Simulation-Based Fault Propagation Analysis of Process Industry Using Process Variable Interaction Analysis

By

Amir Hossein Hosseini

A Thesis Submitted in Partial Fulfillment
Of the Requirements for the Degree of
Master of Applied Science
In

Faculty of Engineering and Applied Science
Program
University of Ontario Institute of Technology

Jan, 2013

To My Dear Parents

&

To My Lovely Wife, Fatemeh

&

To My Sister & Little Brother

Abstract

There are increasing safety concerns in chemical and petrochemical process industry. The huge explosion of Nowruz oil Field platform that happened in Persian gulf-IRAN at 1983, along with other disastrous events have effected chemical industrial renaissance and led to high demand to enhance safety. Oil and chemical Industries involve complex processes and handle hazardous materials that may potentially cause catastrophic consequences in terms of human losses, injuries, asset lost and environmental stresses. One main reason of such catastrophic events is the lack of effective control and monitoring approaches that are required to achieve successful fault diagnosis and accurate hazard identification. Currently, there are aggressive worldwide efforts to propose an effective, robust, and high accuracy fault propagation analysis and monitoring techniques to prevent undesired events at early stages prior to their occurrence. Among these requirements is the development of an intelligent and automated control and monitoring system to first diagnose faulty equipment and process variable deviations, and then identify hazards associated with faults and deviations. Research into safety and control issues become high priority in all aspects. To support these needs, predictive control and intelligent monitoring system is under study and development at the Energy Safety and Control Laboratory (ESCL) – University of Ontario Institute of Technology (UOIT). The purpose of this research is to present a real time fault propagation analysis method for chemical / petrochemical process industry through fault semantic network (FSN) using accurate process variable interactions (PV-PV interactions). The effectiveness, feasibility, and robustness of the proposed method are demonstrated on simulated data emanating from a well-known Tennessee Eastman (TE) chemical process. Unlike most existing probabilistic approaches, fault propagation analysis module classifies faults and identifies faulty equipment and deviations according to obtained data from the underlying processes. It is an expert system that identifies corresponding causes and

consequences and links them together. FSN is an integrated framework that is used to link fault propagation scenarios qualitatively and quantitatively. Probability and fuzzy rules are used for reasoning causes and consequences and tuning FSN.

Acknowledgment

My deep appreciation goes to many people whose advice, assistance and encouragement have enabled me to get to this stage in my life. I am really fortunate to have met so many great people in my life. To those who are missing in this brief list, and were supposed to be here, I sincerely apologize.

First, I wish to express my sincere thanks to my advisor, Prof. Hossam A. Gabbar. His knowledge, valuable guidance, and unlimited patience inspired the completion of this dissertation. Also I would like to thank Mr. Sajid Hussain for his brotherhood help, guidance and willingness to spend his precious time with me are beyond my appreciation.

I do not have the words to express my gratitude to my parents and wife for their endless love, care, and support. Finally, I would like to especially thank to my lovely wife, whose patience and compassion during the final stages of my study, was priceless.

Amir Hossein Hosseini

Ontario, CANADA
Jan 2013

Table of Contents

Chapter 1 : Introduction 1

1.1 Motivation 3

1.2 Problem Definition 4

1.3 Research Objective 5

1.4 Simulation and Data Collection 6

1.5 Thesis Overview 7

Chapter 2: Background 9

2.1 Traditional Technique of Fault Diagnosis. 10

2.2 Artificial Intelligence Methods in Fault Diagnosis 11

2.2.1 Artificial Neural Network 12

2.2.2 Genetic Programing 13

2.2.3 Fuzzy Logic 13

2.3 Probabilistic Approach in Fault Diagnosis 14

Chapter 3 : Methodology 16

3.1 Wavelet De-noising Technique 20

3.1.1 De-noising Process Data Using Wavelet Technique 20

3.1.2 Application of Wavelet Technique in the Study 21

3.2 Genetic Programing 22

3.2.1 Steps of Genetic Programing 25

3.2.2 Initialization 25

3.2.2.1 Grow Initialization Method 26

3.2.2.2 Full Initialization Method 27

3.2.2.3 Ramped Half-and-Half Method 28

3.2.3 Selection 28

3.2.4 Fitness Function 28

3.2.5 Genetic Operation 29

3.2.5.1 Crossover 29

3.2.5.2 Mutation 30
3.2.5.3 Reproduction 31
3.2.6 The GP Application in Fault Propagation 31
3.3 Artificial Neural Network (ANN) 32
3.3.1 Basics of Artificial Neural Network 33
3.3.2 ANN Architecture 34
3.3.3 Learning Modes of ANN 35
3.3.4 The NARX Network 35
3.3.5 The NARX Application in Fault Propagation 37
3.4 Fault Semantic Network (FSN) 38
3.4.1 Semantic Network 38
3.4.2 Basics of Fault Semantic Network 39
3.4.3 Nodes and Process Variables 40
3.4.4 Structure of Fault Semantic Network 42
3.4.5 Relationships Between Nodes 44
3.4.6 Reasoning in Fault Semantic Network 45
Chapter 4 : Chemical Process Plant Simulation 48
4.1 Why Simulation?. 49
4.2 Tennesse Eastman (TE) Process 50
4.2.1 Manipulated and Measured Variables 54
4.2.2 Disturbances 57
4.3 Process Data Simulation 60
4.3.1 Simulation Software 60
Chapter 5 : Results & Discussion 65
5.1 De-noising Process Data Using Wavelet 66
5.1.1 Wavelet Analysis of Data Under Disturbed Condition IDV(1) 67
5.1.2 Wavelet Analysis of Data Under Disturbed Condition IDV(4) 68
5.1.3 Wavelet Analysis of Data Under Disturbed Condition IDV(7) 69
5.1.4 Wavelet Analysis of Data Under Disturbed Condition IDV(15) 70
5.2 Disturbance & Deviation Effect on Process Variables 71
5.3 Genetic Programing Results 72
5.3.1 GPLAB- Genetic Programing Toolbox 72
5.3.1.1 Fitness Function 73

5.3.2 Disturbance IDV(1)-A/C Feed ratio 74
5.3.3 Disturbance IDV(4)-Reactor Cooling Water Inlet Temperature 75
5.3.4 Disturbance IDV(7)- Header Pressure Loss 76
5.3.5 Disturbance IDV(15)-Condensor Cooling Watter Valve 77
5.3.6 Formulation of Relations among Process Variables 78
5.4 NARX Results 79
5.4.1 NARX 79
5.4.2 Disturbance IDV(1)-A/C Feed ratio 80
5.4.3 Disturbance IDV(4)-Reactor Cooling Water Inlet Temperature 81
5.4.4 Disturbance IDV(7)- Header Pressure Loss 82
5.4.5 Disturbance IDV(15)-Condensor Cooling Watter Valve 83
5.5 Comparison of GP & NARX Results 84
5.5.1 Mean Absolute Error & Mean Absolute Percentage Error 87
5.6 Prototype Development and Testing-Fault Semantic Network 89
6. Conclusion 98
6.1 Potential Applications 101
6.1.1 Risk Calculation 101
6.1.2 Safety Performance Indicator (SPI) 101
6.1.3 Automated Hazard Identification 102
6.2 Future Work 103
6.2.1 Application of Automated Hazard Identification on LNG Process Plant 104
6.2.2 Application of Automated Hazard Identification on Hydrogen Production Process Plant Cu-Cl Cycle 106
References 107
Appendixes 112
Appendix 1: GP MATLAB code 112
Appendix 2: FSN- LabVIEW® Block Diagram 113
Appendix 3: List of Research Papers 114

List of Figures

Figure 1.1: Research Framework.....7
Figure 3.1: framework of research methodology 19
Figure 3.2: Genetic Programing 22
Figure 3.3: Tree Structure Representation 23
Figure 3.4: Genetic Programing Flowchart 25
Figure 3.5: Grow Initializing Method 26
Figure 3.6: Full Initialization Method 27
Figure 3.7: Crossover Operator Mates the Parents 30
Figure 3.8: Mutation Operation 31
Figure 3.9: Overall View of Neural Network 32
Figure 3.10: Mathematical Model of an Artificial Neural Network 34
Figure 3.11: Nonlinear Autoregressive Exogenous Inputs Model (NARX) 36
Figure 3.12: Tree-Structured Fault Semantic Network 39
Figure 3.13: A FSN for the Vibration Fault in Pump 43
Figure 3.14: Types of Reasoning.....47
Figure 4.1: Tennessee Eastman (TE) process.....53
Figure 4.2: Schematic View of Simulation in Simulink Environment 61
Figure 5.1: shows the original and denoised data of process variables under disturbed condition IDV (1).....67
Figure5.2: shows the original and denoised data of process variables under disturbed condition IDV (4).....68
Figure 5.3: shows the original and denoised data of process variables under disturbed condition IDV (7).....69
Figure 5.4: shows the original and denoised data of process variables under disturbed condition IDV (15).....70
Figure 5.5: shows a comparison of process variable interaction between disturbed and normal operational condition. 71
Figure 5.6: shows the effect of changes in A/C feed ratio as manipulated variable under disturbance IDV (1) 74

Figure 5.7: Actual vs. Predicted Trend of Reactor Temperature under Disturbance IDV (4).....75
Figure 5.8: Actual Vs. Predicted Trend of Feed A under Disturbance IDV (7).....76
Figure 5.9: Actual Vs. Predicted Trend of Reactor Pressure under Disturbance IDV (15).....77
Figure 5.10: shows the effect of changes in A/C feed ratio as manipulated variable under disturbance IDV (1) 80
Figure 5.11: Actual vs. Predicted Trend of Reactor Temperature under Disturbance IDV (4)....81
Figure 5.12: Actual Vs. Predicted Trend of Feed A under Disturbance IDV (7).....82
Figure 5.13: Actual Vs. Predicted Trend of Reactor Pressure under Disturbance IDV (15).....83
Figure 5.14: Error Histogram of Reactor Feed rate Prediction-IDV (1). 84
Figure 5.15: Error Histogram of Reactor Pressure Prediction-IDV (1).....84
Figure 5.16: Error Histogram of Reactor Level Prediction-IDV (1). 85
Figure 5.17: Error Histogram of Stripper Pressure Prediction-IDV (1). 85
Figure 5.18: Error Histogram of Reactor Temperature Prediction-IDV (4).....86
Figure 5.19: Error Histogram of Feed A Prediction-IDV (7). 86
Figure 5.20: Error Histogram of Reactor Pressure Prediction-IDV (15) 87
Figure 5.21: FSN Module 89
Figure 5.22: P&ID of TE Process-Main Reactor.....90
Figure 5.23: Fault propagation Analysis 95
Figure 5.24: Schematic View of FSN Prototype 97
Figure 5.25: TE process Simulation in AspenHysys 98
Figure 5.26: FDS Designed in ESCL-UOIT 103
Figure 5.27: P&ID of LNG Process (De-methanizer Unit) 104
Figure 5.28: AspenHysys Simulation Environment – LNG (Process) 105
Figure 5.29: Hydrogen Production process – Cu-Cl Cycle 106

List of Tables

Table 3.1: Nodes Specification for the Pump Vibration Example 41
Table 4.1: Physical Properties of Components 51
Table 4.2: Plant Operating Modes 54
Table 4.3: Process manipulated variable 55
Table 4.4: Measured variables 56
Table 4.5: Process Disturbances 60
Table 4.6: Selected Disturbances 62
Table 4.7: Process Manipulated & Measured Variables for Disturbance IDV (1) 62
Table 4.8: Process Manipulated & Measured Variables for Disturbance IDV (4) 63
Table 4.9: Process Manipulated & Measured Variables for Disturbance IDV (7) 63
Table 4.10: Process Manipulated & Measured Variables for Disturbance IDV (15) 64
Table 5.1: GPLAB Parameters Values 72
Table 5.2: The nonlinear Relationship between Xi (Manipulated variable) & Yi (Measured Variable) 78
Table 5.3: MAE & MAPE Values 88
Table 5.4: Process Variables, Nodes & Values 91
Table 5.5: Failures, Mechanism, Symptoms and Description 92
Table 5.6: Sensors and Description 93
Table 5.7: Designed Case Study Implemented by FSN 96

List of Algorithm

Algorithm 3.1: *node* grow (*depth*) 26
Algorithm 3.2: *node* full (*depth*) 27

Chapter 1

Introduction

Operating an industrial process needs a high level of safety and reliability which is possible through making strict regulations and implementing safety culture. Neglecting safety regulations and culture lead to catastrophic accidents followed by serious consequences, human losses and environmental impacts. Therefore designing a proper safety system by considering safety requirements and implementing monitoring strategies to avoid undesired accidents is an important concern of industries.

Petrochemical industries lose an estimated 20 billion dollars every year [1]. The BP Texas refinery accident can be considered as a clear-cut example. In March 2005, an explosion happened in BP Texas refinery (USA) and caused the death of 15 people and injured 180 ones. The economical damage, including the income losses during the reconstruction period from March 2005 to March 2006, was estimated at US$ 1.5 billion [2]. As another example, the release of gas caused the world worst industrial disaster, which is called Bhopal disaster that took place in India in December 1984. It caused 8000 deaths and 558,125 injuries and other environmental consequences [3]. The most recent disaster is the catastrophic event that occurred

in Venezuela's biggest oil refinery that happened in August 2012. A huge explosion unleashed a ferocious fire Saturday, killing at least 39 people and injuring more than 80 [4]. Given these statistics, improving the plants from the point of view of control and monitoring is a challenge for oil, gas and chemical industries.

Over the years, engineers and researchers were focusing on improving industrial processes to increase their proficiency. These efforts led to efficient industrial processes and improved their products both quantitatively and qualitatively. Although these progresses have advantages, however they increase complexity of industrial processes and make them more complicated. Beside the increase in process complexity, monitoring and control approaches of process industries have to be updated. Processes that were controlled manually by operators through a simple procedure now have to be controlled automatically through complex procedures.

Significant progresses are done in the area of hazard identification and safety system design. Recent technologies and automation are key players to achieve such progresses. Although these technologies have advantages such as efficient productivity, lower cost of production and labor, high quality of product and etc., however on the other hand, it increases the complexity of process systems. Complex systems have more chances to be subjected to faults and hazardous conditions. Any of these faults and hazardous conditions may result in costly and serious incidents or accidents such as, explosion, plant shutdown, injuries and etc.

Faults in plant equipment have the potential to affect the performance of the entire process. Even for minor faults, it might lead to a serious degradation in the whole process. Therefore considering the simplest faults seems necessary in industries especially the ones that deal with

continuous processes, where lots of process variable and hazardous materials such as chemical and petroleum industries.

Recent technologies are focusing on controlling and monitoring of processes, and modeling system dynamics, as well as predicting its behavior and the relationships among process variables. This will provide means to detect abnormal events at early stages by understanding the relationship among process variables and their impacts on fault propagation scenarios. These targets are achievable through an accurate fault propagation analysis which results in a successful fault diagnosis. Given this, controlling and monitoring the process operation to implement a well-designed fault diagnosis technique is among the most important concerns in process industries.

1.1. Motivation

Implementing a successful, productive, and safe operation is among the top priorities of plant processes. It results in higher efficiency, lower cost of production and reduction of hazards. These goals are only achievable under well controlled and well monitored process operations. Implementing an efficient monitoring, hazard identification and fault diagnosis system makes it possible to reduce undesired and sudden events that may result in catastrophic accidents or disasters. Fault diagnosis and hazard identification studies are attracting engineers and researchers as the first and foremost priority of industries.

Since 1984, different techniques are introduced in order to implement fault diagnosis and hazard identification for chemical processes [5]. These techniques include Fault Tree Analysis (FTA), Failure Mode and effect Analysis (FMEA), Quantitative Risk Assessment (QRA), Hazard and Operability Analysis (HAZOP) and etc. Although these techniques are implemented in different

industries around the world, however minor and major accidents are still occurring because of their limitation, such as low accuracy of diagnosis mainly because of lack of accuracy in fault propagation analysis and slow response time. Moreover, these techniques have to be implemented by operators and their skill and knowledge are not adequate to monitor process plant and understand faulty conditions. Therefore human errors are also considered as main causes of many faults and hazard scenarios in process industry, where more attention is required.

Accurate and quantitative fault propagation analysis is the basic of successful Fault diagnosis which is a challenging and difficult task because there are many parameters that are involved and may affect the accuracy of diagnosis. These parameters include types of alarming equipment, their response time, quality and sufficiency of obtained process data and of course the time delay in faults. Given these information, implementing an intelligent fault propagation analysis that overcomes the limitations and enhance fault diagnosis and hazard identification seems to be a necessary need and concern of industries [6].

Predictive control and monitoring system is under development at the Energy Safety and Control Laboratory (ESCL) – University of Ontario Institute of Technology (UOIT), which include developing real-time fault propagation analysis and automated solution for chemical process industry. The proposed methods are implemented as part of an integrated and intelligent fault propagation analysis presented in [7-9].

1.2. Problem Definition

According to industrial statistics, the main reason of 70% of accidents is human error. These accidents result in significant economic, safety and environmental effects. Despite the infrequent catastrophic events that result in significant consequences, minor accidents are very common in

chemical industries. These frequent minor accidents and incidents cause many injuries, illnesses and occupational health impacts [1].

Considering the given statistics, in order to implement a successful fault diagnosis and hazard identification system, it is necessary to reduce the human involvement in controlling and monitoring chemical processes, which is only possible through the implementation of quantitative fault propagation analysis and hazard identification system. Existing fault diagnosis techniques and technologies are limited to analysis of specific fault scenarios in case-by-case basis, with the considerations of process data without integrating other state variables that might be involved in any given fault scenario. Typical hazard identification techniques are able to identify major hazard and fault scenarios while they can't guarantee complete identification of all possible hazard scenarios. This means that some fault scenarios might occur for the first time during plant operation, while operator is not aware of such fault propagation scenario. The use of simulation might provide potential support to analyze different fault propagation scenarios that will provide strong knowledgebase to support fault diagnosis during plant operation. In addition, and because there are large number of process variables that might be involved in each fault propagation scenario, it is important to study and develop intelligent algorithms to analyze the relationships among process variables and their impacts on fault propagation scenarios.

1.3. Research Objective

The primary target of this research is to develop an intelligent simulation-based fault propagation analysis and hazard identification systems that consists of two modules of pattern recognition and reasoning. The first module will perform pattern recognition, which will detect inter-relation patterns among process variables, both quantitatively and qualitatively. The second module is to

perform the reasoning that builds and tunes the proposed fault semantic networks, which will support the identification of possible causes and consequences for each fault scenario. The specific summery of the proposal objectives are as follow:

1. Process data analysis using signal processing technique for de-noising
2. PV-PV interaction analysis and error estimation: Identifying the relationships among process variables quantitatively using two different methods of genetic programing and neural network and comparing them.
3. Estimate static probabilities for quantitative fault propagation analysis and identifying causes and consequences through fault semantic network (FSN).

1.4. Simulation and Data Collection

In order to generate sufficiently rich data sets in this research, a case study process P&ID is selected which is proposed by Eastman chemical company at 1990. The process modeled by the eastman chemical company is a defined chemical process called The Tennessee Eastman (TE) process [10]. AspenHysys which is simulation software of chemical processes and MATLAB-Simulink are used to generate data from the TE process.

1.5. Thesis Overview

Figure 1.1 represents the research framework and the thesis is organized as follow.

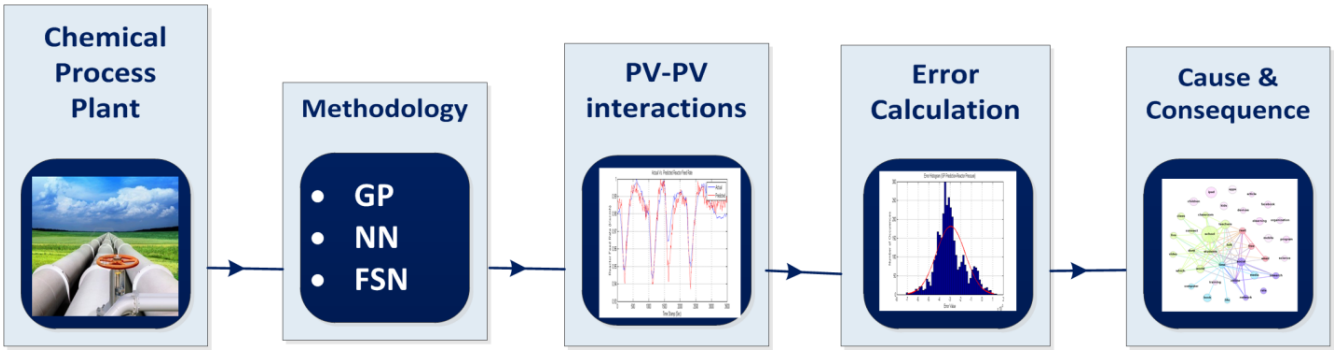


Figure 1.1: Research Framework

- **Chapter2-Background,** discusses the background and literature of the work that is done so far.
- **Chapter 3-Methodology,** discusses the theory behind this research. This chapter reviews the theory of genetic programing (GP) and artificial neural network (ANN)-Nonlinear Auto-Regressive Exogenous (NARX) as techniques to identify the relationships among process variables. Also it introduces fault semantic network (FSN) as a new tool to model a process for the purpose of fault propagation analysis and hazard identification.
- **Chapter 4-Simulation of chemical process,** introduces the simulation software, simulated chemical process plant, manipulated and measured process variables, their values, ranges and applied disturbances in detail.
- **Chapter 5: Results and Discussion,** presents the GP and NARX prediction results in identifying relationships among process variables and compare them by measuring mean

absolute error. Then it discusses implementation of designed FSN according to the case study and defined hazard scenarios.

- **Conclusion and Future works,** summarizes the conclusions, discusses the contributions of this work and proposes further research directions.
- **Appendixes,** are consists of brief MATLAB code of GP and LabVIEW® block diagram of designed FSN and list of research papers derived during the study.

Chapter 2

Background

Chapter Summery

Designing real-time intelligent fault propagation analysis and hazard identification is the purpose of this research. Lacking an accurate and quantitative fault propagation analysis is a problem that fault diagnosis techniques are suffering. This chapter reviews related works and researches done in this area. First, it reviews the traditional methods of fault diagnosis and then focuses on recent approaches such as artificial intelligent and probabilistic approach.

2.1. Traditional Techniques of Fault Diagnosis

As Venkat and his research group [11, 12] reviewed different fault detection and fault diagnosis methods. They classified fault diagnosis methods in three categories as follow:

- Quantitative Model-Based Methods
- Qualitative Model-Based Methods
- Process History-Based Methods

However, the disadvantages that come with these methods are their limitations in modeling nonlinear system and having model errors due to the simplicity of approximation that considerably reduces the effectiveness of these methods.

In another study [13], they tried to design an expert system to implement a real-time fault diagnosis system using computer-aided techniques. When an abnormal situation arises in process, real-time data extracted from sensors is stored in an online database and then the expert system reasons it to find causes and consequences according to the designed knowledgebase. Similar to the work done at [13], Nan and Khan [14] have studied other computer-aided fault diagnosis techniques. They proposed a knowledge-based fault diagnosis method that uses fuzzy-logic as an inference engine to reason according to the extracted real-time data and knowledgebase. These techniques were useful and offered practical ways to perform fault diagnosis. The performance of these fault diagnosis techniques depends on the frequency and presence of fault data and the quality and accuracy of fault and hazard scenarios' knowledgebase.

As part of the researches and works done in the area of computer-aided fault diagnosis, Gelgele and Wang at [15] have developed a software prototype called EXEDS with application of fault diagnosis in automotive engines. The software analyzes failure symptoms to diagnose faults and suggests desired maintenance actions. It consists of knowledgebase and expert system which include a list of symptoms, diagnosis modules, remedies and associated rules respectively.

Principle component analysis (PCA) is another technique used for the purpose of fault diagnosis as Shams and Budman used it in fault detection according a set of data extracted from the TE process [62]. However it has some limitation as Haifeng Chen [63] in a study mentioned the PCA limitations. The PCA has trouble with high dimensional data and huge number of data. In addition it has problem with analyzing incomplete data.

2.2. Artificial Intelligence Methods in Fault Diagnosis

Artificial intelligent techniques are used for different purposes such as pattern recognition, fault diagnosis, etc. This part presents a background of their applications in pattern recognition and fault diagnosis.

Different techniques are introduced in the past to uncover the relationship patterns among process variables [16, 17]. These techniques have some disadvantages over the new ones. The disadvantages mostly deal with variables distribution. The variables are assumed to distribute randomly so they may be analyzed either in a limited range or left undetermined, which means these drawbacks make the technique unable to cover full range of variables. Being unable to analyze non-linear relation between variables is another disadvantage that traditional techniques have limitations. Since these techniques are assumed to analyze the variables that are linearly related; consequently, when it comes to analyzing non-linear relations, the results obtained are

poor. Another issue that can be mentioned for traditional methods is their limited representation way. There are some techniques that overcome these difficulties, such as non-linearity of relation between the variables, however they could not solve the problem of representation. It is considered as a critical issue because it requires substantial efforts for the researchers to translate the analysis results into a suitable representation that is understandable by the human [18]. Moreover, selecting the best combination of variables that result in producing the best solution is one of the critical problems that these techniques are involved in [19].

2.2.1. Artificial Neural Network

Artificial Neural network (ANN) which is one of the most favorable techniques in the era of artificial intelligent. Since it has the ability to detect the interactions between the variables, it has the potentiality to be considered as a robust technique in predicting systems behaviors and modeling the processes. Although ANN gains success in the field of modeling and prediction, it cannot show clearly the relationships among process variables in the form of mathematical equations [20]. Also ANN gained more popularity in variety of other research fields. ANN is commonly used in automation, fault diagnosis, and control began at 1992 [21].

H.N. Kolvo [22] used ANN to implement fault diagnosis and real-time control strategy in chemical process here both static and dynamic fault diagnosis methods are discussed. In addition, real-time controllers are tested where they demonstrated robustness and good performance of ANN.

In another interesting work done by Ruiz and his research group [23], they designed a framework based on artificial neural network for fault diagnosis in batch chemical process. It consists of ANN structure with knowledge based expert system.

Kerpenko, Sepehri and Scuse [24] made further studies on the application of ANN where they used multilayer neural networks to diagnosis actuator faults in a control valve system. Parameters such as peak time, dead time, and percentage overshoot are considered as the roots of faults in the valve. However, the relationships among these parameters are uncovered through implementing multilayer feed-forward neural networks. The results show that neural network is able to estimate the faults level accurately.

There are other studies that used ANN for the purpose of fault diagnosis in industries and their results showed improved capabilities of dynamic neural networks in identifying nonlinear systems and diagnosing faults [25, 26].

2.2.2. Genetic Programing

Genetic Algorithm (GA) and Genetic Programming (GP) are other promising techniques for modeling and prediction of complex systems behavior. They have several advantages over others to overcome their drawbacks and show the relationships among process variables as prediction equations. The GP has a great history of application. Despite it can be implemented parallel with other methodologies, it has been performed for different purposes and problems. The GP has been used in trading strategies [27], and regression applications [28]. It is also used as a tool to predict the system behavior such as its application in sediment load prediction [29] and monthly rainfall prediction [30]. However for the purpose of pattern recognition and fault diagnosis; Genetic Programing has not been used extensively.

2.2.3. Fuzzy Logic

Fuzzy trend analysis is another technique that can be used for the purpose of pattern recognition and trend analysis among process variables. In one study [31], Dash and his researchers tried to use fuzzy approach in order to overcome the imprecise classification in trend analysis. They used fuzzy-logic as a trend-based reasoning process to model and reason about process variables in chemical process industry to detect and diagnosis faults.

In an interesting study, Sampath and Singh [32] compared the accuracy of three methods of artificial intelligent in diagnosing faults: GA, ANN, and fuzzy logic. The results showed that GA has better accuracy in fault detection compared with other methods, while ANN shows better performance than fuzzy.

2.3. Probabilistic Approaches in Fault Diagnosis

Probabilistic approaches are commonly used in fault diagnosis. Belief or Bayesian networks are promising probabilistic techniques that can offer accurate calculation of probabilities of faults. Commonly, the Bayesian Network (BN) technique is applied on fault diagnosis of complex systems. In a study done by Guzman and Kramer [33], they compared belief network with rule-based expert system and showed how these two methods contrast with each other. An expert system consists of a knowledgebase made of heuristic rules that demonstrated successful operation in diagnosing faults. However they have some disadvantages and limitations. The probabilistic approaches not only overcome these drawbacks, but also construct a logical way to reason and model the underlying process in order to detect faults accurately.

Bickford and Malloy [34] used Bayesian approach to develop an online fault diagnosis software prototype to detect and diagnose faults in turbine engine. The proposed software monitors and classifies the source and type of sensor, component, and engine faults. The proposed software provides a mathematical means to integrate data from multiple diagnostic instruments and algorithms in order to detect and classify faults in real time.

In another similar study done by Romessis and his team [35], they proposed a method for building Bayesian Belief Networks (BBN) to diagnose faults in a gas turbine. The results show that the proposed method was successful in fault diagnosis with accuracy of %96, which shows the high reliability of BBN as a probabilistic method in fault detection and diagnosis. Khan and his group at Memorial University has developed probabilistic framework using BBN to analyze faults and apply on hazard identification. Their approach provided new features for accurate prediction of faults and their propagation [5, 14].

