FINAL YEAR PROJECT II: DISSERTATION

Structural Fault Detection Using Weighted Principal Component Analysis (WPCA)

Prepared by:

Muhamad Fazlin bin Abdul Rahim

14655

Supervisor:

Ir Dr Abdul Halim Shah B Maulud

Preliminary report submitted in partial fulfillment of the requirement for the

Bachelor of Engineering (Hons)

(Chemical Engineering)

MAY 2014

Universiti Teknologi PETRONAS

Bandar Seri Iskandar

31750 Tronoh,

Perak Darul Ridzuan.

CERTIFICATION OF APPROVAL

Structural Fault Detection Using Weighted Principal Component Analysis (WPCA)

by

Muhamad Fazlin bin Abdul Rahim
14655

A project dissertation submitted to the
Chemical Engineering Programme
Universiti Teknologi PETRONAS
in partial fulfilment of the requirement for the
BACHELOR OF ENGINEERING (Hons)
(CHEMICAL)

Approved by,

____
(Ir Dr Abdul Halim Shah B Maulud)

UNIVERSITI TEKNOLOGI PETRONAS
TRONOH, PERAK
May 2014

CERTIFICATION OF ORIGINALITY

This is to certify that I am responsible for the work submitted in this project, that the original work is my own except as specified in the references and acknowledgements, and that the original work contained herein have not been undertaken or done by unspecified sources or persons.

MUHAMAD FAZLIN BIN ABDUL RAHIM

ABSTRACT

Fault that occurred in a system is actually affecting the quality of the products produced and as a result, the process monitoring is required to eliminate the fault in the system and eventually increase and met the performance specification. Principal Component Analysis(PCA) is a method that have been introduced in process monitoring to detect the fault in the system and it has been categorized as one of the method of Multivariable statistical process monitoring (MSPM) as its ability to monitor multivariable system. The extension of PCA is proposed which is Weighted Principal Component Analysis (WPCA) to deal with the situation of useful information being submerged and reduced missed detection rate of T^2 statistic. The main idea of WPCA is building conventional PCA model and then using change rate of T^2 statistic along every principal component (PC) to capture the most useful information in process, and setting different weighting values for PCs to highlight useful information.WPCA method will be focusing on how to detect structural fault since most of the literatures only focusing on the variable change. In this paper, structural fault will be simulated using that CSTR model which will be developed using MATHLAB software. Lastly, the process data will be collected and tested with WPCA.

ACKNOWLEDGEMENT

I would like to express my deepest gratitude to the Chemical Engineering department of Universiti Teknologi PETRONAS (UTP) for providing the chance to undertake this remarkable Final Year Project (FYP). My knowledge and skills has been put to a test after completing various kinds of project during my five years intensive chemical engineering course. This course has a good coverage on the overall chemical engineering programme whereby a student with any majors has been assigned with different scope of the study thus contribute the effort and knowledge towards achieving a project goal.

Most important, a very special note thanks to my supervisor Ir Dr Abdul Halim Shah B Maulud, who was always willing to assist and provided good support throughout the project completion. Your excellent support, patience and effective guidance have helped my project to completion. I would like to thank to FYP1 and FYP2 coordinators, Dr Maziyar Sabet and Dr Asna respectively for arranging various seminars as support and knowledge to assist the group in the project. The seminars were indeed very helpful and insightful to us.

` Last but not least my heartfelt gratitude goes to my family and friends for providing me continuous support throughout the duration of this project.

TABLE OF CONTENT

Contents

List of Figures

List of Table

CHAPTER 1: INTRODUCTION

1.1 Background of Study

According to Chiang (2000), “In the process and manufacturing industries, there has been a large push to produce higher quality products, to reduce product rejection rates, and to satisfy increasingly stringent safety and environmental regulations. Process operations that were at one time considered acceptable are no longer adequate. To meet the higher standards, modern industrial processes contain a large number of variables operating under closed loop control. The standard process controller (PID controllers, model predictive controllers, etc) are designed to maintain satisfactory operations by compensating for the effects of disturbances and changes occurring in the process. While these controllers can compensate for many types of disturbances there are changes in the process which the controller cannot handle adequately. These changes are called faults. More precisely, a fault is defined as an unpermitted deviation of at least one characteristic property or variable of the system.” The process fault that happened in a system could be divided into two which are variable change and structural change. Variable is a typical form of disturbance trajectories include step changes and exponential variations usually observed in the variables themselves. Structural change happens when the governing characteristics of the process changes.

Over the past 20 years, the chemical industry has made a concerted effort to streamline operations. Their goal was simply to produce products as many as possible. Nowadays, as the market is highly competitive worldwide, production efficiency and product consistency become essential to success. Even though many chemical processes have been around for years and engineers have acquired lots of experience, many operational problems and inefficiencies still go undiagnosed for a prolonged period of time. Therefore, process monitoring and diagnosis are strongly required to produce the product and maintain the process equipment. For example, a heat exchanger that becomes fouled over a period of time may be unnoticed because it has no effect on the final product. Yet the incremental amount of the steam needs to be adjusted for fouling costs a significant amount of money. Process problems like this one should be monitored, detected and diagnosed (Chen, 2002).

Nowadays, industrial processes are more and more complex for that reason they include a lot of sensors. Consequently, an important amount of data can be obtained from a process. A process dealing with many variables can be named multivariate process. However, the monitoring of a multivariate process cannot be reduced to the monitoring of each process variable because the correlations between the variables have to be taken into account. Process monitoring is an essential task. The final goal of the process monitoring is to reduce variability, and so, to improve the quality of the product (Montgomery, 1997).

In ensuring the operation of the system met the performance specification, process monitoring is essential so that the fault in the operation could be detected, diagnosed and eliminated (Chiang, 2000). The four procedures associated with process monitoring are: fault detection, fault identification, fault diagnosis and fault recovery (Chiang, 2000).Univariate stastical monitoring is one of the methods used in process monitoring to detect changes or fault in the industrial system where it is used to monitor only small number of process variable. As this method caused difficulties in monitoring multivariable system so, Multivariable statistical process monitoring (MSPM) was introduced (Tatara, 2002).

Multivariate statistical process monitoring approaches have progressed significantly in recent years and among them principal component analysis (PCA) as a classical method is the most widely used (Jiang, 2012).In general, Principal component analysis (PCA) is a reliable and simple technique for capturing variable relation and allows extension of principles of univariate statistical process monitoring (SPM) to multivariate process monitoring.Jiang (2012) added that currently, many extensions of PCA, such as Kernel PCA (KPCA), Dynamic PCA (DPCA), Probabilistic PCA (PPCA) and Multiway PCA (MPCA), and so on, have been proposed to improve the performance of process monitoring and solve more problems. Weighted Principal Component Analysis (WPCA) is also one of the advanced-PCA methods that will be studied in by the author.

This project will be focusing on the detecting the fault that occur in a system especially in structural fault which is happens when the governing characteristics of the process changes. In this project, the weighted principal component analysis (WPCA) was proposed to be one of method to improve the performance of process monitoring and this method was compared to Principle Component Analysis (PCA) and Squared Prediction Error (SPE) also known as Q statistic.

1.2 Problem Statement

Large amounts of data are collected in many industrial processes. The task of fault detection is to use this data to determine when abnormal process behavior has occurred, whether associated with equipment failure, equipment wear, or extreme process faults (Russell, 2000).Different kind of methods have been used in detecting the process fault by using those data. One of the familiar methods is Principal Component Analysis (PCA) which is part of multivariate statistical monitoring techniques. One of the extensions of PCA is Weighted Principal Component Analysis (WPCA).However based on recent literatures, these kind of methods mostly focusing on the variable changes that occur on the system and the structural fault that occurred in a system was ignored. For instant based on one literature by Qingchao Jiang(2012), WPCA has simulated 21 faults and from that literature only one structural fault has been tested. As the result, further studies to improve in structural fault detection will be done in this project.

1.3 Objectives

There are several objectives have been identified for the purpose of this project. The two main objectives for this project are:

1. To develop CSTR model and generate structural fault.

2. To investigate the performance of Weightage Principal Component Analysis (WPCA) compared to PCA T^2 and Q statistic.

1.4 Scope of Study

The scope of study is as the following:

- Develop model which is CSTR simulation model using the Mathlab software.
- Structural fault will be simulated using that model.
- Finally, the process data will be collected and tested with advanced PCA method which is WPCA.

CHAPTER 2: LITERATURE REVIEW AND THEORY

2.1 Univariate Statistical Monitoring

Statistical methods for detecting changes in industrial processes are included in a field generally known as statistical process control (SPC) or statistical quality control. The most widely used and popular SPC techniques include univariate methods that involve observing a single variable at a given time, obtaining the mean and variance of the variable, and checking its value against upper and lower control limits. A univariate approach may indeed work for monitoring a small number of process variables that are not correlated (Eric Tatara, 2002).

Chiang (2000) stated that "a univariate statistical approach to limit sensing can be used to determine the threshold for each observation variable (a process variable observed through a senor reading), where these thresholds define the boundary for in-controlled operations and a violation of these limit with on-line data would indicate a fault. This approach is typically employed using a Shewhart chart (Figure 1) and has been referred to as limit sensing and limit value checking. The values of the upper and the lower control limits on the Shewhart chart are critical to minimizing the rate of false alarms and the rate of missed detections. A false alarm is an indication of a fault, when in actuality a fault has not occurred; a missed detection is no indication of a fault, though a fault has occurred.

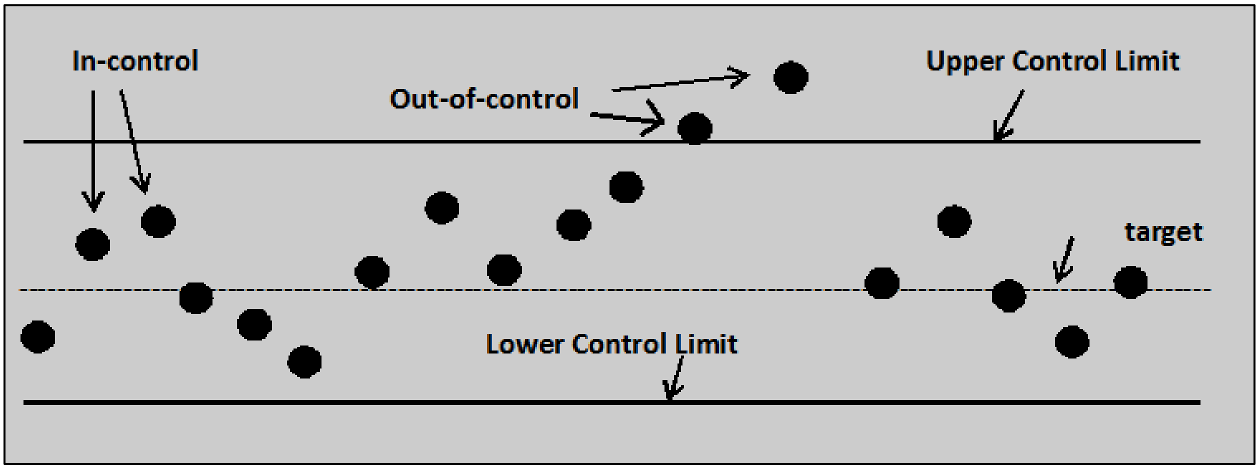


Figure 1 : An illustration of the Shewhart chart. The black dots are observation (Chiang, 2000)

2.2 Multivariate Statistical Monitoring

Eric Tatara(2002) stated that "application of univariate statistical process monitoring (SPM) methods to larger multivariable systems becomes difficult, if not impossible, and is often erroneous. This simplified approach to process monitoring requires an operator to continuously monitor perhaps dozens of different univariate charts, which substantially reduces the ability of plant personnel to make accurate assessments about the state of the process". As a result, Multivariable statistical process monitoring (MSPM) techniques was introduced. He added that multivariable statistical process monitoring (MSPM) techniques offer the proper theoretical framework for monitoring multivariable processes.MSPM techniques reduce the amount of raw data presented to an operator and provide a concise set of statistics that describes the process behavior. Many of the current MSPM techniques are only valid for data that are independent and identically distributed.

