APPLICATION OF PROCESS MONITORING BASED ON INFERENTIAL MEASUREMENT APPROACH

ZAIDI BIN SALIM

UNIVERSITI MALAYSIA PAHANG

APPLICATION OF PROCESS MONITORING BASED ON INFERENTIAL MEASUREMENT APPROACH

by

ZAIDI BIN SALIM

Thesis submitted in partial fulfilment of
the requirements for the award of the degree of
Bachelor of Chemical Engineering

Faculty of Chemical and Natural Resources Engineering
UNIVERSITI MALAYSIA PAHANG

February 2013

SUPERVISOR'S DECLARATION

I hereby declare that I have checked this thesis and in my opinion, this thesis is adequate in terms of scope and quality for the award of the degree of Bachelor of Chemical Engineering.

Signature:
Name of Supervisor: Dr. Mohd Yusri Mohd Yunus
Position:
Date:

STUDENT'S DECLARATION

I hereby declare that the work in this thesis is my own except for the quotations and summaries which have been duty acknowledged. The thesis has not been accepted for any degree and is not concurrently submitted for award of other degree.

Signature:
Name: Zaidi Bin Salim
ID Number: KA09168
Date:

DEDICATION

Special dedication to my beloved father and mother,

brothers and sisters,

professor and friends…..

Special Thanks for all of your Care, Love, Encouragement and Best Wishes.

ACKNOWLEDGEMENTS

I am very grateful to my research advisor, Dr. Mohd Yusri Mohd Yunus for all the knowledge he provided and his direction in this project. I would like to thank him for the initial help he provided in introducing me to Matlab program. His work ethic is one to be emulated. He made time available for his students and his timely response to inquiries were truly appreciated. I would like to thank the staff and professors in the Department of Chemical Engineering at Universiti Malaysia Pahang who helped during the whole period of my research.

I would also like to thank my group members for their help and insight on my project. Also to my fellow friends, thank you for providing a tending environment to work in, and I wish you all the best in the completion of your thesis work and beyond. Finally, I would like to thank my family and friends for all their love and support.

ABSTRACT

In this study, a new multivariate method to monitor continuous processes is developed based on the Process Control Analysis (PCA) framework. The objective of the study is to develop A new MSPM method and analyze the monitoring performance of system A and B. In industrial practice, monitoring process are usually performed based on an approximate model. As the number of variables increases, the fault detection performance tends to be slow in progression, as well as, introduce greater complexity in the later stages especially in fault identification and diagnosing. These research implements and analyzes Multiple Linear Regression (MLR) method to a continuous process which simplify the number of variables used. This research also based on the conventional MSPM technique. After that, the developed method was analyzed and finally, all the performance result of the developed method was compared with the conventional method. The monitoring results clearly demonstrate the superiority of the proposed method. The MLR methods show that the fault detection performance improved and better than the conventional method.

ABSTRAK

Dalam kajian ini, satu kaedah baru multivariat untuk memantau proses yang berterusan dibangunkan berdasarkan rangka kerja Analisis Kawalan Proses (PCA). Objektif kajian ini adalah untuk membangunkan satu kaedah MSPM baru dan menganalisis prestasi pemantauan antara sistem A dan B. Dalam amalan industri, memantau proses biasanya dilakukan berdasarkan model anggaran. Apabila bilangan pembolehubah meningkat, prestasi pengesan kesalahan cenderung untuk menjadi perlahan semasa proses berjalan, serta memperkenalkan kerumitan yang lebih besar di peringkat akhir terutama dalam mengenal pasti kerosakan dan mendiagnosis. Penyelidikan ini melaksanakan dan menganalisis kaedah Regresi Linear Berganda (MLR) untuk proses berterusan yang meringkaskan bilangan pembolehubah yang digunakan. Kajian ini juga berdasarkan teknik MSPM yang lazim. Selepas itu, kaedah yang direka telah dianalisis dan akhirnya, semua hasil prestasi kaedah yang direka dibandingkan dengan kaedah yang lazim. Keputusan pemantauan jelas menunjukkan keunggulan kaedah yang dicadangkan. Kaedah MLR menunjukkan bahawa prestasi pengesanan kerosakan meningkat dan lebih baik daripada kaedah yang lazim.

TABLE OF CONTENTS

LIST OF FIGURE

LIST OF TABLE

LIST OF SYMBOLS

Y	Response variable
X	Independent variable
$\mathbf{C}_{mxm}$	Variance-covariance matrix
ε_i	Error variable
y_i	Dependent variable
T^2	Hotelling T-squared distribution
σ^2	Variance
^{T}	Transpose
β	Unknown parameter

LIST OF ABBREVIATIONS

SPC	Statistical Process Control
SQC	Statistical Quality Control
PCA	Principle Component Analysis
MSPC	Multivariable Statistical Process Control
MSPM	Multivariate Statistical Process Monitoring
MLR	Multiple Linear Regression
CSTRwR	Continuous-stirred tank with recycle
ICA	Independent Component Analysis
MDS	Multidimensional Scaling
QR	Quantile Regression
NOC	Normal operating condition
SPE	Squared Prediction Error
OLS	Ordinary Least Square

CHAPTER 1

INTRODUCTION

1.1 Background of Study

In process monitoring, the main objective is always to detect changes or departure behavior from the normal process characteristics. Due to the nature of the processes that always change over time and are affected by several sources, process monitoring become more challenging in any chemical based industries.

Application of statistical methods in monitoring and control of industrial processes are included in a field generally known as statistical process control (SPC) or statistical quality control (SQC) (Damarla, 2011). The most widely used and popular

SPC techniques involve univariate methods, that is, observing and analyzing a single variable at a time. Statistical Process Control (SPC) is an effective method of monitoring a process through the use of control charts. By collecting data from samples at various points within the process, variations in the process that may affect the quality of the end product or service can be detected and corrected, thus reducing waste as well as the likelihood of passing down to the customer. Thus, early detection and prevention of the problems are both crucial in this respect.

However, industrial quality problems are multivariate in nature, since they involve measurements on a number of variables simultaneously, rather than depending on one single variable. As a result, Multivariable Statistical Process Control (MSPC) system (Kano et al., 2001) is introduced, where a set of variables which are the manipulated variables and controlled variables are identified and the jointly monitored. In conclusion, early detection and diagnosis of process faults while the plant is still operating in a controllable region can help avoid abnormal event progression and reduce productivity loss.

1.2 Problem Statement

Monitoring and controlling a chemical process is a challenging task because it involves a huge number of variables. Usually, process monitoring is executed based on

the principal-component analysis (PCA) technique, nevertheless it has its own limitation. As the number of variables grows, the number of PCs also becomes larger. Thus to overcome the problem, it is desirable to reduce those variables, while embedding them into a single measurement model, where the original variations can still be preserved.

1.3 Research Objectives

The main purpose of this research is to propose a new MSPM technique, where the original variables are modeled into linear composites in order to reduce the number of variables in monitoring, where eventually it may also monitoring the performances. Hence, the objectives are:

a) To develop the conventional MSPM method for the original set of variables (System A).
b) To develop a new MSPM method which applies Multiple Linear Regression (MLR) technique. (System B).
c) To analyze the monitoring performances between System A and System B.

1.4 Research Questions

1.4.1 Can the MLR technique sufficiently be used to model the original variables?

1.4.2 How are the generic monitoring performances of the proposed method as compared to the traditional scheme?

1.4.3 What is the optimized condition which must be complied in order to improve the new technique?

1.5 Scopes of Study

The research is based on multivariate statistical process monitoring (MSPM) where in this research, multiple linear regression method is used. The method will relate the variables of the process with the process itself and also it will relate certain variables on the controller in the system. The scopes of the study are:

a) Mainly focus to select only certain variable.
b) A continuous-stirred tank reactor with recycle (CSTRwR) system is used for demonstration, whereby the faults are consisting of abrupt and incipient.
c) Shewhart control chart is chosen to show the progression of the monitoring statistics.

d) All algorithms are developed and run based on Matlab version 7 platform.

1.6 Expected Outcomes

By doing this research, the developed MSPM method will able to be justified that it is comparatively better than conventional PCA-based MSPM method in monitoring the multivariate of non-linear process. The research also will show the development of advanced multivariate way of process monitoring in terms of variables points besides of samples scores.

1.7 Significance of Study

This study produces a new idea on how to reduce the complexity of monitoring analysis by using MLR technique in modeling all the variables involved. The MLR method will lessen the number of variables used. As the number of variables used decrease, the number of principle component used is also decrease. The method is expected to have similar or improve the monitoring progressions especially in terms of fault detection sensitiveness.