Lacking an intelligent, accurate and quantitative fault propagation analysis is an issue that fault diagnosis techniques are facing with. Integrating artificial intelligent technique with probabilistic approach would be a novel solution for this issue. Fault semantic network (FSN) can be considered as a technique to overcome these limitations and implement a successful and accurate fault propagation analysis.

The proposed fault propagation analysis method is an expert system that identifies corresponding causes and consequences and links them together. FSN is an integrated framework that is used to link fault propagation scenarios qualitatively and quantitatively. Bayesian belief network (BBN) is used to estimate probability of all fault propagation paths, and fuzzy expert system (FES) is used for reasoning causes and consequences and tuning FSN.

Chapter 3

Methodology

Chapter Summery

The goal of this work is to implement a real-time and intelligent fault propagation analysis and hazard identification system. In this study, two types of variables considered as manipulated (Independent) and measured (Dependant) variables where measured ones are affected by manipulated variables. To reach the aim of this thesis, the relationships among dependant and independent process variables have to be identified to see how they affect each other. This chapter is going to discuss the methodologies proposed in this work. The first step is the extraction of process data from sensors installed on different equipment in the underlying plant process. This chapter includes three main parts discussed as follow.

The ultimate goal of the study is to develop a real-time fault propagation analysis and hazard identification method through fault semantic network (FSN). Once process data are ready to be analyzed, the FSN module starts identifying the quantitative relationships among them and tracing deviations in following steps. The overview of research methodology is presented in Figure 3.1.

1. **Extracting real-time process data:** Flow of many process data come from process plant. These real-time data are extracted from sensors and controllers installed on all equipment in the underlying plant to monitor the process conditions.
2. **Data Processing:** Real-time process data are typically associated with noises. Therefore data processing is needed prior to the application of fault propagation algorithms. In order to have better data analysis the noise has to be removed through the application of de-noising techniques. Therefore once these data are extracted from sensors and prior to start the analysis, all of them are de-noised separately and simultaneously. To this, wavelet de-noising technique is performed in the study. To be able to compare process variables, they should be in unified scale. Hence, normalization techniques are proposed. Now process data are ready to go under analysis.
3. **Relationship among process variables through GP & NARX:** In order to uncover the relationship among process variables, GP and NARX are used as pattern recognition techniques to identify their relationship quantitatively. This step analyzes propagation of faults among process variables, therefore any small deviation in a process variable will be detected and main cause process variable will be identified. Also, when process goes outside its defined safety limits causes will be analyzed. Although the FSN can trace the process deviations from real-time data that dynamically update it, however formulating

the relationships among process variables is necessary steps. If extracted data are not complete for any reason such as faulty sensor, sensor failure and etc. the FSN can recognize the issue through the formulation obtained form process variables inter-actions. Hence, the FSN will be able to compensate lack of complete data.

4. **Fault propagation analysis through FSN:** Once the inter-action among process variables recognized, it is time to trace deviations between process variables and failure mode in equipment. The FSN represents process variables and failure modes as separate nodes. To analyze propagation of faults, a database is designed that includes all process variables, all possible relations among them, different hazard scenarios, associated failure modes and all possible causes and consequences. FSN is constructed based on ontology structure of fault models on the basis of process object oriented model (POOM) where failure mode (FM) is described using symptoms, enablers, process variables, causes, and consequences and they are represented by separate nodes. The FSN update itself through real-time process data received from the plant. If any process variable is outside of defined range (as defined in database), it will be detected. Considering the probability associated with each node, *if-then* rules and the defined scenarios in database, the FSN traces deviations and reason between nodes.

The framework of the proposed research methodology is presented in Figure 3.1.

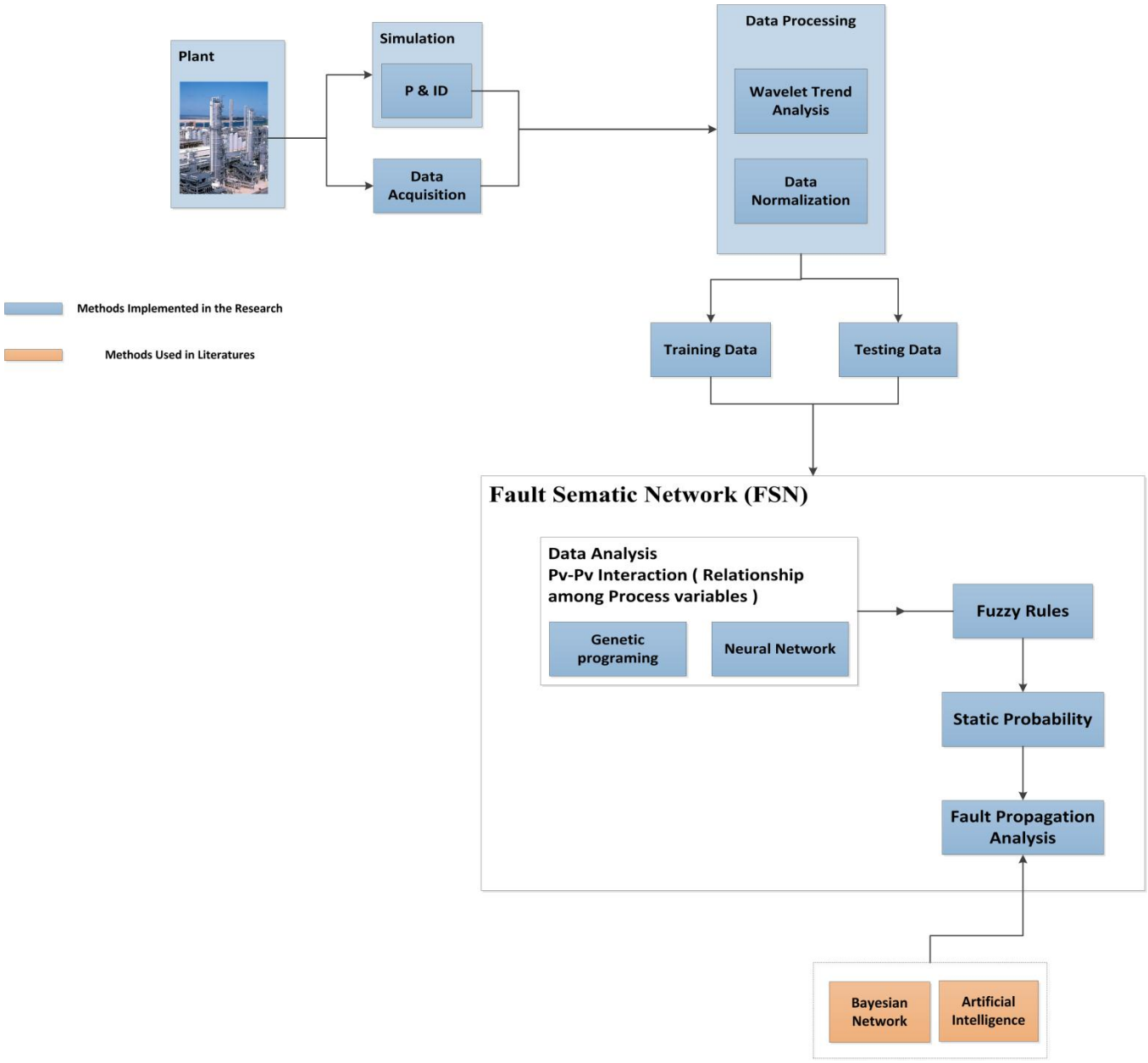


Figure 3.1: Framework of Research Methodology

3.1. Wavelet De-noising Technique

One of the foremost difficulties that industries are facing is the challenges to extract pure data from process plants. The extracted data might be incomplete, indirect, or noisy. In order to perform data analysis and get reliable results, it is a necessity to process data before the analysis. Processing data makes it possible to enhance the relevant signal characteristics of process variables [36]. One of the most famous data processing methods is de-noising data. Different techniques have been introduced [37, 38] to obtain clean process data. Wavelet trend analysis is one of most promising techniques for de-noising variety of process data sets. It provides robust capabilities to obtain clean and reliable data. David Donoho has worked on wavelet techniques for several years [39].

Wavelet technique compresses and de-noises process data without degrading them. It operates like a filter to assigns coefficients correspond to specific details of process data set. If the detail is small, it will be omitted without affecting main features of data set. In other word, it sets all coefficients that are less than a particular threshold to zero [40]. Therefore, the results will be cleaned-up signals that still show important details.

3.1.1. De-noising Process Data using Wavelet Technique

Wavelet analysis follows the Fourier transform manner. It uses scales and translated version of original wavelets $\psi(x)$ defined as follow.

$$\psi_{j,k}(x) = \delta\psi\left(2^j x - k\right). \qquad (3.1) \quad [41]$$

Where δ is a constant and k is the wavelet`s translation. Given this, $f(x)$ can be calculated as follow.

$$\hat{f}(x) = \sum_{\forall j, \forall k} c_{j,k} \psi_{j,k}(x). \qquad (3.2) \quad [41]$$

Where cj, k is the wavelet coefficient determined from wavelet transform.

$$c_{j,k} = \int_{-\infty}^{\infty} f(x) \psi_{j,k}(x) dx. \qquad (3.3) \quad [41]$$

Considering the functions used in the wavelet transform and coefficients, it is possible to find different signals frequency in a timely manner.

Wavelet analysis basis functions are finite and limited to one size. This makes wavelet analysis a useful technique for detecting local features in a signal like discontinuities and spikes. Wavelet analysis is a joint time-frequency analysis technique. This feature also makes wavelet analysis to tell about when in time a particular event took place.

De-noising the signal is one of the most effective applications of wavelets in signal processing. The wavelet transform-based de-noising methods can produce much higher de-noising quality than conventional methods. Furthermore, the wavelet transform-based methods retain the details of a signal after de-noising [41, 42]. Wavelet function assigns coefficients of $C_{j,k}$ corresponds to signals of data by wavelet transform. These coefficients indicate the contribution of wavelet $\psi(x)$ in processing data. If the coefficient is small, it means that the contribution is small therefore it has to be omitted form the processing [43].

3.1.2. Application of Wavelet Technique in the Study

In this study, wavelet technique is used to remove the noise from extracted process data. The results of its performance in discussed in chapter 5 in detail.

3.2. Genetic Programing

As it is mentioned in the literatures, evolutionary methods are under many studies for the purpose of pattern recognition. Genetic programing (GP) is one of these methods used by many researchers [27-30]. GP is a promising technique for modeling and prediction of complex systems behavior. This capability of GP in pattern recognition and prediction of system behaviour can meet the requirement of this study in identifying the inter-relation pattern among process variable.

Genetic Programing is an evolutionary model and one of the machine learning techniques. It is an evolutionary algorithm-based methodology inspired by natural evolution to use computation in solving problems. The GP is first introduced by Koza (1992) [44]. It provides a framework to automatically create a computer program from variety of functions and terminals. To this aim, The GP uses Darwin principle of natural selection to genetically produce a population of computer programs. The process is illustrated in figure 3.2.

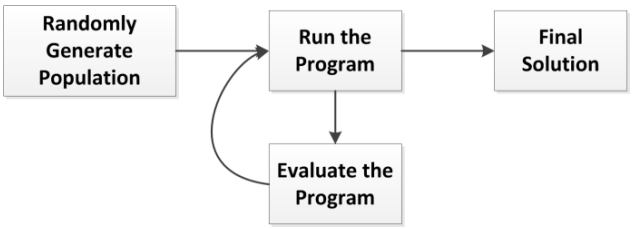
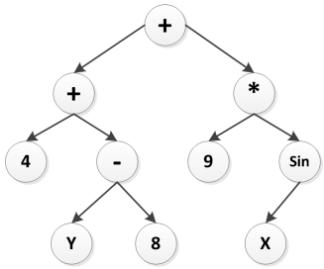


Figure 3.2: Genetic Programing

The GP starts the program by a randomly generated population from computer programs. It creates new generation by automatically breeding the population. Genetic programing use different genetic operation to apply on the selected individuals where they are selected based on

probability of participation and their fitness value. In fact, the process of solving problems in GP is to find the fittest individual.

As it is shown in Figure 3.3 [9], genetic programing represents the functions in tree structure where any mathematical function can be represented in this form. In order to interpret the tree structure, the interpretation starts from the root node located in sub-trees.



$$(\mathbf{4} + (\boldsymbol{Y} - \mathbf{8})) + (\mathbf{9} \times \mathbf{sin}(\boldsymbol{x}))$$

Figure 3.3: Tree Structure Representation

The tree shows a function and the leaves are called terminals. The tree includes nodes and links to connect the node together where nodes show the mathematical operation or any executive operation.

3.2.1. Steps of Genetic Programing

There are 5 major preparatory steps for implementation of Genetic Programing.

1. Set of process variables to get involved in the program
2. The objective function
3. Some criterion to control the pattern recognition of process data
4. Some criterion to terminate the pattern recognition of process data

Figure 3.4 shows the flowchart of Genetic Programing. It specifies the basic sequence of executional steps that are as follow:

1. Randomly create an initial population of individuals composed of functions and terminals.
2. Creating new generation with the following steps:

   2.1. Run the program with each population and assign the corresponding fitness value.

   2.2. Select two pair of individuals to be implemented by genetic operations according to the probability of participation and fitness value.

3. Creating new individual by applying the following operation:

   3.1. Reproduction: Copy the selected individual to the new population.

   3.2. Crossover: Create new individuals by recombining the randomly selected part of individuals.

   3.3. Mutation: Create new individuals by randomly mutating the randomly selected part of individuals.

4. Selecting the best-so-far individual during the run [44].

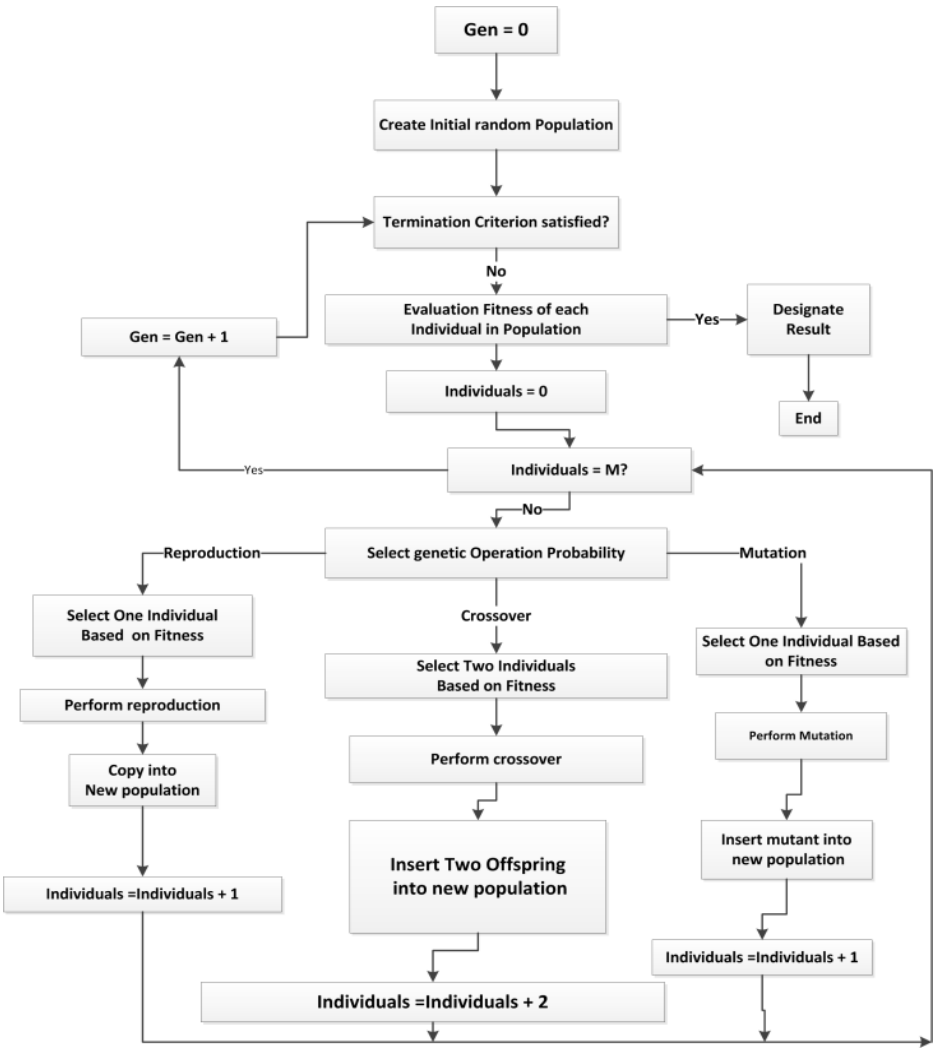


Figure 3.4: Genetic Programing Flowchart

3.2.2. Initialization

Initialization of a population is the first step of the GP. In GP, the initialization is defined as process of randomly creating an initial population of computer programs consist of functions and terminals. Initialization continues till the maximum assigned depth for trees met. In other words, maximum tree depth is a constraint for initialization process. There are 2 main initializing methods called *full method* and *grow method*. There is another method which is a combination of full and grow methods, which is called Ramped Half-and-Half Method.

3.2.2.1. Grow Initialization Method

The search space for the GP to create a population is composed of functions represented by tree structure. A tree is made of terminals and mathematical operations and it continues to grow till meet the maximum tree depth value. Tree starts initially with depth = 0. If the depth of tree is less than the assigned maximum tree size, the nodes are chosen randomly from terminals and internals nodes. Then the grow method has to be called to create child nodes. The algorithm 1 shows how the grow algorithm generate child nodes.

Algorithm 3.1 *node* grow (*depth*): (3.1) [45]

if *depth* < *maximum tree depth*
  node ← *random*($T \cup I$)
  for *i* = *1 to number of children* **of** *node* **do**
    $child_i$ = *grow(depth+1)*
  od
else
  node ← *random*(T)
fi
return *node*

The grow method starts with creating a root node from non-terminal *n1*. Next, the terminal *T1* in created as a child and the other child is an internal node *n2*. The *T1* cannot grow more because it is a termination element however the *n2* grows and the grow procedure repeats for it to produce children. Figure 3.5 shows the grow method in detail.

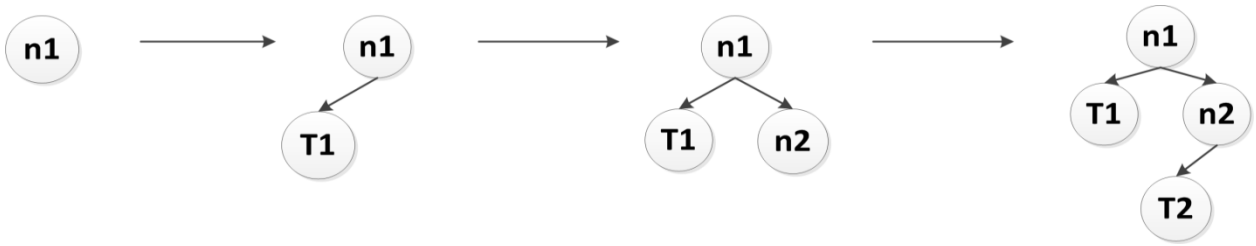


Figure 3.5: Grow Initializing Method

3.2.2.2. Full Initialization Method

The *full initializing method* is another main method of initialization. The full method performs like the grow one with one important difference. In full method, if the tree depth is smaller than the maximum depth of tree, the nodes has to be chosen randomly from nonterminal nodes and terminals are chosen just beyond the depth size. Algorithm 2 shows the detail of full method as follow:

Algorithm 3.2 *node* full (*depth*): (3.2) [45]

```
if depth < maximum tree depth
    node ← random(I)
    for i = 1 to number of children of node do
        child_i = grow(depth+1)
    od
else
    node ← random(T )
fi
return node
```

As figure 3.6 shows, *n1* has been selected as the root node and full method starts to generate children. The *n2* which is internal node is created as the first child and it terminates by *T1*. Same procedure has been done for the second child is *n3* which is again an internal node and terminates by *T2*.

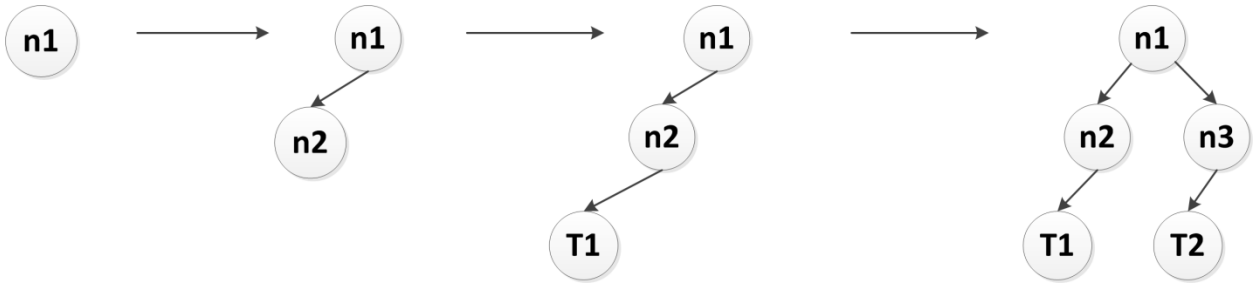


Figure 3.6: Full Initialization Method

3.2.2.3. Ramped Half-and-Half Method

The ramped half-and- half method is a combination of full and grow methods. If *D* is the assigned value for maximum tree depth, the ramped half-and-half method divides the population to *D-1* parts. Each part has the tree depth value of *2 to D.* Finally the resultant trees are mixture of full and grow methods [45].

3.2.3. Selection

Initialization creates the initial population by different methods of initialization. Then, it is the time to select some individuals from the population to breed a new generation. Each individual consists of a pair of manipulated and measure process variables. Selection is a procedure to select individuals according to their fitness values. Fitness value is obtained from fitness function defined for procedure of selection.

3.2.4. Fitness Function

Fitness function is a particular type of objective function that generally measures how much the solution is close to the desired target result. In genetic programing, the fitness function is used to find the best individual between the generated populations. It assigns a fitness value for each individual and the pair which is closer to the set aim; it would be the best-so-far solution.

In GP, for many problems, the fitness measure is calculated by the error created by computer program. The computer program is better that has smaller error and the error tends to zero. The criterion to measure fitness can be any and it differs according to environment.

The general form of an objective function is as follow:

$$\text{Maximize or Minimize } Z = \sum_{i=1}^{n} Ci\ Xi \qquad (3.4)$$

Where C_i = the objective function coefficient corresponding to the ith variable

X_i = the ith decision variables.

The coefficients of the objective function indicate the contribution to the value of the objective function of one unit of the corresponding variable. For example, if the objective function is to maximize the present value of a project, and *Xi* is the ith possible activity in the project, then *ci* (the objective function coefficient corresponding to *Xi*) gives the net present value generated by one unit of activity *i*.

3.2.5. Genetic Operation

In the first step of genetic programing, which is the initialization, the initial population has been created and evaluated by fitness function to find the best individual parents. Since genetic variation is a necessity for the process of evolution, In order to have a diversity of genetic in the process the next step is to create offspring from parents. So operators have to be used to shuffle the genetics and keep the diversity in the population by creating offspring. The operators are called genetic operators that include crossover, mutation, and reproduction.

3.2.5.1. Crossover:

Crossover is one of the basic genetic operators used to create new generation by swapping the chromosomes between the generations. The crossover exchanges the taken gens from parents to create child. The swapping process should be done in crossover point where the point has to be

selected randomly based on probability. The point can be either a node or a link between nodes in parent individuals and swapping the sub-tree between them. Figure 3.7 shows the procedure of crossover operation.

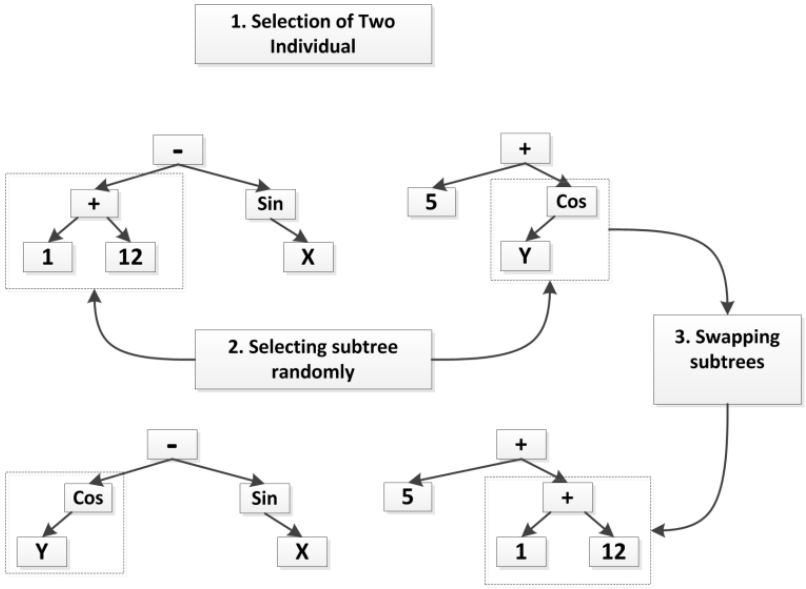


Figure 3.7: Crossover Operator Mates the Parents

3.2.5.2. Mutation:

Mutation is another basic genetic operator. It is used to generate new population from the initial one. The mutation consists of a randomly mutation point in the tree and moving some sub-tree from their initial state. In the process of evolution, mutation selects sub-trees according to their probabilities. The procedure of mutation operation is shown in Figure 3.8.

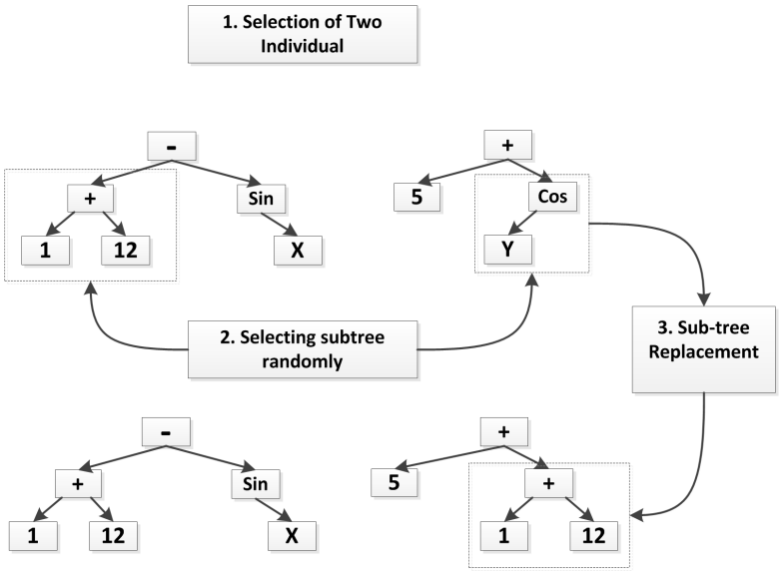


Figure 3.8: Mutation Operation

3.2.5.3. Reproduction

The third genetic operation is reproduction that can be considered as the easiest one. The reproduction operator acts in simple way where it just copies selected individuals to the new generation. Like other genetic operators, the process of individual selection in the reproduction is done according to probability [46].

3.2.6. GP Applications in Fault Propagation

The purpose of using GP in this study is identifying the relationships among process variables. As mentioned in this section, GP starts the analysis process by creating individuals where each individual includes manipulated and measured process variables which are input and output respectively. First, it selects the best individuals according to assigned fitness function then it identifies the inter-relation patterns between input and output. In other word, GP formulates relationships among process variables. The results of GP application in this research in discussed in chapter 5 in detail.

3.3. Artificial Neural Network (ANN)

As it is mentioned in the literatures, artificial intelligent methods are under many studies for the purpose of pattern recognition and fault propagation analysis. Artificial Neural Network (ANN) as one of these methods is used by many researchers [20-26]. As GP, ANN is also a well-known technique for modeling of complex systems behavior and fault propagation. Therefore these characteristics of ANN can meet the requirement of this study in identifying the inter-relation pattern among process variable.

ANN refers to artificial neural network which consists of processing elements, inter-connecting artificial neurons, links between neurons with assigned values called weights. In artificial neural network, all elements operate in parallel.

In the architecture of neural network, neurons create a brain-like structure to represent mathematical model of a system. They are constructed to mimic the properties of biological neurons. In summery neural network is a computational system inspired processing and learning ability of biological brain. Figure 3.9 shows an overall view of neural network.

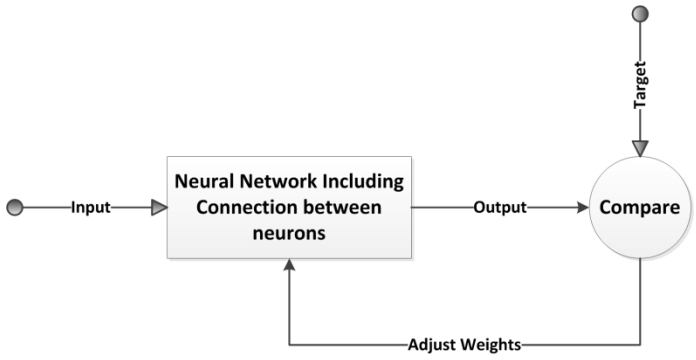


Figure 3.9: Overall View of Neural Network

3.3.1. Basics of Artificial Neural Network

An artificial neural network is a biological inspired of computational model composed of layers of neurons. Layers of neurons consists of input layer, hidden layer and output layer are connected together to build the whole structure of neural network. A brief description of ANN components are as follow.

