ENHANCEMENT OF PCA-BASED FAULT DETECTION SYSTEM THROUGH UTILISING DISSIMILARITY MATRIX FOR CONTINUOUS-BASED PROCESS

NUR AFIFAH BINTI HASSAN

BACHELOR OF CHEMICAL ENGINEERING

UNIVERSITI MALAYSIA PAHANG

NUR AFIFAH BINTI HASSAN   BACHELOR OF CHEMICAL ENGINEERING   2013   UMP

ENHANCEMENT OF PCA-BASED FAULT DETECTION SYSTEM THROUGH UTILISING DISSIMILARITY MATRIX FOR CONTINUOUS-BASED PROCESS

NUR AFIFAH BINTI HASSAN

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

Faculty of Chemical Engineering and Natural Resources
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:

Position:

Date:

STUDENT'S DECLARATION

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

Signature:

Name:

ID Number:

Date:

Dedicated to my parents

ACKNOWLEDGEMENTS

First and foremost, I would like express my sincerest gratitude to my beloved supervisor, Dr. Mohd Yusri Bin Mohd Yunus for their ideas, encouragement, invaluable guidance, critics, and friendship in overseeing the progress of my research until the completion of this thesis. Without his continued support, this thesis would not have been completed as presented here. I believe that all of their advices and comment are for the benefit of this thesis.

Furthermore, I would like to express our deepest gratitude to my family especially my parents for their love and sacrifice throughout my life. In particular, the patience and understanding shown by my mum, dad and sister during the honours year are greatly appreciated. Last but not least, special thanks to my friend and everyone who given a lot of moral support and encouragement to me in making of this thesis.

Praise to God for His help and guidance that I am finally able to complete this thesis.

ABSTRACT

This research is about enhancement of PCA-based fault detection system through utilizing dissimilarity matrix. Nowadays, the chemical process industry is highly based on the non-linear relationships between measured variables. However, the conventional PCA-based MSPC is no longer effective because it only valid for the linear relationships between measured variables. Due in order to solve this problem, the technique of dissimilarity matrix is used in multivariate statistical process control as alternative technique which models the non-linear process and can improve the process monitoring performance. The conventional PCA system was run and the dissimilarity system was developed and lastly the monitoring performance in each technique were compared and analysed to achieve aims of this research. This research is to be done by using Matlab software. The findings of this study are illustrated in the form of Hotelling's T^2 and Squared Prediction Errors (SPE) monitoring statistics to be analysed. As a conclusion, the dissimilarity system is comparable to the conventional method. Thus can be the other alternative ways in the process monitoring performance. Finally, it is recommended to use data from other chemical processing systems for more concrete justification of the new technique.

ABSTRAK

Kajian ini adalah tentang peningkatan PCA berasaskan sistem pengesanan kesalahan melalui perbezaan matrik. Kini, proses industri kimia adalah berdasarkan hubungan bukan linear antara pembolehubah yang diukur. Walaubagaimanapun, konvensional PCA berasaskan MSPC adalah tidak lagi berkesan kerana ia hanya sah untuk hubungan linear antara pembolehubah yang diukur. Oleh kerana dalam usaha untuk menyelesaikan masalah ini, teknik perbezaan matrik yang digunakan dalam kawalan proses multivariat statistik sebagai alternatif teknik model proses bukan linear dan boleh meningkatkan prestasi proses pemantauan. Sistem PCA konvensional telah dijalankan dan sistem perbezaan telah dibangunkan dan akhir sekali pemantauan prestasi dalam setiap teknik dibandingkan dan dianalisis untuk mencapai matlamat kajian ini. Kajian ini adalah untuk dilakukan dengan menggunakan perisian Matlab. Dapatan kajian ini digambarkan dalam bentuk "Hotelling's T^2" dan " Squared Prediction Errors" (SPE) statistik pemantauan untuk dianalisis. Sebagai kesimpulan, sistem perbezaan adalah setanding dengan kaedah konvensional. Oleh itu boleh menjadi cara alternatif lain dalam proses pemantauan prestasi. Akhirnya, ia adalah disyorkan untuk menggunakan data daripada sistem pemprosesan kimia lain untuk justifikasi yang lebih konkrit untuk teknik baru ini.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