1.8 Report Organization

This thesis is divided into five chapters which are the introduction, literature review, methodology, result and discussion, and also conclusion and recommendation. The first chapter renders an overview of statistical process control (SPC), multivariate process and their use in process monitoring. This chapter also presents the objectives of the present work, scope, the expected outcome and significance of the research project. The second chapter emphasizes on fundamentals and theory of the study, limits and extension of PCA, inferential measurement and also multiple linear regression (MLR). In chapter three, multiple linear regression method will be presented. Chapter four discuss on the results and some discussion on the research. Finally, conclusion and recommendation will be discussed in chapter five.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Since the last decade, design and development of data based model control has taken its momentum. This very trend owes an explanation. Identification and control of chemical process is a challenging task because of their multivariate, highly correlated and non-linear nature. Very often there are a large number of process variables are to be measured thus giving rise to a high dimensional data base characterizing the process of interest. To extract meaningful information from such a data base; meticulous preprocessing of data is mandatory. Otherwise those high dimensional dataset maybe

seen through a smaller window by projecting the data along some selected fewer dimensions of maximum variability. This chapter will emphasize on the fundamental and theory, limits and extension of PCA, inferential measurement and also about multiple linear regression.

2.2 Fundamentals and Theory

Research by Smith (2002) illustrated that Process Control Analysis is a way of identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences. Since pattern in data can be hard to find in data of high dimension, where the luxury of graphical representation is not available, PCA is a powerful tool for analyzing data. According to Bakshi (1998), PCA is a MSPC technique used for the purpose of data compression without losing any valuable information. Alvarez (n.d.) in his thesis mentions that PCA can be described as a method to project a high dimensional measurement space onto a space with significantly fewer dimensions. PCA finds linear combinations of variables that describe major trends in data set. Mathematically, PCA is based on an orthogonal decomposition of the covariance matrix of the process variables along the directions that explain the maximum variation of the data.

Principal components (PCs) are transformed set of coordinates orthogonal to each other. The first PC is the direction of largest variation in the data set. The projection of original data on the PCs produces the score data or transformed data as a linear combination of those fewer mutually orthogonal dimensions. PCA technique was applied on the auto-scaled data matrix to determine the principal eigenvectors, associated eigen values and scores or the transformed data along the principal components. The drawbacks are that the new latent variables often have no physical meaning and the user has a little control over the possible loss of information. Generally, PCA is a mathematical transform used to find correlations and explain variance in a data set.

2.3 Limits and extension of PCA

Although PCA is good for linear or almost linear problems, it fails to deal well with the significant intrinsic nonlinearity associated with real-world processes. Juricek et. al (n.d.) in their paper state most industrial processes, and almost all found in the chemical industry, are multivariable which has two or more inputs and outputs, nonlinear and are constantly responding to disturbances that are cannot be measured and occurring at unknown times. Hence, nonlinear extensions of PCA have been investigated by different researchers (Zhao and Xu, 2004).Both the strength and weakness of PCA is

that it is a non-parametric analysis. PCA is also commonly viewed as a Gaussian model; that is, the data is assumed to come from a Gaussian distribution.

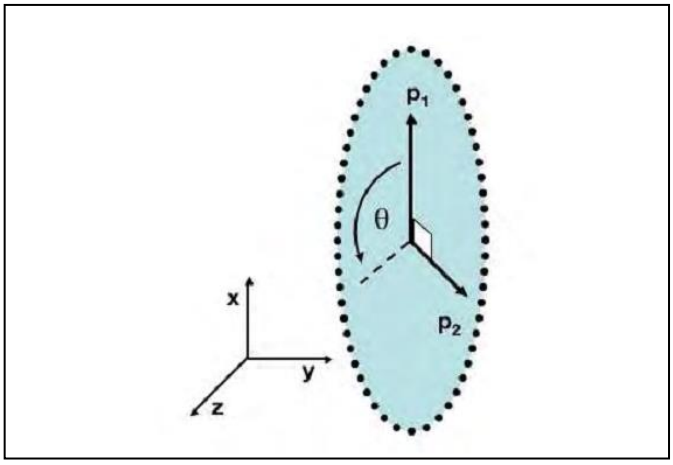


Figure 2.1 Data point which track a person on a Ferris wheel

For example, from Shlens (2005) study, we can consider the recorded positions of a person on a Ferris wheel over time in Figure 2.1. The probability distributions along the axes are approximately Gaussian and thus PCA finds (p1, p2), however according to Shlens, this answer might not be optimal. The most concise form of dimensional reduction is to recognize that the phase or angle along the Ferris wheel contains all dynamic information. Thus, the appropriate parametric algorithms are to first convert the data to the appropriately centered polar coordinates and then compute PCA.

This prior non-linear transformation is sometimes termed a kernel transformation and the entire parametric algorithm is termed kernel PCA. Other common kernel transformations include Fourier and Gaussian transformations. This procedure is parametric because the user must incorporate prior knowledge of the structure in the

selection of the kernel but it is also more optimal in the sense that the structure is more concisely described. One might envision situations where the principal components need not be orthogonal. Furthermore, the distributions along each dimension (xi) need not be Gaussian. If we are using a probabilistic interpretation of PCA, we might want to assume that the data is Gaussian because uncorrelated Gaussian random variables are also independent. Because PCA decorrelates the data, the resulting encodings in the basis of the principal components are independent. The random processes generating the encodings might then be thought of as the underlying independent causes of the data.

According to Nikolov (2010), in Independent Component Analysis (ICA), it is assumed that there are independent, non-Gaussian random variables which is traditionally called sources, and that they are transformed by a mixing matrix, for example matrix W, to give a measurement $x = (x^1, \ldots, x^d)$. This gives the relationship between the measurements x and the sources y:

$$x = Wy \tag{2.1}$$

The goal of ICA then is to recover W and y given x, only by looking at the statistical structure of x. There is a lot to be said about this problem, and there are many techniques for solving it including maximization of nongaussianity, maximum likelihood, minimization of mutual information between components of the encoding, maximization of mutual information between the data and the encodings, nonlinear decorrelation, and diagonalizing higher order cumulant tensors. Compared to indepedence, uncorrelatedness is a relatively weak statement to make about a set of random variables. However, uncorrelatedness can mean something stronger if we first pass the transformed

data y through a nonlinearity and then decorrelate it. Whereas decorrelating the components of y involves only second-order statistics such as making the covariances zero, this nonlinearity brings higher order statistics into play when modeling the data.

Multidimensional Scaling (MDS) is another technique popularly used for data exploratory purpose. The underlying concept of MDS is that it utilizes the inter-objects dissimilarity measures of a set of data to find a configuration that could represent the data as precisely as possible. One significant advantage of MDS over PCA is that it never looks on to identify a single model, whether linear or non-linear trend that could represent the data on the whole, but appreciate every single distance between those objects in the data set, in order to build the point configurations intended.

2.4 Inferential Measurement

Many chemical products are sold for their effect rather than their chemical composition. In these cases, it is often difficult to provide reliable, fast, on-line measurements to control product quality. The quality measure may only be available as a laboratory analysis or very infrequently on-line. This can lead to excessive off-specification products, especially when changing from one operating region to another. Inferential Measurement is a powerful and increasingly used methodology that allows

process quality, or a difficult to measure process parameter, to be inferred from other easily made plant measurements such as pressure, flow or temperature.

2.4.1 Benefits

Inferential measurements have some benefits. First, there is faster return of information. What this means is that process upsets can be detected quicker and remedial action can be taken before it is too late. The inferential estimates usually carry a fair degree of feed forward information. For instance, disturbances affecting tray temperatures in a distillation column may show up much later in the product compositions because of the location of the tray and because of the dynamics of the system. However, if tray temperatures are used to estimate product compositions, then any disturbances on temperatures will immediately be reflected in the composition estimates.

If the estimates can be generated at a reasonable accuracy and at a fast enough frequency, then it can be used as the feedback signal to an automatic feedback controller. By reducing human involvement in the control loop, more consistent production can be achieved. As a result of the improvements, better process regulation can be achieved which also means that there will be increased scope for process optimization. With better process regulation, the process operator's time will be better

spent in carrying out higher level supervisory tasks. All these means that plant productivity is increased, leading to higher profitability.