- Input connections that are $y1, y2, \dots, yn$ and they are linked through connections with weights $w1, w2, \dots, wn$.
- Input function $u = f(y, w)$ which is a function of $y\ and\ w$ that are inputs and weights respectively.
- Activation function calculates the activation level of a neuron.
- Output function calculates the output value of a neuron.

A neuron is a function of the output vector $(y1, \dots, yk)$ and the output can be derived from the formula 3.5 where f is the function typically the sigmoid function [47].

$$f(xj) = f\left[aj + \sum_{i=1}^{k} wij\ yj\right] \qquad (3.5) \qquad [47]$$

First mathematical model of a neuron is proposed by McCulloch and Pits at 1943 [57]. Figure 3.10 represents a mathematical model of an artificial neuron where,

$$xj = \sum_{i=1}^{k} wij\ yi \qquad (3.6) \qquad [57]$$

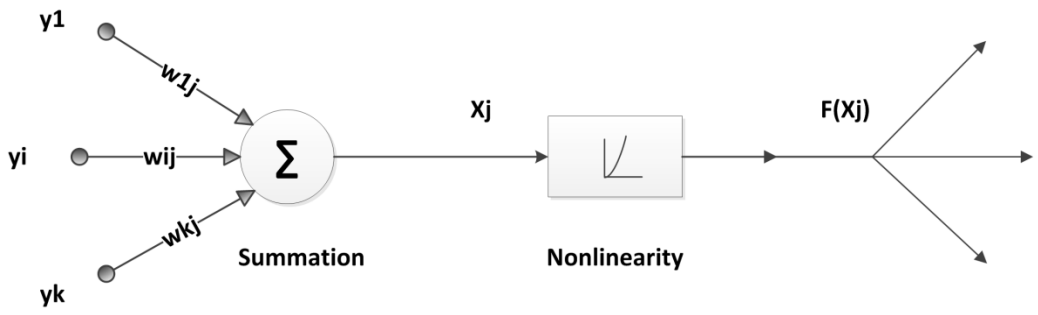


Figure 3.10: Mathematical Model of an Artificial Neural Network

As it is shown in Figure 3.10, artificial neural network receives inputs with their assigned weights. These inputs with their weights have to be summed and transferred to produce output.

3.3.2. ANN Architecture

As mentioned at the beginning of the chapter, neural network composed of elements made a brain-like structure. Basically the structure of neural network is a multi-layer structure. It consists of one input layer, many hidden layers located between input and output layer, and one output layer. There are several types of neural network architectures described briefly as follow.

- **Feed Forward Architecture:** In the feed forward architecture, the flow of information is from input neurons toward output ones. In other word, there is no backward direction and neurons don't have any information from previous ones.
- **Recurrent Architecture:** Recurrent or feedback architecture has same structure as feed forward. However there are connections from output neurons toward input ones. In other words, information flows in both direction of forward and backward.

3.3.3. Learning Modes of ANN

Although ANN has promising features, however the most interesting feature of ANN is its learning capability. Learning capability makes ANN able to adapt itself with any environment. The network can be trained according to input data in a way to produce desired output. The learning capability of ANN is possible through some learning algorithm. Learning algorithms are mainly categorized in three categories.

1. **Supervised Learning:** Procedure of supervised learning includes input and desired output. Network is trained according to both input and its corresponding output. In fact training process continues until network associate input to its corresponding output.
2. **Unsupervised Learning:** Unlike the supervised learning in this type of learning, algorithm just includes input. All inputs have to be supplied to network so it would be able to be trained to identify the system structure.
3. **Reinforcement Learning:** This type of learning is a combination of supervised and unsupervised learning. Inputs are supplied to the network and it is trained with a look at output calculated by neural network [48].

3.3.4. The NARX Network

As mentioned in previous part of this chapter, in recurrent structure of neural network, the application of feedback can have variety of forms. It can be either from output neurons to input ones called global feedback or it can be output to hidden layers [49].

Mapping input to its corresponding output in one of the applications of the recurrent network. In order to implement the mapping technique, input-output recurrent network has been used.

Figure 3.11 shows the structure of a nonlinear autoregressive with exogenous inputs (NARX) model. It is a multiple input, single output model where the input $u(n)$ is applied to the system through tapped-delay-line memory of q units. The output $y(n+1)$ is also fed to a tapped-delay-line memory of q units. These memories are used to fed the input of the multilayer perception.

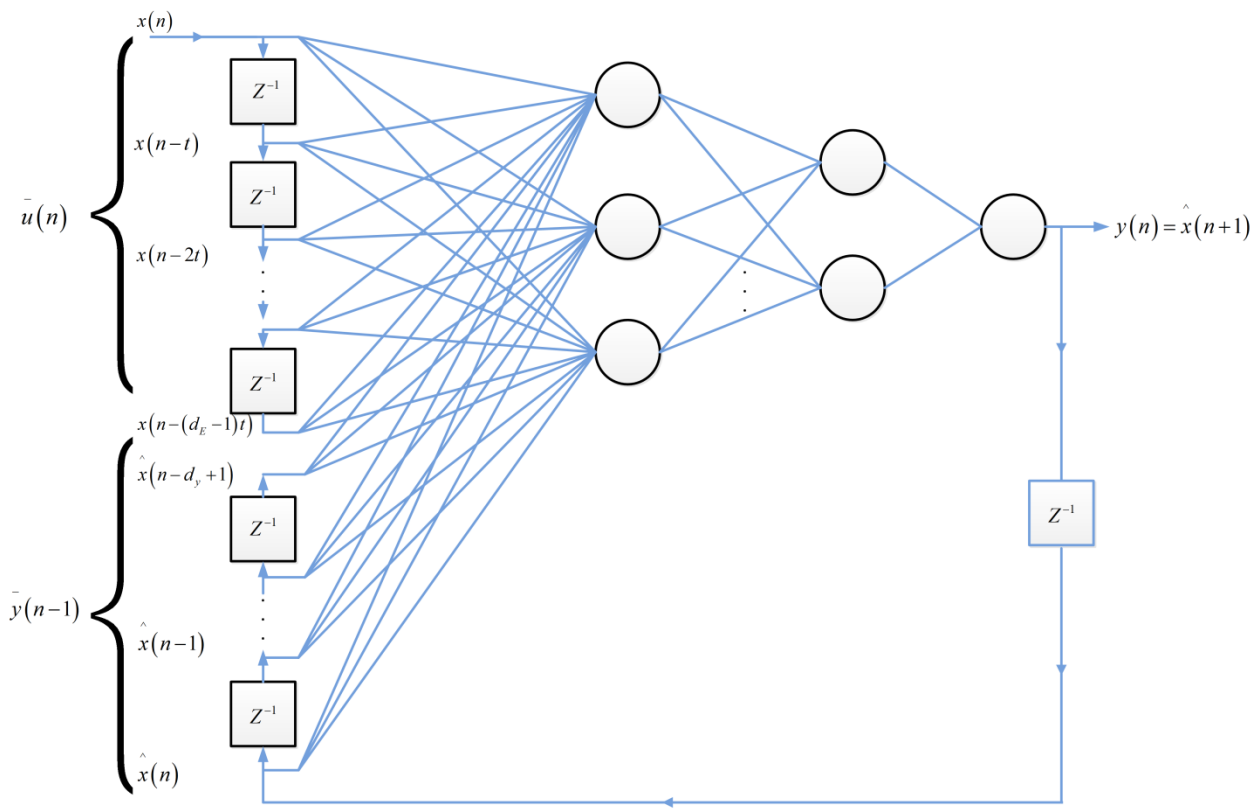


Figure 3.11: Nonlinear Autoregressive Exogenous Inputs Model (NARX)

As shown in the Figure 3.11, the output $y(n+1)$ is one step ahead of input $u(n)$ by one unit of time. The other vectors applied to input layer of multilayer perception are as follow.

- Inputs that include $u(n), u(n-1), \dots, u(n-q+1)$ represent exogenous inputs come from outside of the network
- Outputs that include$y(n), y(n-1), \dots, y(n-q+1)$.

Given this, the dynamic behaviour of NARX is as follow.

$$y(n+1) = F(y(n), \dots, y(n-q+1), u(n), \dots, u(n-q+1)) \qquad (3.7)$$

Where F is a nonlinear function and $y(n)$ and $u(n)$ denote output and input respectively.

The nonlinear mapping $f(.)$ is an unknown function and can be approximated by a standard multilayer Perceptron (MLP) network. The resulting connectionist architecture is then called a NARX network. Training the NARX network, it can be carried out in one out of two modes:

- **Series-Parallel (SP) Mode:** In this case, the output's regressor is formed only by actual values of the system's output.
- **Parallel (P) Mode:** In this case, estimated outputs are fed back and included in the output's regressor.

3.3.5. ANN-NARX Application in Fault Propagation

NARX is a special mode of neural network and as it is shown in Figure 3.11, it consists of multiple inputs and one output. For the application of NARX for this study, the manipulated variables are inputs and measured one is output. According to the module described for NARX, it identifies the inter-relations patterns among inputs and the output which are manipulated and measured process variables respectively. Relationships among process variables are shown by weights. The results of NARX application are described in chapter 5 in detail.

3.4. Fault Semantic Network (FSN)

This section is going to introduce a computer technique which is a tool of representing knowledge based on the relationship between variables called Fault semantic network (FSN). Originally, FSN described by Gabber at 2007 [50] is a graphical model used for reasoning under uncertainty. The FSN consists of nodes and arcs where nodes represent variables and arcs show the connection between variables. Connections represent qualitative or quantitative relationship between nodes. FSN has the ability to represent the relation between variables both quantitatively and qualitatively. In addition it is able to dynamically change according to different pre-defined scenarios and their probabilities.

The theory behind the FSN is discussed briefly in this part and explained that how FSN can create a semantic network of process variables to interpret and model behaviour of any system dynamically. The process we are focusing on is a part of chemical process, so the nodes are representing process variables such as, temperature, pressure, flow rate and etc.

3.4.1. Semantic Network

Semantic network is a network structure that represents relations between concepts. The earliest attempt has been done by Collins and Quillian at 1969 [51] when they introduce semantic network in a tree structure (directed or undirected graph) consists of nodes and arcs where nodes represent concepts and connections show relations between nodes. Figure 3.12 shows a tree-structured semantic network.

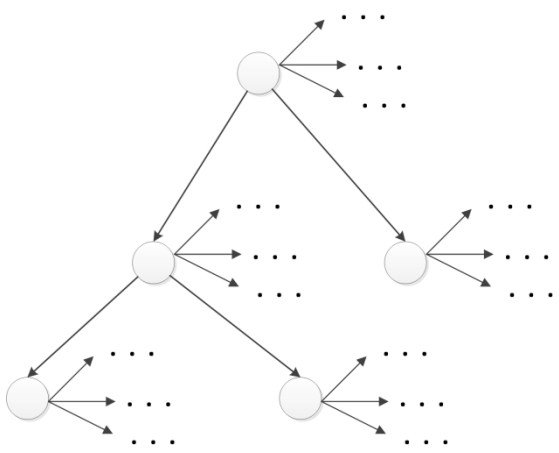


Figure 3.12: Tree-Structured of Fault Semantic Network

Semantic network models systems in order to represent different concepts such as, statistical, taxonomic and industrial and specify relations between them. Specifying relations between concepts is possible by implementing a special set of procedure that performs reasoning between them.

Semantic network could get success and gain more attention in different industries because of its wide application such as fault recognition, root cause analysis and etc. Fault semantic network is one of the promising applications of semantic network to find the causes and consequences which is applicable in variety of industrial processes.

3.4.2. Basics of Fault Semantic Network (FSN)

Fault Semantic Network is a graphical model that represents a system as a semantic network in an uncertain domain. The FSN is composed of nodes and arcs where nodes are selected from a set of random variables represented by$\{\boldsymbol{X} = X1, \dots, Xi, \dots Xn\}$ and directed arcs that represent connections between two different nodes, $Xi \rightarrow Xj$. In other words, the directed arc shows the

relation from node Xi to node Xj and it is not possible to return to node Xi by following the same directed arc. In the FSN, the strength of the relationship between variables is assigned by different reasoning methods that will be discussed in next parts.

Implementing the FSN for a process with lots of process variables is not an easy job especially when you are asked to do it in real world. There are steps that should be undertaken by engineers and researchers to let them create a dynamic modelling of a process as a semantic network that contains all the possible faults and possible relationships between variables. The steps are going to discuss in this chapter briefly. In order to describe clearly application of semantic networks in real industries, Vibration in a pump has been used as an example.

Example: Analysis of technical inspector in a refinery records some vibration in a pump that pumps water from a separator in down-stream to an up-stream tank. There are different causes that can results in pump vibration. According to relevant information and historical data it is possible that pressure increase and existence of vapor in pump are 2main causes of vibration in the pump. Finally vibration test certainly shows existence of vibration in pump.

3.4.3. Nodes and process variables

The first thing that has to be considered for the theory behind nodes in a semantic network is to answer the following question:

- What variables are nodes representing?
- What values should be assigned for each node?
- What is the state of each node?

The value assigned for each node may change as the semantic network in changing dynamically, however a node should take exactly one value at a time. Nodes are categorized in different types according to their assigned value. Different types of node are as follow:

- **Boolean Nodes:** This type of nodes taking binary values of true (T) or false (F). For instance the node that is showing vibration represents the vibration fault in pump.
- **Ordered Values:** In this type, the node is given some qualitative values such as, low, medium and high. For example, the node represent pressure can get any values of medium.
- **Integral Values:** It is a quantitative value given to a node. For example the node representing age of the pump can take any values between 1 to 10 years. Values should be chosen in a way to represent the domain effectively.

In order to make the example clearer, the nodes name, type of nodes and their values are shown in Table 3.1.

Table 3.2: Nodes Specification for the Pump Vibration Example

Node Name	Node Type	Node Value
Pressure	Binary	$\{Low, High\}$
Pump Age	Integral	$\{1 - 10\ years\}$
Vapor Inside	Boolean	$\{T, F\}$
Vibration	Boolean	$\{T, F\}$
Vibration Test	Boolean	$\{T, F\}$

The provided example is an example just to show early stages to making nodes in a semantic network. The example provides limited number of nodes, for instance it just cover vibration fault in a pump however there are other fault that should be covered in the semantic network to be complete. In addition, the assigned values should be exact and cover all ranges. For instance the pressure may have a value between low and high that is necessary to be considered [52, 60].

3.4.4. Structure of Fault Semantic Network

The structure of a semantic network should represent relations between variables qualitatively and quantitatively. In the FSN, nodes correspond to different faults/causes/consequences and directed arcs are links between them describe the dependencies. Initially, FSN is constructed based on ontology structure of fault models on the basis of process object oriented model (POOM) where failure mode (FM) is described using symptoms, enablers, process variables, causes, and consequences. Considering the example provided at the beginning of this section, the fault is vibration in a pump. The next thing that should be considered is factors that affect pump vibration that are pressure and existing vapor inside the pump. So directed arcs has to be added from nodes *pressure* and *vapor inside* toward the node *vibration*. Same procedure happens between the node *vibration* and next nodes. Since vibration in pump results in having vibration test, so a directed arc has to be drawn from node *vibration* to the node *vibration test*. Figure 3.13 shows the complete structure of the FSN for the pump vibration example.

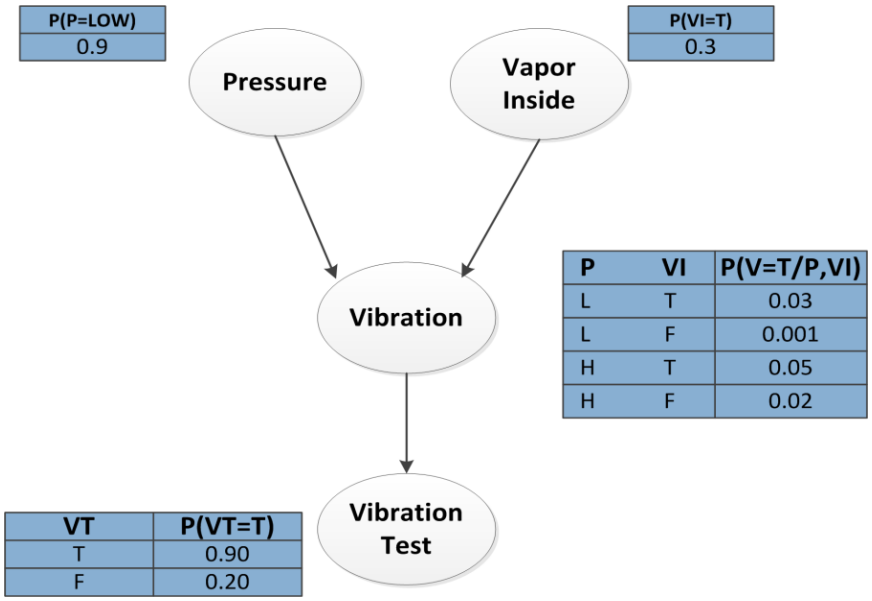


Figure 3.13: A FSN for the Vibration Fault in Pump

Since the structure of semantic network is similar to tree, they have common terminology. A node that doesn't have parent called *root node* and the node that has no child called *leaf*. Any other nodes in between called an *intermediate node*. Given this, in a semantic network the root node represent the root cause and the leaf represents consequence. Considering the given example, *pressure* and *vapor inside* can be considered as root causes.

3.4.5. Relationships between Nodes

After the brief introduction about semantic network, its structure and its components, in this section the connections between nodes will be described. This can be done through different approaches, such as assigning probability distribution for each node.

- **Probabilistic Approach:** After setting up the structure of semantic network, next step is to consider all the possible combination of a node with its neighbors both as parent and child. Then a probability value has to be assigned for each combination separately. For instance, consider the *pump vibration* node. The *pressure* and *vapor inside* nodes have four possible combinations of $\{H-T, H-F, L-T, L-F\}$ and each one has its own probability value$\{0.05, 0.02, 0.03, 0.001\}$.

- **Fuzzy Approach:** In FSN, nodes correspond to different faults/causes/consequences and connections between them describe the dependencies. In fuzzy approach, Rules are associated with each transition of the causation model within FSN. In other word, relations between variables are described by specifying some *if-then-else* statement. For example, vibration in pump might be associated with rules such as (*IF* Pressure = Low *AND* Vapor inside = True *THEN* FM = Pump vibration). These rules are initially defined in generic form based on domain knowledge, i.e. regardless of plant specific knowledge and then explained for plant specific.

- **Mathematical Formulation and Activation Function:** in this type of reasoning, relations between concepts are specified by mathematical equations. Mathematical equations are composed mathematical functions such as, Sin, Logarithmic, and arithmetic functions like, subtraction, division and etc. For example the relation between pressure and pump vibration can be formulated as,$y = [(x + cosx) - \cos(x \times x)] \times x$ where x stand for pressure and y represent vibration in pump. The formula which is a sample of genetic programing result shows the relation between pressure, as a cause, and pump vibration, as consequence, and the link between them.

3.4.6. Reasoning in Fault Semantic Network

Fault semantic network has the ability to change dynamically. It means that as the system is running, semantic network is changing as well and new data are replacing with old one. In order to get proper result from semantic network, it has to reason in a way to follow the changes in system appropriately. It is necessary to reason the network so that it can be dynamically adapted according to changes in operational conditions [52].

Fault semantic network maps process variables and reason their qualitative and quantitative relationships in systematic way. There are different types of reasoning that help engineers and researchers in process industry. Figure 3.14 shows four types of reasoning.

- **Diagnostic Reasoning:** In this type, reasoning starts from symptoms and results in causes so the reasoning direction is opposite to the arc direction. In the pump example, vibration test results can be considered as a symptom. Having positive result of vibration test shows existence of vibration in pump. In other word, it is possible to recognize the fault and then see whether pressure and vapor inside can be causes.
- **Predictive Reasoning:** This type of reasoning predicts faults in advance even without assuming symptoms. For instance increasing pressure in down-stream unit of the pump may not be a cause of pump vibration; however it increases probability of occurrence of vibration. In this case, engineers should expect more symptoms such as vibration test results.
- **Inter-causal Reasoning:** This type of reasoning is used when multiple causes result in one consequence. For instance, both pressure and existence of vapor in pump can result in one consequence which is vibration in pump.

- **Combined Reasoning:** Sometimes by using the three mentioned type of reasoning you can't end up with satisfactory results. So it is needed to combine the above reasoning methods to get acceptable results. Figure 3.14(d) shows a combination of predictive and diagnostic reasoning.

What are discussed and what are going to be discussed:

In this chapter the methodology behind this work includes:

- *Data processing (Wavelet Technique)*
- *Pattern recognition techniques (GP & NARX)*
- *Fault semantic network*

are discussed in detail.

Chapter 4 is going to describe the chemical process plant selected for the purpose of data extraction and identifying faults and hazard associated with the applied deviations and disturbances.

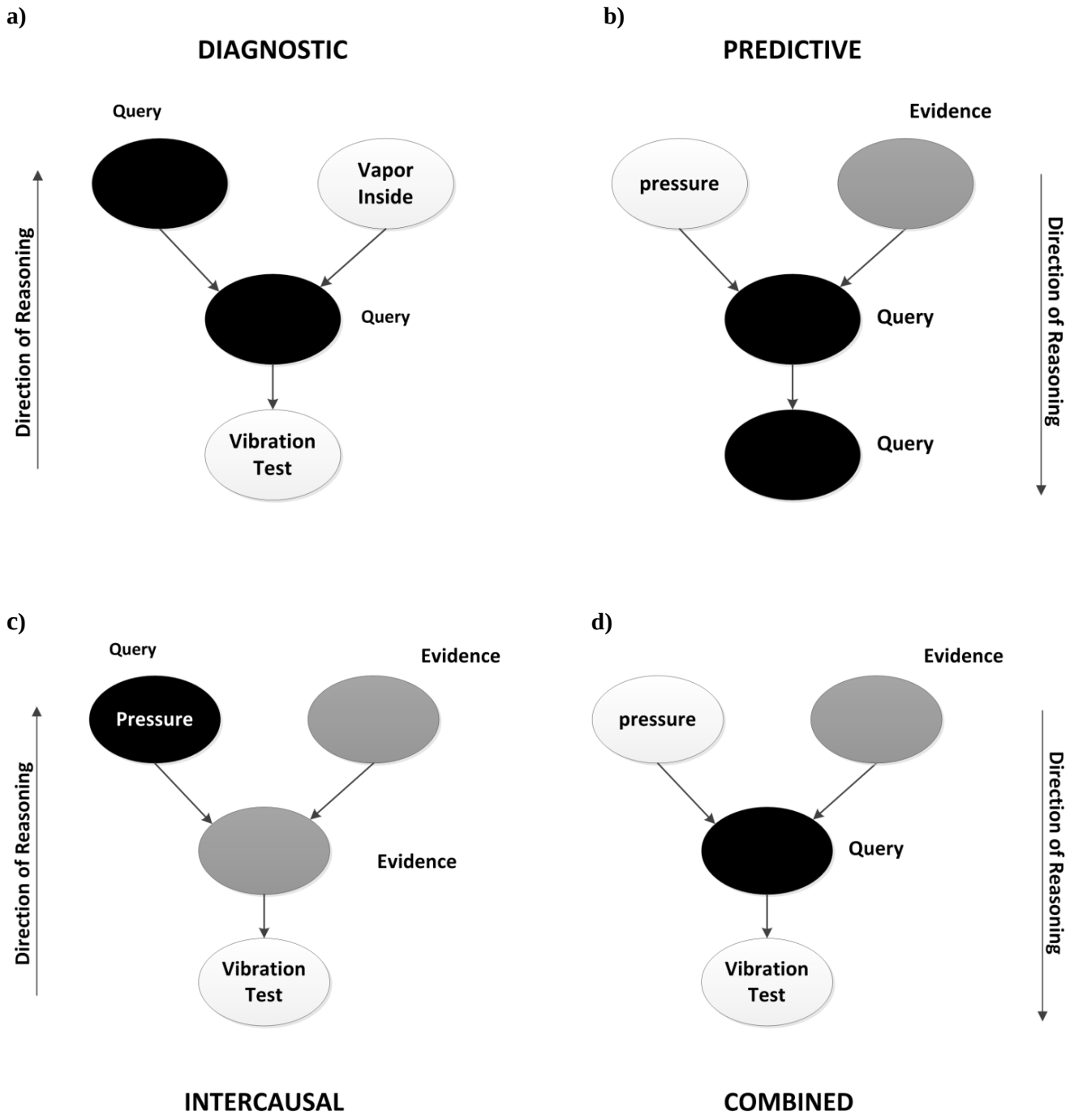


Figure 3.14: Types of Reasoning. Figure 3.14(a) diagnostic reasoning. Figure 3.14(b) shows predictive reasoning. Figure3.14(c) shows inter-causal reasoning and Figure 3.14(d) shows the combined reasoning [52].

Chapter 4

Chemical Plant Process Simulation

Chapter Summery

Chapter 4 first, discusses the necessity of using simulation in different industries. Then it describes the selected chemical process plant, manipulated (independent) and measured (dependant) process variables, designed deviations and disturbances for the purpose of this research in detail. Deviations are considered as manipulating dependant variables.Despite the deviations, some disturbances are also applied to affect the operation of process plant. In summary, there are some deviations that applied to the process under four certain disturbed operational conditions.Two softwares of ***AspenHysys*** *and* ***MATLAB*** *are used to simulate chemical processes in the study.*

4.1. Why Simulation?

Over the yerars, the process involved in industries considerably changed and incearisingly get more complex. Processes and operations that were controlled by operatore before, now it is getting very difficult to control them by conventional control strategoies. It is getting more serious in industries that are dealing with reactions especially the chemical ones, hazardous materials and more complex processes.

Generally, it is easy and reasonable for simple processes to take the risk of implementing different control strategies, changes in process variable on real plant to chesk feed back of system. However, when it comes to a complex process like a chemicla process that dealis with hazardous material, it is a risky option to implement different tests, changes and strategies on real plant. Nowadays, with the developed technology in simulation and control strategies, it is highly recommended to use simulation for this aim.

It is very important to Creat a model of a system or simulate a process in advance. The reason for this are:

- To protect the human, equipment and environment from unsafe condition.
- To be able to test and analyze system or process in a short period of time.

Using modeling and simulation gives operators and engineers the opportunity to analyze the process from different points of view such as, changing process variables in the process, testing the tolerances of equipments, checking the margines for changes, implementing variaty of control strategy in the process to see the system feedback. In other words, simulation environment givesresearchers and operators the possibility of implementing all the options you

can do in real industrial environment without being worried about hazards and unsafe operational condition. For instance in a simulation of plant shutdown, there is no potential of damage or or accident but it helps designers to dsign the process more efficient and safe.

Different tools including simulation software and modelling tools has been introduced and developed to model variaty of processes and systems in different industries. Matlab, Aspen are two powerful software used for this purposes in oil and gas industries.

4.2. Tennessee Eastman (TE) Process

One important modeling tools proposed for the researchers and studies is launched by Eastman chemical company at 1990. The process modeled by the eastman chemical company is a defined chemical process alled The Tennessee Eastman (TE) process.

The Tennessee Eastman (TE) process is a designed chemical process that has been introduced by Downs and Vogel in early 1990 [53]. Afterward, it became as a standard process to be used by industries and researchers in different experiments with different purposes such as comparison of control strategies, fault detection and etc. Since the TE process represents a real chemical process with real faults, many authors have used the TE process as a valid simulated process and published their results based on that [54, 55]. Number of parameters, disturbances, complexities and control strategies implemented in the TE Process make it more similar to a real process and reliable to be applicable in different perspectives.

The TE process is a well design process to model and analyze a nonlinear and open-loop unstable chemical process from different point of views. The process consists of five major units of operation including, a vapor-liquid separator, a compressor,two reactors, a condenser and a product stripping column. The TE diagram and 8 components are shown in Table 4.1 [53]. The

nonlinearity characteristic of the process in mainly because of chemical reactions in the reactors. Figur 4.1 shows a schematic view of the TE process.

The process produces two products and two by-products from four reactions as follow:

$A(g) + C(g) + D(g) \rightarrow G(liq)$, Product 1, (1)

$A(g) + C(g) + E(g) \rightarrow H(liq)$, Product 2, (2)

$A(g) + E(g) \rightarrow F(liq)$, By-product, (3)

$3D(g) \rightarrow 2F(liq)$, By-product (4)

Where in the process there are four feed streams of A, D, E, and C, one product stream, and one purge stream. Almost all of the inert (B) enters in the largest feed (C), which actually contains almost 50% of component [54]. Table 4.1 shows the component physical properties.

Table4.1: Physical Properties of Components

Component	Molecular Weight	Liquid Density (kgm^{-3})	Liquid Heat Capasity ($kj\ kg^{-1}\ C^{-1}$)	Vapor Heat Capacity ($kj\ kg^{-1}\ C^{-1}$)	Heat of Vaporization ($kj\ kg^{-1}$)
A	2.0	----	----	14.6	----
B	25.4	----	----	2.04	----
C	28.0	----	----	1.05	----
D	32.0	299	7.66	1.85	202
E	46.0	365	4.17	1.87	372
F	48.0	328	4.45	2.02	372
G	62.0	612	2.55	0.712	523
H	76.0	617	2.45	0.628	486

The reactant A, D and E directly enter the reactor while the reactant C firstly enter the product stripper and then through a rcycle steam process reaches the reactor. Products of the reactor exit from the upper exit of the reactor and enter the condesor for condensation process. The condenser converts vapor to liquid and then the condensor products passed through the vapor-liquid separator. The separator has to separate vapor from liquid where the liquid which is heavier exits from down and the vapor which is lighter exits from up. The liquified product form separator trasfer to product stripper to purified and the vapor product transfer to compressor to recycle reactor as feed or exit in purge stream.