According to Sankar Mahadevan(2009),over the past few years several multivariate statistical process monitoring (MSPM) data based tools such as principal components analysis (PCA) , dynamic principal components analysis (DPCA), canonical variate analysis

(CVA)(Russel et al.,2000) , modified independent component analysis (MICA)(Lee et al.,2004) , kernel principal component analysis (KPCA)(Lee et al.,2007), kernel independent component analysis (KICA) and correspondence analysis (CA) have been developed. These techniques mainly consist of the following preliminary steps:

- developing a model based on the normal operating data;
- proposing a distance metric and setting appropriate thresholds based on a predefined confidence measure;
- projecting a new test data onto this model, calculating the distance metric and appropriately classify it as normal or faulty data;
- in the event of a fault, identifying variables that are related to the fault using appropriate contribution measures; and
- Identifying the root cause of the fault.

2.3 Process Monitoring Procedures

Chiang (2000) mentioned that the types of faults occurring in industrial system include process parameter changes, disturbance parameter changes, actuator problems and sensor problems. He added that to ensure the process operations satisfy the performance specifications, the faults in the process need to be detected, diagnosed and removed. These tasks are associated with process monitoring.

According to Verron (2010), the process monitoring includes four procedures: fault detecting (decide if the process is under normal condition or out-of-control); fault identification (identify the variables implicated in an observed out-of-control status); fault diagnosis (find the root cause of the disturbance); process recovery (return the process to a normal status).

In the other words, Chiang (2000) stated that “Fault detection is determining whether a fault occurred. Early detection may provide invaluable warning on emerging problems, with appropriate actions taken to avoid serious process upsets. Fault identification is identifying the observation variables most relevant to diagnosing the fault. The purpose of this procedure is to focus the plant operator's and engineer's attention on the subsystems most pertinent to the diagnosis of the fault, so that the effect of the fault can be eliminated in a more efficient manner. Fault diagnosis is determining which fault occurred, in other words, determining the cause of the observed out-of-control status. The fault diagnosis is procedure is essential to the counteraction or elimination of the fault. Process recovery, also called intervention, is removing the effect of the fault, and it is the procedure needed to close the process monitoring loop (Figure 2). ”

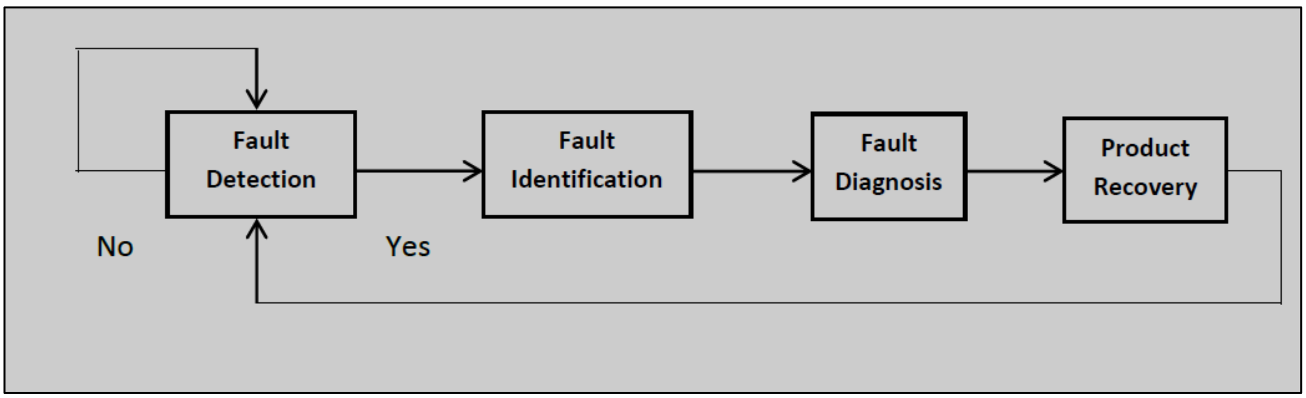


Figure 2 : Scheme of the Process Monitoring Loop (Chiang, 2000)

2.4 Types of Process Faults

Process disturbance or fault in a monitoring can be classified into two types which are:

I. Variable change
   - Variable is a typical form of disturbance trajectories include step changes and exponential variations usually observed in the variables themselves. For example, changes in feed composition, temperature, pressure or impurity levels. This type of faults can be effectively detected if suitable univariate process monitoring techniques are properly implemented.

II. Structural change

- Structural change happens when the governing characteristics of the process changes. For example, a drift in reaction kinetics which might due to catalyst deactivation or a change in heat transfers due to a fouling in heat exchanger. This results a change in a process relationship between variables in a process.

2.5 Principal Component Analysis (PCA)

Industrial process data are usually multivariate in nature and are highly correlated. PCA mainly aims at decorrelating this data and projection of the data in a relatively lower dimensional subspace (Sankar Mahadevan, 2009).In using PCA model, two statistics are constructed to interpret the mean and variance information of process, known as T^2 statistic and Q (also known as Squared Prediction Error, SPE) statistic (Chen et. al, 2004).

Basically, PCA which is one of the multivariate statistical analysis techniques have long been used for detection and diagnosis of abnormal operating situations in many industrial processes. In general, they build a model from normal process data and then compare the abnormal process status against the predefined monitoring model. The major advantages of these multivariate statistical analysis methods are their ability to handle larger numbers of highly correlated variables and reduce the high-dimensional process measurement space into a low-dimensional latent variable space (Zhao, 2014).

$$X = \begin{matrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{matrix} \qquad \text{Eq. (1)}$$

There is a published equation of principle component analysis as a linear dimensionality reduction technique which determines a set of orthogonal vectors, called loading vectors, ordered by the amount of variance explained in the loading vector direction. Given a training set of n observations and m process variables stacked into a matrix X as in Eq. (1), the loading vectors are calculated by solving the stationary points of the optimization problem

$$\max_{v \neq 0} \frac{v^T X^T X v}{v^T v} \quad \text{Eq. (2)}$$

were v $v \in R^m$.Chiang(2000) also stated that the stationary point of Eq.(2) can be computed via the singular value decomposition(SDV)

$$\frac{1}{\sqrt{n-1}} X = U \sum V^T \quad \text{Eq. (3)}$$

where $U = R^{n \times n}$ and $V = R^{m \times m}$ are unitary matrices and the matrix $\sum = R^{n \times m}$ contains the non-negative real singular values of decreasing magnitude along its main diagonal ($\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min(m,n)} \geq 0$) and zero off diagonal elements . The loading vectors are the orthonormal column vectors in the matrix V, and the variance of the training set projected along the i^{th} column of V is equal to σ_i^2.

2.6 Fault Detection

Fault detection is one of the steps in the process monitoring and it is described as the step taken to decide if the process is under normal condition or out-of-control (Verron, 2010).Russell (2000) had published an equation that stated that normal operations can be characterized by employing Hotelling's T^2 statistic

$$T^2 = x^T P \sum_a^{-2} P^T x \quad \text{Eq. (4)}$$

where P includes the loading vectors associated with the a largest singular values, $\sum_a$ contains the first a rows and columns of $\sum$, and x is an observation vector of dimension m. Given a number of loading vectors, a, to include in Eq.(3), the threshold can be calculated for the T^2 statistic using the probability distribution

$$T_{\alpha}^{2} = \frac{(n^2-1)}{n(n-a)} F_a(a, n-a) \qquad \text{Eq. (5)}$$

where $F_a(a, n-a)$ is the upper $100a$ % critical point a of the F-distribution with a and $n-a$ degrees of freedom. The T_2 statistic with Eq.(5) defines the 2 normal process behavior, and an observation vector outside this region indicates that a fault has occurred. Russell (2000) added that the portion of the measurement space corresponding to the lowest $m-a$ singular values can be monitored by using the Q statistic developed by Jackson and Mudholkar:

$$Q = r^T r, \qquad r = (1 - PP^T)x \qquad \text{Eq. (6)}$$

He added that the threshold for the Q statistic can be calculated from its approximate distribution:

$$Q_a = \theta_a \left[\frac{\mathrm{h}_0 c_\alpha \sqrt{2\theta_2}}{\theta_1} + 1 + \frac{\theta_2 \mathrm{h}_0 (\mathrm{h}_0 - 1)}{\theta_1^2} \right]^{1/\mathrm{h}_0} \qquad \text{Eq.(7)}$$

where $\theta_1 = \sum_{j=a+1}^{n} \sigma_j^{2i}$, $h_0 = 1 - (2\theta_1\theta_2)/3\theta_2^2$,and c_α is the normal deviate corresponding to the $(1-\alpha)$ percentile.

2.7 Weighted Principal Component Analysis (WPCA)

WPCA is one of the extend PCA in order to increase the performance of process monitoring and solve problems in industries. Firstly, WPCA uses normal operational data to build conventional PCA model. Secondly, change rate of T^2 statistic along each principal component is constructed to capture the most useful information in process and select the principal components with useful information for online monitoring. Distinct weighting values are then set on different principal components and T^2 and Q statistics are calculated to determine the state of process. (Qingchao Jiang, 2012).He added that the main merit of the proposed WPCA is not only using normal operational process data to build PCA model, but also taking fault information into consideration. It determines the weighting values according to the importance of the PC objectively, to identify the useful components as well as useless ones. In addition, he also mentioned that the idea of WPCA is to adaptively set different weighting values on different principal components; to highlight the importance of principal components with significant information of process variation. The mathematical representative will be shown in the following section.