2.5 Multiple Linear Regressions

In Wise (n.d.) research, he stating that multiple linear regression (MLR) is a multivariate statistical technique for identifying and determining the linear correlations between two or more independent variables and a single dependent variable. In addition, Nathans et. al (2012) stating that in MLR applications, independents variables often intercorrelated, resulting in a statistical phenomenon that is when correlation of predictors are high, means that associations when there are correlations between independent variable. This referred to as multicollinearity. For situations where the data are drawn from reasonably homogeneous populations and the response (Y) is a normally distributed, traditional method such as MLR can yield insightful analyses. The usefulness of MLR can breakdown quickly if these stringent assumptions are not met. According to Young et al (2008), MLR has three important assumptions which are:

a) linearity of the coefficients
b) normal or Gaussian distribution for the response errors
c) the errors have a common distribution. In many industrial settings when modeling a quality characteristic, these assumptions may not be valid.

Quantile Regression (QR) is an approach that allows us to examine the behavior of the response variable (Y) beyond its average of the Gaussian distribution like median (50th percentile), 10th percentile, 80th percentile, 90th percentile, and so on. Examining the behavior of the regression curve for the response variable (Y) for different quantiles with respect to the independent variables (X) may result in very different conclusions relative to examining only the average of Y. Examining the lower percentiles using QR may be more important and be more beneficial for continuous improvement and cost savings.

MLR and MSPC, both techniques involve huge number of variables in a system. MLR is one of multivariate statistical technique which simplifies the number of variables in the system by preserving the original process data. If MLR is applied into PCA which is used in MSPC, it produces less number of dimensions. This certainly improves and makes the calculation process easier which help boosting the process control and monitoring performance. By applying the multiple linear regressions in the system, the estimates of the unknown parameters obtained from linear least squares regression are the optimal. It uses data very efficiently. Good results can be obtained with relatively small data sets. The theory associated with linear regression is well-understood and allows for construction of different types of easily-interpretable statistical intervals for predictions, calibrations, and optimizations.

CHAPTER 3

METHODOLOGY

3.1 Introduction

This chapter explains on how the study has been done. It explains the procedure for the process. The new method is applied to the on-line monitoring of a simulated continuous stirred tank reactor with recycle. However, an off-line modeling and monitoring is applied before the on-line monitoring process.

3.2 Fault Detection and identification

The complete procedures of fault detection and identification comprise of two main phases namely as off-line modelling and monitoring (Phase I) and on-line monitoring (Phase II) as show in Figure 3.1. In Phase I, it is an operation for the normal operating condition while Phase II is for the analysis of fault detection performance.

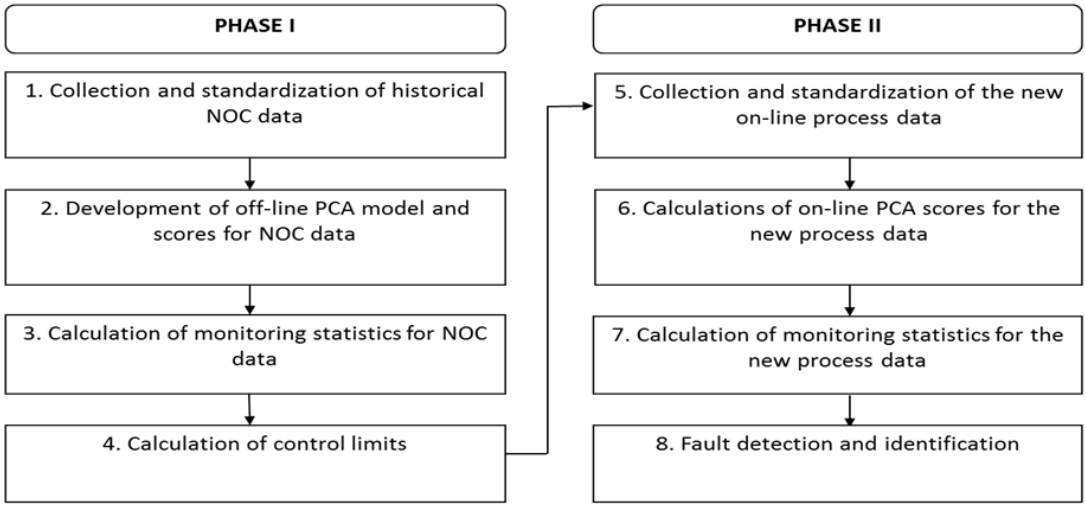


Figure 3.1 Procedures of fault detection and identification

From figure above, in Phase I which is the off-line modeling and monitoring, Firstly, a set of normal operation condition (NOC) data, $\mathbf{X}_{n\times m}$ (*n*: samples, *m*: variables), are identified off-line based on the historical process data archive. X is shown in the form of matrix.

$$\mathrm{X} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,m} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,m} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,m} \end{bmatrix} \tag{3.1}$$

NOC simply implies that the process is operated at the desired setting condition and produces satisfactory products that meet the qualitative as well as quantitative specified standard. Then, the data are then standardized to zero mean and unit variance with respective to each of the variables because PCA results depend on data scales.

$$\breve{x}_{j,i} = \frac{\left(x_{j,i} - \bar{x}_i\right)}{\sigma_i} \tag{3.2}$$

In the second step, the development of PCA model for the NOC data requires the establishment of a set of variance-covariance matrix, $\mathbf{C}_{mxm}$.

$$\mathrm{C} = \frac{1}{n-1}\breve{\mathrm{X}}'\breve{\mathrm{X}} = \begin{bmatrix} c_{1,1} & c_{1,2} & \cdots & c_{1,m} \\ c_{2,1} & c_{2,2} & \cdots & c_{2,m} \\ \vdots & \vdots & \vdots & \vdots \\ c_{m,1} & c_{m,2} & \cdots & c_{m,m} \end{bmatrix} \tag{3.3}$$

C is then transformed into a set of basic structures of eigen-based formula.

$$\mathrm{C} = \mathrm{V}\Lambda\mathrm{V}^{\mathrm{T}} \tag{3.4}$$

Finally, the PCA model of can be simply developed by:

$$\mathbf{P} = \breve{\mathbf{X}}\mathbf{V} \tag{3.5}$$

$$\mathbf{P} = \begin{bmatrix} \mathbf{p_1} & \cdots & \mathbf{p_m} \end{bmatrix} = \begin{bmatrix} \breve{x}_{1,1}v_{1,1} + \cdots + \breve{x}_{1,m}v_{m,1} & \cdots & \breve{x}_{1,1}v_{1,m} + \cdots + \breve{x}_{1,m}v_{m,m} \\ \vdots & \cdots & \vdots \\ \breve{x}_{n,1}v_{1,1} + \cdots + \breve{x}_{n,m}v_{m,1} & \cdots & \breve{x}_{n,1}v_{1,m} + \cdots + \breve{x}_{n,m}v_{m,m} \end{bmatrix} \tag{3.6}$$

The third step basically involves calculation of the Hotelling's T^2 distribution and Squared Prediction Error (SPE) monitoring statistics.

$$T_i^2 = \sum_{j=1}^{a} \frac{p_{i,j}^2}{\lambda_j} \tag{3.7}$$

$$\begin{aligned} \tilde{\tilde{\mathbf{E}}} &= \breve{\mathbf{X}} - \hat{\mathbf{X}} \\ &= \breve{\mathbf{X}} - \mathbf{P}_a \mathbf{V}_a^{\mathrm{T}} \\ &= \breve{\mathbf{X}} - \breve{\mathbf{X}}\mathbf{V}_a\mathbf{V}_a^{\mathbf{T}} \\ &= \breve{\mathbf{X}}\left(\mathbf{I} - \mathbf{V}_a\mathbf{V}_a^{\mathbf{T}}\right) \end{aligned} \tag{3.8}$$

$$SPE_i = \tilde{\tilde{\mathbf{e}}}_{\mathbf{i}} \tilde{\tilde{\mathbf{e}}}_{\mathbf{i}}^{T}$$

The final task in phase I deal with developing the control limits for both of the statistics. For this operation, the following equations are used.

$$T_\alpha = \frac{A(n-1)}{(n-A)} F_{A,n-A,\alpha} \tag{3.9}$$

$$SPE_\alpha = \theta_1 \left(\frac{z_\alpha \sqrt{2\theta_2 h_0^2}}{\theta_1} + \frac{\theta_1 h_0 (h_0 - 1)}{\theta_1^2} + 1 \right)^{\frac{1}{h_0}} \tag{3.10}$$

For the on-line monitoring which is Phase II, all the steps follow similar procedures of steps 1 to 3 in phase I. However for the last step, there are two main

operations which have to be conducted separately which are fault detection and fault identification. For fault detection, if any special event that is not in conformance to the common cause nature occurs, then it is regarded as a fault situation. A fault situation will be declared if either of the monitoring statistics exceeding its respective control limit for a pre-defined successive number of samples.