The TE process produces two main products of G and H through four reactions. All reactions are irreversible and exothermic. The reaction rates are a function of temperature through the Arrhenius expression:

$$k = \alpha \exp(-\frac{E}{RT}) \qquad (4.1) \qquad [53]$$

Where:

$K = Reaction\ rate\ constant$

$\alpha = Constant$

$E = Activation\ energy$

$T = Absolute\ temperature$

$R = Universal\ gas\ constant$

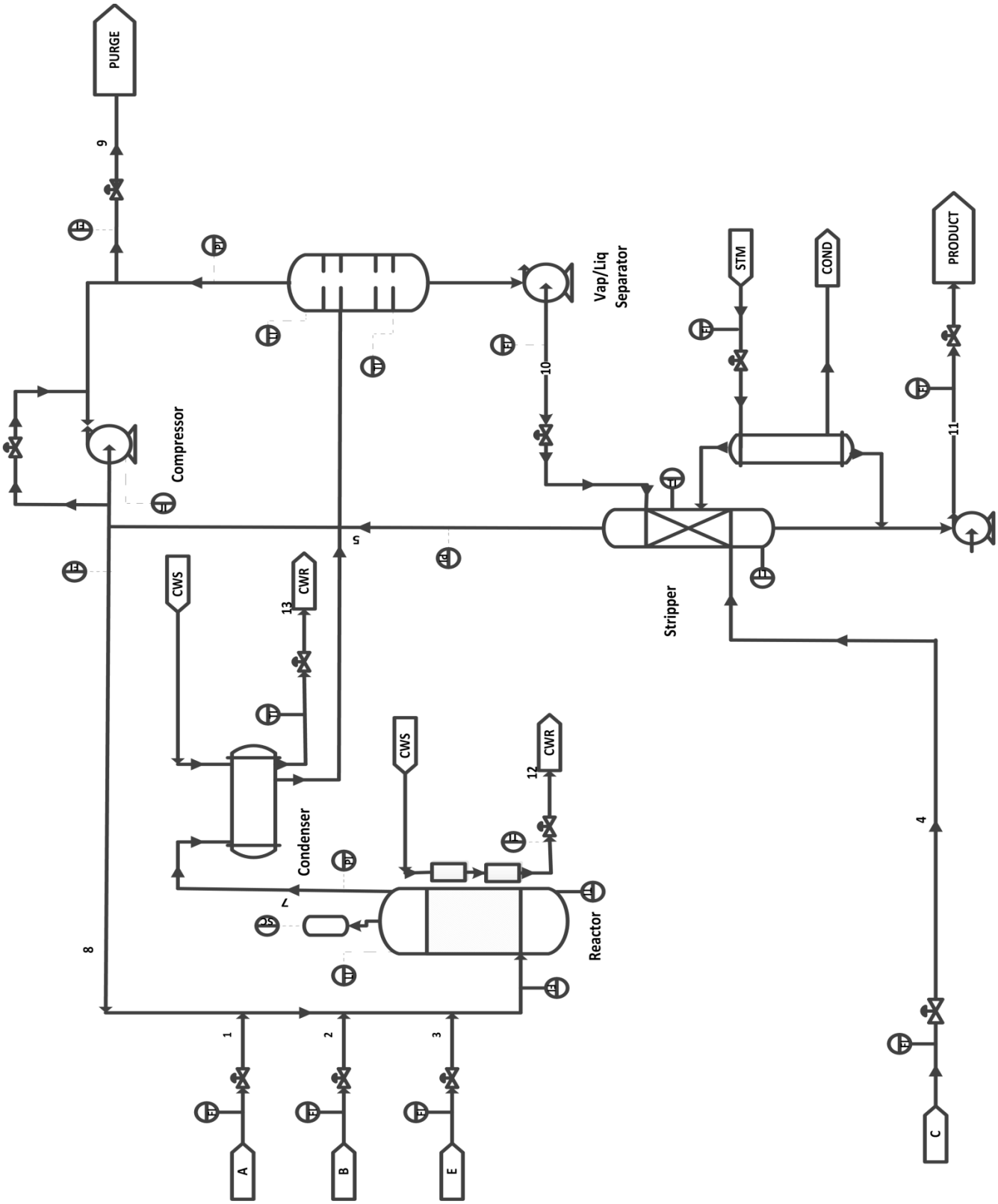


Figure 4.1: Tennessee Eastman (TE) process [54]

There are six different operating mode defined for the TE process according to the plant production rate and ratio between the two main products of G and H. The six operating modes are listed in Table 4.2.

Table 4.2: Plant Operating Modes

Modes	G/H Mass Ratio	Production Rate (Stream 11)
1	50/50	7038 kg hr^{-1} G – 7038 kg hr^{-1} H
2	10/90	1408 kg hr^{-1} G – 12699 kg hr^{-1} H
3	90/10	10000 kg hr^{-1} G – 1111 kg hr^{-1} H
4	50/50	Maximum Production Rate
5	10/90	Maximum Production Rate
6	90/10	Maximum Production Rate

4.2.1. Manipulated and Measured Variables

As discuused before, the TE process gradually gained more attention and became as a reference process to be used by engineers and researchers for different purposes and appliocations. Features like, variaty of process variables, disturbances, complexities and control strategies implemented in the TE Process make it realistic.

The TE process has 41 measured and 12 manipulated variables where manipulated variables are considered as independent and measured ones as dependent. In order to see how manipulated variables affect the measured ones, they are changing either randomly or by steps under normal and disturbed operational condition. The manipulated and measured variables are shown in Table 4.3 and Table 4.4 respectively.

Table 4.3: Process manipulated variable

Variable Name	Variable Number	Base case Value (%)	Low Limit	High Limit	Units
D Feed Flow (stream 2)	XMV (1)	63.053	0	5811	Kgh^{-1}
E Feed Flow (stream 3)	XMV (2)	53.980	0	8354	Kgh^{-1}
A Feed Flow (Stream 1)	XMV (3)	24.644	0	1.017	kscmh
A and C Feed Flow (stream 4)	XMV (4)	61.302	0	15.25	kscmh
Compressor Recycle Valve	XMV (5)	22.210	0	100	%
Purge valve (Stream 9)	XMV (6)	40.064	0	100	%
Separator Pot Liquid Flow (stream 10)	XMV (7)	38.100	0	65.71	m^3h^{-1}
Stripper Liquid Product Flow (stream 11)	XMV (8)	46.534	0	49.10	m^3h^{-1}
Stripper Steam Valve	XMV (9)	47.446	0	100	%
Reactor Cooling Water Flow	XMV (10)	41.106	0	227.1	m^3h^{-1}
Condenser Cooling Water Flow	XMV (11)	18.114	0	272.6	m^3h^{-1}
Agitatopr Speed	XMV (12)	50.000	150	250	rpm

Table 4.4: Measured variables

Variable Name	Variable Number	Base Case Value	Units
A feed (stream I)	XMEAS (1)	0.25052	kscmh
D feed (stream 2)	XMEAS (2)	3664.0	Kgh^{-1}
E feed (stream 3)	XMEAS (3)	4509.3	Kgh^{-1}
A and C feed (stream 4)	XMEAS (4)	9.3477	kscmh
Recycle flow (stream 8)	XMEAS (5)	26.902	kscmh
Reactor feed rate (stream 6)	XMEAS (6)	42.339	kscmh
Reactor pressure	XMEAS (7)	2705.0	Kpa gauge
Reactor level	XMEAS (8)	75.000	%
Reactor temperature	XMEAS (9)	120.40	C
Purge rate (strc:ant 9)	XMEAS (10)	0.33712	kscmh
Product separator tetnperature	XMEAS (11)	80.109	C
Product separator level	XMEAS (12)	50.000	%
Product separator pressure	XMEAS (13)	2633.7	Kpa gauge
Product separator underflow (stream 10)	XMEAS (14)	25.160	m^3h^{-1}
Stripper level	XMEAS (15)	50.000	%
Stripper pressure	XMEAS (16)	3102.2	Kpa gauge
Stripper underflow (stream I I)	XMEAS (17)	22.949	m^3h^{-1}
Stripper temperature	XMEAS (18)	65.731	C
Stripper steam flow	XMEAS (19)	230.31	Kgh^{-1}
Compressor work	XMEAS (20)	341.43	kw
Reactor cooling water outlet temperature	XMEAS (21)	94.599	C
Separator cooling water outlet temperature	XMEAS (22)	77.297	C

4.2.2. Disturbances

Twenty disturbances named by IDV (1) to IDV (20) have been simulated to be implemented in different probability of occurrence in the TE process for generating sets of data for different purposes. The disturbances are chosen in a way to cover all the operational aspects of TE process. A brief description of modelled disturbances are mentioned here also Table 4.5 shows the disturbances and their type of changes.

- **Disturbance 1:** It is caused by an increase of component C in stream 4, then the component A decreases and the component B is constant. Disturbance 1 causes impalnce in reactor feed in the reactor and as a result, reactor pressure increases.
- **Disturbance 2:** It is same as disturbance 1 with the same A/C feed ratio but in the disturbance 2, there is an increase in component B in Feed C.
- **Disturbance 3:** The temperature of feed D increases that causes increase intemperature and pressure of the reactor.
- **Disturbance 4:** Same as disturbance 3, it is a step change increasing of in the reactor cooling water temperature that causes an increase in reactor inlet temperature.
- **Disturbance 5:** Same as disturbance 3, it is a step change increasing of in the condenser cooling water inlet temperature.
- **Disturbance 6:** It is because of loss of feed A which causes imbalace of chemical composition in reactor feed and finnaly as a result the plant shuts down.
- **Disturbance 7:** This disturbance in because of loass of C head pressure and adecrease in reactor pressure is the effect.
- **Disturbance 8:** It happens when there is a random variation on compositions in the reactor feed stream which causes imbalance in components inside the reactor.
- **Disturbance 9:** This disturbance is because of random variation of temperature in feed D.

- **Disturbance 10:** Same as disturbance 9, it is because of random variation of temperature in feed C.
- **Disturbance 11:** It is because of random variation of reactor cooling water inlet temperature that causes temperature variation in the reactor.
- **Disturbance 12:** Rapid fluctuation in condenser cooling water inlet temperature which results in a decrease in condenser output temperature.
- **Disturbance 13:** It is due to imbalances in reactor kinetics. It may be slower than normal operation that affects the reaction products.
- **Disturbance 14:** Stickiness in the reactor cooling water valve is the cause disturbance 14. Fluctuation in flow rate, pressure and temperature are the consequences of sticky valve.
- **Disturbance 15:** Stickiness in the condenser cooling water valve is the cause disturbance 15. Fluctuation in flow rate, pressure and temperature are the consequences of sticky valve.
- **Disturbance 16:** Although Downs and Vogel didn't explain the details of disturbances 16 to disturbance 20 but other references [56] shows that It is because of slow drift in temperature of the utility stream.
- **Disturbance 17:** According to [56], it is a sinusoidal variation of utility stream's temperature. As a result it is expected to have sinusiodal variation in other process variables such as reactor pressure and temperature.
- **Disturbance 18:** It is probably due to extra noises during the process of sampling measurements values.
- **Disturbance 19:** Probably existance of noise in striiper is the cause of the disturbance 19. Fluctioation in production rate is expected as a consequence.
- **Disturbance 20:** It is blockage in condenser tube and it cosequently causes fluctuation in the temperture of cooling water [56].

Table 4.5: Process Disturbances

Process Variable	Variable Number	Type
A/C feed ratio, B composition constant (stream 4)	IDV (1)	Step
B composition, A/C ratio constant (stream 4)	IDV (2)	Step
D feed temperature (stream 2)	IDV (3)	Step
Reactor cooling water inlet temperature	IDV (4)	Step
Condenser cooling water inlet temperature	IDV (5)	Step
A feed loss (stream I)	IDV (6)	Step
C header pressure loss--reduced availability (stream 4)	IDV (7)	Step
A, B, C feed composition (stream 4)	IDV (8)	Random variation
D feed temperature (stream 2)	IDV (9)	Random variation
C feed temperature (stream 4)	IDV (10)	Random variation
Reactor cooling water inlet temperature	IDV (11)	Random variation
Condenser cooling water inlet temperature	IDV (12)	Random variation
Reaction kinetics	IDV (13)	Slow drift
Reactor cooling water valve	IDV (14)	Sticking
Condenser cooling water valve	IDV (15)	Sticking
Unknown	IDV (16)	Unknown
Unknown	IDV (17)	Unknown
Unknown	IDV (18)	Unknown
Unknown	IDV (19)	Unknown
Unknown	IDV (20)	Unknown

4.3. Process Data Simulation

In order to implement fault semantic network on a process, the first step is generate data from the process. To train the GP and designed FSN perfectly, Data must be from different types as follow:

- Normal operational condition
- Disturbed operational condition

Also to get better results and modelling, the data should be captured from variety of spots in the plant. To this aim, a simulation of TE process in Simulink environment has been used.

4.3.1. Simulation Software

- **AspenHysys,** is a comprehensive process modelling and simulation software used to simulate the processes in oil, gas, and chemical industries [58].
- **MATLAB-Simulink,** developed by MathWorks, is a commercial tool for modeling, simulating and analyzing systems dynamically. Simulink is widely used in simulation of processes and modelling systems that are going to be analyzed for different purposes such as simple cases like data generation and complicated cases like control theory. Its great ability in simulation and modeling satisfy set of engineering requirement.

Simulation of the TE process is an important and necessary step in reaching the goal of this study. In this research, we used the simulation of TE process which is done by Simulink software. Figure 4.2 shows a schematic view of simulated TE process in Simulink used to generate data.

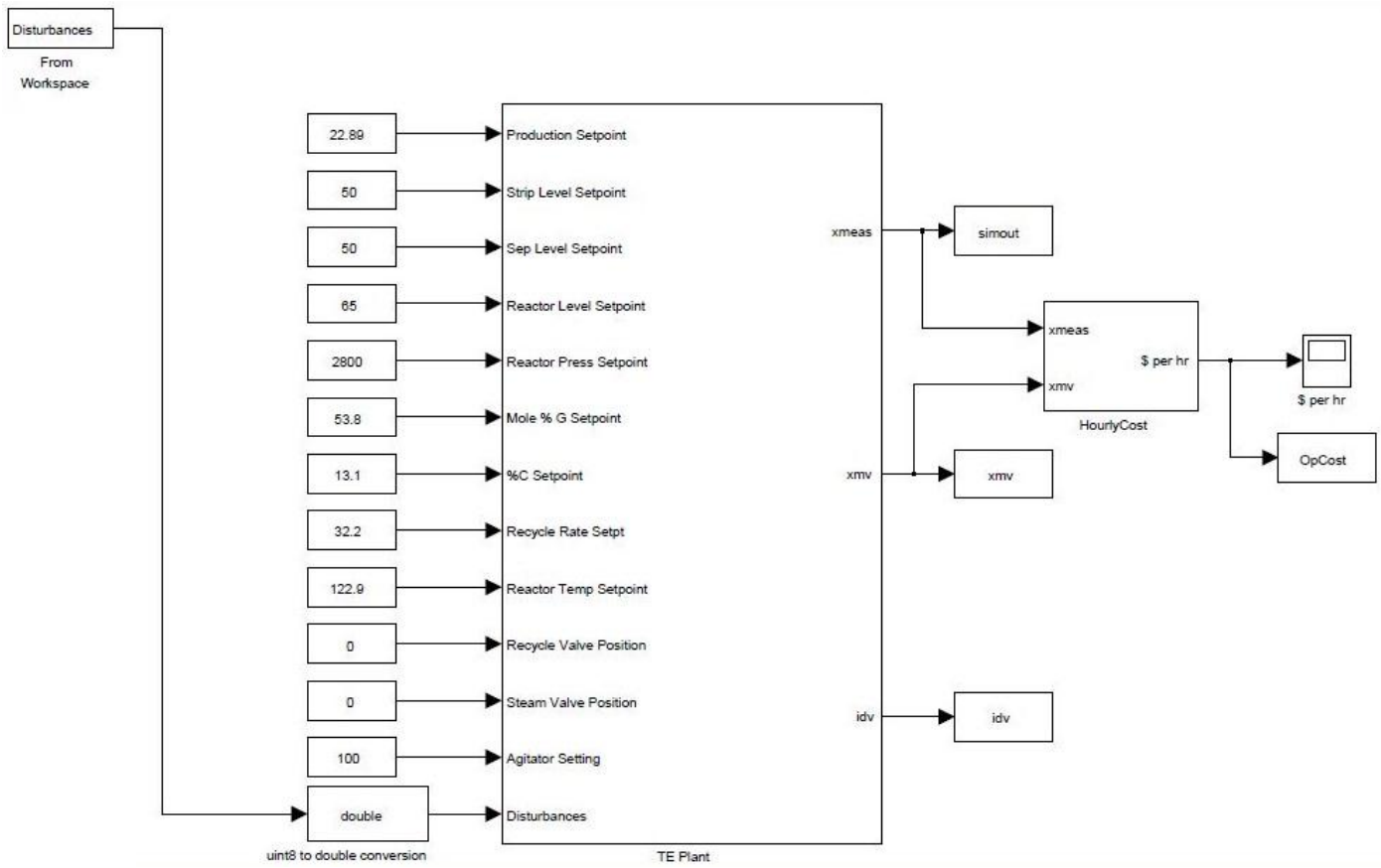


Figure 4.2: Schematic View of Simulation in Simulink Environment

As mentioned, the TE process has 41 measured and 12 manipulated variables where manipulated variables are considered as independent and measured ones as dependent. In order to see how manipulated variables affect the measured ones, they are changes either randomly or by steps under different disturbed operational condition.

In order to simulated disturbed operational condition, twenty disturbances named by IDV (1) to IDV (20) have been simulated to be implemented with different probability of occurrence in the TE process.

For the purpose of this research, four disturbances of IDV (1), IDV (4), IDV (7) and IDV (15) with different probabilities chosen by experience. Table 4.6 shows the disturbances, their probabilities and type of changes.

Table 4.6: Selected Disturbances

Variable Number	Process Variable	Probability of Occurrence	Type
IDV (1)	A/C feed ratio, B composition constant (stream4)	10%	Step
IDV (4)	Reactor cooling water inlet temperature	10%	Step
IDV (7)	C header pressure loss-reduced availability (stream 4)	50%	Step
IDV (15)	Condenser cooling water valve	5%	Sticking

For the process that operates under the disturbance of IDV (1), the dependent and independent variables, their values, ranges and units are described in Table 4.7. Where the A/C feed ratio is considered as the manipulated variables and the reactor feed rate, reactor pressure, level and stripper pressure are chosen as the measured variables.

Table 4.7: Process Manipulated & Measured Variables for Disturbance IDV (1)

Process Manipulated Variables			
Variable Name	**Base Case Value (%)**	**Low & High Limit**	**Unit**
A/C Feed Flow rate (Stream4)	61.302	0 - 15.25	kscmh
Process Measured Variables			
Variable Name	**Base Case Value**	**Unit**	
Reactor Feed rate (Stream6)	42.339	kscmh	
Reactor pressure	2705.0	Kpa (gauge)	
Reactor Level	75.000	%	
Stripper Pressure	3102.2	Kpa (gauge)	

The dependent and independent variables for the process that operates under disturbance IDV (4) are described in Table 4.8 where the reactor cooling water flow is chosen as independent and the reactor temperature as dependent.

Table 4.8: Process Manipulated & Measured Variables for Disturbance IDV (4)

Process Manipulated Variables			
Variable Name	**Base Case Value (%)**	**Low & High Limit**	**Unit**
Reactor Cooling water Flow	41.106	0 - 227.1	m3h-1
Process Measured Variables			
Variable Name	**Base Case Value**	**Unit**	
Reactor Temperature	3102.2	C	

The IDV (7) disturbs the process by loss of pressure in Header C. The manipulated and measured variables are shown in Table 4.9.

Table 4.9: Process Manipulated & Measured Variables for Disturbance IDV (7)

Process Manipulated Variables			
Variable Name	**Base Case Value (%)**	**Low & High Limit**	**Unit**
D Feed (Stream 2)	63.053	0 – 5811	Kg h^{-1}
Process Measured Variables			
Variable Name	**Base Case Value**	**Unit**	
A feed (stream I)	0.25052	Kscmh	

The IDV (15) disturbs the process by simulating the sticky valve at the condenser cooling water stream. The sticky valve does not respond instantaneously to changes in the actuator. A first order lag can be modeled in the response of the actual valve position to changes in the actuator

position. The behavior of the valve percent opening as a function of the actuator position is shown as follow.

$$\tau \frac{d(Valve\ \%)}{dt} + Valve\ \% = Act\ \% + Offset. \qquad (4.2) \qquad [58]$$

Where τ is the valve stickiness time constant, the valve percentage is current opening percentage of valve. Table 4.10 shows the manipulated and measured variable under disturbed condition IDV (15).

Table 4.10: Process Manipulated & Measured Variables for Disturbance IDV (15)

Process Manipulated Variables			
Variable Name	**Base Case Value (%)**	**Low & High Limit**	**Unit**
Separator pot liquid Flow (stream 10)	38.100	0 - 65.71	m^3h^{-1}
Process Measured Variables			
Variable Name	**Base Case Value**	**Unit**	
Reactor pressure	2705.0	Kpa (gauge)	

What are discussed and what are going to be discussed:

In this chapter the process plant and simulation software are described in detail.

Chapter 5 is going to present results as follow:

- *GP & NARX pattern recognition results.*
- *Comparison of GP & NARX results through mean absolute error.*
- *Calculate the relationship among process variables quantitatively.*
- *Building and tuning FSN for defined hazard scenario.*
- *Identifying causes and consequences through FSN.*

Chapter 5

Results & Discussion

Chapter Summery

Chapter 5 presents the results and discuss them in detail. It starts with wavelet de-noising results. Then the GP and NARX results are presented and discussed. Comparison of GP and NARX results is the next part of this chapter followed by implementing the FSN prototype for the defined Hazard scenarios its results.

After simulating different types of data, then process data has to be analyzed to for the purpose of this study which is determining relationships between Process variables. In order to have a complete analysis, it is necessary to process data in advance. Data processing is any process that a computer program does to enter data and summarise, analyses or otherwise convert data into usable information.

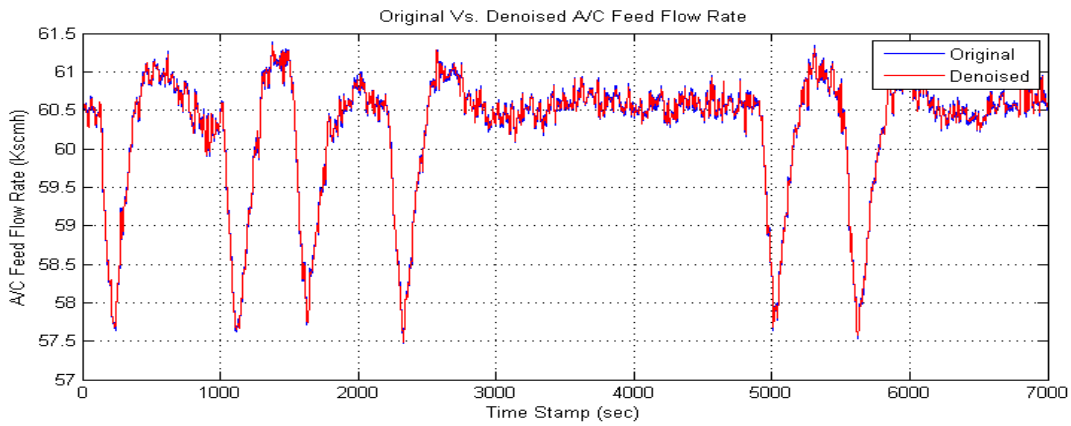
It is a necessity to have complete and less-noised data in one scale. To meet these aims, the extracted data firstly de-noised using wavelets to have complete and less-noised data. Then normalization technique is applied to treats data from different scale in a way to be comparable in one scale. Therefore we would be able to compare corresponding normalized values from different sets of process data.

5.1. De-noising Process Data Using Wavelets

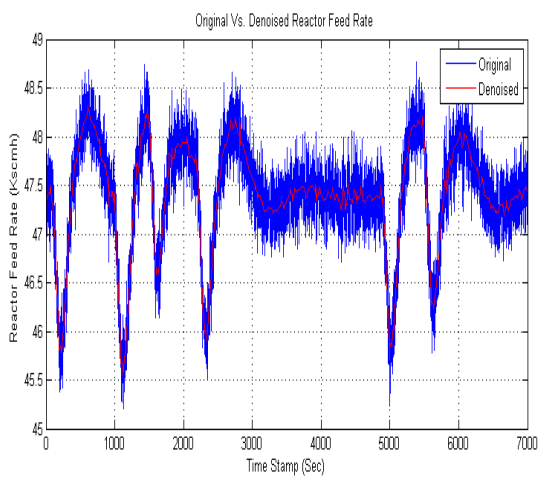
In order to have a complete and reliable analysis on process data, it is necessary to have complete, direct and less-noise data. De-noising is the process of removing noises from a signal. Wavelet techniques are used for the purpose of this research In order to de-noise process data extracted from the TE process. Results are developed using equation models [41]. The equations models are developed in LabVIEW® programing software. Results for different process variables and different disturbed conditions are presented in figures 5.1, 5.2, 5.3 and 5.4. Although these figures shows results of de-noising process, however the relation among process variables and its reason is discussed in this part.

5.1.1. Wavelet Analysis of Data under Disturbed Condition IDV (1)

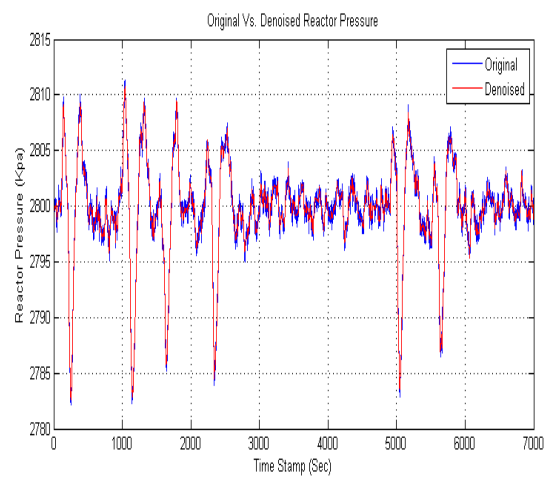
a)



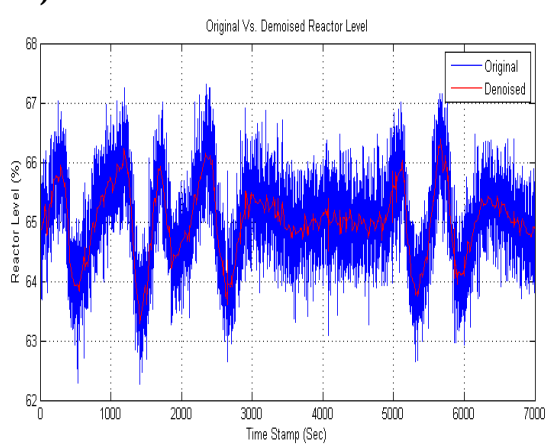
b)



c)



d)



e)

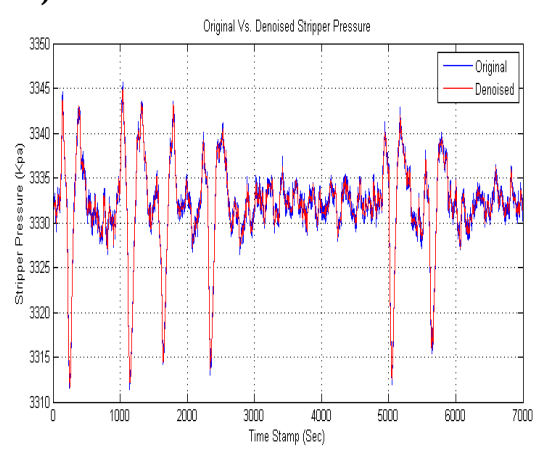


Figure5.1 shows the original and denoised data of process variables under disturbed condition IDV (1) where Figure 5.1(a) shows A/C feed ratio as manipulated variable and measure variables are: Figure 5.1(b) shows original and denoised data of reactor feed rate. Figure 5.1(c) shows original and denoised data of reactor pressure. Figure 5.1(d) shows original and denoised data of reactor level. Figure 5.1(e) shows original and denoised data of stripper pressure.

In this part of the study, A/C feed ratio as the independent variable is manipulated to affect four measured variable. The process of data extraction done in a period of 7000 sec. Figure 5.1 (a) shows the manipulation in A/C feed ratio where six picks in the graphs represents disturbances or manipulations. The picks happened approximately at T=200 sec, 1100 sec, 1600 sec, 2300 sec, 5000 sec and 5600 sec. As figures 5.1 (b), (c), (d) and (e) show, decrease in A/C feed flow directly affect reactor feed rate, reactor pressure, reactor level and stripper pressure respectively.