X	Normal operating data
X^T	Normal operating data transpose
$\check{X}$	Standardised data
C_{mxm}	Variance-covariance matrix
λ	Eigen values
V	Eigenvectors
P	PC scores
T	Matrix of non-linear PC scores
F(.)	Non-linear PC loading function
E	Residual matrix
I	Batch samples
J	Process variables
K	Time
i	Row
j	Column
R_i	Range of the variable Z_i
B	Scalar product matrix
q_i	Loading vector of PCA
x	Data
$\bar{x}$	Data means
σ	Standard deviation
k	Principal component

A	Number of PCs retained in the PCA model
n	Number of nominal process measurements per variable
$p_{i,j}$	i^{th} score for Principal Component j
λ_j	Eigenvalue corresponds to Principal Component j
z_α	Standard normal deviate corresponding to the upper (1-α) percentile
$\mathbf{X}_\mathbf{z}$	Standardized matrix of original matrix, X
E	Residual matrix (n x m)
I	Identity matrix
$V_{\boldsymbol{A}}$	Eigenvector matrix contains up to A eigenvectors
e_i	i^{th} row in residual matrix
Q_i	SPE statistics
$\{\delta_{rs}\}$	Dissimilarity
Λ	Diagonal matrix
V^T	Normalized orthogonal matrix

LIST OF ABBREVIATIONS

PBR	Packed bed reactor
PFR	Plug flow reactor
CA	Canonical correlation analysis
CSTRwR	Simulated continuous stirred tank reactor with recycle
CVA	Canonical variate analysis.
FA	Factor analysis
F1	Fault 1
F2	Fault 2
ICA	Independent component analysis
IT-net	Input-training neural network
MDS	Multidimensional scaling
MPCA	Multi-way PCA
MSPC	Multivariate statistical process control
MSPCA	Multi-scale PCA
MSPM	Multivariate statistical process monitoring
NOC	Normal operating data
PARAFAC	Parallel factors analysis
PC	Principal component
PCA	Principal component analysis
PLS	Partial least square
SD	Singular decomposition
SVD	Singular value decomposition
SPC	Statistical process control

SPE	Squared prediction errors

CHAPTER I

INTRODUCTION

1.1 Research Background

In general, there are two typical types of process monitoring schemes applied widely in chemical-based industry, which are individual-based monitoring also known as Statistical Process Control (SPC) and multivariate-based monitoring that also synonymous to Multivariate Statistical Process Control (MSPC) or Multivariate Statistical Process Monitoring (MSPM).

Traditionally, SPC performs a toolkit for managing process malfunction by way of providing early warning through fault detection (Montgomery, 1985; Grant and Leavenworth, 1988; Wetherill and Brown, 1991). The control chart is one of the key tools which are used to monitor the processes that are in control by using mean and range. According to Cinar, Palazoglu and Kayihan (2007), the generic purpose of statistical process control (SPC) is to detect the nature of faults in the process that lead to disastrous deviation from the desired goal. Among others, the main procedures should include data collection, control chart development, and followed by control chart progression analysis. The next step involves process diagnosis, which is to find the root cause of the changes as well as execute corrective actions corresponding to the nature of the faults. Thus, on-line monitoring and diagnosis are important to ensure that high quality product can be maintained over the period of operations (MacGregor, 1994).

Unfortunately, SPC has its own weaknesses and as a result MSPM is introduced. The main limitation of SPC is that it ignores the correlations among the monitored variables (Cinar, et al., 2007). This limitation is addressed by MSPM for further enhancement in the quality control mechanisms.