3.3 Application of Multiple Linear Regression Method

In this research, the new algorithm was implemented between step 1 and step 2 in Phase I which is for off-line modeling and monitoring. Those steps are based from Figure 3.1 which shows the procedure of fault detection. Similarly, this method was also being implemented in Phase II which is for the on-line monitoring, between step 5 and step 6 which is the collection and standardization of the new on-line process data and the calculation of the on-line PCA scores for the new process data. The algorithm will reduce the number of variables used but on the same time preserving the original data, which is need to be done before the calculation process. MLR is a statistical technique that uses several explanatory variables to predict the outcome of a response variable. The goal of multiple linear regressions (MLR) is to model the relationship between the explanatory and response variables. MLR takes a group of random variables and tries to find a mathematical relationship between them. The model creates a relationship in the form of a straight line (linear) that best approximates all the individual data points. MLR

is often used to determine how many specific factors such as the price of a commodity, interest rates, and particular industries or sectors, influence the price movement of an asset. For example, the current price of oil, lending rates, and the price movement of oil futures, can all have an effect on the price of an oil company's stock price. MLR could be used to model the impact that each of these variables has on stock's price.

A linear regression model assumes that the relationship between the dependent variable y_i and the p-vector of regressors x_i is linear. This relationship is modelled through a disturbance term or error variable, ε_i which is an unobserved random variable that adds noise to the linear relationship between the dependent variable and regressors. The classical first-order simple linear regression model has the form (Young et al.,2008),

$$y_i = \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i = x_i{}^T \beta + \varepsilon_i, \quad i = 1, \ldots, n, \qquad (3.11)$$

Where y_i is the value of the response variable in the i^{th} observation, β_1 is a slope parameter, x_{i1} is the value of the independent variable in the i^{th} observation, ε_i is a random error term of the i^{th} observation with mean $E(\varepsilon_i) = 0$ and variance $\sigma^2\{\varepsilon_i\} = \sigma^2$, with the error terms being independent and identically distributed, $i = 1, \ldots, n$. ^{T} denotes the transpose, so that $x_i{}^T\beta$ is the inner product between vectors x_i and β. Often these n equations are stacked together and written in vector form as

$$y = X\beta + \varepsilon \qquad (3.12)$$

where

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{x}_1^{\mathrm{T}} \\ \mathbf{x}_2^{\mathrm{T}} \\ \vdots \\ \mathbf{x}_n^{\mathrm{T}} \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ x_{21} & \cdots & x_{2p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

The least squares method is a common method in simple regression and MLR and is used to find an affine function that best fits a given set of data. Ordinary least squares (OLS) is the simplest and thus most common estimator. It is conceptually simple and computationally straightforward. OLS estimates are commonly used to analyze both experimental and observational data. The OLS method minimizes the sum of squared residuals, and leads to a closed-form expression for the estimated value of the unknown parameter β:

$$\beta = (X^T X)^{-1} X^T y \tag{3.13}$$

The estimator is unbiased and consistent if the errors have finite variance and are uncorrelated with the regressor.

CHAPTER 4

RESULT AND DISCUSSION

4.1 Introduction

This chapter shows the results of work that has been presented. The results emphasize the integration of process monitoring algorithm based on conventional PCA and CMDS and also the Multiple Linear Regression-based MSPM method. Firstly, evidence on the case study used for this analysis is described briefly and then the PCA and MLR results are also discussed. At the end, a summary is given in short.

4.2 Case Study

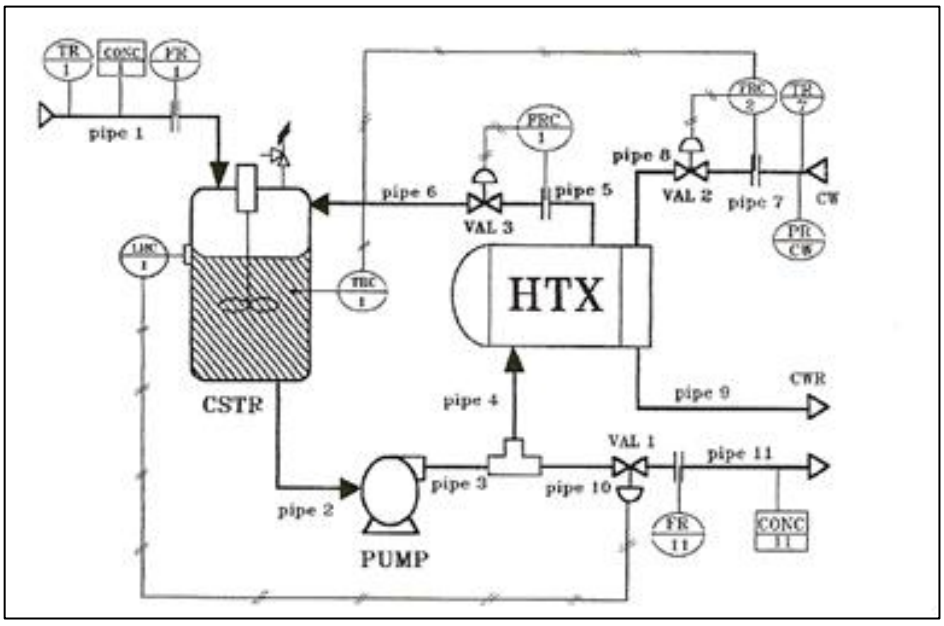


Figure 4.1 A CSTRwR system

A continuous stirred tank reactor (CSTRwR) system is used for the case study. The CSTRwR system is presented in Fig. 4.1. In this system, an irreversible heterogeneous catalytic exothermic reaction from reactant A to product B takes place in the reactor vessel. The process objective is to indirectly maintain the product concentration at a desired level by controlling three parameters which are temperature, residence time and mixing conditions in the CSTR. Temperature in the reactor is controlled by manipulating the flow rate of the cold water fed to the heat exchanger via a cascade control system. The residence time is controlled by maintaining the level in the reactor, and the mixing condition is controlled by maintaining the recycle flow rate. A

recent study shows that a set of multiple neural networks algorithms has been developed to enhance the reliability of fault diagnosis operation for this system (Zhang, 2006). In this process, there are ten on-line measured process variables and three controller outputs. As a result, thirteen on-line information sources are considered as listed in Table 4.1.

Table 4.1 List of variables in the CSTRwR system

Process			Instruments		
No.	Variables	Variable Names	No.	Variables	Variable Names
1	V1	Tank temperature	11	V11	Controller 1
2	V2	Tank level	12	V12	Controller 3
3	V3	Feed temperature	13	V13	Controller 2
4	V4	Inlet flow rate			
5	V5	Recycle flow rate			
6	V6	Outlet flow rate			
7	V7	Cooling water flow rate			
8	V8	Product concentration			
9	V9	Feed concentration			
10	V10	Heat exchanger entrance pressure			

4.3 Overall Monitoring Performance

4.3.1 First Phase (Off-line Modeling and Monitoring)

At the beginning, the standardized NOC data were taken to be modeled based on two approaches which are the conventional PCA method and also MLR-PCA based method. Subsequently, the accumulated data variances explained by those two approaches are shown in Figure 4.2. Figure 4.2 show the accumulated data variance explained by different principle components for PCA method (left) and MLR-PCA method (right).

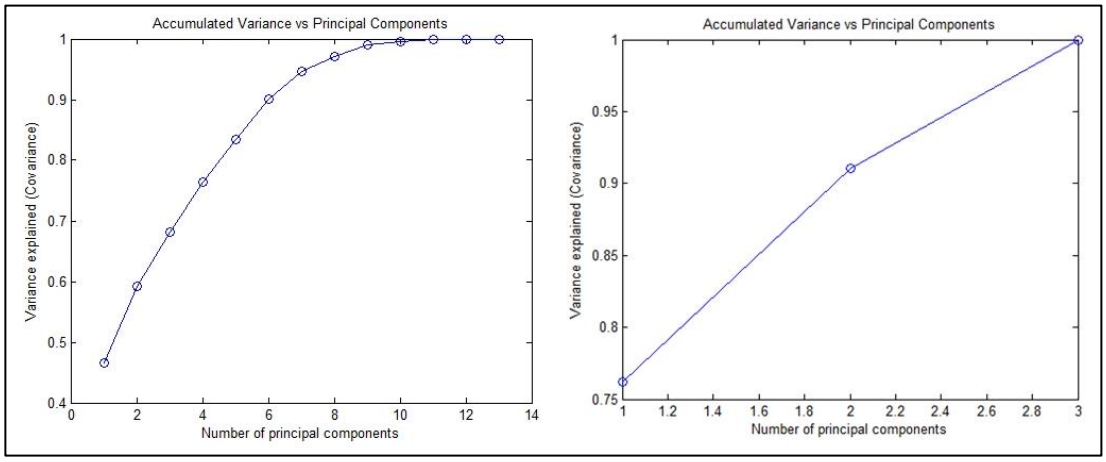


Figure 4.2 Accumulated data variance explained by different PCs for PCA method and MLR-PCA method.