5.1.2. Wavelet Analysis of Data under Disturbed Condition IDV (4)

Figure 5.2 presents de-noised and noised data in red and blue respectively. Despite the effect of de-noising, process variables affects also going to be discussed in this section. Reactor cooling water flow rate is selected as the independent variable to go under some deviations. Effect of applied deviations on reactor temperature as the dependent variable which was under monitoring is as follow. Figure 5.2 (b) shows the direct effect of reactor cooling water flow on the reactor temperature.

a) **b)**

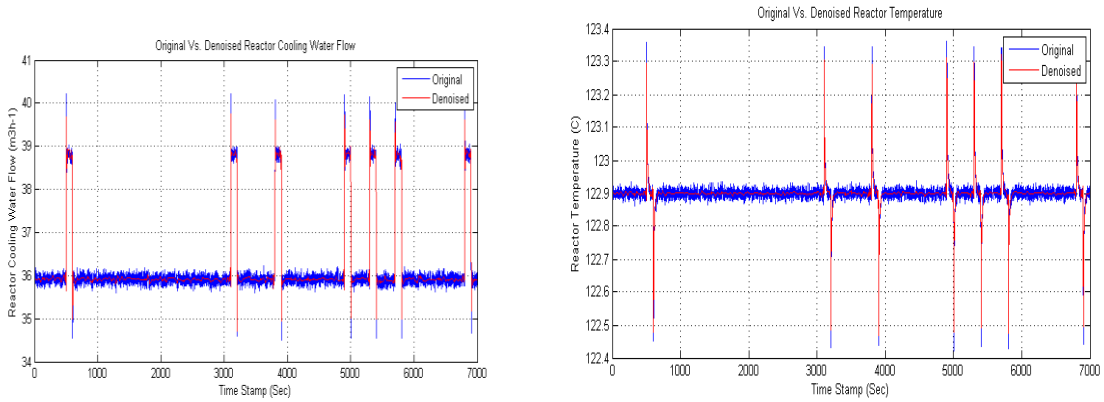


Figure5.2 shows the original and denoised data of process variables under disturbed condition IDV (4) where Figure 5.2(a) shows reactor cooling water flow as manipulated variable and Figure 5.2(b) shows original and denoised data of reactor temperature as measured variable.

5.1.3. Wavelet Analysis of Data under Disturbed Condition IDV (7)

Figure 5.3 presents de-noised and noised data in red and blue respectively. Despite the effect of de-noising, process variables affects also going to be discussed in this section. Changes in Feed D also affect other variables such as Feed A. Some deviations applied on Feed D, and Feed A was under monitoring is a period of 7000 sec. The picks at T=5000 sec in Figure 5.3(b) proof this effect.

a) **b)**

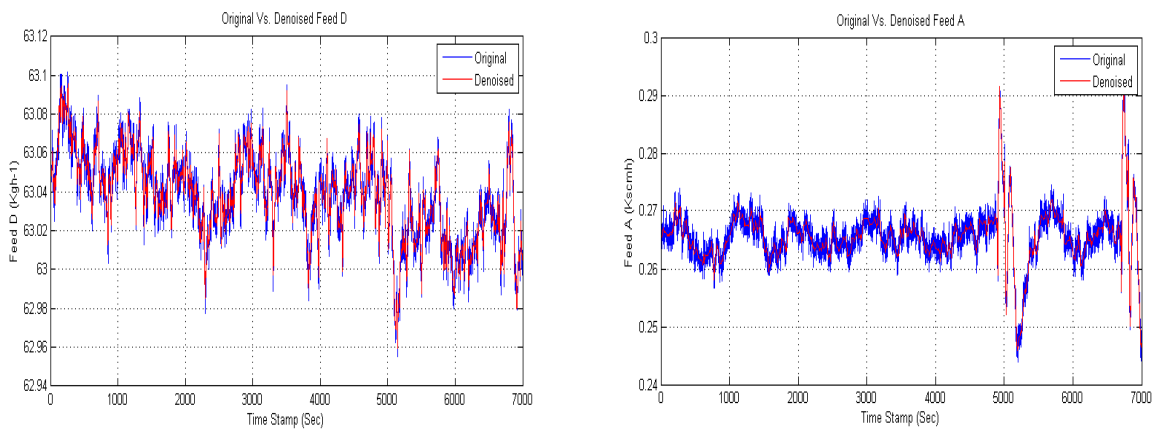


Figure 5.3 shows the original and denoised data of process variables under disturbed condition IDV (7) where Figure 5.3(a) feed D flow rate as manipulated variable and Figure 5.3(b) shows original and denoised data of feed A flow rate as measured variable.

5.1.4. Wavelet Analysis of Data under Disturbed Condition IDV (15)

Figure 5.4 presents de-noised and noised data in red and blue respectively. Despite the effect of de-noising, process variables affects also going to be discussed in this section. Changes in separator liquid flow also affects reactor pressure. Again some deviations applied on manipulated variable which is separator liquid flow and reactor pressure is monitored is a period of 7000 sec. The records don't show Considerable effects that can be because of the considered range of deviation which is small, however those picks at T=2000 sec in Figure 5.4 (b) can be considered as the effect of applied deviation.

a) **b)**

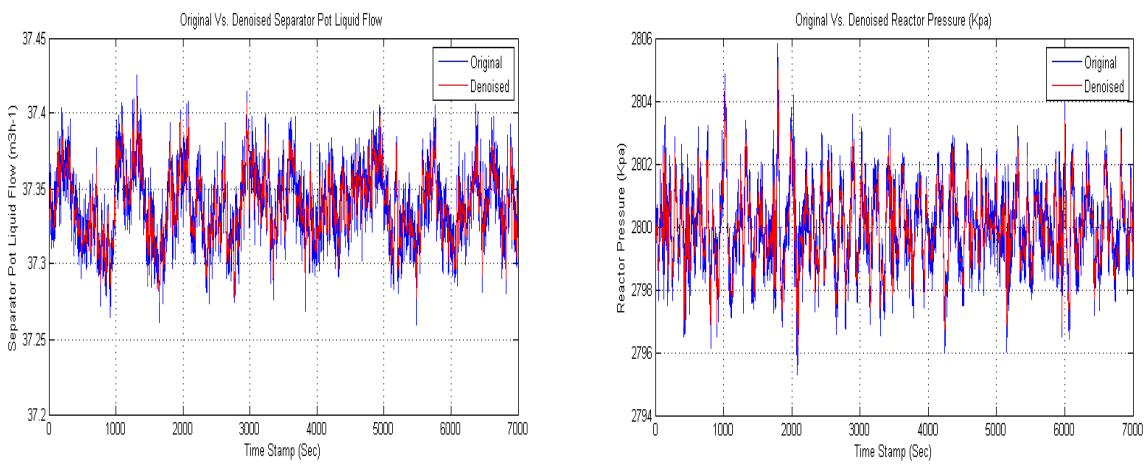


Figure 5.4 shows the original and denoised data of process variables under disturbed condition IDV (15) where Figure 5.4(a) separator pot liquid flow as manipulated variable and Figure 5.4(b) shows original and denoised data of reactor pressure as measured variable.

5.2. Disturbance & Deviation Effect on Process Variables

In order to briefly illustrate how the manipulation of a process variable affects the other ones under normal and disturbed condition, the effect of changes in the A/C feed flow rate (stream4) as the manipulated variable on reactor feed rate (stream6) as the measured one are shown in Figure 5.5. The noise appears in Figure 5.5 (b) & (d) shows the effect of applied disturbances. Also the picks in Figure 5.5 (c) & (d) shows the applied deviations effects.

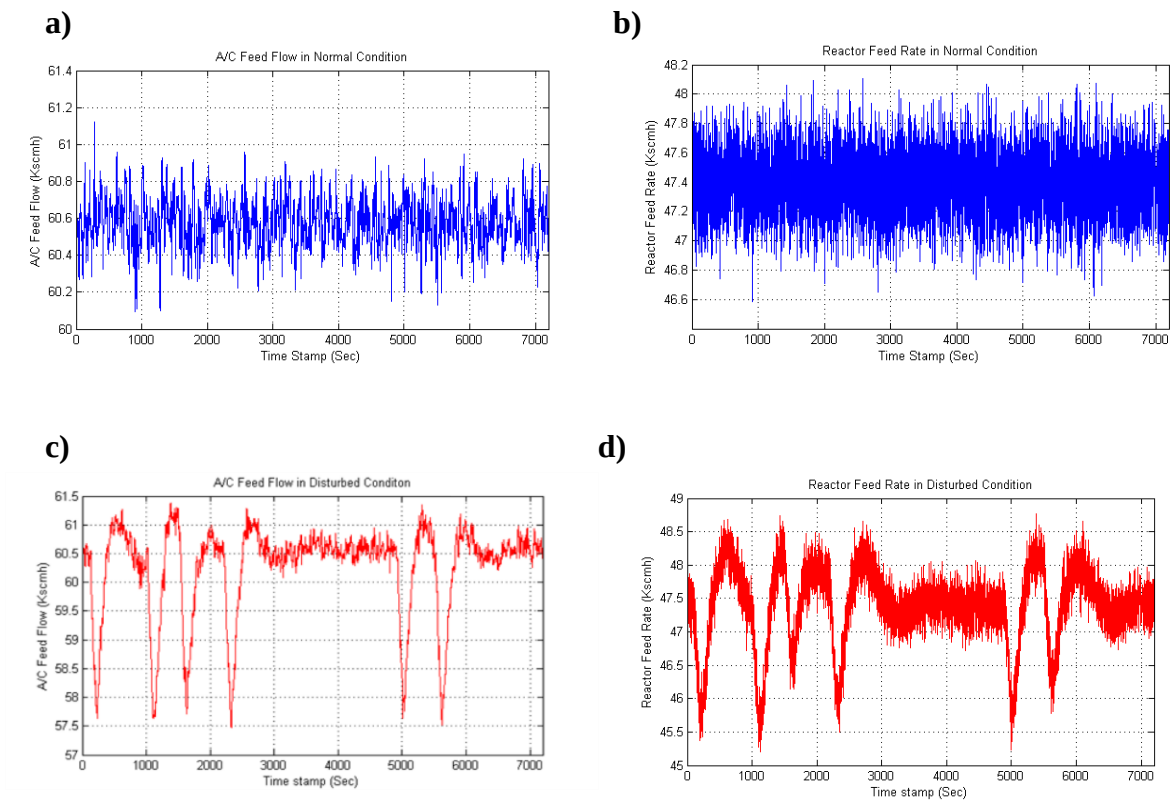


Figure 5.5 shows a comparison of process variable interaction between disturbed and normal operational condition. The blue graphs show their effect under normal and the red ones show under disturbed condition. Figure 5.5(a)&(c) show A/C feed ratio as manipulated variable in normal and disturbed condition respectively, Figure 5.5(b)&(d) shows reactor feed rate as measured variable in normal and disturbed condition respectively.

5.3. Genetic Programing Results

5.3.1. GPLAB - Genetic Programing Toolbox

GPLAB is a genetic programing toolbox in MATLAB. The last version of GPLAB lunched at 2007 is used in this research [59]. There are parameters involved in each algorithm of GPLAB that affect its behaviour. Some parameters automatically set or has its own default values and some of them as to be assigned by user. Generally, the only parameters that have to be assigned by user are the maximum number of generations, the population size, and the names of the files that contain the data set to be used. The parameters value is shown in Table 5.1.

Table 5.1: GPLAB Parameters Values

Parameter	Value
Maximum number of generations	25
Population size	50

Once generating data and data processing finished, data is divided into two parts to be used as training and testing data to be used as the input of the toolbox. GPLAB uncovers the relations among variables and shows their relationship as mathematical formulation.

5.3.1.1. Fitness Function

Selection of individuals among the selected ones is varied till the best fitness measures are produced. The fitness criterion is defined as

$$\min f(x) = \sum |xi - xj|. \qquad (5.1) \qquad [9]$$

Where, min f(x) is the minimization of the sum of absolute differences between the obtained and the expected results xi and xj respectively. A GP evolved candidate would be selected as a better solution if it has a lower fitness value.

All the available data for the manipulated and measured variables are used to be analyzed by the GP. As mentioned before, 4 sets of data has been simulated by TE process in four different disturbed conditions. The points that the study is focusing on is how independent variables affect the dependent ones and the relation between them.

5.3.2. Disturbance IDV (1) – A/C Feed Ratio

For the process that operates under the disturbance of IDV (1), the dependent and independent variables are described in Table 4.7. Figure 5.6 shows the actual and predicted trend of four measured variables under effect of A/C feed ratio changes as manipulated variable in disturbed condition done by GP.

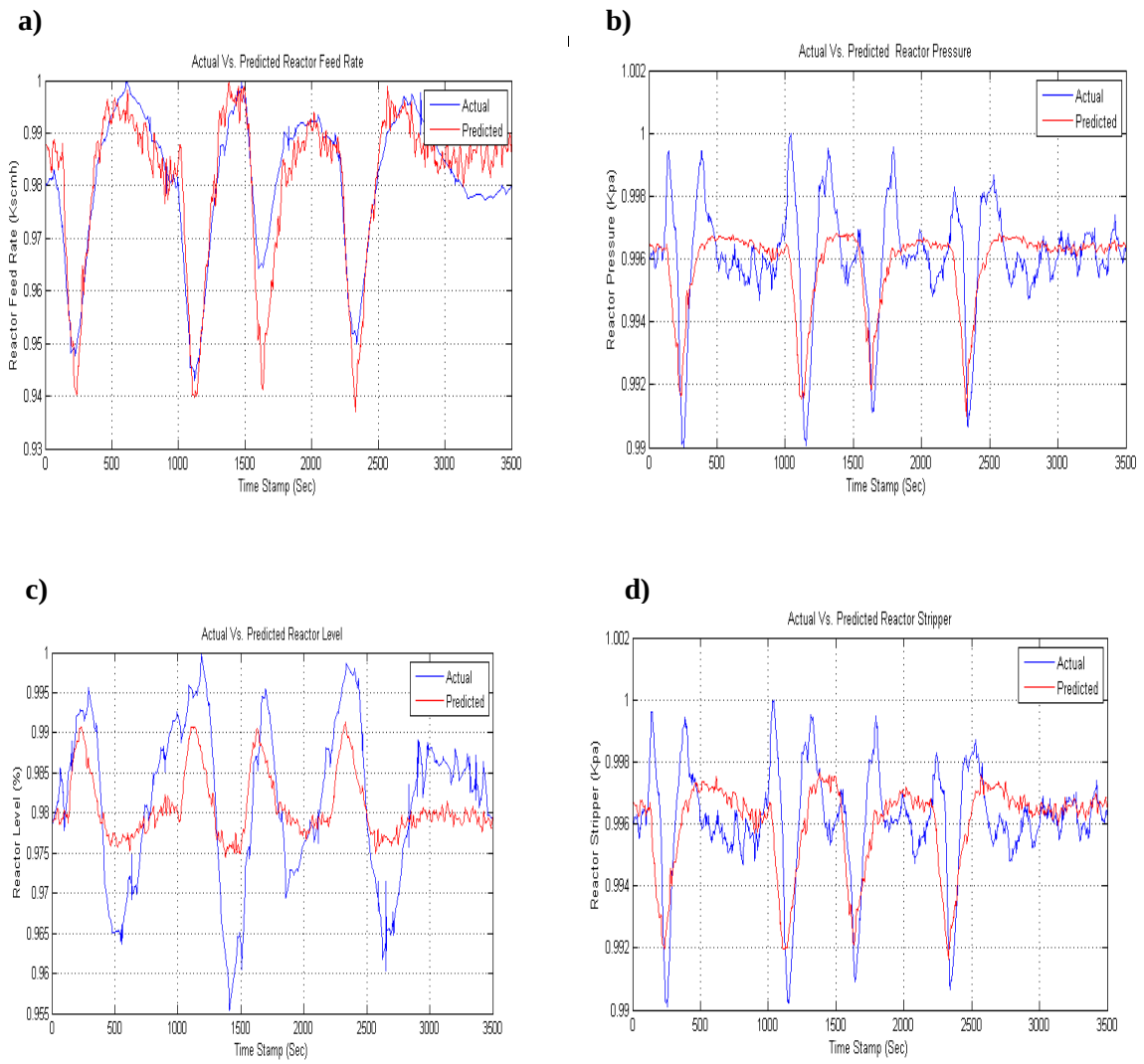


Figure 5.6 shows the effect of changes in A/C feed ratio as manipulated variable under disturbance IDV (1) on measure variables where: Figure 5.6(a) shows the actual and predicted trend of reactor feed rate. Figure 5.6(b) shows the actual and predicted trend of reactor pressure. Figure 5.6(c) shows the actual and predicted trend of reactor level. Figure 5.6(d) shows the actual and predicted trend of stripper pressure.

5.3.3. Disturbance IDV (4) – Reactor Cooling Water Inlet Temperature

The dependent and independent variables for the process that operates under disturbance IDV (4) are described in Table 4.8. The Figure 5.7 shows actual and predicted trend of reactor temperature as measured variable in TE process. These are developed using equation models [59] done by GP.

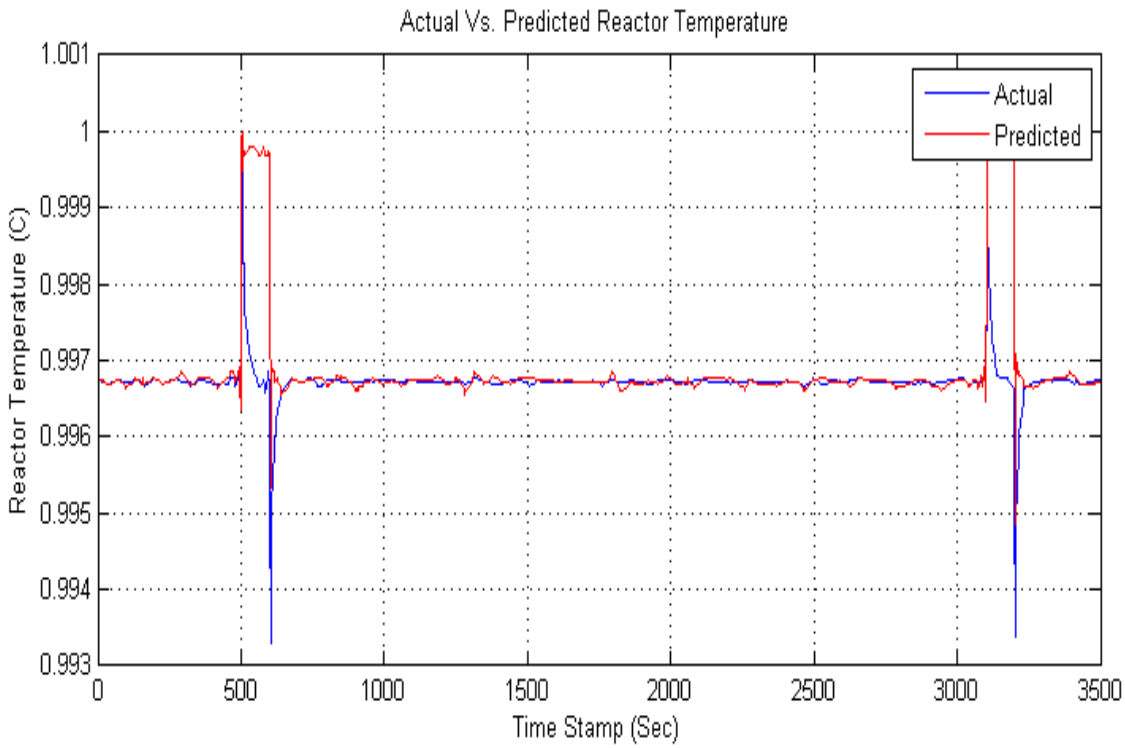


Figure 5.7: Actual vs. Predicted Trend of Reactor Temperature under Disturbance IDV (4) Where the Manipulated Variable is Reactor Cooling Water Flow.

5.3.4. Disturbance IDV (7) – C header pressure loss-reduced availability (stream 4)

The IDV (7) disturbs the process by loss of pressure in Header C. The manipulated and measured variables are shown in Table 4.9. The actual feed D as measured variable and GP predicted trend has been shown in Figure 5.8.

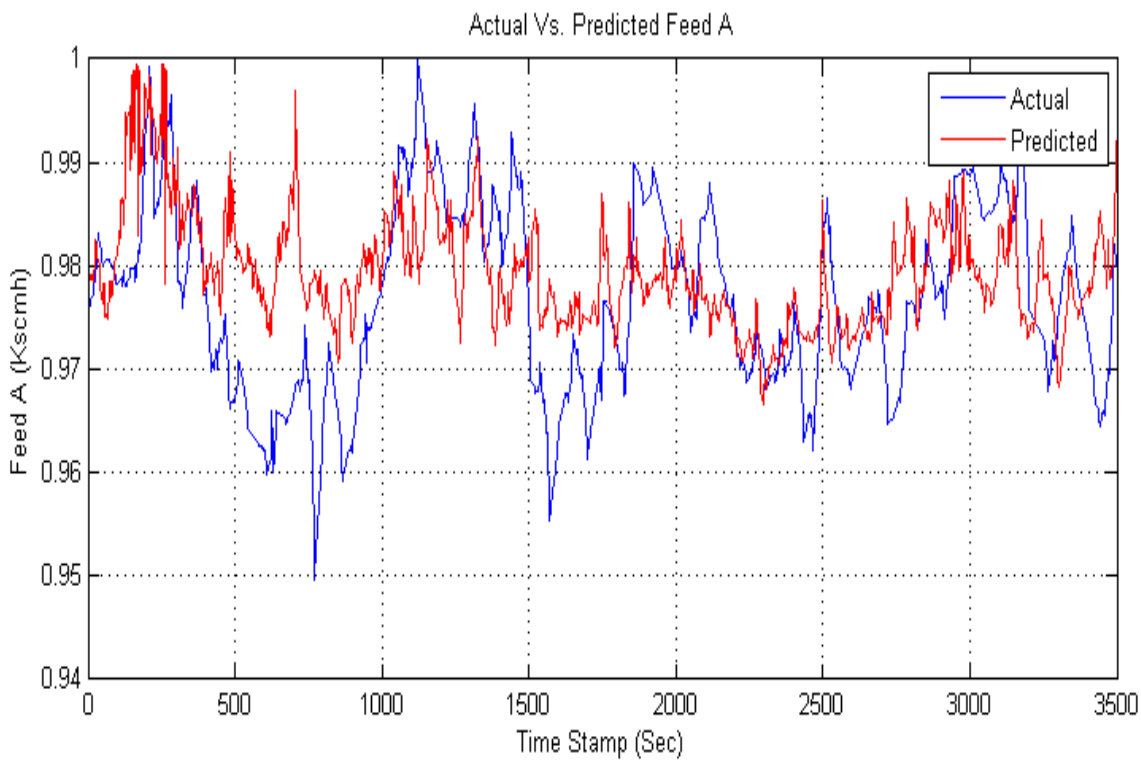


Figure 5.8: Actual Vs. Predicted Trend of Feed A under Disturbance IDV (7) Where the Manipulated Variable is Feed D.

5.3.5. Disturbance IDV (15) – Condenser cooling water valve

The IDV (15) disturbs the process by simulating the sticky valve at the condenser cooling water stream. Table 4.10 shows the manipulated and measured variable under disturbed condition IDV (15). The actual and predicted behavior of reactor pressure under disturbance IDV (15) is shown in Figure 5.9.

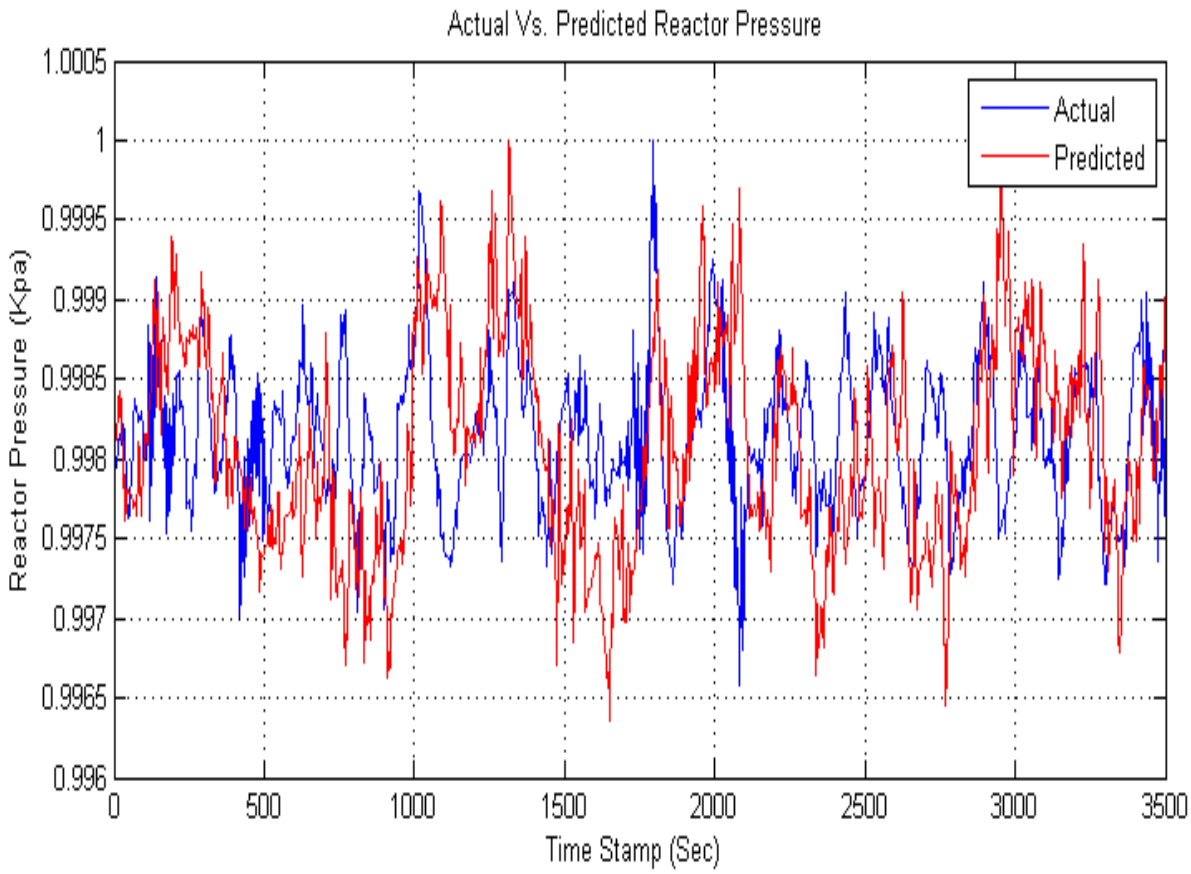


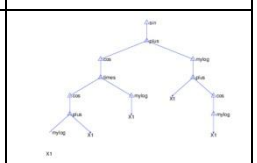
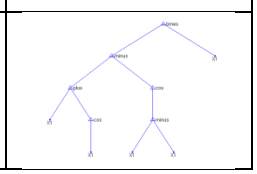
Figure 5.9: Actual Vs. Predicted Trend of Reactor Pressure under Disturbance IDV (15) Where the Manipulated Variable is Separator Pot Liquid Flow.

5.3.6. Formulation of Relation among Process Variable

The relation among variables can be shown in tree format as well. For instance, the formulated relationship among A/C feed ratio as manipulated variable and its corresponding measured variables where the process is under disturbance IDV (1) are presented in Table 5.2.

Where the variable X_i is considered as a dependent variable which is the reactor feed ratio in the equations. Subtraction, addition, sin, cosine and log are the functions that make the relationship.

Table 5.2: The nonlinear Relationship between Xi (Manipulated variable) & Yi (Measured Variable)

$\boldsymbol{yi}$ $\boldsymbol{= Measured\ variable}$	$\boldsymbol{The\ relation\ between\ yi}$ & $\boldsymbol{xi}$ $\boldsymbol{= \frac{A}{C} feed\ ratio\ as\ manipulated\ variable}$	$\boldsymbol{Tree\ structure}$
$\boldsymbol{y1}$ **= Reactor Feed rate (Stream6)**	$y1 = \sin[\cos[\cos(\log x1 + x1) \times logx1] + \log(x1 + \cos(logx1))]$	
$\boldsymbol{y2}$ **= Reactor pressure**	$y2 = [(\log(x1 \times x1) - x1) \times (\sin(x1) - x1)]$	
$\boldsymbol{y3}$ **= Reactor Level**	$y3 = sinx1 + [(\log(x1) + x1) - \sin(\log(x1))]$	
$\boldsymbol{y4}$ **= Stripper Pressure**	$y4 = [(x1 + cosx1) - \cos(x1 \times x1)] \times x1$	

5.4. NARX Results

5.4.1. NARX

MATLAB R2012a contains a neural network toolbox that includes all types of neural network. Feedback structure used for different purposes of pattern recognition, prediction and system modelling is selected for the aim of this research. In this type of learning, network is trained according to input data and desired output value. Input data and desired output are the only information fed to neural network in this learning mode. During the training process, weights are assigned randomly to each input then by summation of inputs and their weights, output will produce. Since weights are assigned randomly, the resultant output may differ from the desired output, therefore the error will be calculated according to the difference between the resultant and desired output. Error value will be sent to input layer as a feedback to update weights value. This process has to be repeated until the error approaches zero and desired output is achieved.

5.4.2. Disturbance IDV (1) – A/C Feed Ratio

For the process that operates under the disturbance of IDV (1), the dependent and independent variables are described in Table 4.7. Figure 5.10 shows the actual and predicted trend of A/C feed ratio and the other four measured variables in disturbed condition done by GP.