2.7.1 Mathematical Representative

1. Suppose the loading matrix $P = [p_1, p_2, \ldots, p_k] \in R^{s\times k}$,where s is the number of principals components retained. $p_k \in R^{s\times k}$ is the loading vector corresponding to the kth principal component.
2. Set a weighting matrix $W = \begin{bmatrix} W_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & W_k \end{bmatrix} \in R^{s\times k}$ on P,then the weighted loading matrix:

   $$P_w = PW = [p_1, p_2, \ldots\ldots p_k]\begin{bmatrix} W_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & W_k \end{bmatrix} = [w_1 p_1, w_2 p_2, \ldots, w_k p_k]$$
3. The weighted principal components:

   $$T_W = XP_W$$

4. The T^2 statistic after weighted becomes :

$$T^2 = x^T(\Lambda_k)^{-1}W^TP^Txx^TP\begin{bmatrix}\frac{W_1}{\lambda_1} & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \frac{W_k^2}{\lambda_k}\end{bmatrix}P^Tx$$

CHAPTER 3: METHODOLOGY

3.1 Develop Model-CSTR Simulation Model

Jana(2011) stated that "the continuous stirred tank reactor (CSTR) or backmix reactor is a very common processing unit in chemical and polymer industry. The name suggests that it is a tank type reactor in which the contents are well stirred and it runs with continuous flow of reactants as well as products. The CSTR is normally run at steady state. The main feature of this type of reactor is the complete uniformity of concentration and temperature throughout the reactor due to the perfect mixing. Also, the concentration and temperature of the material leaving the tank must be exactly the same as those of the material in the tank. The CSTR is widely used for large-scale production. The continuous operation results in more consistent product properties, an improved energy consumption (for example, the exothermic heat can be utilized to heat feed streams) and a higher productivity through the reduction of inactive periods (filling, heating, cooling and emptying)".

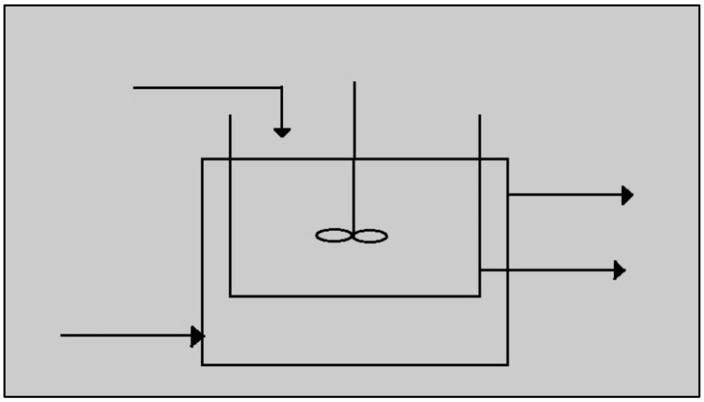


Figure 3 Schematic representation of CSTR

He added that there are some assumptions have been mode in developing CSTR model using MATLAB software:

1. The heat losses from the process are negligible (well-insulated).
2. The mixture density and heat capacity are assumed constant.
3. There are no variations in concentration, temperature, or reaction rate throughout the reactor as it is perfectly mixed.
4. The exit stream has the same concentration and temperature as the entire reactor liquid.
5. The overall heat transfer coefficient is assumed constant.
6. No energy balance around the jacket is considered. Indeed, the jacket temperature can directly be manipulated in order to control the desired reactor temperature.
7. The reactor is a flat-bottomed vertical cylinder and the jacket is around the outside and the bottom.

The CSTR simulation model in MATLAB will be built using these predefined parameters and operating conditions:

Table 1: Parameters and Operating Conditions For CSTR Simulation Model

Operating Parameter	Value
Cross-sectional area of the reactor, ft^2	10.36
Concentration of reactant A in the exit stream, lb-mol/ft^3	0.05
Concentration of A in the feed stream, lb-mol/ft^3	0.9

Diameter of the cylindrical reactor, ft	3.6319
Activation energy, BTU/ lb-mol	30000
Volumetric feed flow rate, ft^3/h	20
Height of the reactor liquid, ft	3.8610
Heat of reaction, BTU/ lb-mol	-30000
Universal gas constant, BTU/ (lb-mol)(R)	1.987
Frequency factor, h–1	7.08×10^{10}
Multiplication of mixture density and heat capacity, BTU/(ft^3)(R)	37.5
Reactor temperature, R	650
Feed temperature, R	600
Jacket temperature, R	70.0
Overall heat transfer coefficient, BTU/(ft^2)(R)(h)	150

(Jana, 2011)

3.1.1 Model Development

Total Continuity Equation:

Mass inflow rate = Fi

Mass outflow rate = Fo

Rate of mass accumulation within reactor = $\frac{d(\rho V)}{dt} = \frac{d(\rho A_c h)}{dt}$

Eq. (3.1)

Ac is cross-sectional area of reactor and h is the height of the reactor liquid.

$$\frac{d(\rho V)}{dt} = (F_i - F_o)\rho$$

$$\frac{dV}{dt} = F_i - F_o$$

Eq. (3.2)

The reactor holdup, V and the exit flow rate Fo can be related as:

$F_o \propto \sqrt{V}$

For this CSTR, $F_o = \sqrt{10A_c h}$ Eq. (3.3)

Combining equations 3.2 and 3.3:

$$\frac{dh}{dt} = \frac{F_i}{A_c} - \sqrt{\frac{10h}{A_c}}$$

Eq. (3.4)

Component Continuity Equation:

Mass inflow rate component A = FiCAf,

Mass outflow rate component A = FoCA,

Rate of generation of component A = – (–rA)V

Rate of accumulation of component A within the reactor = $\frac{d(VC_A)}{dt}$

where –rA is the rate of consumption of chemical species A. The basic balance equation then becomes,

$$\frac{d(VC_A)}{dt} = F_i C_{Af} - F_o C_A - (-r_A)V$$

$$C_A \frac{dV}{dt} + V\frac{dC_A}{dt} = F_i C_{Af} - F_o C_A - (-r_A)V$$

Eq. (3.5)

Substituting equation 3.2 into 3.4 and simplifying,

$$\frac{dC_A}{dt} = \frac{F_i}{A_c h}(C_{Af} - C_A) - (-r_A)$$

Eq. (3.6)

For the given first-order reaction,

$$
\begin{aligned}
-r_A &= kC_A \\
&= \alpha \exp\left(\frac{-E}{RT}\right) C_A
\end{aligned}
$$

Eq. (3.7)

Combining equations 3.5 and 3.6,

$$\frac{dC_A}{dt} = \frac{F_i}{A_c h}\left(C_{Af} - C_A\right) - \alpha \exp\left(\frac{-E}{RT}\right) C_A$$

Eq. (3.8)

Energy Balance Equation:

Energy input rate = FiCpTf

Energy output rate = FoCpT + UiAh(T-Tj)

Energy added by exothermic reaction $= \left(-\Delta H\right) V \alpha \exp\left(\frac{-E}{RT}\right) C_A$

Energy accumulation rate:

$$\frac{d\left(V\rho C_p T\right)}{dt} = F_i \rho C_p T_f - F_o \rho C_p T - U_i A_h \left(T - T_j\right) + \left(-\Delta H\right) V \alpha \exp\left(\frac{-E}{RT}\right) C_A$$

Eq. (3.9)

Using equation 3.2 and further simplifying:

$$\frac{dT}{dt} = \frac{F_i}{A_c h}\left(T_f - T\right) + \left(\frac{-\Delta H}{\rho C_p}\right) \alpha \exp\left(\frac{-E}{RT}\right) C_A - \frac{U_i A_h}{\rho C_p A_c h}\left(T - T_j\right)$$

Eq. (4.0)

And therefore, this is the final form of the energy balance equation.

3.2. Simulation of Structural Faults

As stated before, CSTR simulation model will be developed and will be used to generate structural fault. Then, fault will be analyzed by PCA, WPCA and SPE methods and the data that are analyzed by WPCA methods will be compared with PCA and SPE method. The structural faults that will be simulated in this project are as the following:

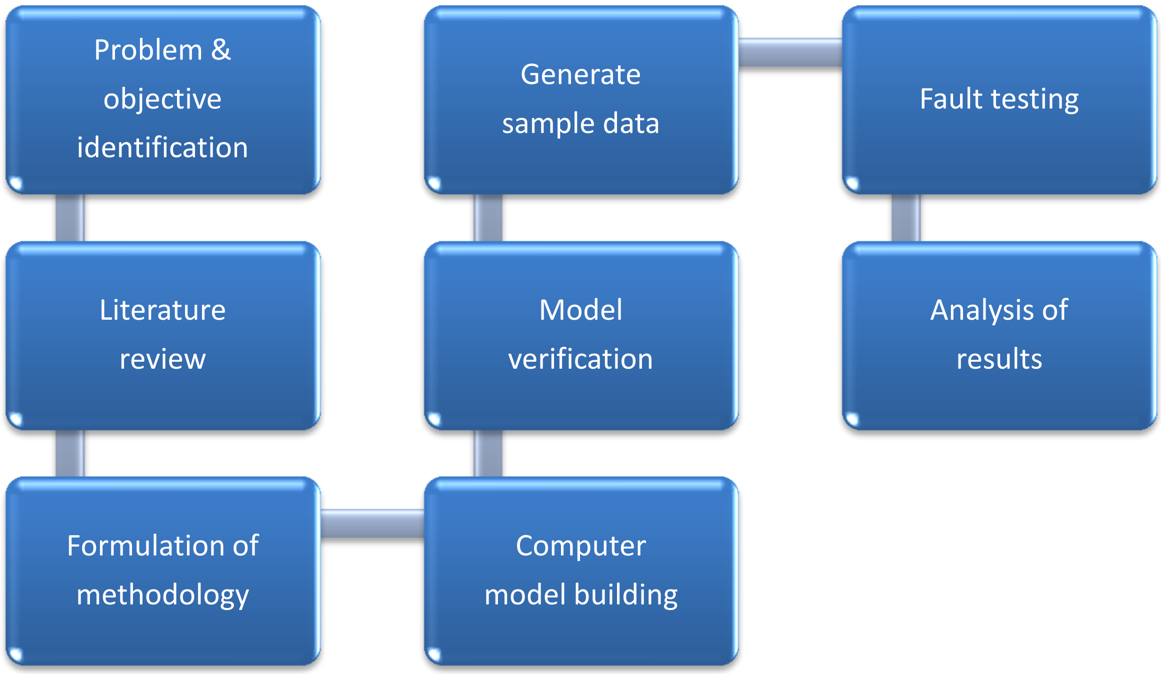
I. Drift in reaction kinetics.
   - e.g. Activation energy
   - Drift ranges:1%, 5% and 20%

3.3 Gantt Chart

No	Detail Work	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	Project Work Continuation															
2	Progress Report Submission							●								
3	Project Work Continuation															
4	Pre-SEDEX										●					
5	Draft Final Report Submission											●				
6	Dissertation Submission (Soft Bound)												●			
7	Technical Paper Submission												●			
8	Viva													●		
9	Dissertation Submission (Hard Bound)															●

Process ● **Key Milestone**

3.4 Project Flowchart & Key Milestone



3.5 Detailed Project Flowchart

i. Gathering of information from journals, research papers and etc. was conducted which were to study the fundamental knowledge and concepts of Principal Component Analysis (PCA) and focused on advanced method of PCA which is Weighted Principal Component Analysis (WPCA).

ii. CSTR simulation model was developed as shown in the Figure 1.The model then been tested by adding disturbances to the main inputs through the sine wave function and random number function. Then graph of sine wave and noise wave of the output was observed.

iii. The structural fault which is drift in kinetic energy was generated in the CSTR model and tested by T^2 statistic and Q statistic. The result is then been analyzed to compare WPCA method and PCA method.

CHAPTER 4: PRELIMINARY RESULT

4.1 Simulink Model

The diagram shows the computer model built using Simulink for the dynamic simulation of a CSTR using the given parameters. This model is then used to generate a sample set of baseline data (without faults) to be tested and used as benchmark later on.