From the analysis, the statistic outcomes based on MLR-PCA did not give the same result as PCA. This indicates that MLR-PCA results are not equivalent to PCA results. In the case of MLR-PCA based MSPM approach, the trend of the accumulated data variation explained by MLR-PCA based MSPM approaches is not similar to that in PCA. From the graph, by using 3 principle components, the variance explained is about 100% instead of about 67% in PCA. Variance explained by the MLR-PCA method is higher than the PCA method. This shows that the new method is better than the conventional method.

4.3.1.1 Monitoring Outcome Based on Three PCs

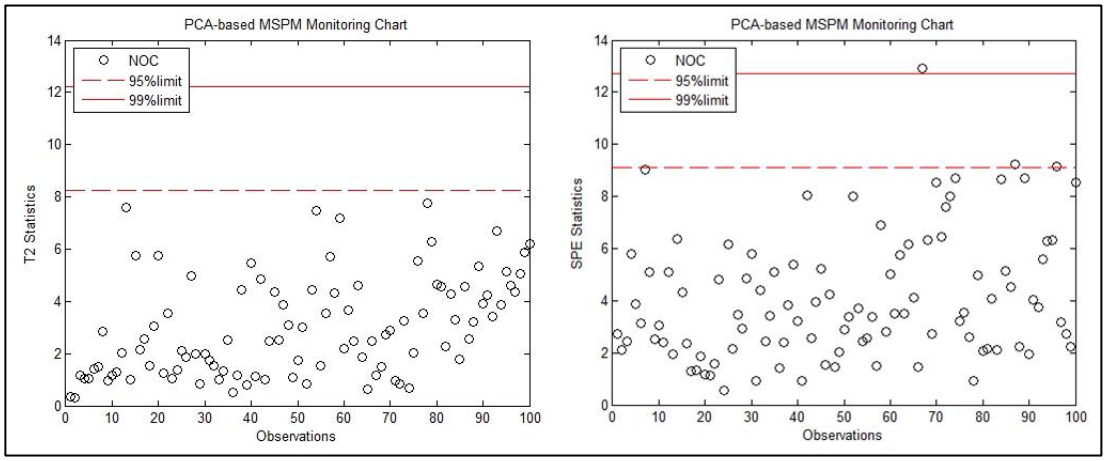


Figure 4.3 PCA based MSPM Monitoring Chart for NOC Data

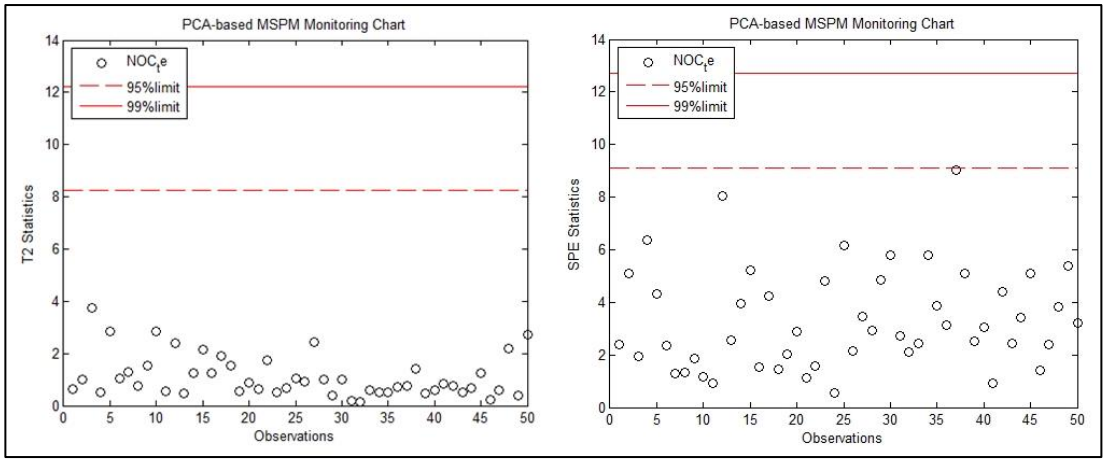


Figure 4.4 PCA based MSPM Monitoring Chart for NOC Test Data

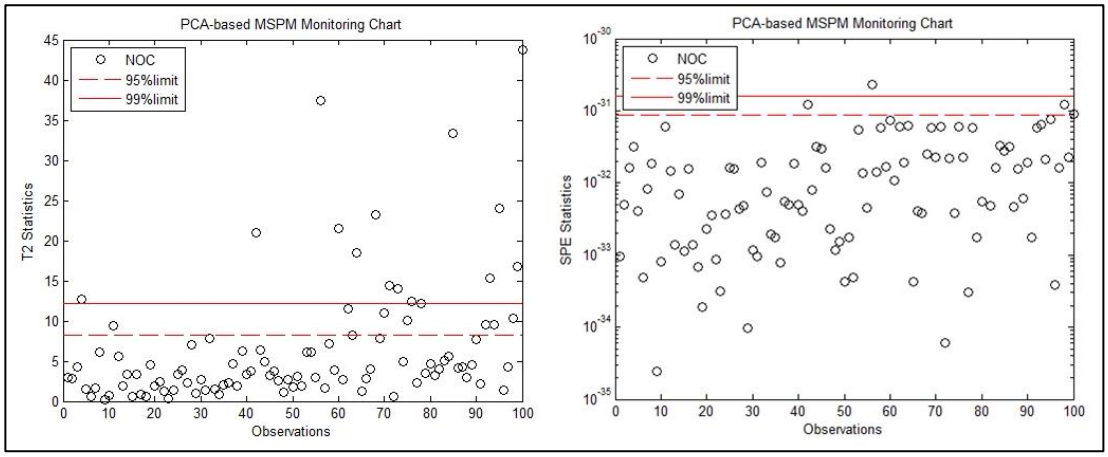


Figure 4.5 MLR-PCA based MSPM Monitoring Chart for NOC Data

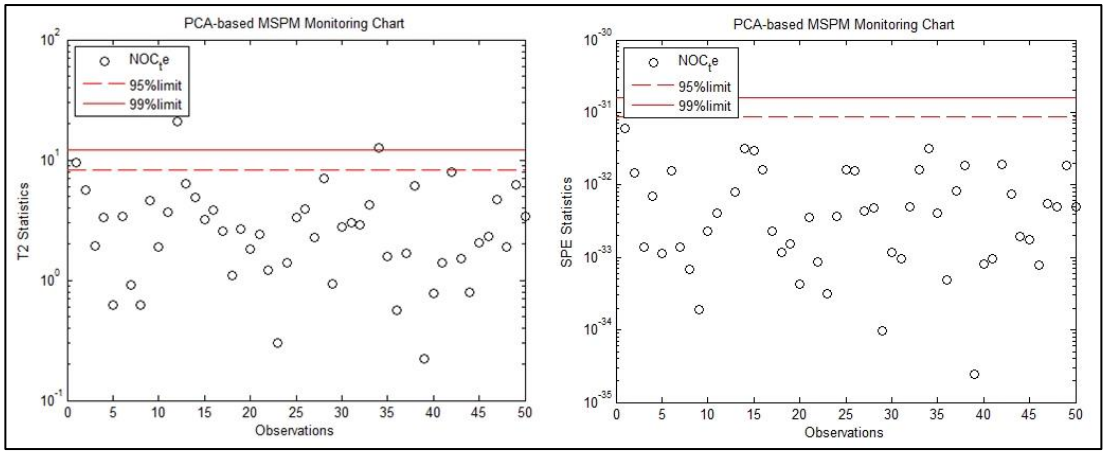


Figure 4.6 MLR-PCA based MSPM Monitoring Chart for NOC Test Data

The T^2 and SPE statistics for the NOC data as well as the NOC test data were calculated. The Hoteling's T^2 statistic and SPE statistics were then used to calculate the confidence limits of the values. After plotting those values with respect to their confidence limits which are fixed at 95% and 99% respectively, eight figures were produced, and categorized as shown individually from Figure 4.3 to Figure 4.6. Figure 4.3 and Figure 4.4 show the monitoring progression of T^2 (left) and SPE (right) based on PCA models for NOC data and NOC Test data. Figure 4.5 and Figure 4.6 show the monitoring progression of T^2 (left) and SPE (right) based on MLR-PCA models for NOC data and NOC test data.