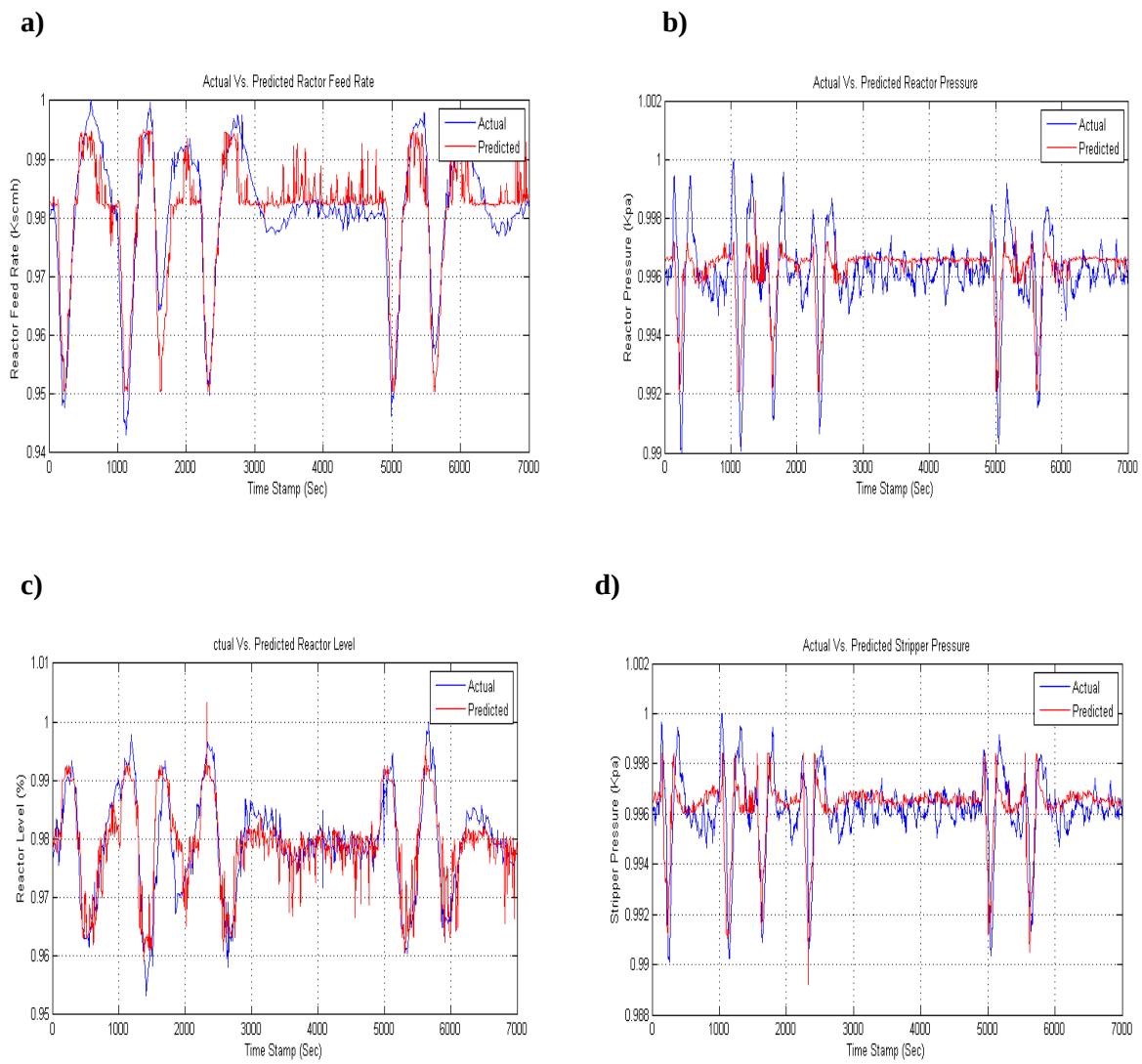


Figure 5.10 shows the effect of changes in A/C feed ratio as manipulated variable under disturbance IDV (1) on measure variables where: Figure 5.10(a) shows the actual and predicted trend of reactor feed rate. Figure 5.10(b) shows the actual and predicted trend of reactor pressure. Figure 5.10(c) shows the actual and predicted trend of reactor level. Figure 5.10(d) shows the actual and predicted trend of stripper pressure.

5.4.3. Disturbance IDV (4) – Reactor Cooling Water Inlet Temperature

The dependent and independent variables for the process that operates under disturbance IDV (4) are described in Table 4.8. The Figure 5.11 shows actual and predicted trend of reactor temperature as measured variable in TE process.

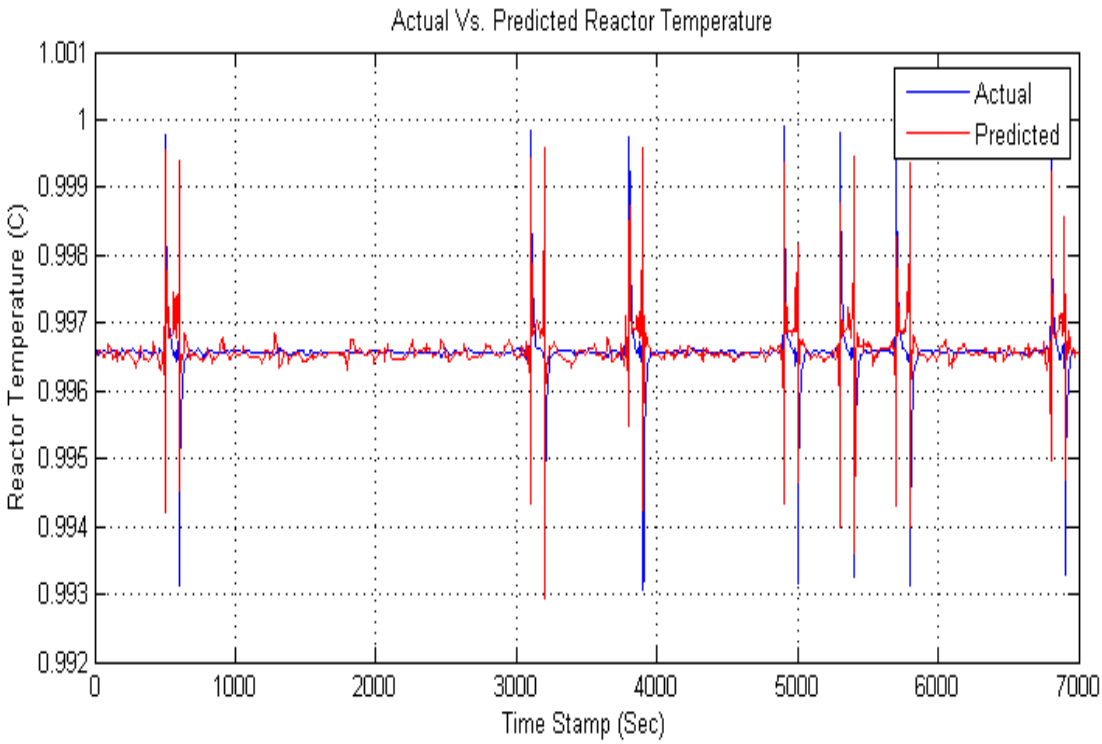


Figure 5.11: Actual vs. Predicted Trend of Reactor Temperature under Disturbance IDV (4) Where the Manipulated Variable is Reactor Cooling Water Flow.

5.4.4. Disturbance IDV (7) – C header pressure loss-reduced availability (stream 4)

The IDV (7) disturbs the process by loss of pressure in Header C. The manipulated and measured variables are shown in Table 4.9. The actual feed D as measured variable and GP predicted trend has been shown in Figure 5.12.

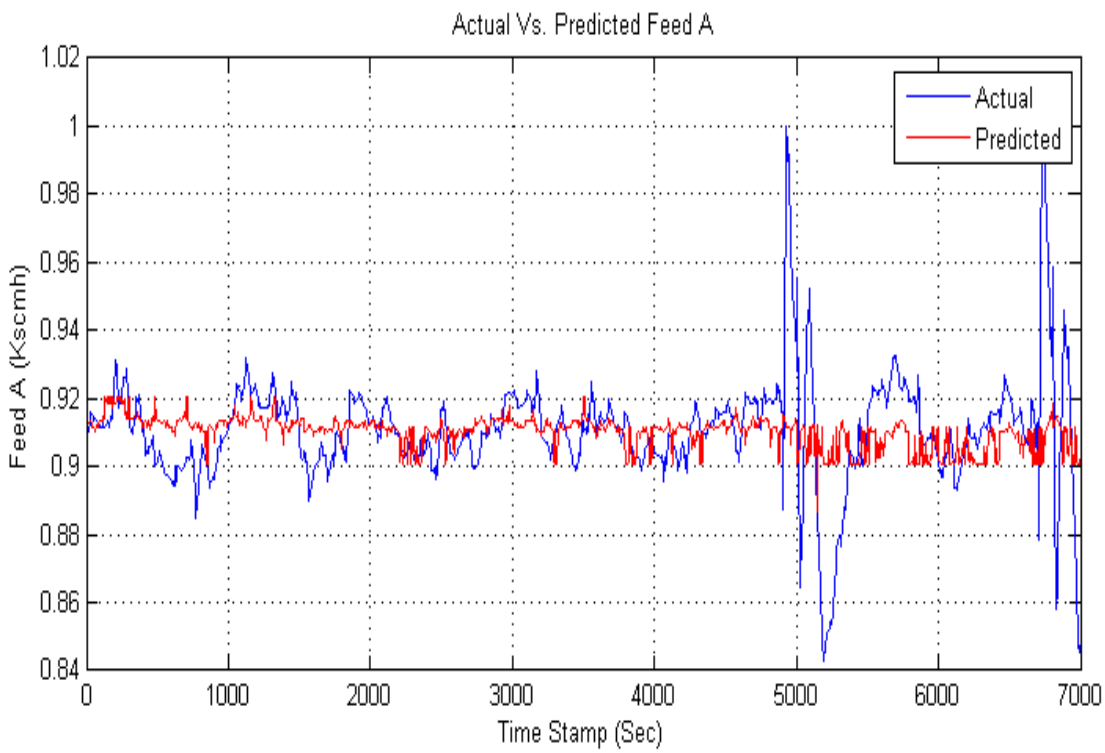


Figure 5.12: Actual Vs. Predicted Trend of Feed A under Disturbance IDV (7) Where the Manipulated Variable is Feed D.

5.4.5. Disturbance IDV (15) – Condenser cooling water valve

The IDV (15) disturbs the process by simulating the sticky valve at the condenser cooling water stream. Table 4.10 shows the manipulated and measured variable under disturbed condition IDV (15). The actual and predicted behavior of reactor pressure under disturbance IDV (15) is shown in Figure 5.13.

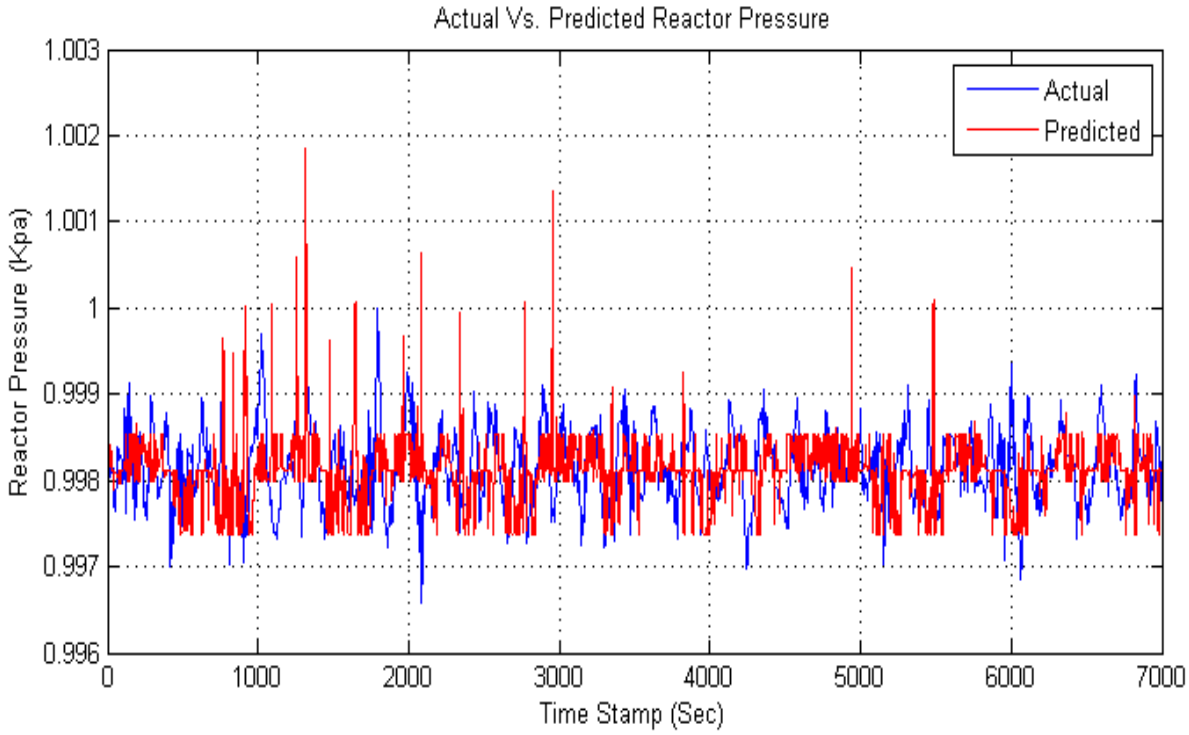


Figure 5.13: Actual Vs. Predicted Trend of Reactor Pressure under Disturbance IDV (15) Where the Manipulated Variable is Separator Pot Liquid Flow.

5.5. Comparison of GP & NARX Results

Two methods of genetic programing and neural network are used to identify the relationship among process variables. In order to compare performance of these two techniques graphically, the error histogram diagram is drawn according to their performances in predicting the behaviour of process variable.

a) **b)**

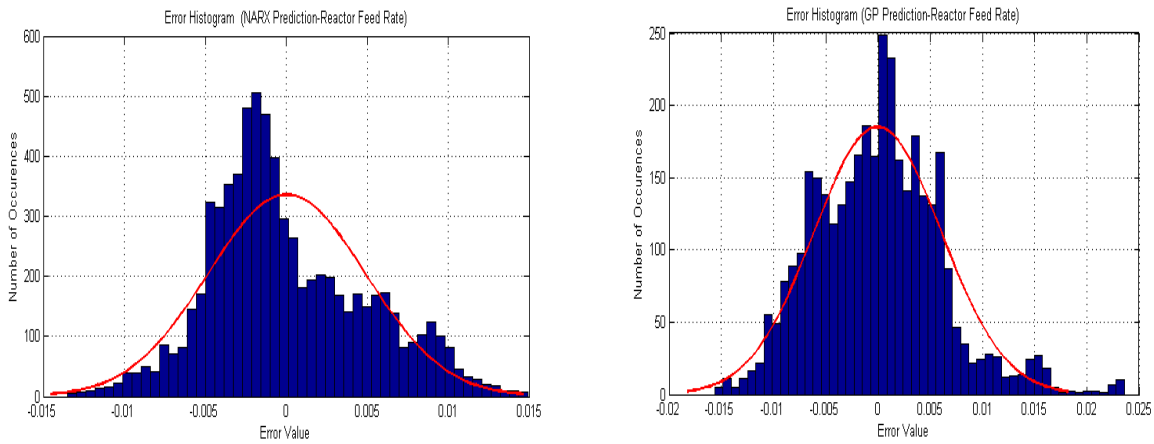


Figure 5.14: Error Histogram of Reactor Feed rate Prediction-IDV (1). 5.14(a) & (b) show NARX & GP Errors Respectively.

a) **b)**

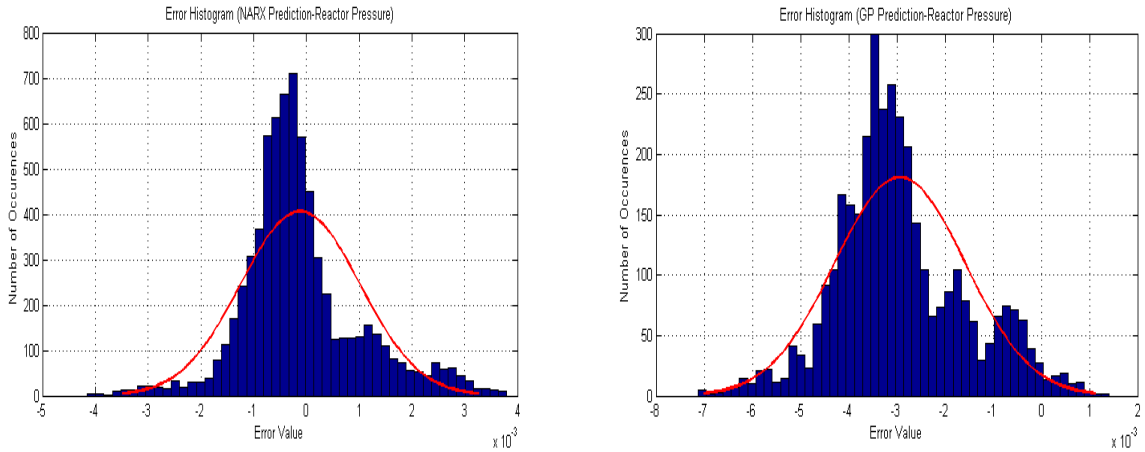


Figure 5.15: Error Histogram of Reactor Pressure Prediction-IDV (1). 5.15(a) & (b) show NARX & GP Errors Respectively.

a) **b)**

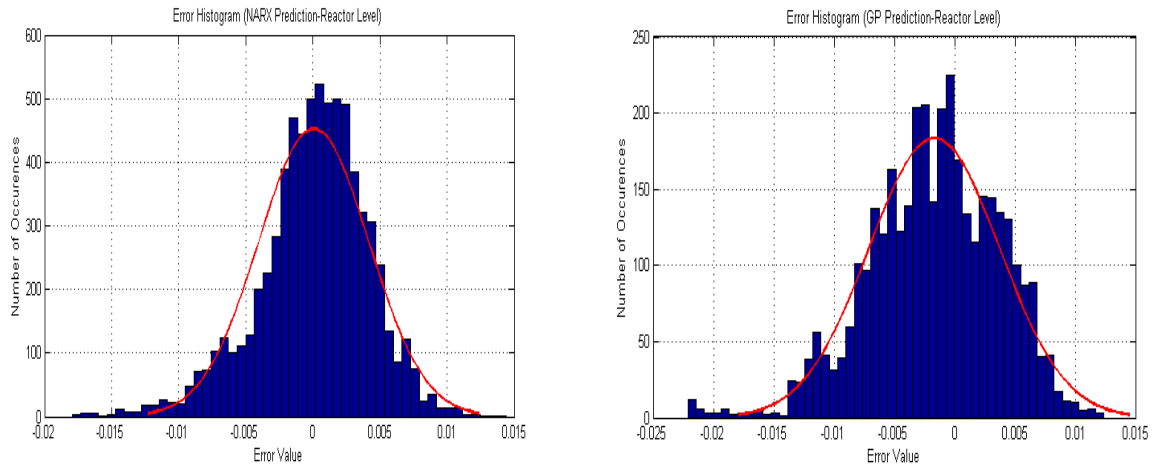


Figure 5.16: Error Histogram of Reactor Level Prediction-IDV (1). 5.16(a) & (b) show NARX & GP Errors Respectively.

a) **b)**

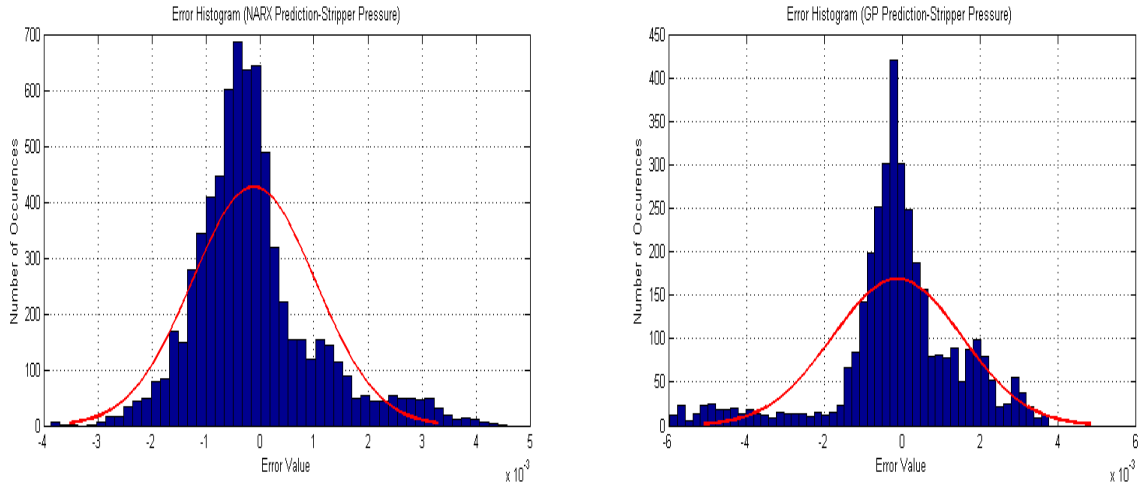


Figure 5.17: Error Histogram of Stripper Pressure Prediction-IDV (1). 5.17(a) & (b) show NARX & GP Errors Respectively.

a) **b)**

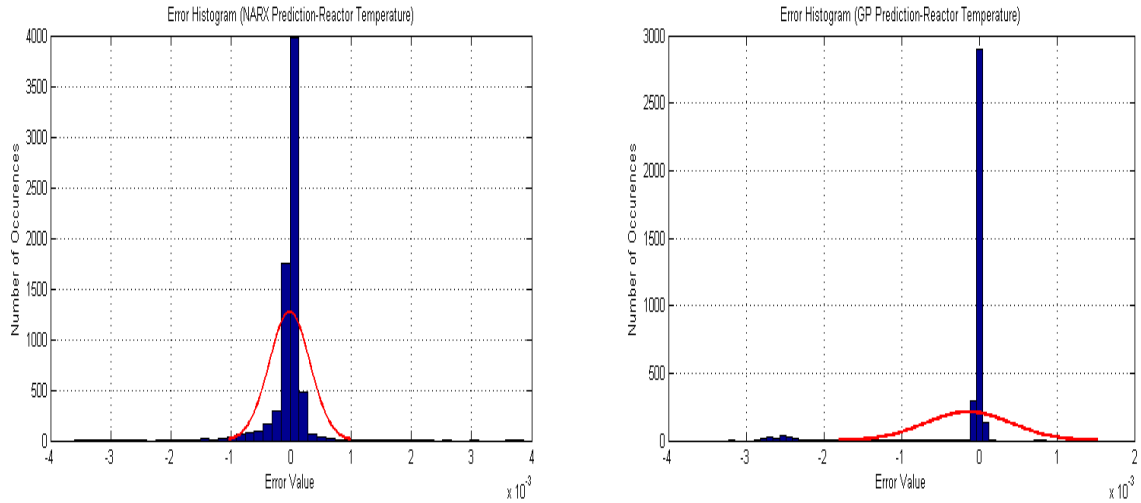


Figure 5.18: Error Histogram of Reactor Temperature Prediction-IDV (4). 5.18(a) & (b) show NARX & GP Errors Respectively.

a) **b)**

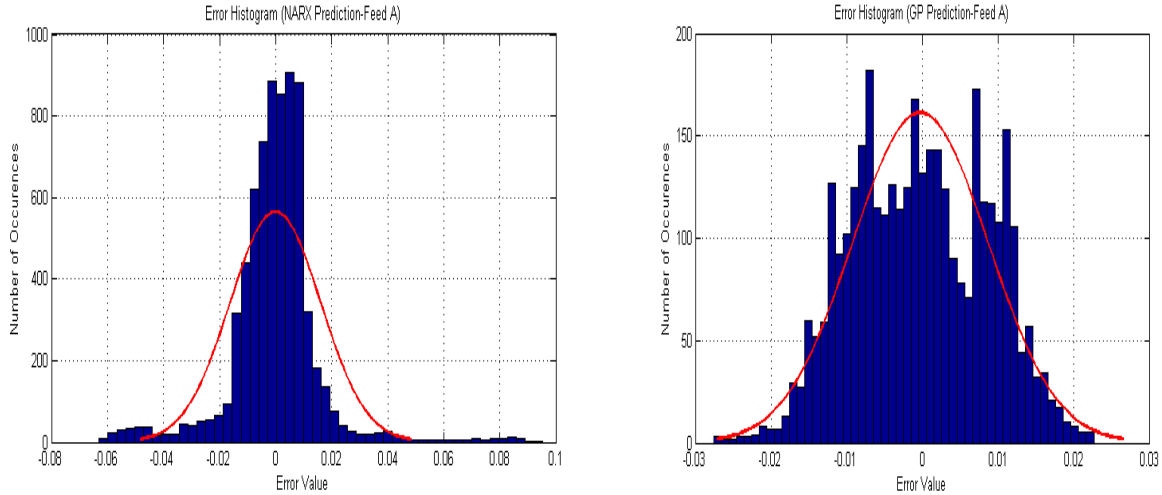


Figure 5.19: Error Histogram of Feed A Prediction-IDV (7). 5.19(a) & (b) show NARX & GP Errors Respectively.

a) **b)**

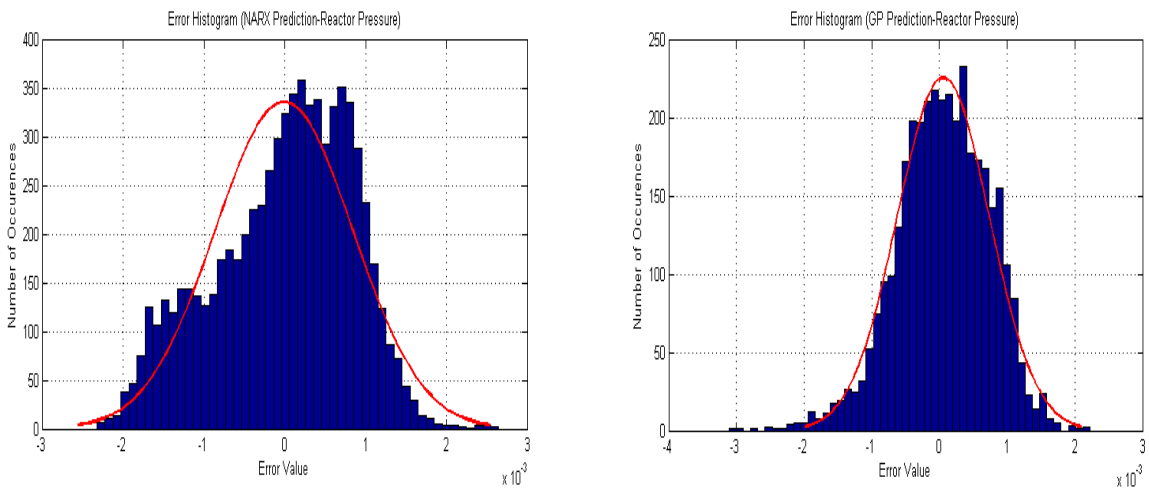


Figure 5.20: Error Histogram of Reactor Pressure Prediction-IDV (15). 5.20(a) & (b) show NARX & GP Errors Respectively.

5.5.1. Mean Absolute Error (MAE) & mean Absolute Percentage Error (MAPE)

In order to measure the accuracy of GP and NN in trend estimation, mean absolute error and mean absolute percentage error is used through Formula 5.2 and 5.3 respectively.

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |fi - y| \qquad (5.2)$$

Where f_i and y_i are the predicted and true values of process variables respectively.

$$MAPE = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{At - Ft}{At} \right| \qquad (5.3)$$

Where f_t and A_t are the predicted and true values of process variables respectively. Calculated error value for each process variable is presented in Table 5.3.

Table 5.3: MAE & MAPE Values

Deviation	Process Variables		GP Prediction		NARX Prediction	
	Manipulated	**Measured**	**MAE**	**MAPE**	**MAE**	**MAPE**
IDV (1)	A/C Feed Flow rate (Stream4)	Reactor pressure	0.0047	0.4855	0.0038	0.3987
		Reactor Feed rate (Stream6)	0.0011	0.1090	0.0007	0.0739
		Reactor Level	0.0045	0.4589	0.0032	0.3198
		Stripper Pressure	0.0013	0.1288	0.00075	0.0741
IDV (4)	Reactor Cooling water Flow	Reactor Temperature	0.00027	0.0280	0.00008	0.0081
IDV (7)	D Feed (Stream 2)	A feed (stream I)	0.0072	0.7416	0.0012	0.1295
IDV (15)	Separator pot liquid Flow (stream 10)	Reactor pressure	0.00054	0.0546	0.0003	0.0312

As it the error values are presented in Table 5.3, accuracy of NARX in prediction of selected process variables behaviour and identifying their inter-relationship pattern is better in comparison with GP.

5.6. Prototype Development and Testing - Fault Semantic Network

Design of a real-time monitoring prototype to analyze process data and find the causes and consequence for the purpose of fault propagation analysis is discussed in this section. To this, LabVIEW® software developed by National Instruments Corporation is used to model the system as a FSN. The LabVIEW® block diagram of the developed prototype is included as Appendix 2. Figure 5.21 presents main steps of building the FSN.