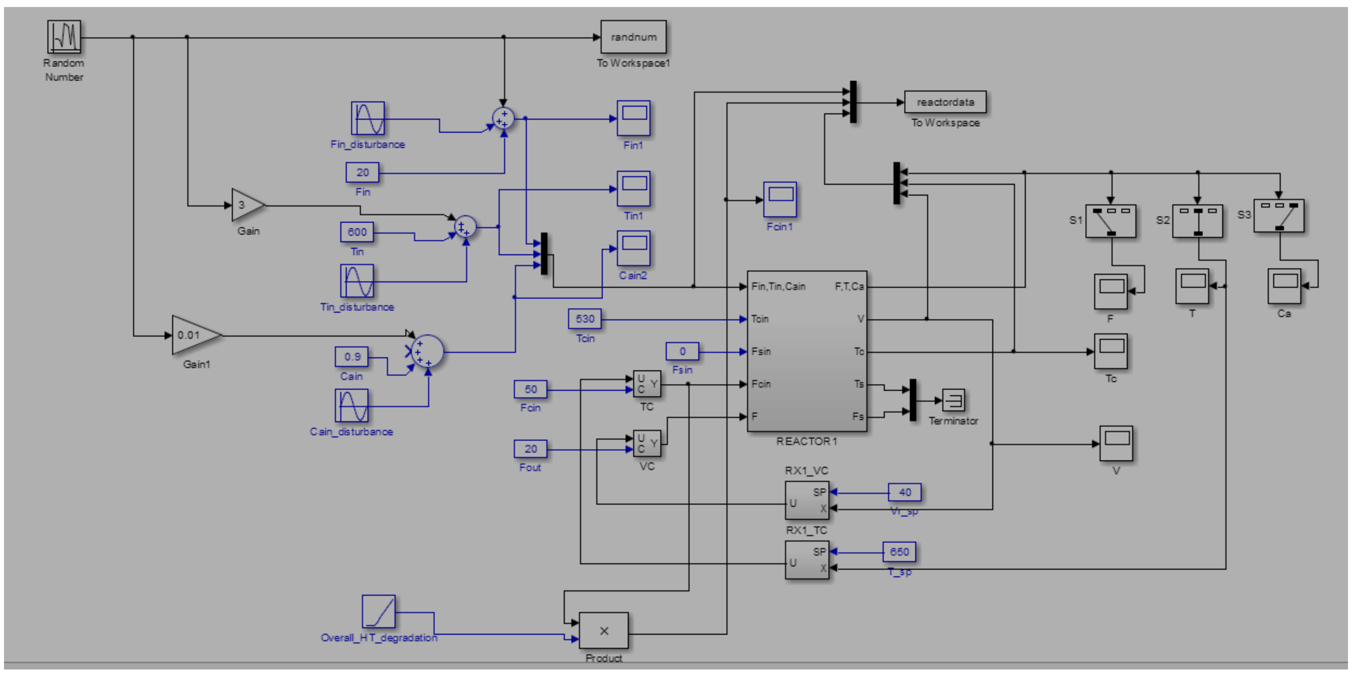


Figure 4 : Simulink Model

The model shows that there are three different inputs which are feed flow rate (Fin), temperature (Tin) and concentration (Cain).By using the sine wave function and random number function, disturbances are added to the main inputs. This is to simulate a non-ideal operating condition. Based on the model, the three main output which are product flow rate (F), temperature (T) and concentration(C) are generated after being inputted to the reactor.

4.2 Collected Data

4.2.1 Flowrate

Input:

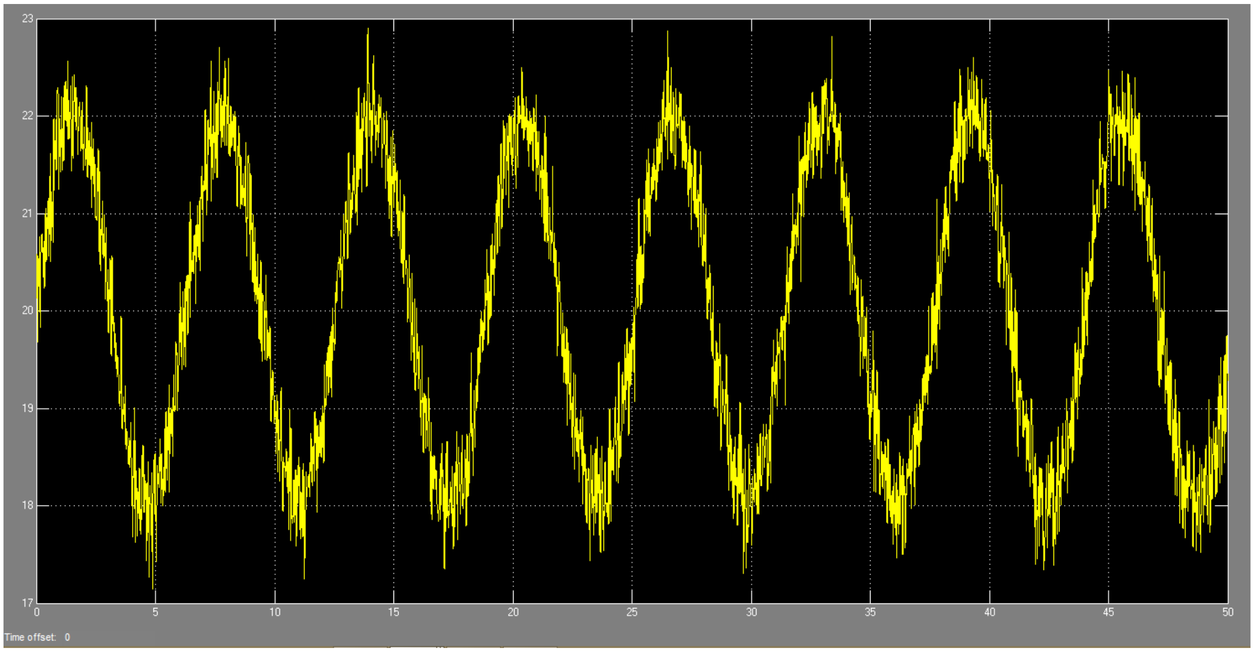


Figure 5 : Feed flowrate

Output:

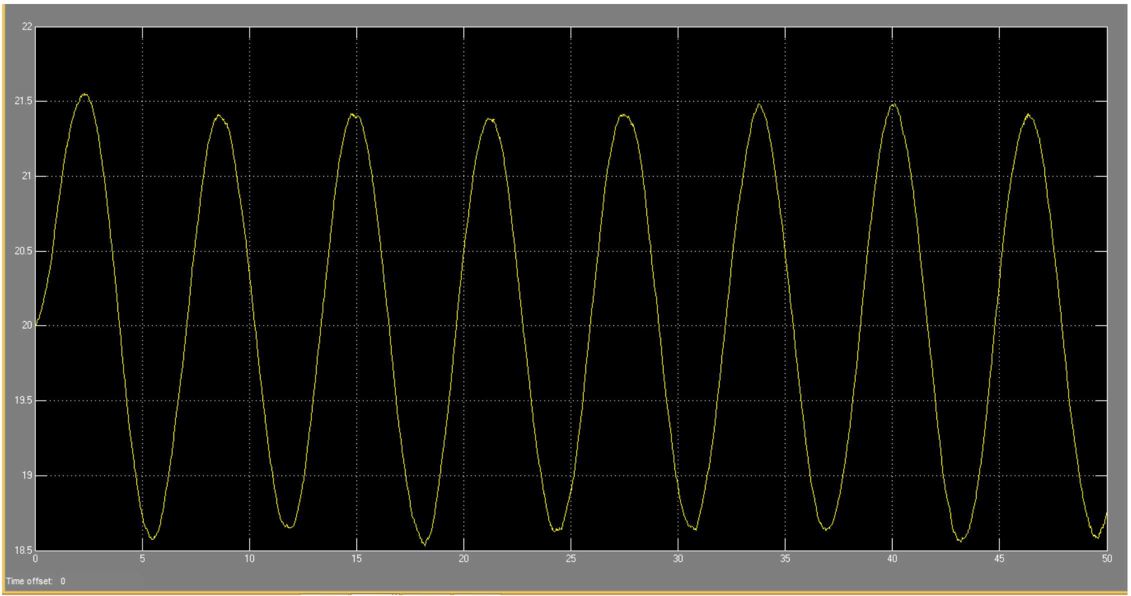


Figure 6 : Product flowrate

4.2.2 Temperature

Input:

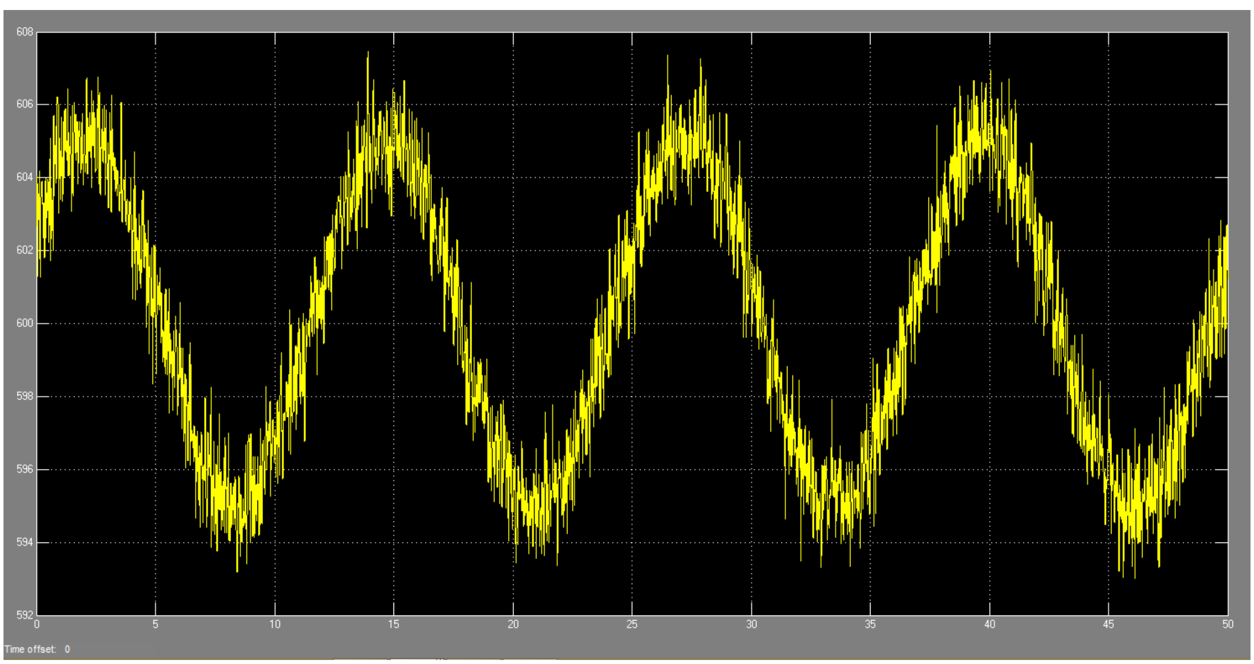


Figure 7 : Feed stream temperature

Output:

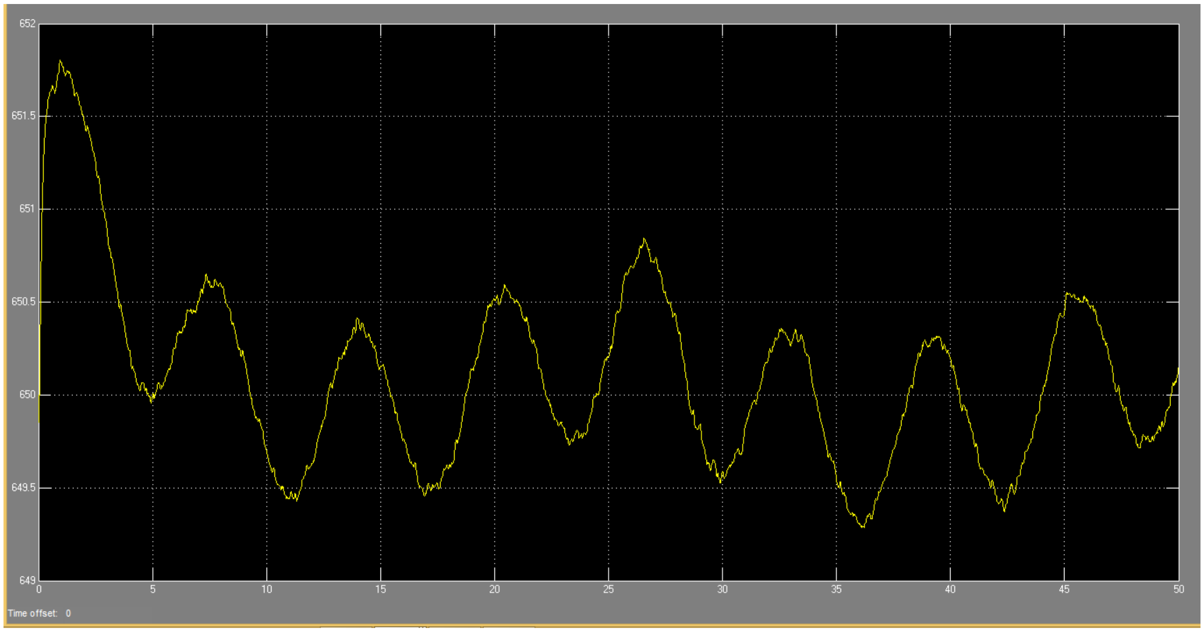


Figure 8 : Product stream temperature

4.2.3 Concentration

Input:

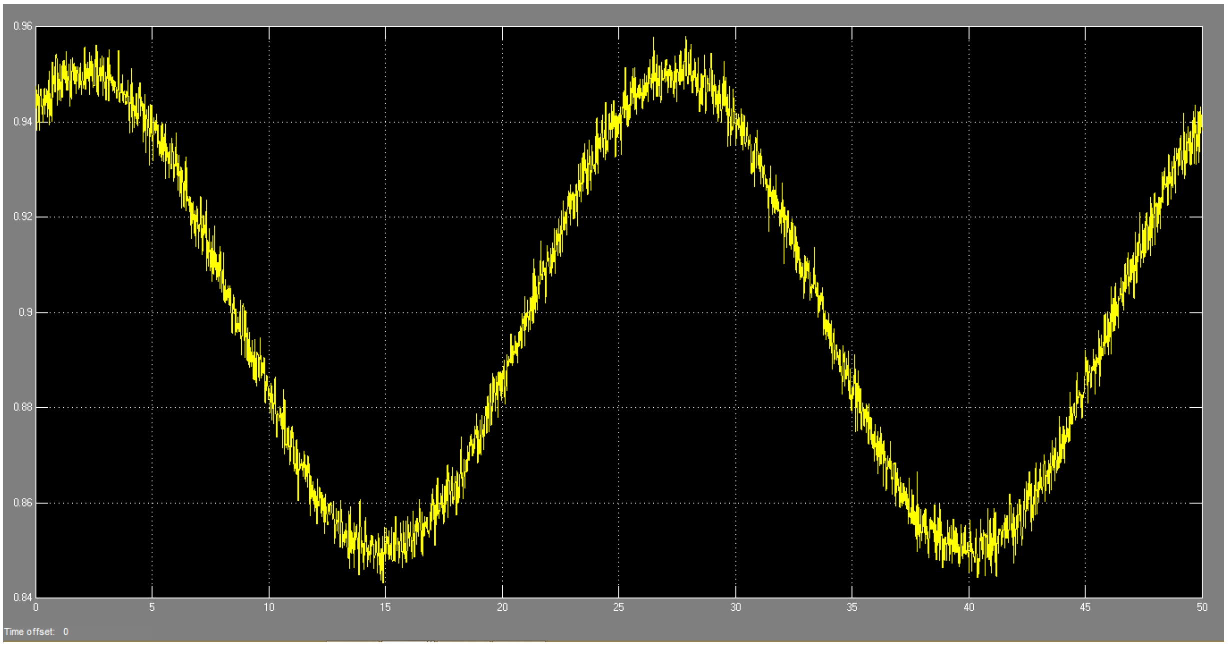


Figure 9 : Reactant concentration in feed

Output:

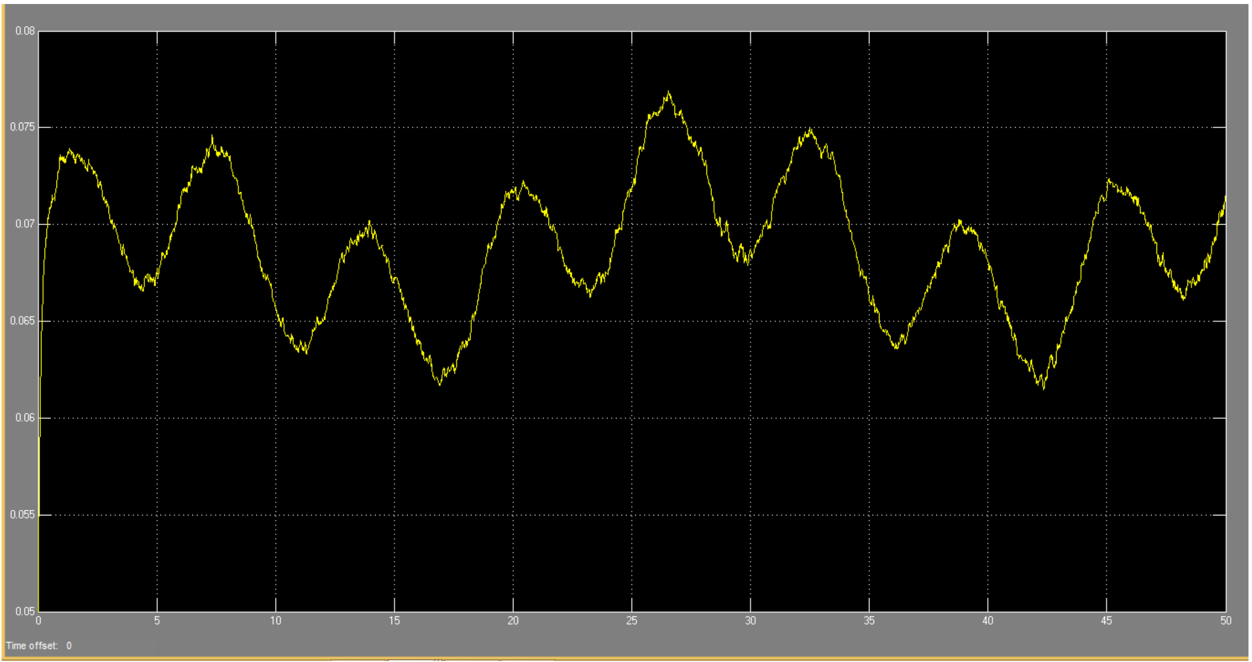


Figure 10 : Reactant concentration in product

The data obtained in the graphs shown in the previous pages are the sample baseline data set which is generated without structural faults. As seen in all the input graphs, the noise and sine wave disturbances are evident with fluctuating values along the plot. However, the output graphs show a rather smooth profile as if the noise is cancelled with only the sine wave profile.

The previous graphs which are the sample baseline data set have been generated after running the simulation without structural fault. Based on the all the input graph, the noise and sine wave disturbance are clearly shown with fluctuating values along the axis of the graph.However,in all the output graph, it is noticed that the only sine wave profile clearly shown while the noise is only small distortion along the plot line especially for the output flow rate. This is because the outputs have been controlled by the control structure that already added to the model to compensate for the differences which is to further simulate a non-ideal real life condition. This model will be further studied by testing it with structural faults that will be created which are deviations in reaction kinetics and heat transfer.

CHAPTER 5 : RESULT AND DISCUSSION

<u>Case 1 (1 % drift in activation energy)</u>

a)

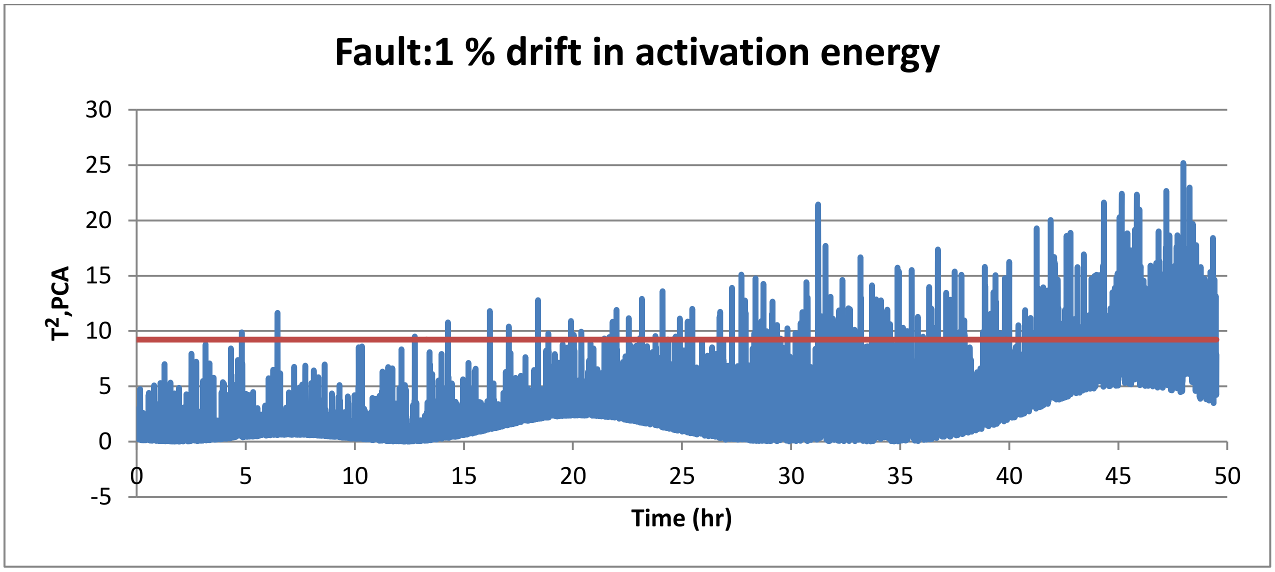


Figure 11 : T^2 statistic of 1% drift in activation energy using PCA.