From the observations, for Figure 4.3, the results from the figure show that the T^2 values for NOC data are within the control limits whereas approximately 1 sample out of 100 measurements for SPE values are placed outside the control limits. For Figure 4.4, both the T^2 values and SPE values for NOC test data are within the control limits. For Figure 4.5, about 14 samples out of 100 samples of the T^2 value and 1sample out of 100

samples of the SPE values are within the control limits while all the balance are placed outside the control limits. For Figure 4.6, **i**t can be seen that the SPE statistics for the NOC test data are located well below the confidence limits, whereas about 2 samples of the NOC test data for T^2 value are placed far beyond the confidence limits.

4.3.2 Second Phase (On-line Monitoring)

The process is described by 13 process variables, which actually are various sensor indicators. A set of historical data known to be as NOC data in this thesis, represented by the 100 samples was obtained from simulation. Each sample represents a time point where sensor indices were fixed. In order to evaluate the robustness of the monitoring limits, another set of NOC data (the second set of NOC) containing of 50 samples were also collected. The system also subjects to be affected from several malfunction conditions as summarized in Table 4.2.

Table 4.2 List of abnormal operations in CSTRwR

Fault Cases	Fault Causes
9	Pipe 4, 5, or 6 is blocked or control valve 3 fails low
10	Control valve 3 fails high

Table 4.2 shows the possible fault list that can be occur in the system. For each fault presented in Table 4.2, both abrupt and incipient faults are considered. An abrupt fault indicates a sudden change (or step change) in a process variable or parameter and typically it maintains over the operation time until the cause is completely removed. Detecting this kind of malfunctions should be easy for any multivariate monitoring system as the deviations are usually very obvious. On the other hand, an incipient fault depicts a kind of fault that gradually deviates from the normal setting. Thus, the monitoring system typically takes a while to detect these particular abnormal behaviors. In particular, all the faults were introduced at sample 2 and the sampling time was fixed at 4 seconds.

4.3.2.1 Monitoring Outcomes Based on Three PCs

Table 4.3 Fault detection performance

Fault No.	Conventional PCA			MLR Method		
	T^2	SPE		T^2	SPE	
9a	0	1	1	1	1	1
9i	0	29	29	5	8	5
10a	1	1	1	1	1	1
10i	0	30	30	6	9	6

Table 4.3 show the simplified table for the fault detection analysis based on the conventional PCA based MSPM method and also the MLR-PCA based MSPM method. The fault is either detected by the T^2 statistic or the SPE statistic. The fault is detected based on the confidence limits that have been calculated. First, explanation on conventional method is discussed. From these observations, for fault 9, abrupt faults, it could be said that the faults were initially started after sample 1 up to 20. The faults started after sample 29 up to 51 for the insipient fault. For fault 10, the abrupt fault was detected at first sample after the fault is introduced. However, it takes longer time for the insipient fault to be detected which is at sample 30. Next is the MLR-PCA based MSPM method. From the results, for both abrupt fault of fault 9 and fault 10, the fault is detected at the first sample as soon as the fault is introduced. The faults started after sample 5 up to 51 for the insipient fault of fault 9. As for the insipient fault of fault 10, the fault is detected at sample 6. These indicate that the abnormal process behavior can

be detected by the MLR-PCA based MSPM model quite efficiently. Explanation more on the fault detection analysis is discussed below. Fault 9 is used for the explanation.