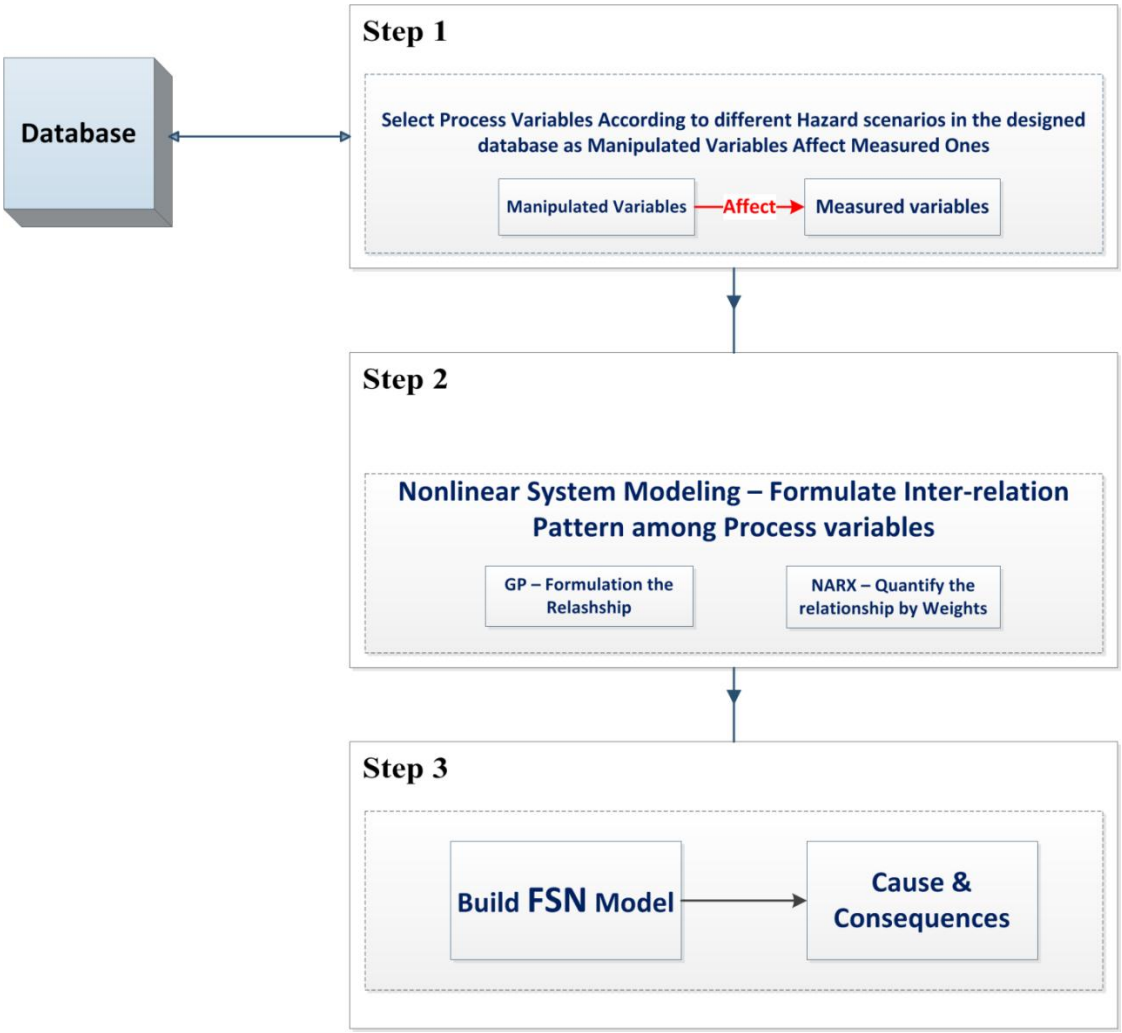


Figure 5.21: FSN Module

Considering the FSN module presented in Figure 5.21, steps of implementing fault propagation analysis through FSN are explained with a case study as follow.

Step 1: Selecting P&ID

The TE chemical process is described in chapter 4 in detail and the extracted data are analyzed. In order to test the FSN module, soma hazard scenarios related to the main reactor of the TE process are defined. The reactor and its input and output are shown in Figure 5.22.

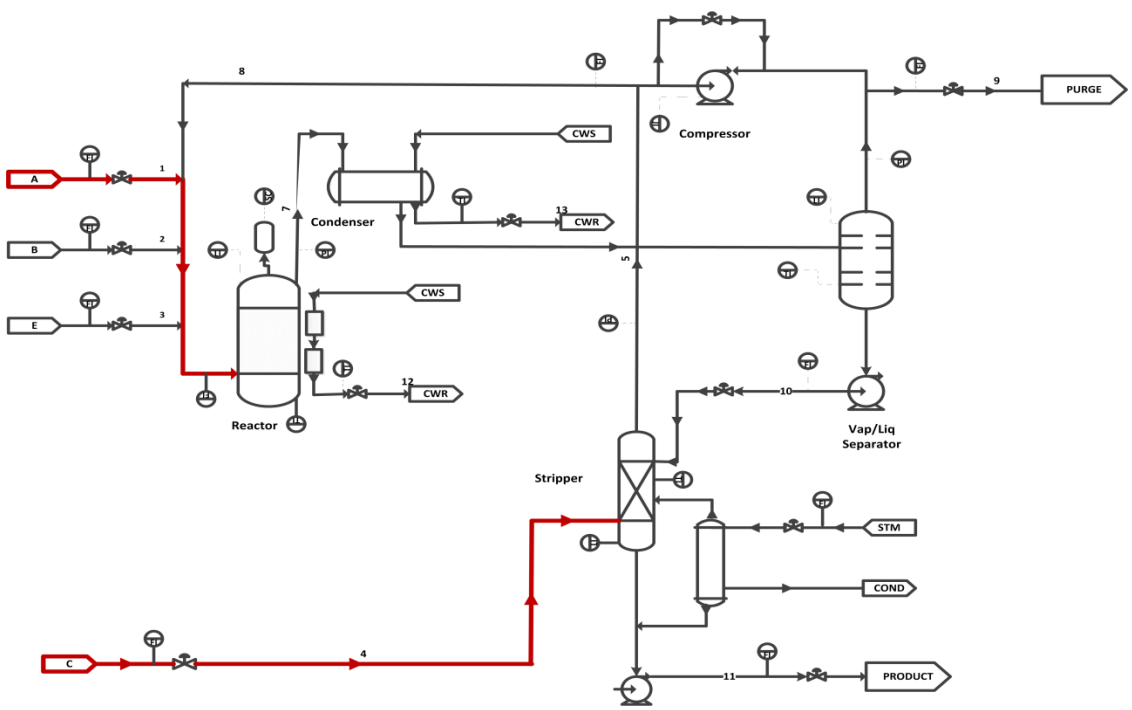


Figure 5.22: P&ID of TE Process-Main Reactor

In the P&ID, There are 3 streams that feed the reactor through feeds A, D and E. Feed C goes to the stripper. Potential faults that might happen are any deviation in reactor feed ratio that may result in deviation in pressure, level of the reactor followed by deviation in reactor in-flow and out-flow. The whole process is discussed in detail in chapter 4.

Step2: Designing Database & Hazard Scenario

As shown in the FSN module, designing a database according to the defined hazard scenarios, equipment, failure modes, process variables and etc. is first step in building the FSN. To this, Hazard scenarios should be defined for each failure mode. In order to demonstrate the application of the FSN in fault propagation analysis and identifying hazard, a hazard scenario related to the reactor of TE process consists of initial events, top event and consequences is designed in MS-Access. The designed database is shown in Tables 5.4, 5.5 and 5.6.

Table 5.4 describe process variables, types of nodes, ranges. Where the Boolean nodes just accept *True /False* value and binary get qualitative values. This table is used for the purpose of probability calculations.

Table 5.4: Process variables, Nodes & Values

Node Name	Node Type	Node Value
A/C Feed Flow Rate (Fr1)	Binary	$\{Low, High\}$
Reactor Feed rate (Stream 6) (Fr2)	Binary	$\{Low, High\}$
Reactor Pressure (RP)	Binary	$\{Low, High\}$
Reactor Liquid Level (LR)	Binary	$\{Low, High\}$
Vibration Test (VT)	Boolean	$\{T, F\}$
Corrosion Test (CT)	Boolean	$\{T, F\}$
Leakage Test (LT)	Boolean	$\{T, F\}$

Table 5.5 presents symptoms, mechanisms and corresponding failures associated with the hazard scenarios.

Table 5.5: Failures, Mechanism, Symptoms and Description

F-ID	Failure Name	Description	Mechanism	Symptoms	
			Final cause	**Med Cause**	**Primary Cause**
F1	Reactor-Leakage Through Wall Crack	Propagating crack on the reactor wall causes leakage	Vibration	Operation at High Pressure	High Flow
					Valve
				High Level	High flow
					Valve
			Corrosion		
			Improper Design		
			Improper repair		
			Structural damage		
			…		
F2	Fracture/Catastrophic Rupture	Fracture may happens and results in a catastrophic event	Long-term embrittlement		
			Loos or missing fasteners		
			Welding defects		
			Structural Damage		
			Improper design		
			Improper Repair		
			..		
F3	Valve Failure	Fails to Operate	Fail to Open/Close	Valve Leakage	
				External / Internal Corrosion	
				Seal problem	
				Design Defects	
				Material defects	
				Human Error	
				…	

Table 5.6 describes installed sensors to extract data from the process plant.

Table 5.6: Sensors and Description

Sensor ID	Description
S1	**Pressure Sensor** – Monitor the Pressure Inside the Reactor
S2	**Level Sensor** - Monitor the Liquid Level of the Reactor
S3	**Flow Meter** - Monitor the Flow of Reactant and Product of the Reactor
S4	**Vibration Sensor** – Monitoring the Vibration of the Reactor
S5	**Thickness Sensor** – Monitoring the Wall Thickness of the Reactor to Prevent Corrosion

Step 3: Fault Propagation and Hazard Identification Using FSN

LabVIEW® software is used to develop a program for the purpose of real-time fault propagation analysis and hazard identification. The FSN program queries the database to dynamically models the faults and hazards scenarios. Figure 5.23 represents a schematic view of the model in the LabVIEW® environment. Also a modeled scenario by the FSN is described in Table 5.7.

There are 2 main phases, the FSN follows to analyze faults propagation and their associated hazards. Integration of these two phases builds the prototype where phase 1 uses inter-relation pattern recognition and phase 2 uses fuzzy expert system (FES) and Bayesian Belief Network (BBN) for reasoning causes and consequences and tuning FSN.

- **Phase 1 - Inter-relation pattern recognition:** In order to uncover the relationship among process variables, GP and NARX are used as pattern recognition techniques to identify their relationship quantitatively. This step analyzes propagation of faults among process variables, therefore any small deviation in a process variable will be detected and main cause process variable will be identified. Also, when process goes outside its defined safety limits causes will be analyzed. Although the FSN can trace the process deviations from real-time data that dynamically update it, however formulating the relationships among process variables is necessary steps. If extracted data are not complete for any reason such as faulty sensor, sensor failure and etc. the FSN can recognize the issue through the formulation obtained form process variables inter-actions. Hence, the FSN will be able to compensate lack of complete data.
- **Phase2 – Reasoning based Probabilities:** Once the inter-action among process variables recognized, it is time to trace deviations between process variables and failure mode in equipment. The FSN represents process variables and failure modes as separate nodes. To analyze propagation of faults, a database is designed that includes all process variables, all possible relations among them, different hazard scenarios, associated failure modes and all possible causes and consequences. In the FSN all failure modes and process variables are represented by nodes separately. The FSN received the real-time process data from the plant and update the nodes dynamically. If any process variable is outside of defined range (as defined in database), it will be detected. In this phase The FSN reasons between nodes according to received process data from process plant and their associated probabilities. The reasoning is done through the queries defined in database and static probability distribution for a set of query nodes.

Figure 5.23 shows fault propagation analysis of the selected case study. As shown, Phase 1 and phase 2 are running simultaneously where phase 1 takes care of process variables and their relationships and phase 2 takes care of reasoning and build the whole propagation scenario.

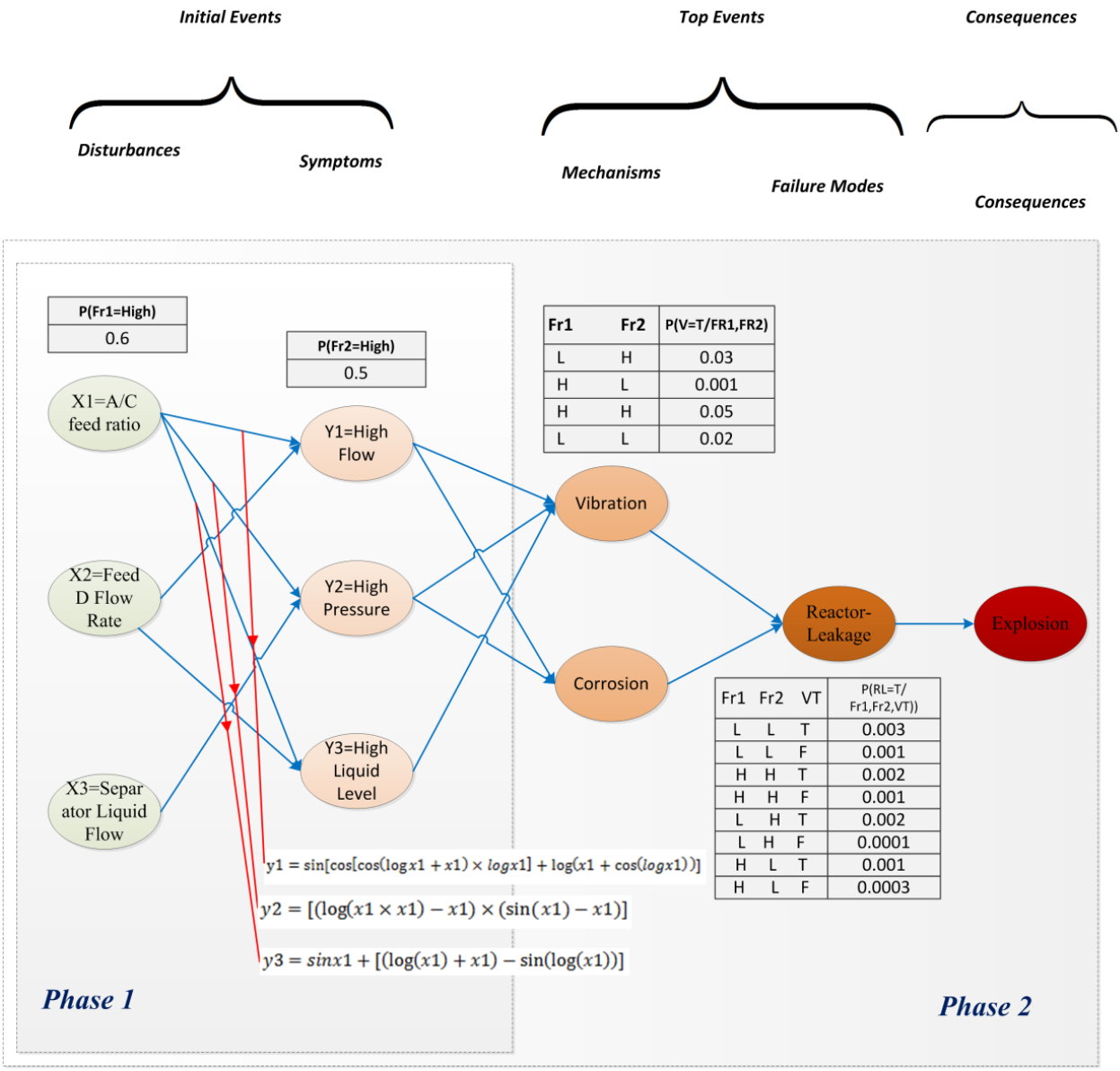


Figure 5.23: Fault Propagation Analysis

Implementing these two phases result in a fault propagation analysis done by the FSN as shown in Table5.7 and Figure 5.23 and 5.24.

According to Table 5.7,

- SY1=Deviation in pressure can be a cause for vibration and corrosion
- SY2=Deviation in reactor liquid level can be cause of vibration
- SY3=Deviation in flow rate into the reactor can be cause of vibration
- Probability of occurring F1=Leakage will increase when S4=vibration sensor confirm the M1=vibration, S1=pressure sensor and S2=level sensor show deviations of pressure and level and S4=wall thickness sensor confirm M2=corrosion.
- The FSN provides some remedies related to failure mode as R1=Welding offered for F1=Leakage.

Table 5.7: Designed Case Study Implemented by FSN

Component	**Reactor**				
Location	**TE Process**				
Symptom	**SY-ID**	**Symptom**	**Wt**	**Semantic Net**	**Mechanism**
	1	High pressure	0.25	SemNet1	M1, M2
	2	High level	0.25		M1
	3	High Flow	0.25		M1,M2
	4	Valve fail to Operate	0.25		M3
Hypothesis / Mechanism	**M-ID**	**Mechanism**	**Wt**	SemNet1	**Diagnosis**
	1	Vibration	0.4		D1
	2	Corrosion	0.3		D1
	3	Valve Failure	0.3		D1
Failure	**F-ID**	**Diagnosis**	**Wt**	SemNet1	**Repair**
	1	D1: [(M1,S4),(M1,S1),(M1,S3),(M2,S5)]	1		R1
Repair	**R-ID**	**Repair**	**Wt**	SemNet1	
	1	Welding	1		
	2				

Figure 5.24 shows a snapshot of FSN prototype while it was running.

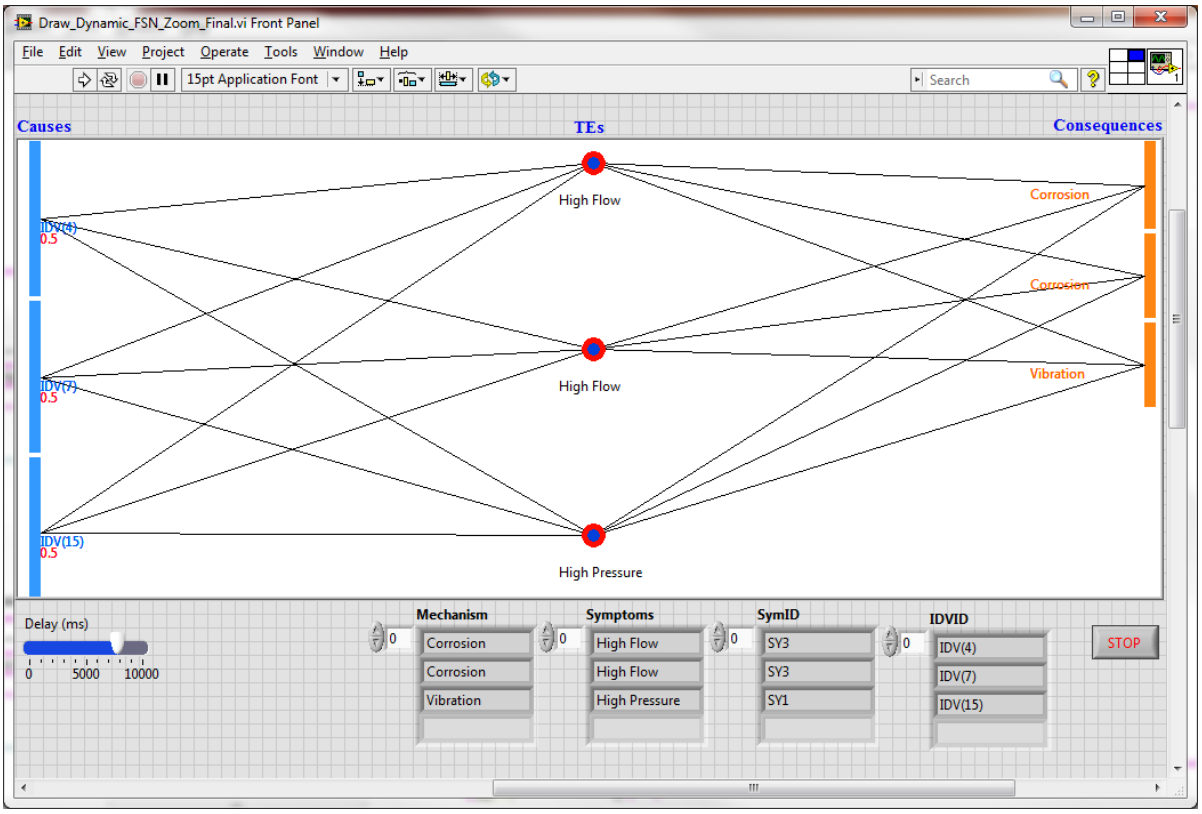


Figure 5.24: Schematic View of FSN Prototype

4.1. Conclusions

In this study, process variable interaction analysis is used to develop a simulation-based fault propagation analysis. The investigation has adopted two approaches that consist of simulation of chemical process plant along with identifying relationships among process variables quantitatively for the purpose of finding causes and consequences and tracing faults associated with them.

In the first phase of this research, Aspenhysys and Simulink are used to simulate the selected chemical process plant for the purpose of extracting process data. The data obtained from both simulators was in an acceptable range however the Simulink data was more exact.

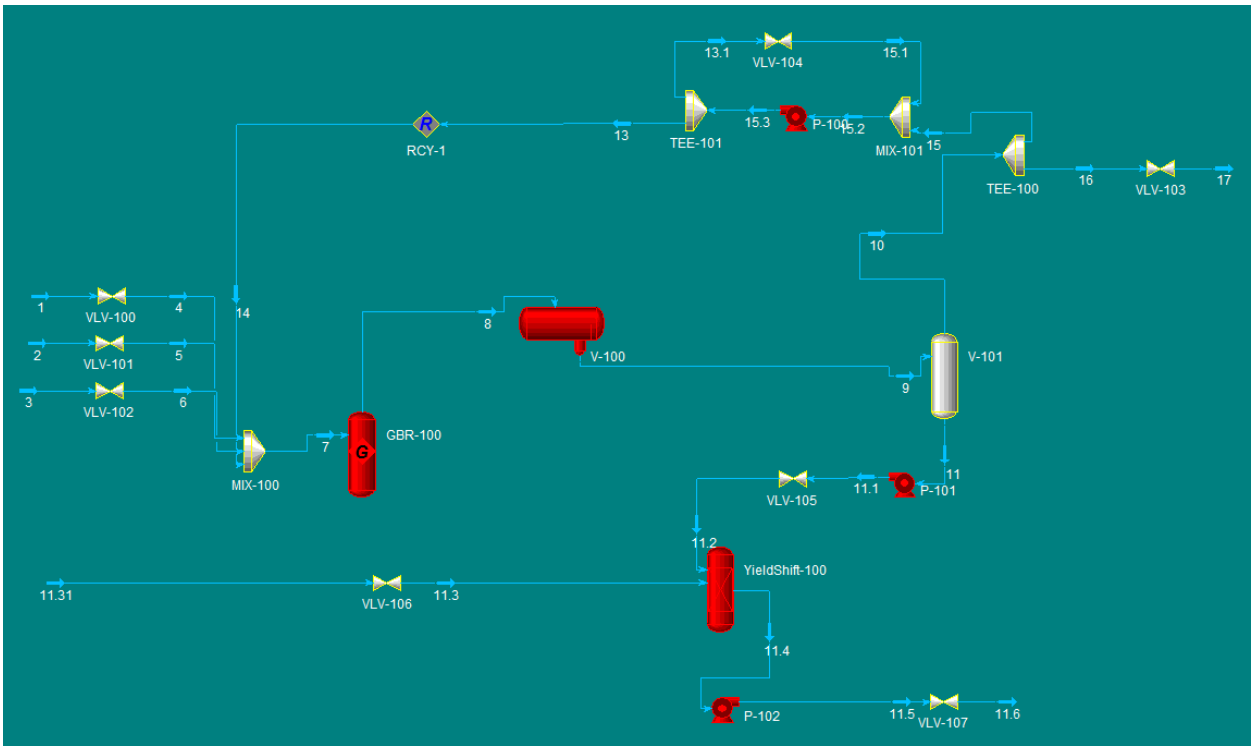


Figure 5.25: TE Process Simulation In AspenHysys

Then in order to have direct and de-noised data, wavelet de-noising technique is used as a data processing technique. Process data analysis is the second phase of the study. In this phase, two methods of genetic programming and artificial neural network as multivariate analysis techniques are applied to identify the quantitative relationship among process variables. As a result, the promising ability of both methods of GP and NARX in pattern recognition is demonstrated and the result compared through calculation mean absolute error. Although prediction of NARX was more accurate in comparison with GP, however both methods show an accurate prediction and analysis which is considerable.

In the last phase of the study, application of GP and NARX as methods of process variable interaction analysis is discussed for the purpose of finding causes and consequences followed by tracing faults associate with them. To this aim, Fault semantic network (FSN) is introduced as a great tool with considerable abilities. In the FSN, nodes correspond to different faults/causes/consequences and directed arcs are links between them describe the dependencies.

Initially, FSN is constructed based on ontology structure of fault models on the basis of process object oriented model (POOM) where failure mode (FM) is described using symptoms, enablers, process variables, causes, and consequences. Implementing the FSN for a process with lots of process variables is not an easy job especially when you are asked to do it in real world. There are steps that should be undertaken by engineers and researchers to let them create a dynamic modeling of a process as a semantic network that contains all the possible faults and possible relationships between variables. In FSN, the strength of the relationship between variables can be assigned both qualitatively and quantitatively through different reasoning approaches such as, probabilistic approach, Fuzzy approach and mathematical formulation.

In a process, there are many of process variables that affect the process operation. In order to implement complete analysis of the process, it is necessary to consider all process variables. In FSN, number of process variables affects the accuracy of analysis. The analysis can be done by using just a manipulated and a measured variable however, it will be incomplete. In summary, the more process variable selected the more accuracy will be obtained.

It is a difficult task to analyze some mechanisms such as corrosion. There are many of parameters that affect corrosion. To implement corrosion related hazard scenarios in FSN, thickness sensors has to be installed according to the standards. Extracted data from the sensor has to be analyzed to uncover the inter-relation pattern among wall thickness and other flow parameters such as pressure, temperature, flow rate and etc.

Considering different accidents related to defined hazard scenarios is another interesting part of the study. The FSN shows the probability of accidents as consequences of hazard scenarios.

In addition, in order to implement FSN for the purpose of fault propagation analysis and hazard identification a software prototype is developed in energy, safety and control laboratory (ESCL) at university of Ontario institute of technology (UOIT).

6.1. Potential Applications

During the development of the proposed method, it has been tried to maintain the generality as much as possible. As a consequence of this effort, the proposed method is envisioned to solve a wide range of Problems. Application of this study is not limited to chemical process plants; however it solves variety of problems in different industrial process, plants and mechanical systems. Some of these applications are already started and implemented in ESCL at UOIT presented respectively as follow.

6.1.1. Risk Calculation

In the FSN, nodes correspond to different Process variables/faults/causes/consequences and directed arcs are links between them describe the dependencies and any node associate with its risk value. Therefore it is possible to calculate risks of occurring failure or accident.

6.1.2. Safety Performance Indicator (SPI)

A considerable work done during the study was developing an indicator that shows safety performance index. It gives this possibility to include the SPI in the FSN, so it can dynamically shows the effectiveness of safety system in a plant as The SPI consists of two main phases that analyze effectiveness of safety system.

- **Predictive actions analysis,** is the first phase that analyze installed safety system in a plant. It includes a database consists of all possible safety barriers (best practice) for each equipment, its associated failure modes and risk reduction factor. It shows how effective the safety system can reduce probability of occurrence an undesired event.

- **Proactive actions analysis,** is the second phase of SPI that analyze consequences of an undesired event from the occupational safety point of view.

6.1.3. Applications in Automated Hazard Identification

Analysis of industrial data sets is an important issue in the industries especially when it comes to be seen from the point of view of safety issues. Identifying hazards and predicting the incident and accident ensure the processes to operate under safe condition and prevent any deficiencies or mal-operation. Detecting inter-relation pattern between variables quantitatively followed by an accurate fault propagation analysis in an intelligent way would be a big step in reaching the aim of Automated Hazard Identification. FSN is a great technique to model system and processes as a semantic network. Once the inter-relation pattern uncovered and FSN built, the basic needs of hazard identification implementation will be met as it allows finding the cause which is manipulated variable, root cause which is the disturbances and consequence which is the measured variable and failure modes. Such information helps operators to act effectively in any faulty and critical situation.

6.2. Future Work

Application of FSN in fault propagation analysis and hazard identification on chemical process is demonstrated through TE process. To further demonstrate its application, ESCL at UOIT is going to preform FSN on LNG, hydrogen production process plant through the simulation. Moreover, to demonstrate its application with real-time data, a process is designed and the prototype is built in the lab as shown in Figure 5.24.