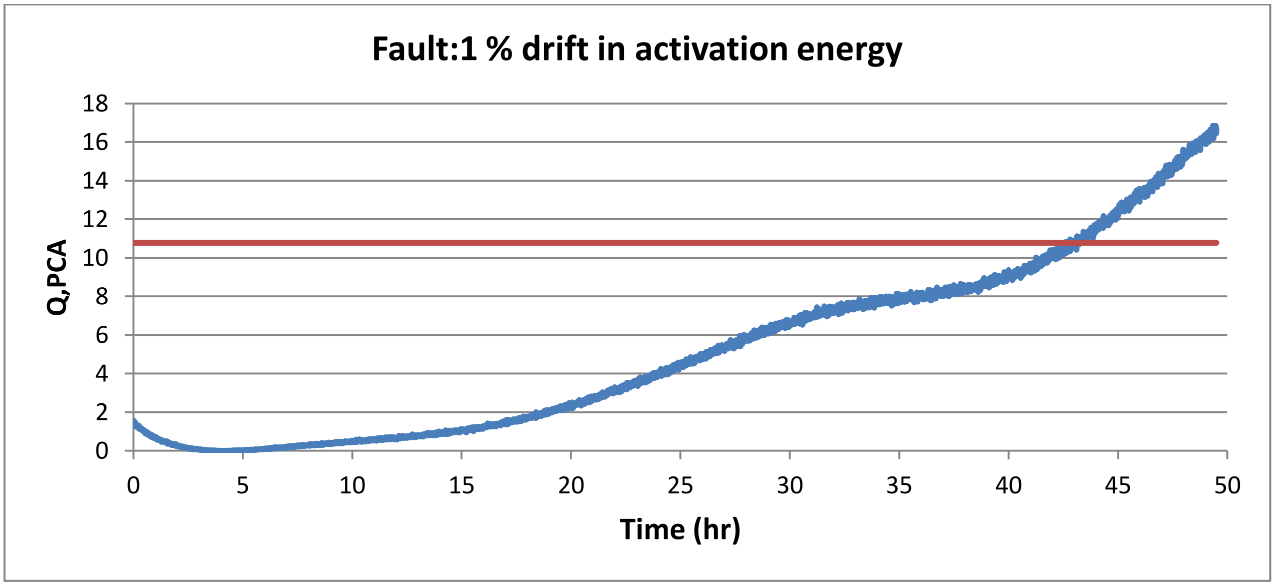


Figure 12 : Q statistic of 1% drift in activation energy using PCA.

b)

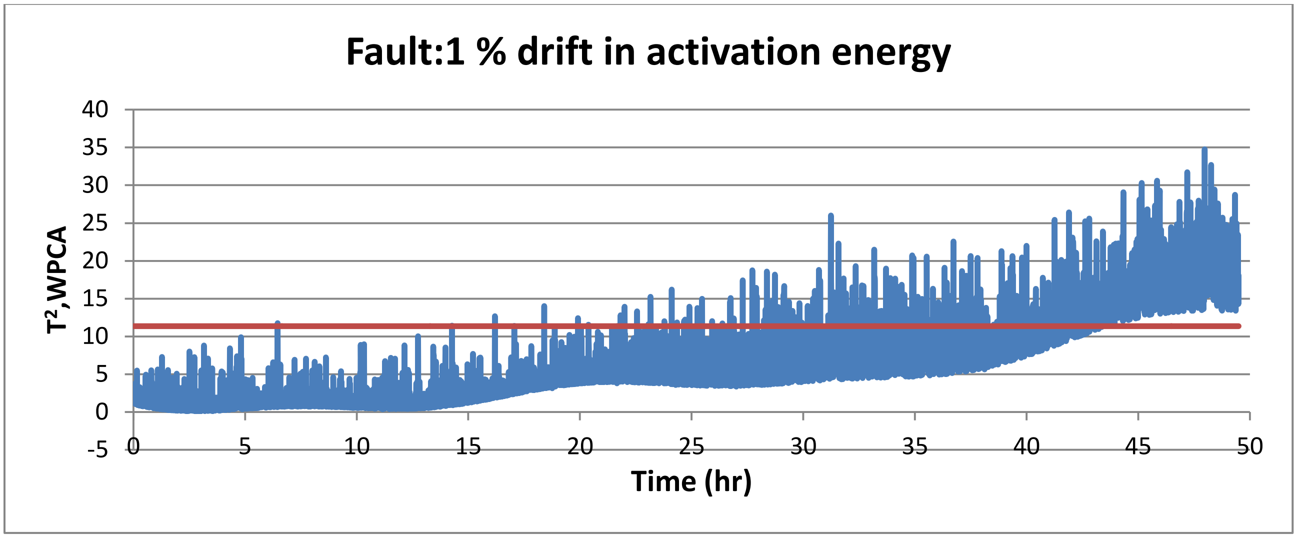


Figure 13 : T^2 statistics of 1% drift in activation energy using WPCA.

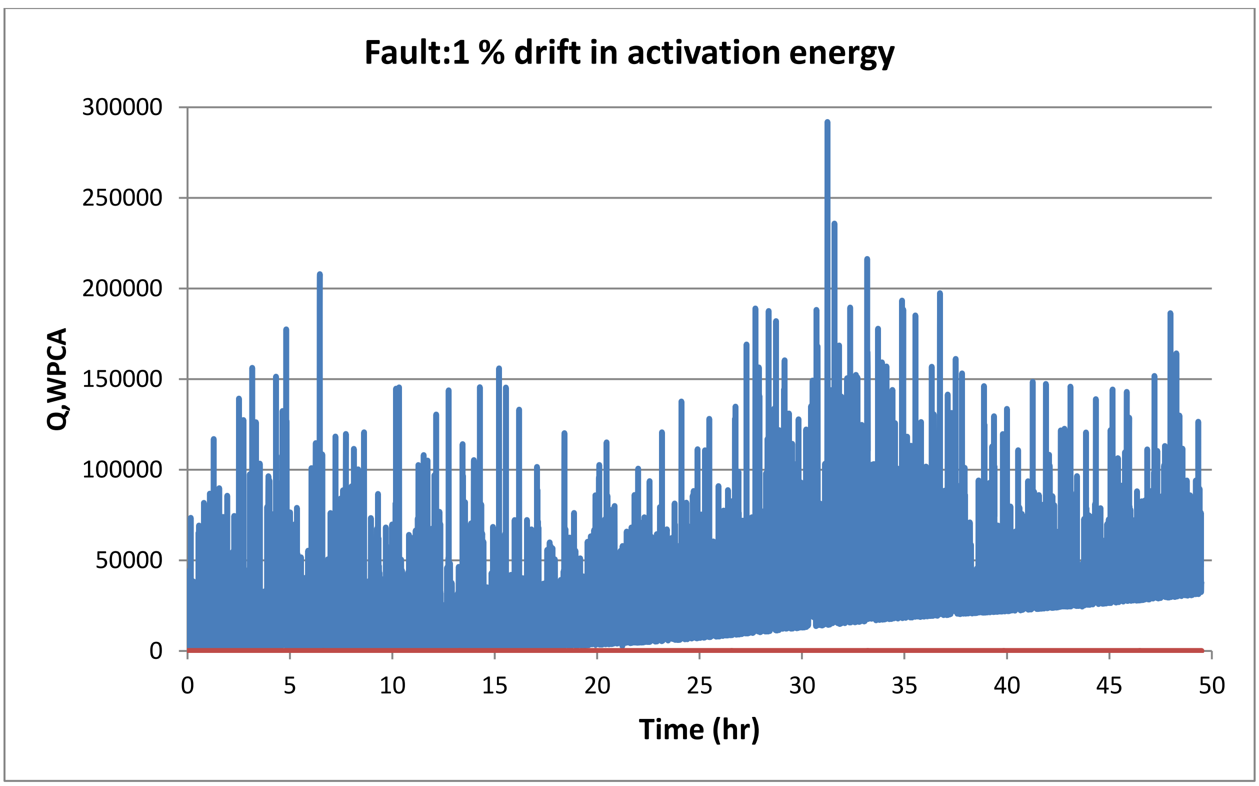


Figure 14 : Q statistics of 1% drift in activation energy using WPCA.

Case 2(5 % drift in activation energy)

a)

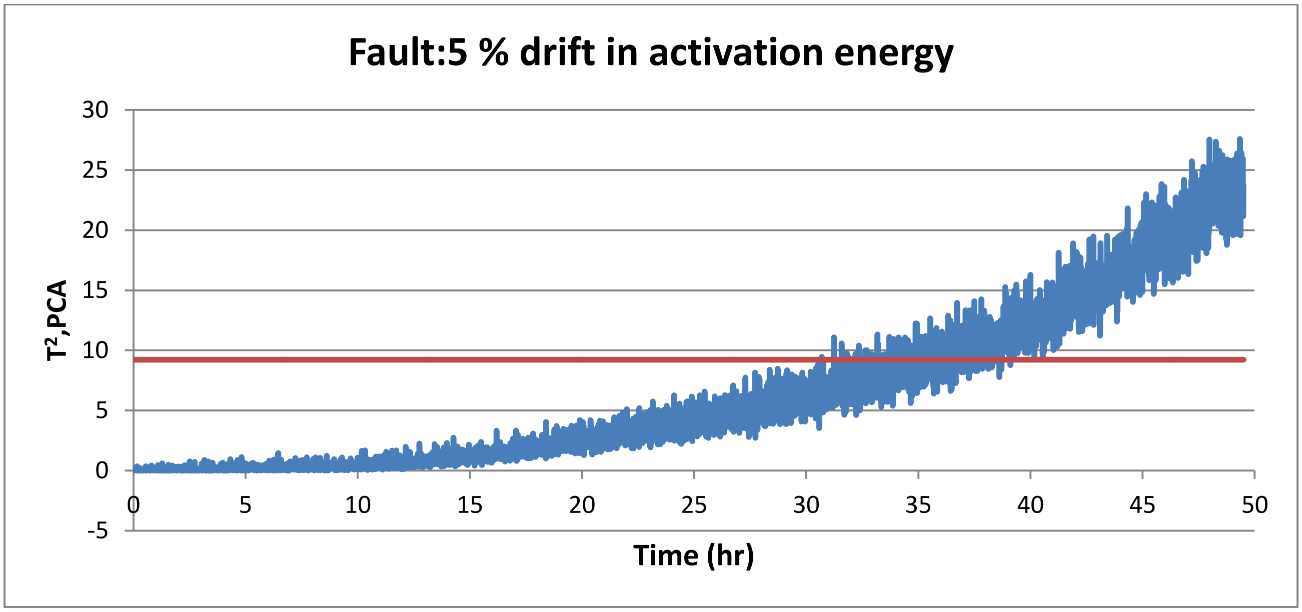


Figure 15: T^2 statistics of 5% drift in activation energy using PCA.

