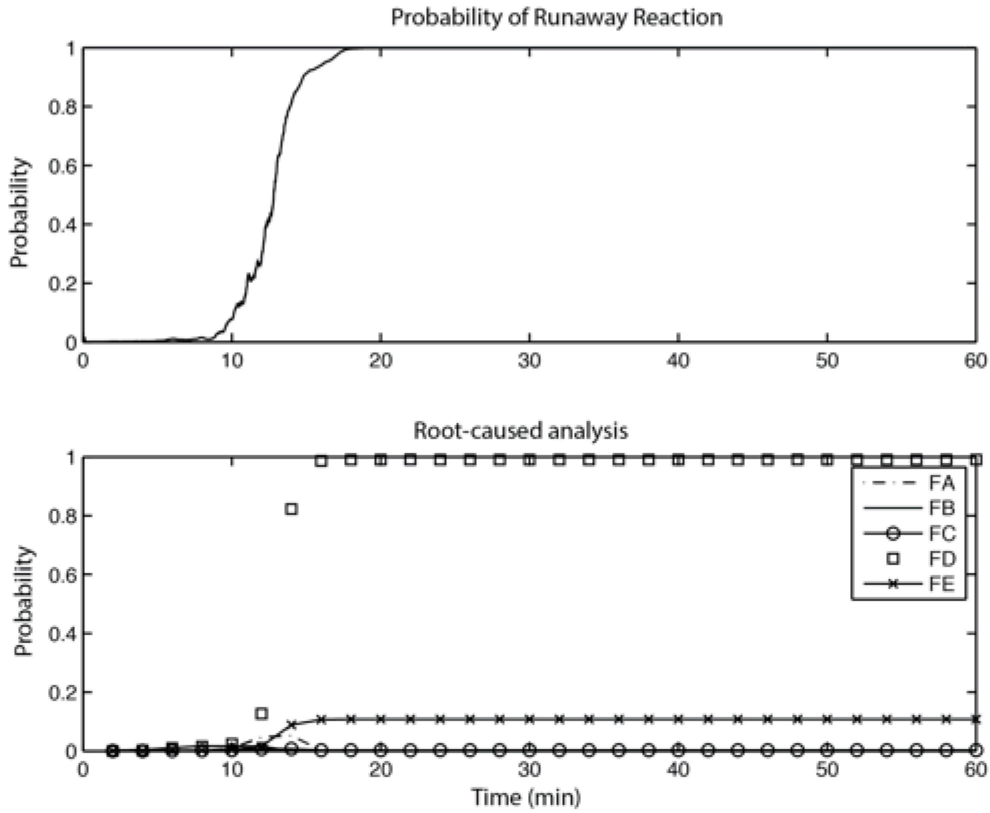


Figure 8-33: Coolant pump failure root-cause analysis

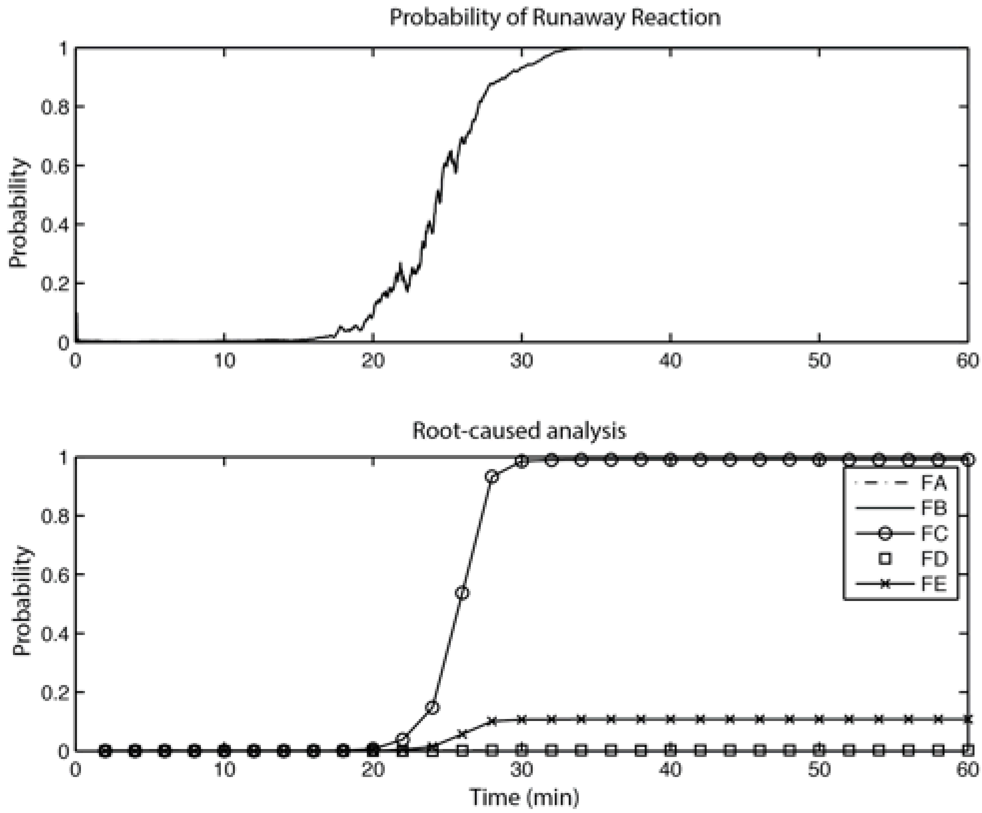


Figure 8-34: Reactant feeding system failure root-cause analysis

APPENDIX C: Publications

12. List of publications

1. Dalpatadu, P., Ahmed, S., & Khan, F. I. (2013). Event based alarm system, *63rd Canadian Chemical Engineering Conference*, Fredericton New Brunswick.

2. Dalpatadu, P., Ahmed, S., & Khan, F. I. (2013). Alarm allocation for event-based process alarm systems. In *IFAC International Symposium on Dynamics and Control of Process Systems* (pp. 815–820). Mumbai, India.

3. Dalpatadu, P., Ahmed, S., & Khan, F. I. (2014). Design of an event-based early warning system for process operations. *Industrial & Engineering Chemistry Research.* (Pending)