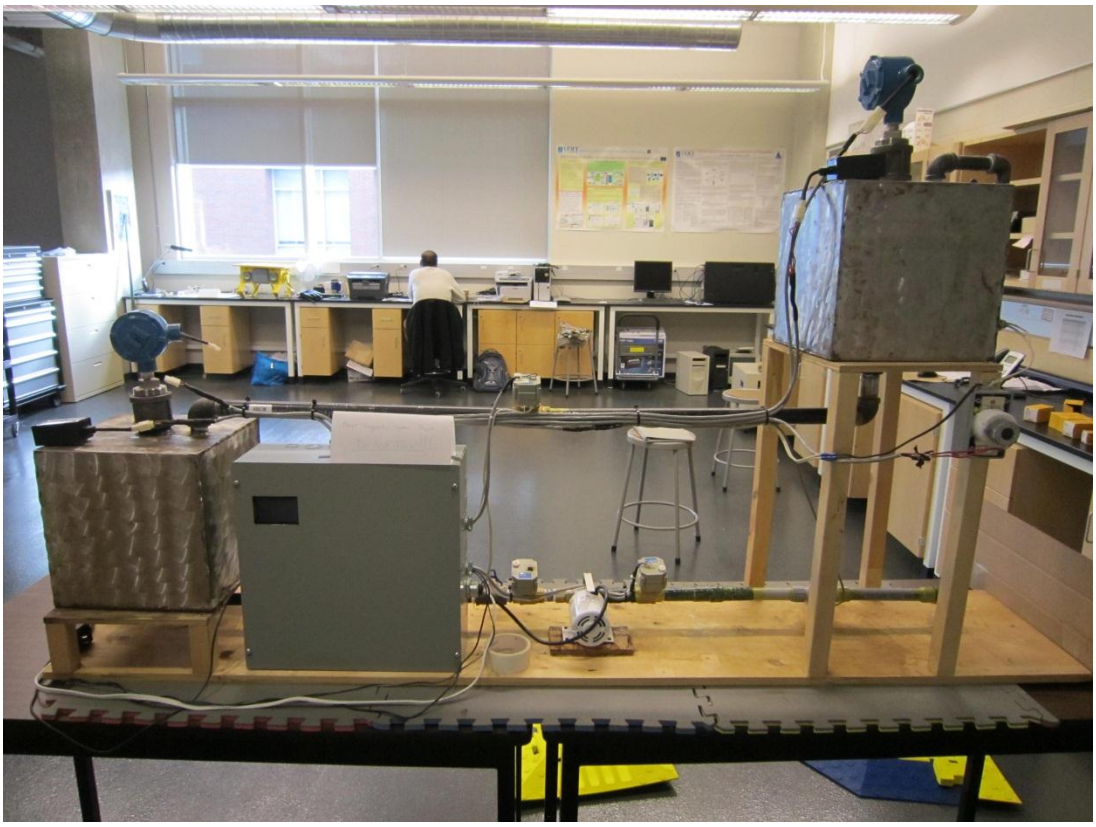


Figure 5.26: FDS Designed in ESCL-UOIT

6.2.1. Application of Automated hazard Identification on LNG process Plant

Application of FSN in fault propagation analysis in demonstrated in this study. As described in previous section, FSN is the main module of automated hazard identification. Given this implementing automated hazard identification in chemical processes is the work that ESCL considered as future work as some parts of the work is discussed in [8]. To this aim, a part of the LNG process plant is selected and simulated by AspenHysys simulation software. The P&ID and simulated process in AspenHysys are presented in Figures 5.25 and 5.26 respectively.

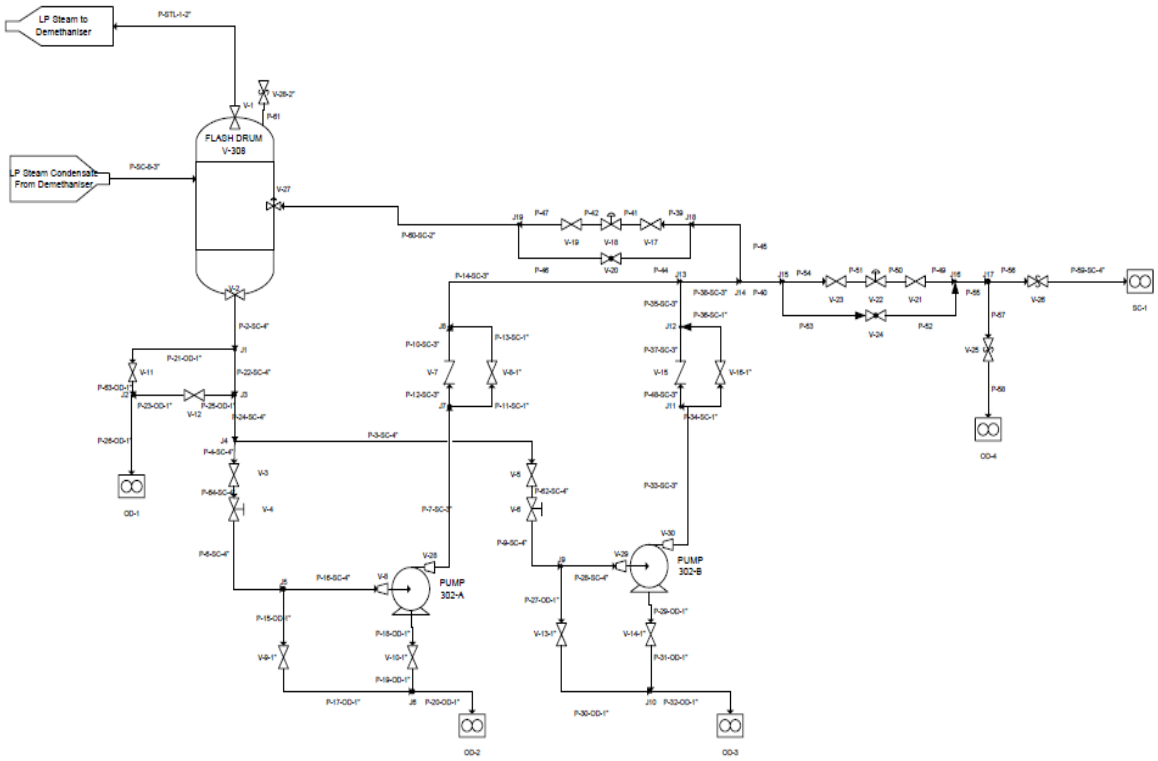


Figure 5.27: P&ID of LNG Process (De-methanizer Unit)

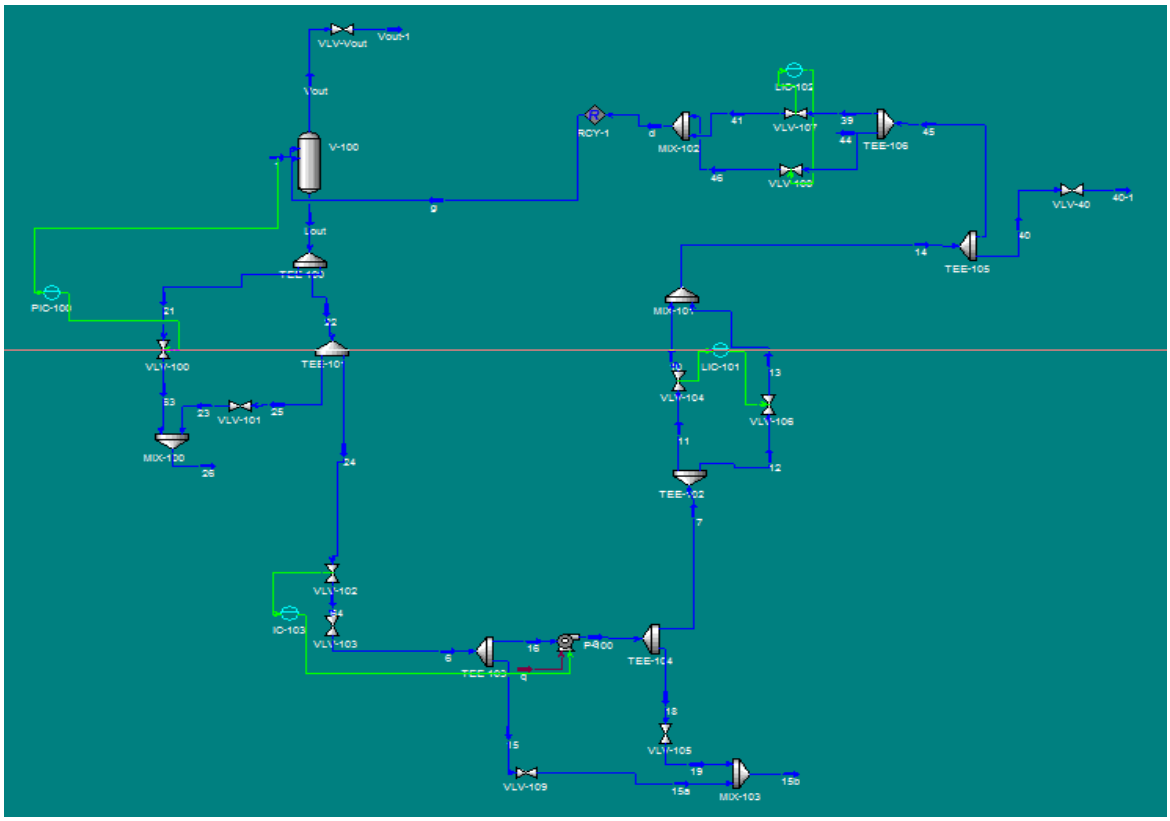


Figure 5.28: AspenHysys Simulation Environment – LNG (Process)

6.2.2. Application of Automated hazard Identification on Hydrogen Production process Plant CU-Cl Cycle

Implementing automated hazard identification on hydrogen production process plant built in Clean Energy Research Lab (CERL) at UOIT is another goal of ESCL as some parts of the work is discussed in [7]. The schematic view of the cycle is shown in Figures 5.27.

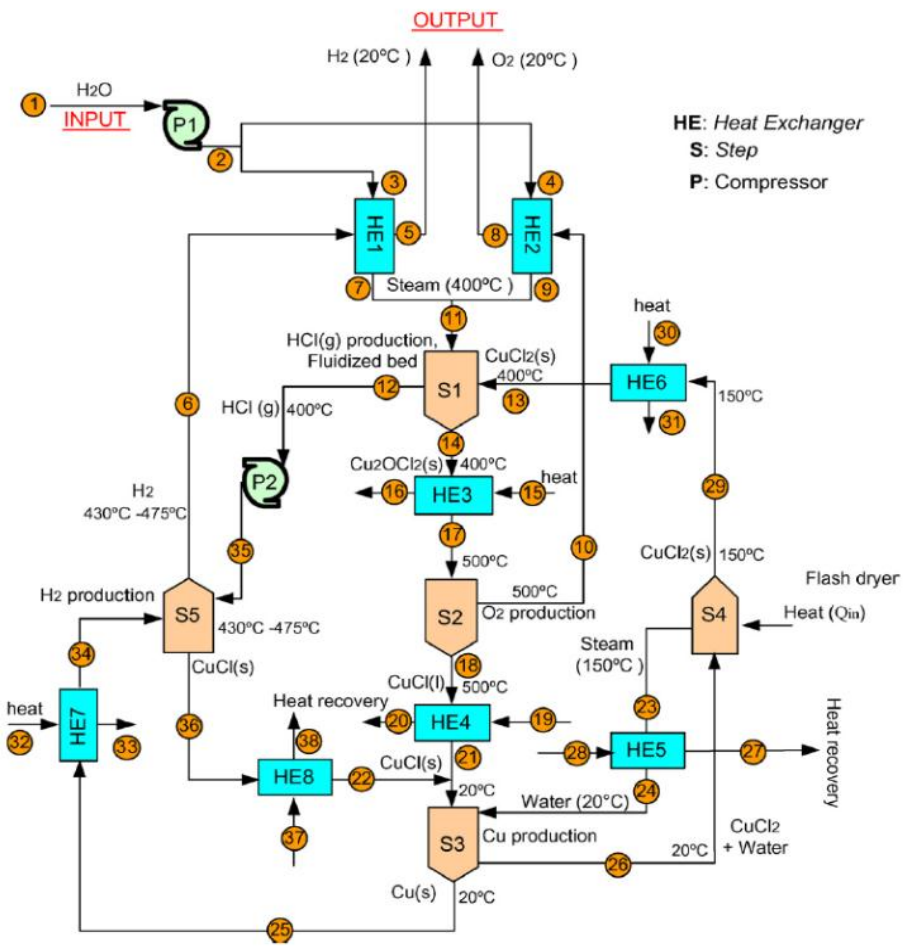


Figure 5.29: Hydrogen Production process – Cu-Cl Cycle [61]

References

[1] Venkat Venkatasubramanian, Raghunathan Rengaswamy, Kewen Yin, Surya N. Kavuri, A review of process fault detection and diagnosis Part I: Quantitative model-based methods, Computers and Chemical Engineering 27 (2003) 293_/311.

[2] Davide Manca, Sara Brambilla, Dynamic simulation of the BP Texas City refinery accident, Journal of Loss Prevention in the Process Industries 25 (2012) 950e957.

[3] en.wikipedia.org/wiki/Bhopal_disaster

[4] CBC News, Canada, Aug 2012

[5] Maryam Kalantarnia, Faisal Khan, Kelly Hawboldt, Modelling of BP Texas City refinery accident using dynamic risk assessment approach, Process Safety and Environmental Protection 88, PP.191-199, 2010

[6] A. Michael Knemeyer, Walter Zinn, Cuneyt Eroglu, Proactive planning for catastrophic events in supply chains, Journal of Operations Management 27 (2009) 141–153.

[7] A.H. Hosseini, H. A.Gabbar, Development of Simulation Based Automated Hazard Identification – Application on Hydrogen Production Plant, IEEE International Workshop on Real Time Measurement, Instrumentation & Control [RTMIC], Oshawa, Canada, June 2011

[8] A.H. Hosseini, H. A.Gabbar, Development of Simulation Based Automated Hazard Identification – Application on Application on Natural Gas Processing Facilities, IEEE International Workshop on Real Time Measurement, Instrumentation & Control [RTMIC], Oshawa, Canada, June 2011

[9] Amir Hossein Hosseini, Sajid Hussain, Hossam A.Gabbar, Detecting Nonlinear Interrelation Patterns among Process Variables Using Genetic Programming and its application in Fault semantic Network, Submitted to Soft Computing journal, Springerlink.

[10] J. J. DoWNS, E. F. VOGEL, a plant wide industrial process control problem, Computers chem. Engng. Vol. 17, No.3, pp. 245-255, 1993.

[11] Venkat Venkatasubramanian, Raghunathan Rengaswamy, Surya N. Kavuri, Kewen Yin, A review of process fault detection and diagnosis, Part III: Process history based methods, Computers and Chemical Engineering 27 (2003) 327_/346

[12] Venkat Venkatasubramanian, Raghunathan Rengaswamy, Surya N. Kavuri, Kewen Yin, A review of process fault detection and diagnosis Part I: Quantitative model-based methods, Computers and Chemical Engineering 27 (2003) 293_/311

[13] Yu Qian, Xiuxi Li, Yanrong Jiang, Yanqin Wen, An expert system for real-time fault diagnosis of complex chemical processes, Expert Systems with Applications 24 (2003) 425–432

[14] Cen Nan, Faisal Khan, M. Tariq Iqbal, Real-time fault diagnosis using knowledge-based expert system, p r o c e s s sa f e t y and e n v i ronmental p rot e c ti on 8 6 (2 0 0 8) 5 5– 7 1.

[15] Gelgele, H. and Wang, K., "An Expert System for Engine Fault Diagnosis: Development and Application," Journal of Intelligent Manufacturing, vol. 9, pp. 539–545, 1998.

[16] Anil K. Jain, Robert P.W. Duin, Jianchang Mao, Statistical Pattern Recognition: A Review, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 22, NO. 1, JANUARY 2000

[17] Luciano Nieddu, Giacomo Patrizi, Invited Review Formal methods in pattern recognition: A review, European Journal of Operational Research 120 (2000) 459±495.

[18] Chris Westbury, lori Buchanan, Michael Sanderson, Mijke Rhemtulla, leah Phillips, Using genetic programming to discover nonlinear variable interactions, Behavior Research Methods, Instruments, & Computers 2003, 35(2), 202-216 .

[19] R. Leardi, R. Boggia, M. Terrile, Genetic Algorithm as a strategy for feature selection, Journal of Chemometrics, Vol.6, PP. 267-281, 1992.

[20] Adil Baykasoglu, Hamza Gullu, Hanifi, anakc, Lale Ozbakkz, Prediction of compressive and tensile strength of limestone via genetic programming, Expert Systems with Applications 35 (2008) 111–123.

[21] Miller, III, W.T., R.S. Sutton, P.J. Werbos, Neural Network in Control, MIT press, Cambridge, 1990.

[22] H.N.Kolvo, Artificial neural network in fault diagnosis and control, Control Eng. Practice, Vol.2, No. 1, PP.89-101, 1994.

[23] Diego Ruiz, Jose Maria Nougues, Zuly Calderon, Antonio Espuna, Luis Puigjaner, Neural Network Based Framework for Fault Diagnosis in Batch Chemical Process, Computers and Chemical Engineering 24, PP.777-784, 2000.

[24] M. Karpenko, N. Sepehri, D. Scuse, Diagnosis of process valve actuator faults using a multilayer neural network, Control Engineering Practice 11 (2003) 1289–1299

[25] Simani, S. and Fantuzzi, C., “Fault Diagnosis in Power Plant Using Neural Networks,” Information Sciences, vol. 127, pp. 125–136, 2000.

[26] R. Mohammadi, E. Naderi, K. Khorasani, and S. Hashtrudi-Zad, Fault Diagnosis of Gas Turbine Engines by Using Dynamic d their results Neural Networks, 2011 IEEE ICQR

[27] Mouloud Oussaidkne, Bastien Chopard,O livier V. Pictet, Marco Tomassini, Parallel genetic programming and its application to trading model induction, Parallel Computing 23 (1997) 1183-I 198.

[28] A. Salhi, H. Glaser, D. De Roure, Parallel implementation of a genetic-programming based tool for symbolic regression, Information Processing Letters 66 (1998) 299-307.

[29] Abdüsselam Altunkaynak, Sediment load prediction by genetic algorithms, Advances in Engineering Software 40 (2009) 928–934.

[30] Satishkumar S. Kashid, Rajib MaityPrediction of monthly rainfall on homogeneous monsoon regions of India based on large scale circulation patterns using Genetic Programming, Journal of Hydrology 454–455 (2012) 26–41.

[31] Sourabh Dash, Raghunathan Rengaswamy, Venkat Venkatasubramanian, Fuzzy-logic based trend classification for fault diagnosis of chemical Processes, Computers and Chemical Engineering 27 (2003) 347_/362

[32] Sampath, S. and Singh, R., “An Integrated Fault Diagnostics Model Using Genetic Algorithm and Neural Networks,” Journal of Engineering for Gas Turbines and Power, vol. 128, pp. 49–56, 2006.

[33] Carlos Rojas-Guzman, Mark A. Kramer, Comparison of belief network and rule-based expert system for fault diagnosis of chemical processes, Engng. applic. Artif. Intell, Vol. 6, No.3, PP.191-202, 1993.

[34] Bickford, R. and Malloy, D., “Development of A Real-Time Turbine Engine Diagnostic System,” in 38th AIAA/ASME/SAE/ASEE Joint Propulsion Conference, (Indianapolis, Indiana), American Institute of Aeronautics and Astronautics, July 7-10 2002.

[35] Romessis, C., Stamatis, A., and Mathioudakis, K., “Setting Up a Belief Network for Turbofan Diagnosis with the Aid of an Engine Performance Model,” in 15th International Symposium on Air Breathing Engines, (Bangalore, India), International Society for Air Breathing Engines, September 2001. 164

[36] Amid Bakhtazad, Ahmet Palazoglu, Jose Romagnoli, Process trend analysis using wavelet-based de-noising, Control Engineering Practice 8 (2000) 657-663

[37] Sarita Dangeti, DENOISING TECHNIQUES - A COMPARISON, A Thesis Submitted to the Graduate Faculty of the Louisiana State University, USA, 2003.

[38] A. BUADES, B. COLL, J. M. MOREL, A REVIEW OF IMAGE DENOISING ALGORITHMS, WITH A NEW ONE, MULTISCALE MODEL. SIMUL. Vol. 4, No. 2, pp. 490–530

[39] D. Donoho, Nonlinear Wavelet Methods for Recovery of Signals, Densities, and Spectra from Indirect and Noisy Data," Different Perspectives on Wavelets, Proceeding of Symposia in Applied Mathematics, Vol. 47, I. Daubechies ed. Amer. Math. Soc., Providence, R.I., 1993, pp. 173-205.

[40] Graps, A. , An introduction to wavelets, Computational Science & Engineering, IEEE, Volume: 2 , Issue: 2 , PP. 50 – 61.

[41] Edmundo G. de Souza e Silva, Luiz F.L. Legey and Edmundo A. de Souza e Silva, "Forecasting oil price trends using wavelets and hidden Markov models", Energy Economics 32 (2010) 1507–1519.

[42] Graps, A., 1995. An introduction to wavelets, IEEE Computational Science and Engineering 2 (2), 50–61.

[43] Vidakovic, B., Muller, P., 1994. Wavelets for Kids. Duke University. <http://www2.isye.gatech .edu/~brani/wp/kidsA.pdf>.

[44] J.R. Koza, Genetic Programming: on the Programming of Computers by Means of Natural Selection, MIT Press, Cambridge, 1993.

[45] Jeroen Eggermont, Genetic Programming, Handout, Leiden University Medical Center, The Netherlands.

[46] Matthew Walker, Introduction to Genetic Programming, 2001.

[47] Nikola K. Kasabov, Foundations of Neural Networks, Fuzzy Systems, andKnowledge Engineering, Massachusetts Institute of Technology, 2nd printing, 1998

[48] Girish Kumar Jha, Artificial neural network and its application, Indian agricultural statistics research center.

[49]Simon Hykin, Neural network a comprehensive foundation, 2nd edition, Pearson education, Inc. 1999

[50] Hossam A. Gabbar, Fault Semantic Networks for Accident Forecasting of LNG Plants, Knowledge-Based and Intelligent Information and Engineering Systems, Lecture Notes in Computer Science, 2010, Volume 6277/2010, 427-437

[51] Collins, A. M., & Quillian, M. R. Retrieval time from semantic memory. Journal of Verbal Learning and Verbal Behavior, 8, 240–248, 1969.

[52] Kevin B. Korb, Ann E. Nicholson, Bayesian Artificial Intelligence, Taylor & Francis e-Library, 2005

[53] J. J. DoWNS, E. F. VOGEL, a plant wide industrial process control problem, Computers chem. Engng. Vol.17, No.3, pp.245-255, 1993.

[54] Truls Larsson, Kristin Hestetun, Espen Hovland, and Sigurd Skogestad, Self-Optimizing Control of a Large-Scale Plant: TheTennessee Eastman Process, Ind. Eng. Chem. Res. 2001, 40, 4889-4901

[55] McAvoy, T. J.; Ye, N. Base control for the Tennessee Eastman problem. Comput. Chem. Eng. 1994, 18 (5), 383-413.

[56] Arvid Sukumaran, Mater thesis, Neural Network Based Control Strategy For the Two Phase Reactor of the Tennesse Eastman Process, University of Masachusetts Lowell, 2003.

[57] McCulloch, W. S. and Pitts, W. H, A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics, 5:115-133, 1943.

[58] AspenTech Operation Guide, Aspen Technology, Inc, 2005.

[59] Sara Silva, GPLAB A Genetic Programming Toolbox for MATLAB, Version.3, 2007.

[60] Karl-Rudolf Koch, Introduction to Bayesian Statistics, 2nd Edition, Springer.

[61] Mehmet F. Orhan, Ibrahim Dincer∗, Marc A. Rosen, Thermodynamic analysis of the copper production step in a copperpper o Bayesian Statistics, 2005.. BulleThermochimica Acta 480, PP.22PP., 2008.

[62] M.A. Bin Shams, H.M. Budman, T.A. Duever, Fault detection, identification and diagnosis using CUSUM based PCA, Chemical Engineering Science 66, PP. 4488–4498, 2011.

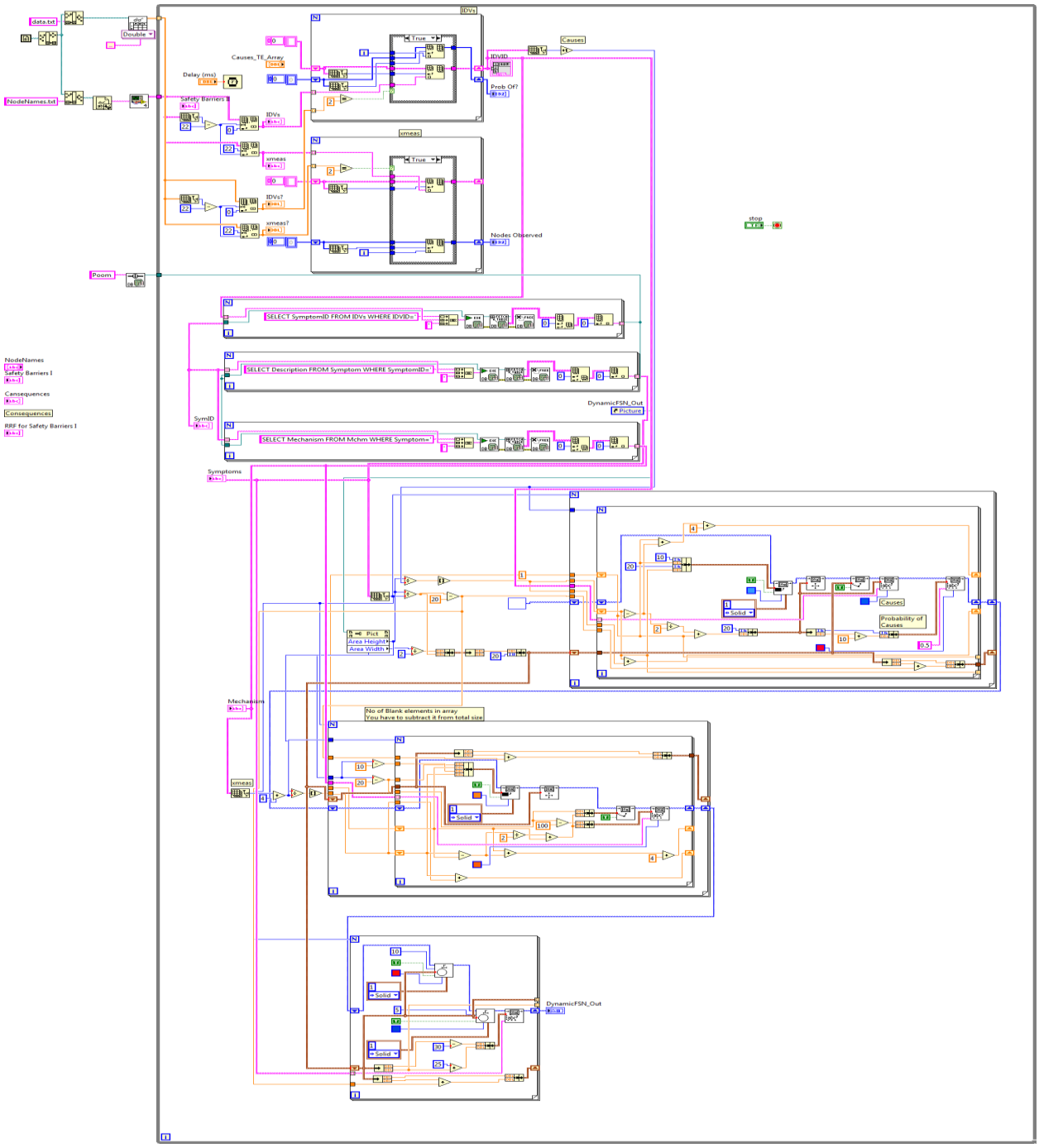
[63] Haifeng Chen, Principal Component Analysis With Missing Data and Outliers, Electrical and Computer Engineering Department Rutgers University, 2002.

APPENDIXES

Appendix 1: GP MATLAB Code:

```
function [v,b]=Hysys

fprintf('Running symbolic regression demo...');
p=resetparams;

p=setoperators(p,'crossover',2,2,'mutation',1,1);
p.operatorprobstype='variable';
p.minprob=0;

p.datafilex='….txt';
p.datafiley='….txt';

p.usetestdata=1;
p.testdatafilex='…..txt';
p.testdatafiley='…..txt';

p.calcdiversity={'uniquegen'};
p.calccomplexity=1;
p.graphics={'plotfitness','plotdiversity','plotcomplexity','plotoperators'};
p.depthnodes='2';

[v,b]=gplab(25,50,p);

desired_obtained(v,[],1,0,[]);
accuracy_complexity(v,[],0,[]);

figure
plotpareto(v);

drawtree(b.tree)[59];

…
```

Appendix 2: FSN – LabVIEW® Programing



Appendix 3: List of Papers

- A.H. Hosseini, H. A.Gabbar, Development of Simulation Based Automated Hazard Identification – Application on Hydrogen Production Plant, IEEE International Workshop on Real Time Measurement, Instrumentation & Control [RTMIC], Oshawa, Canada, June 2011

- A.H. Hosseini, H. A.Gabbar, Development of Simulation Based Automated Hazard Identification – Application on Application on Natural Gas Processing Facilities, IEEE International Workshop on Real Time Measurement, Instrumentation & Control [RTMIC], Oshawa, Canada, June 2011

- Amir Hossein Hosseini, Sajid Hussain, Hossam A.Gabbar, Detecting Nonlinear Interrelation Patterns among Process Variables Using Genetic Programming and its application in Fault semantic Network, Submitted to Soft Computing journal, Springerlink.

- Amir Hossein Hosseini, Sajid Hussain, Hossam A.Gabbar, Simulation-Based Fault Diagnosis of Process Industry using Process Variable Interaction. In progress.