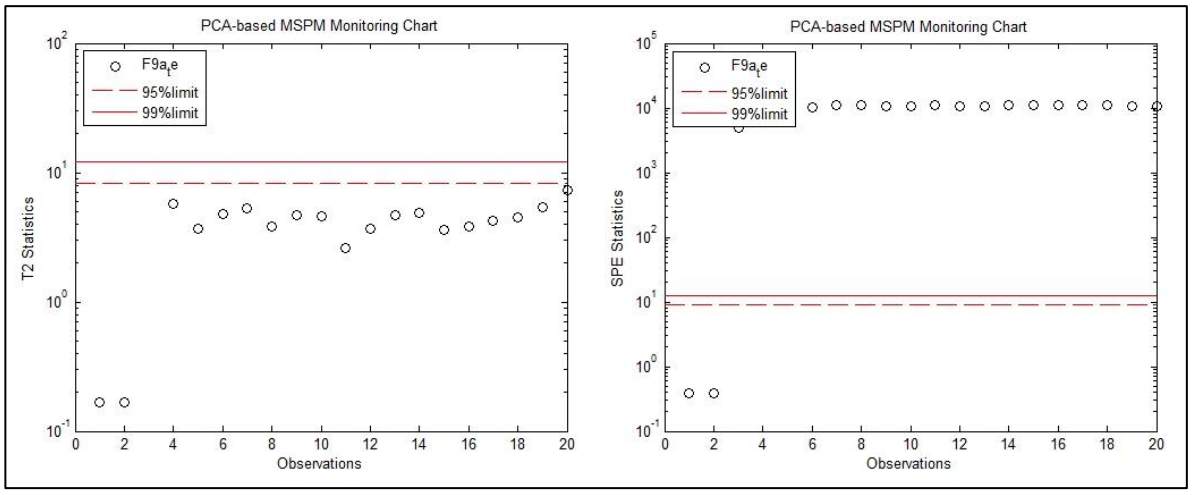


Figure 4.7 PCA based MSPM Monitoring Chart for Fault 9a

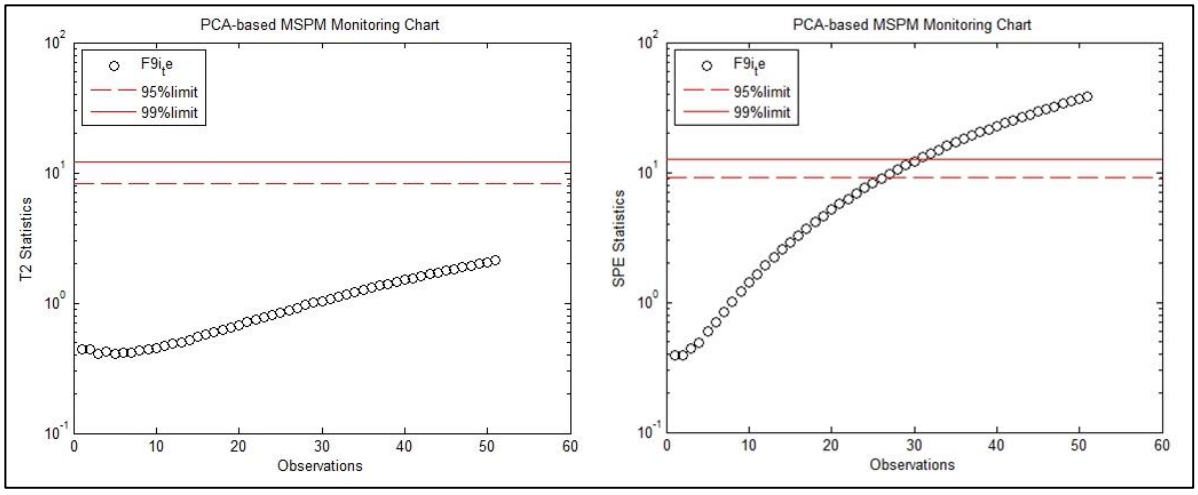


Figure 4.8 PCA-based MSPM Monitoring Chart for Fault 9i

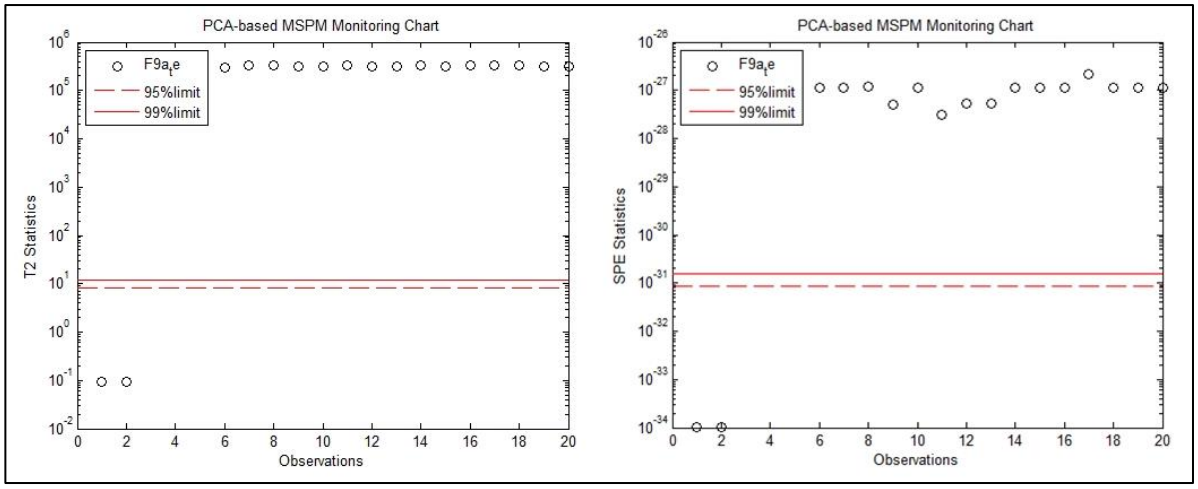


Figure 4.9 MLR-PCA based MSPM Monitoring Chart for Fault 9a

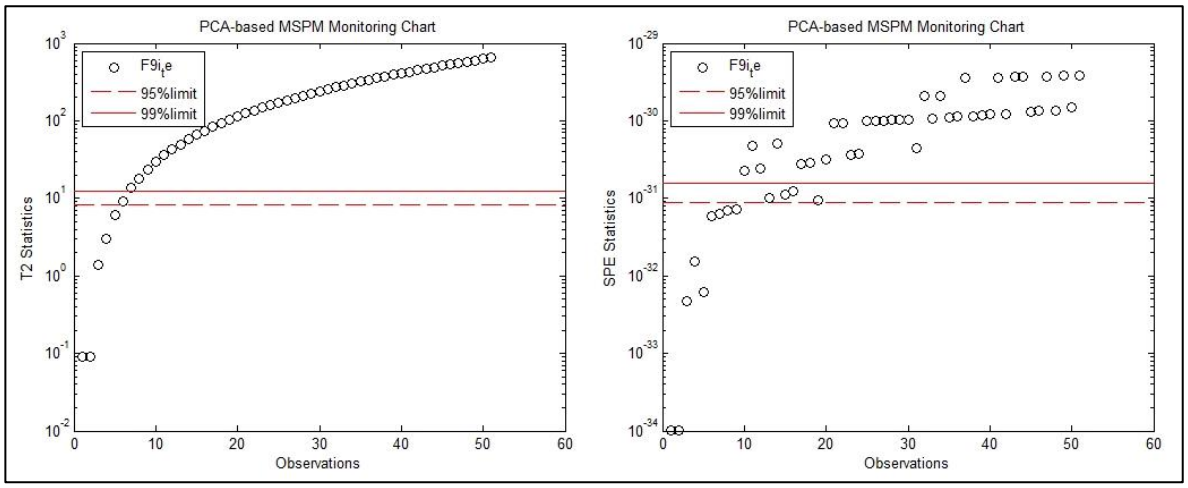


Figure 4.10 MLR-PCA based MSPM Monitoring Chart for Fault 9i

The T^2 statistics for the NOC data as well as the fault data were calculated. Similarly, the SPE statistics for both NOC and fault data were also calculated as well. Figure 4.8 show the monitoring progression of T^2 (left) and SPE (right) on F9 based on

MLR-PCA models for incipient fault while Figure 4.7 show the monitoring progression of T^2 (left) and SPE (right) on Fault 9 based on MLR-PCA models for abrupt fault.

From the observations, for Figure 4.7, the results from the figure show that the T^2 values for NOC data are within the control limits whereas approximately 18 samples out of 20 measurements for SPE values are placed outside the control limits. For Figure 4.8, T^2 values for NOC data are also within the control limits whereas approximately 20 samples out of 51 measurements are placed outside the control limits. For Figure 4.9, approximately, about 2 samples out of 20 samples for both T^2 and SPE value are within the control limits whereas the others are outside the control limits. Lastly, for Figure 4.10, the fault gradually develops with time, about 7 samples out of 51 samples of the T^2 value and 10 samples out of 51 samples of the SPE values are within the control limits while all the balance are placed outside the control limits.

In comparison with the results of PCA depicted in Figure 4.7 and Figure 4.8, the MLR-PCA based MSPM in Figure 4.9 and Figure 4.10 shows comparable outcomes with respect to the point locations. These signify that the faults can also be detected based on the T^2 and SPE statistics calculated from the current PCA model. These results show that MLR approaches is really effective to calculate the multivariate scores or values when the data is subjected to different model other than PCA model. However, more analyses have to be performed in order to validate the confidence limits calculation for all the values.

4.4 Summary

Conventional PCA and MLR-PCA based MSPM algorithm were used in process monitoring through the application on a simulated CSTRwR process. Lastly, both NOC data as well as fault data for MLR-PCA based MSPM results were compared to the PCA results and several points have been highlighted. From the results above, it can be said that by using the new algorithm, the fault detection performance is better and can be improve.

CHAPTER 5

CONCLUSION AND RECOMMENDATION

5.1 Introduction

This chapter concludes the entire chapter in this proposed research about the application of process monitoring based on inferential approach. The conclusion consists of the three chapters in these researches which are the introduction, conclusion and recommendation.

5.2 Conclusion

Previously, the research objective for this research is to implement and analyze Multiple Linear Regression (MLR) method in the system. From that, the performance of the system is analyzed and it is shown that it is possible for the new method to be implemented to the system. At the beginning, the conventional MSPM method of process monitoring was firstly studied. The results were determined based on the data that have been collected from the process. In other word, by using this new method, this research study was analyzed and was implemented to the system. The results show that a new method was developed which can improve the fault detection based on MSPM by applying Multiple Linear Regression (MLR) method to the system.

5.3 Recommendation

There are several further research works identified to enhance this study more effectively. Firstly, the study was implemented on several fault cases only. In this study, the only fault cases used just only two cases out of 11 fault cases. . All these fault data will be applied accordingly with the newly designed process monitoring algorithm to really investigate the degree of sensitiveness during the fault detection operation. Thus,

to get a better result, which can give support to the conclusion that has been made, further study on all fault cases need to be made.

Next, the other variables in the system should also be tested by using the MLR-based MSPM method to know the performance of fault detection in the system. This need to be made to predict the best variable needed to be implemented in the simulation. Thus, we know the exact performance of the algorithm.

REFFERENCES

Alvarez, D. G. (n.d.). Fault detection using Principal Component Analysis (PCA) in a Wastewater Treatment Plant (WWTP).

Bakshi, B. R. (1998). Multiscale PCA with Application to Multivariate Statistical Process Monitoring. AIChE Journal.

Chen, J. C., Chen, J. C. (2004). A Multiple-Regression Model for Monitoring Tool Wear with a Dynamometer in Milling Operations. Journal of Technology Studies.

Damarla, S. K. (2011). Multivariate Statistical Process Monitoring and Control. Retrieved from *http://ethesis.nitrkl.ac.in/2941/1/dvenugopalarao_Thesis.pdf*

Delijaicov, S., Fleury, A. T., Martins, F. P. R. (2009). Application of multiple regression and neural networks to synthesize a model for peen forming process planning. Journal of Achievements in Materials and Manufacturing Engineering.

Holland, S. M. (2008). Principle Component Analysis (PCA).

Juricek, B. C., Seborg, D. E., Larimore, W. E. (n.d.). Process Control Applications of Subspace and Regression-based Identification and Monitoring Methods.

Kano, M., Hasebe, S., Hashimoto, I., Ohno, H. (2001). A New Multivariate Statistical Process Monitoring Method Using Principle Component Analysis. Computers & Chemical Engineering.

Kruger, U., Zhang, J., Xie, L. Developments and Applications of Nonlinear Principal Component Analysis – a Review. Retrieved from *http://pca.narod.ru/1MainGorbanKeglWunschZin.pdf*

Nathans, L. L., Oswald, F. L., Nimon, K. (2012). Interpering Multiple Linear Regression: A Guidebook of Variable Importance. Practical Assessment, Research and Evaluation Journal.

Nguyen, M. H., De la Torre, F. Robust Kernel Principal Component Analysis. Retrieved from *http://www.andrew.cmu.edu/user/minhhoan/papers/RKPCA_NIPS08.pdf*

Nikolov, S. (2010). Principle Component Analysis: Review and Extensions. Retrieved from *http://web.mit.edu/snikolov/Public/pca.pdf*

Wise, B.M. (n.d.).Monitoring and Fault Detection with Multivariate Statistical Process Control (MSPC) in Continuous and Batch Processes. Eigenvector Research.

Smith, L. I. (2002). A tutorial on Principle Component Analysis.

Venkatasubramanian , V., Rengaswamy , R., Yin , K., Kavuri , S. N. (2002). A review of process fault detection and diagnosis. Part I: Quantitative model-based methods. Journal of Computers and Chemical Engineering.