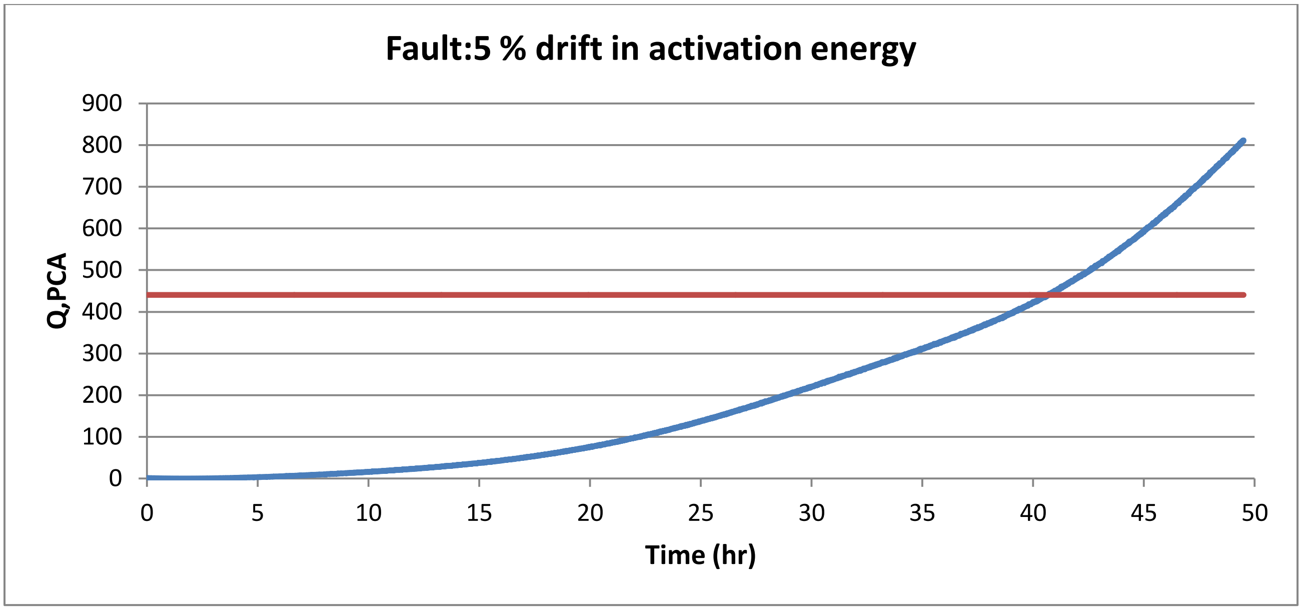


Figure 16 : Q statistics of 5% drift in activation energy using PCA.

b)

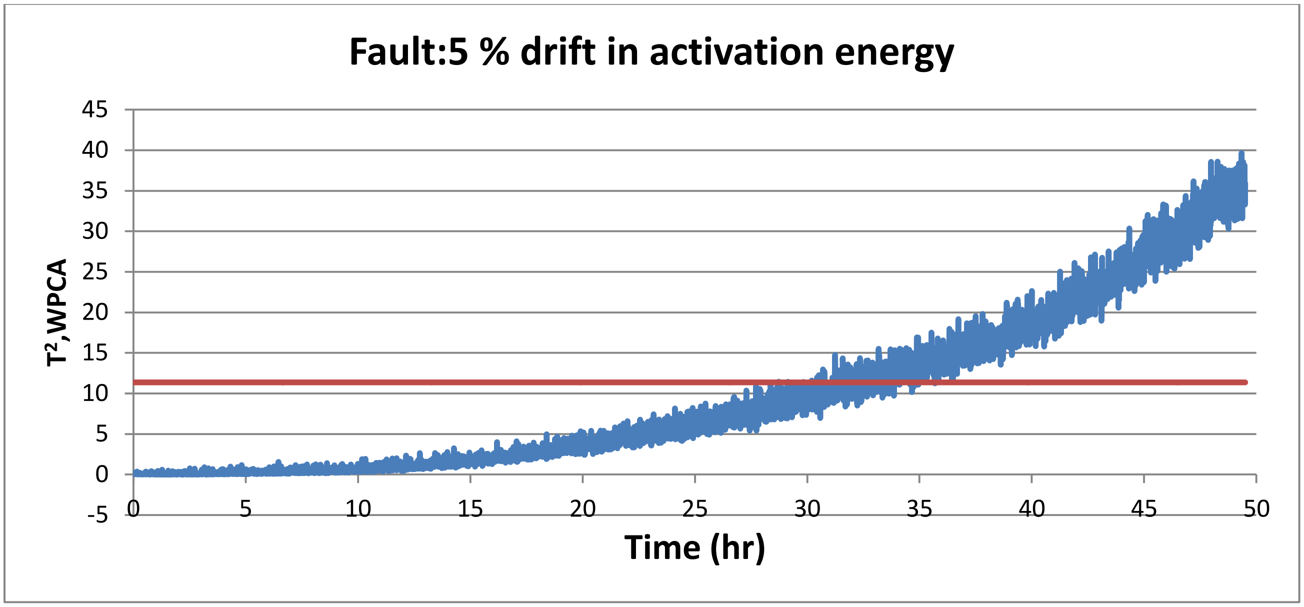


Figure 17 : T^2 statistics of 5% drift in activation energy using WPCA.

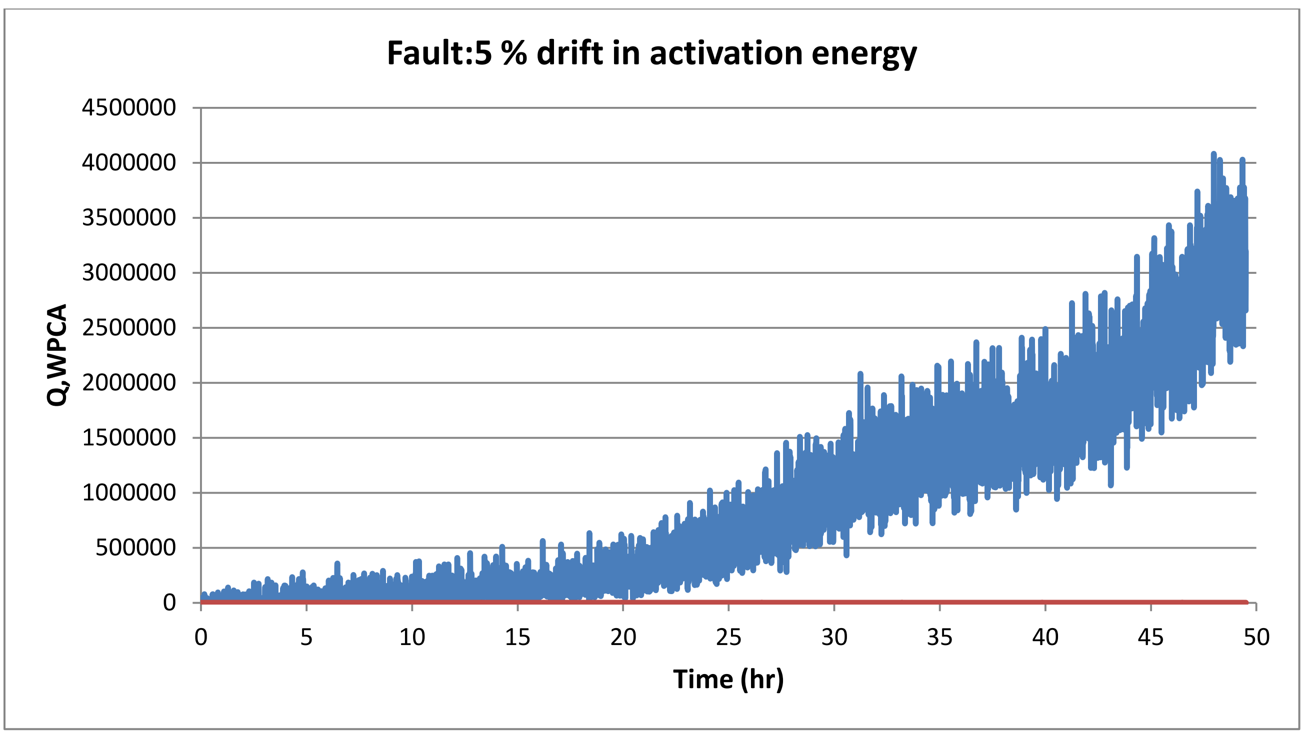


Figure 18 : Q statistics of 5% drift in activation energy using WPCA.

Case 3(20 % drift in activation energy)

a)

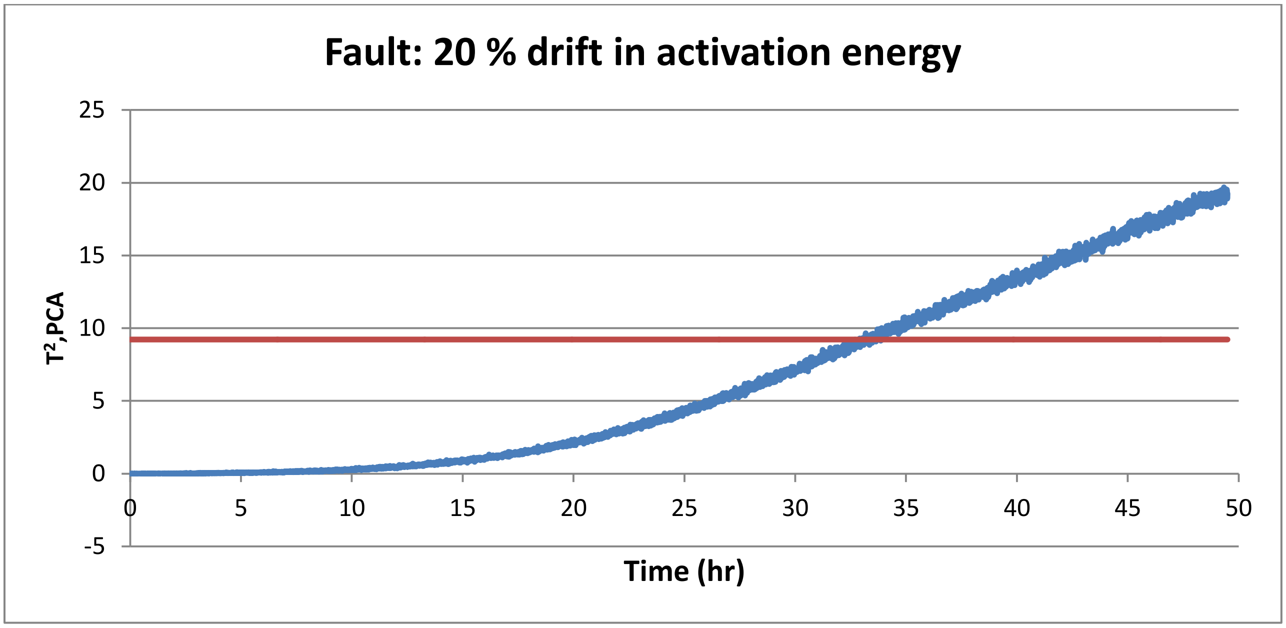


Figure 19 : T^2 statistics of 20% drift in activation energy using PCA.

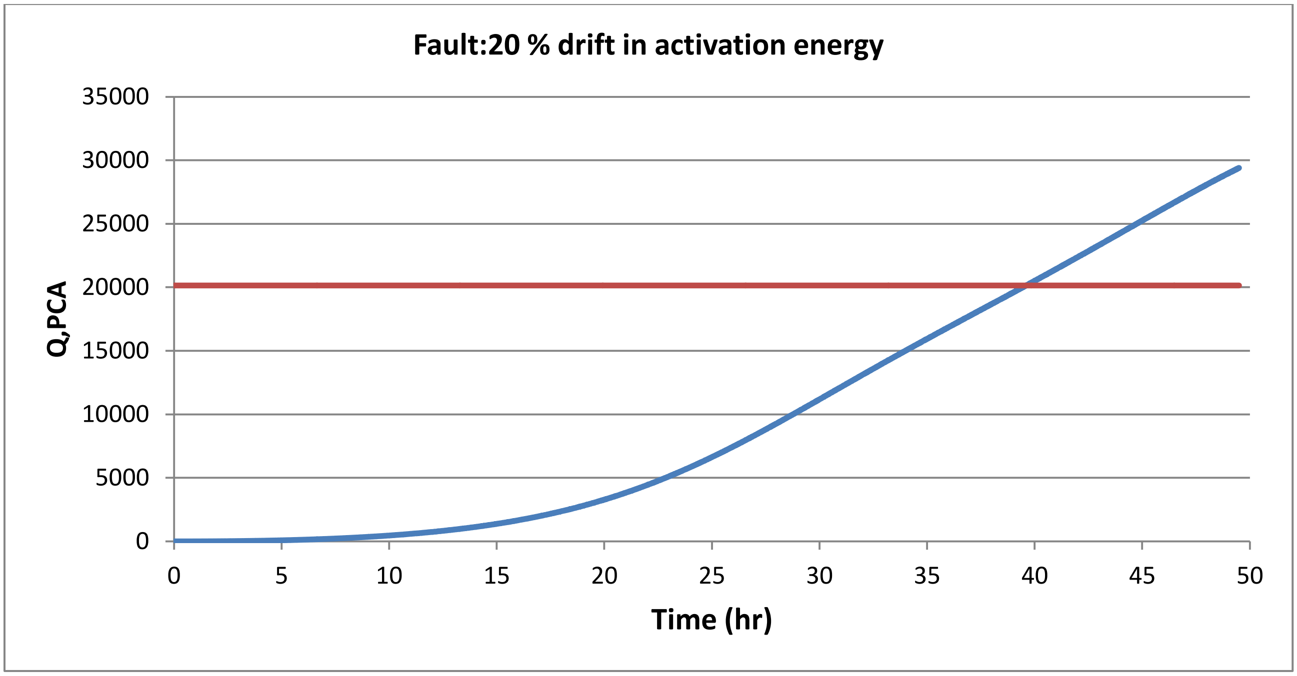


Figure 20 : Q statistics of 20% drift in activation energy using PCA.

b)

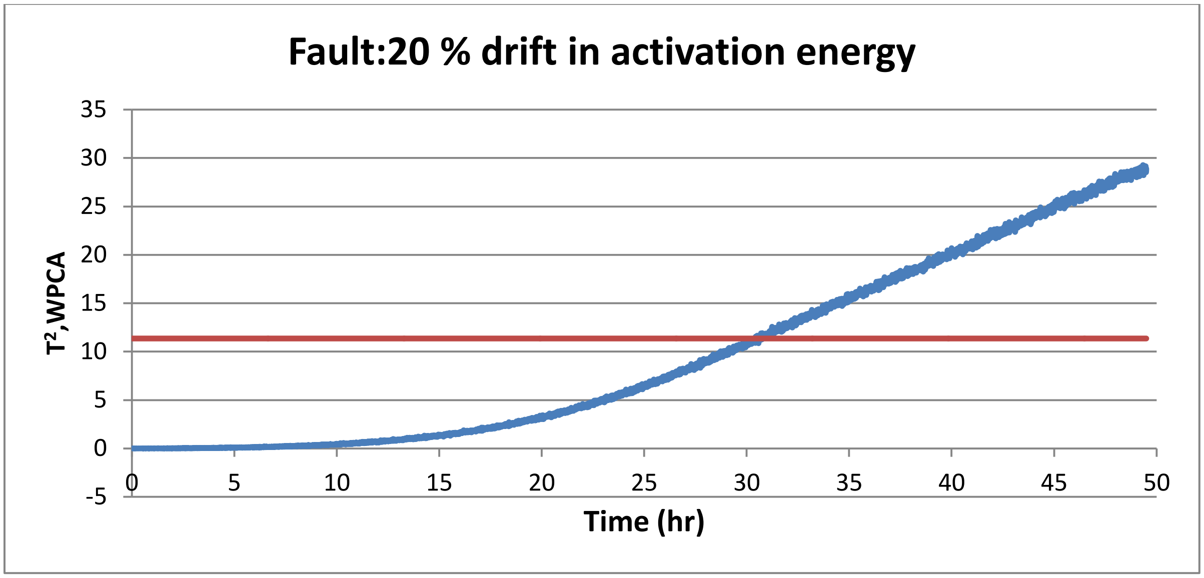


Figure 21 : T^2 statistics of 20% drift in activation energy using WPCA.

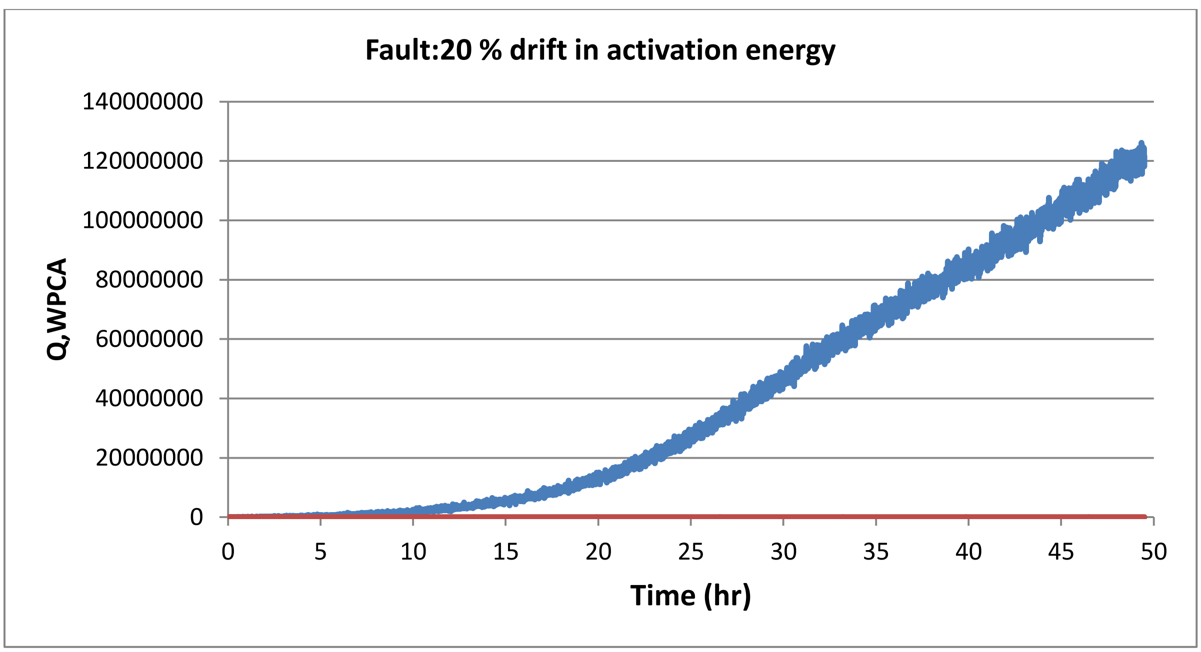


Figure 22 : Q statistics of 20% drift in activation energy using PCA.

The graph above shows the result obtained after the T^2 statistic and Q statistic been constructed for both PCA and WPCA method. The fault generated by using the CSTR simulink model involves the drift in the kinetic energy with three different cases which are different in percentage value of drift in kinetic energy which is 1%,5% and 20 %. From all of the cases, we can see that the T^2 statistic and Q statistic using WPCA method perform better for fault detection compared to by using PCA method.

CONCLUSION

As a conclusion, this project has fulfilled its objective which is to develop CSTR model and generate structural fault and to investigate the performance of (WPCA) compared to PCA T^2 and PCA Q statistic. CSTR model has been successfully developed and been tested with a sample baseline data set has been generated.

Basically, this paper focuses on the improvement of the extension of PCA method which is WPCA method. In WPCA method, it is based on the building conventional PCA model and then using change rate of T^2 statistic along every principal component (PC) to capture the most useful information in process, and setting different weighting values for PCs to highlight useful information. From the results obtained, it indicates that WPCA give better performance compared to PCA method.

REFERENCES

Zhao,C. & Gao, F.(2014). Fault-relevant Principal Component Analysis (FPCA) Method for Multivariate Statistical Modeling and Process Monitoring. *Chemometrics and Intelligent Laboratory Systems*, 1–16.

Tatara,E. & Cinar,A. (2002). An Intelligent System for Multivariate Statistical Process. *ISA Transactions* , 255–270.

Evan L. Russell, L. H. (2000). Fault Detection in Industrial Processes Using Canonical Variate. *Chemometrics and Intelligent Laboratory Systems* , 81–93.

Russell,E.L.,Chiang,L.H.,& Braatz,R.(2000). Fault detection in industrial processes using canonical variate analysis and dynamic principal component analysis,Chemometrics and Intelligent Laboratory Systems 81–93.

Jana, A. K. (2011). *Chemical Process Modelling and Computer Simulation.* PHI Learning Private Limited.

J. Lee, C. Yoo, S.W. Choi, P.A. Vanrolleghem, I.-B. Lee, Nonlinear process monitoring using kernel principal component analysis, Chemical Engineering Science 59 (1) (2004) 223–234.

J. Lee, S.J. Qin, I.-B. Lee, Fault detection and diagnosis based on modified independent component analysis, AIChE Journal 52 (2006) 3501–3514.

Chen,J. & Liao,C.-M. (2002). Dynamic Process Fault Monitoring Based on Neural Network and PCA. *Journal of Process Control*, 277–289.

Chiang,L.H.,Russel,E.L.,& Braatz,R.D.(2000). *Fault Detection and Diagnosis in Industrial system.* Springer.

Montgomery, D.C., 1997. Introduction to Statistical Quality Control, third ed. John Wiley and Sons, New York.

Q. Chen, U. Kruger, M. Meronk, A.Y.T. Leung, Synthesis of T-2 and Q statistics for process monitoring, Control Engineering Practice 12 (2004) 745–755.

Jiang,Q. & Yan, X.. (2012). Chemical Processes Monitoring Based on Weighted Principal Component Analysis and its Application. 11–20.

Mahadevan,S. & Shah,S. L. (2009). Fault Detection and Diagnosis in Process Data Using One-Class Support. *Journal of Process Control*, 1627–1639.

Verron,S.,Tiplica,T.,&Kobi,A. (2010). Fault Diagnosis of Industrial Systems by Conditional Gaussian Network. *Engineering Applications of Artificial Intelligence*, 1229–1235.

Lin,W.,Qian,Y.,& Li,X. (2000). Nonlinear Dynamic Principal Component Analysis for On-line Process Monitoring and Diagnosis. *Computers and Chemical Engineering*, 423-429.