Venkatasubramanian , V., Rengaswamy , R., Yin , K., Kavuri , S. N. (2002). A review of process fault detection and diagnosis. Part II: Qualitative models and search strategies. Journal of Computers and Chemical Engineering.

Venkatasubramanian , V., Rengaswamy , R., Yin , K., Kavuri , S. N. (2002). A review of process fault detection and diagnosis. Part III: Process history based methods Journal of Computers and Chemical Engineering.

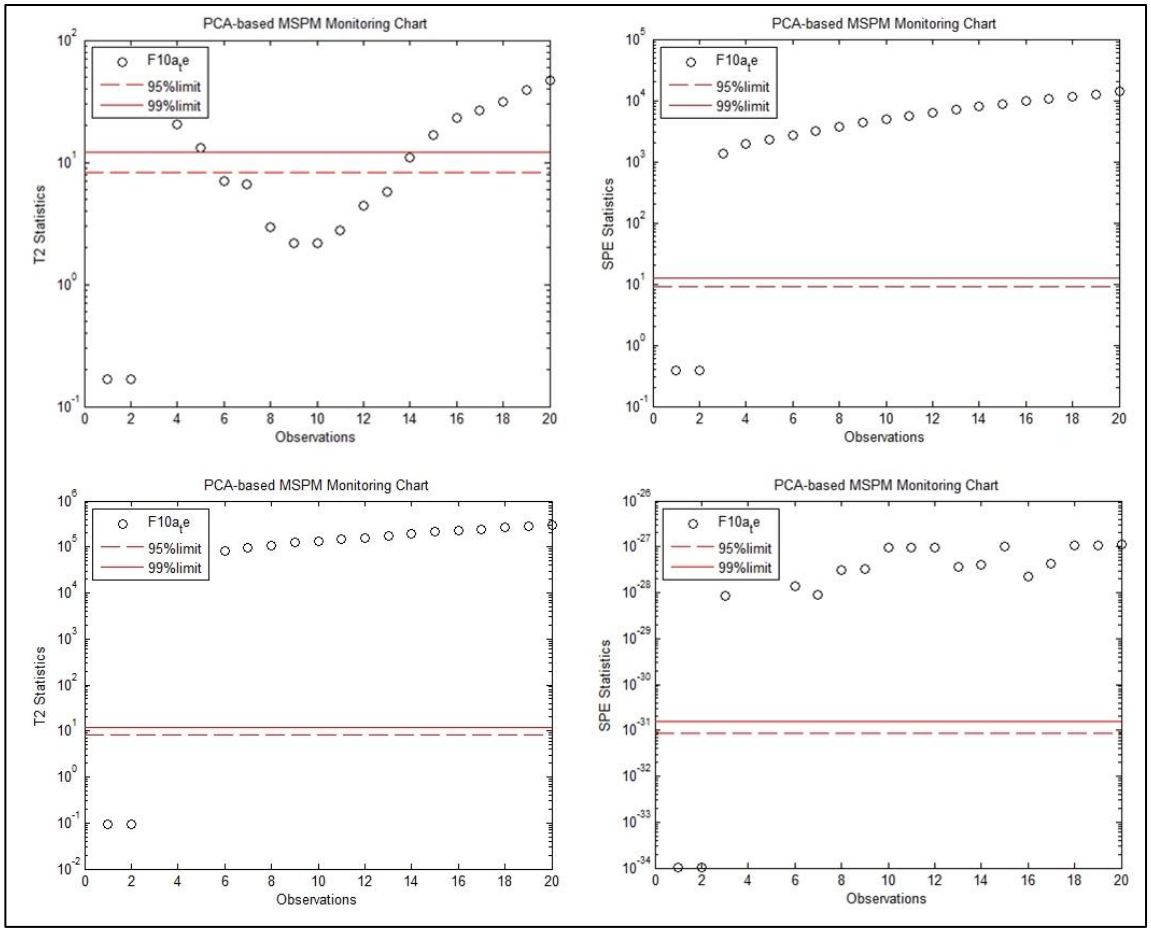
Young, T. M., Shaffer, L. B., Guess F.M., Bensmail, H., Leon, R. V. (2008). A Comparison of Multiple Linear Regression and Quantile Regression for Modeling The Internal Bond of Medium Density Fiberboard. Forest Products Journal .VOL. 58, NO. 4.

Zhang, J. (2006). Improved On-line Process Fault Diagnosis Through Information Fusion in Multiple Neural Networks. Computers and Chemical Engineering.

Zhao, S., Xu, Y. (2005). Multivariate Statistical Process Monitoring Using Robust Nonlinear Principal Component Analysis. Tsinghua Science and Technology.

APPENDIX A

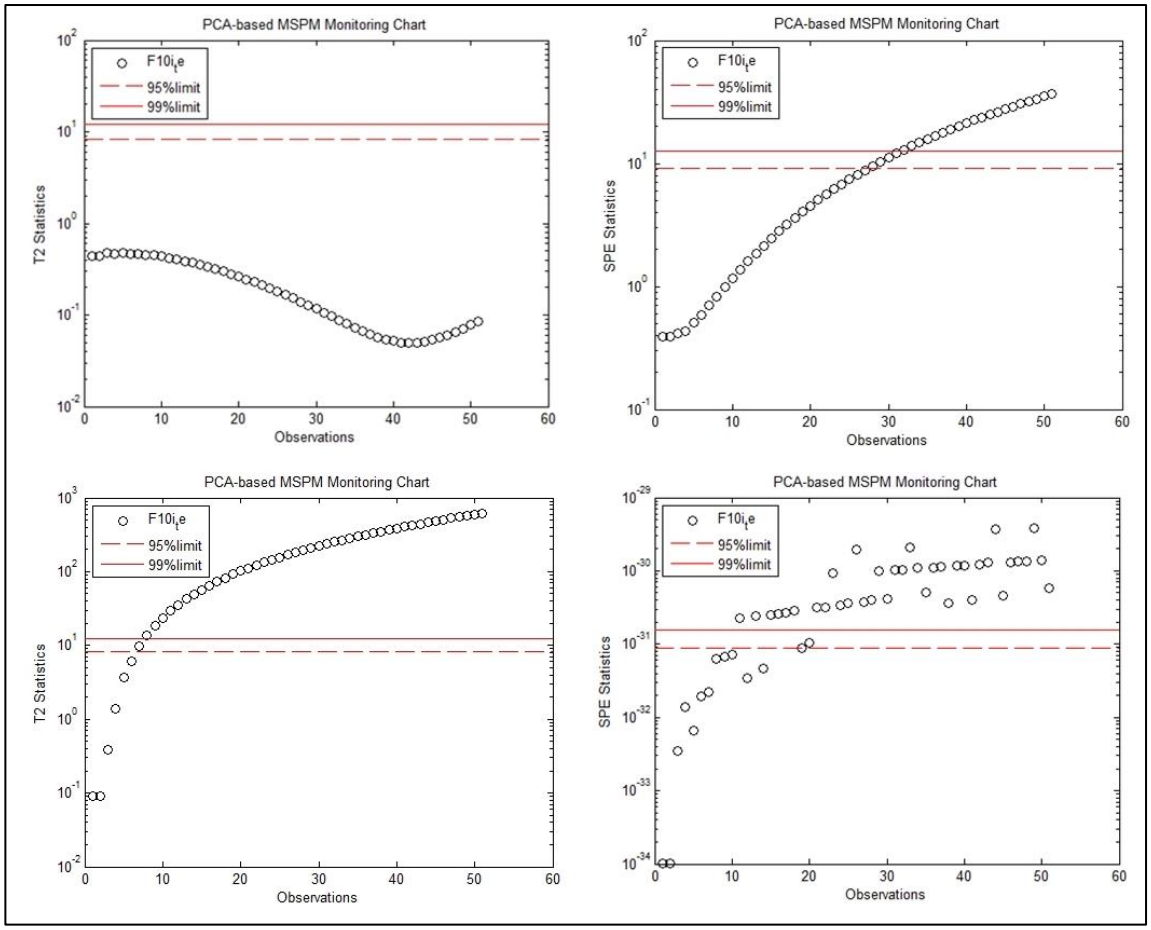
MSPM Monitoring Chart for Fault 10a



Monitoring progression of T^2 (left top) and SPE (right top) based on PCA model. Monitoring progression of T^2 (left bottom) and SPE (right bottom) based on MLR-PCA model.

APPENDIX B

MSPM Monitoring Chart for Fault 10i



Monitoring progression of T^2 (left top) and SPE (right top) based on PCA model. Monitoring progression of T^2 (left bottom) and SPE (right bottom) based on MLR-PCA model.