1.2 Problem Statement and Motivation

Over the last decade, the field of the process monitoring performance and fault diagnosis in chemical process industry has used MSPM as an alternative method based on the existing knowledge. One of the tools multivariable statistical techniques is Principal Component Analysis (PCA) and its extension which can indicate the strong correlations of the data set through a set of empirical orthogonal function (Cinar, et al., 2007). However, "conventional PCA-based MSPM is only valid for the non-autocorrelated data with linear relationships between measured variables. Often, inefficient and unreliable process performance monitoring schemes can materialize as a consequence of the underlying assumptions of PCA-based MSPM being violated" (Choi, Morris and Lee, 2008). Furthermore, based on the study of Choi, Martin and Morris (2005), a large amount of the principal components are retained to clarify a large proportion of the sample variance when dealing with the non-linear relationship between measured variable. Simultaneously, this leads to an increase in the probability of false alarms which happen especially in the T^2 statistic and result of the decrease in order of the components that only explain minimum level of variability.

Recently, the chemical process industry is highly based on the non-linear relationships between measured variables. Nowadays, the conventional PCA-based MSPM is no longer effective for the field of the process monitoring performance and fault diagnosis in a chemical process industry. Therefore, engineer has to find another alternative technique which can solve the current problem of the process monitoring performance and fault diagnosis in a chemical process industry to achieve quality control expectation as the goal to produce the maximum amount of highly quality product that requested and specified by the customer. Perhaps the technique of dissimilarity-based MSPM used in multivariate statistical process control can solve the current problem which models the non-linear process. Fundamentally, dissimilarity technique is used inter distance measures which can cope either linear or non-linear process. Simultaneously, it can improve the process monitoring performance by using MSPM procedures. Thus, this research is to study and explore about the dissimilarity and perhaps can introduce it as another alternative to process monitoring.

1.3 Research Aims and Objectives

The main aim of this research is to propose a new technique in process monitoring which applies dissimilarity-based MSPM. The dissimilarity is based on the process monitoring for non-linear multivariate processes through the application of MSPM. Therefore, the main objectives of this research are:

i. To run the conventional PCA-based MSPM system.
ii. To develop the dissimilarity-based MSPM system.
iii. To compare and analyse the monitoring performance between the conventional PCA and dissimilarity techniques.

1.4 Research Questions

1.4.1 What are the types of scales which can be used by the new system in achieving consistent process monitoring performance?

1.4.2 How effective and efficient the new system may improve the process monitoring performance as compared to the conventional MSPM?

1.4.3 Do the outcomes support the research aim?

1.5 Research Scopes

The research scopes of this research are listed as follow:

i. To develop the conventional MSPM procedure in which the linear PCA algorithm is used for lowering the multivariate data dimensions.

ii. To study and explore about the dissimilarity matrix for constructing the core correlation structure.

iii. Using Matlab software platform version 7 as a tool to achieve the objectives stated earlier.

iv. Focusing on the fault detection scheme only.

v. The nature of the fault in this research includes incipient and abrupt.

vi. Using Shewhart chart to monitor the process performance.

vii. Using CSTRwR system as a case study.

viii. To develop NOC data model using one operating mode.

1.6 Expected Research Contributions

The main expected contribution of this research is to introduce dissimilarity as a new technique for modelling the variable correlation instead of applying PCA method. This study also examines the comparative performance between the proposed approach and the traditional PCA-based MSPM scheme especially in monitoring the multivariate non-linear process.

1.7 Chapter Organizations

The thesis is divided into five main chapters. The first chapter introduces the background of the research which includes the problem statement and motivation, objectives, scopes and contributions. The literature review is presented in chapter II, where it describes the fundamental of MSPM, process monitoring issues and extension and multidimensional scaling in the MSPM framework. Chapter III explains the proposed methodology. Chapter IV demonstrates the case study as well as explained the results of analysis, which cover the performance of conventional PCA-based MSPM system and dissimilarity-based MSPM system and finally, conclusion is presented in Chapter V.

CHAPTER II

LITERATURE REVIEW

2.1 Introduction

The aim of statistical process monitoring is to detect the occurrence and the nature of the operational change that cause the process to deviate from their main objective. The statistical technique is the method for detecting the changes on occurrence. The techniques include collection, classification, analysis and interpretation of data (Cinar, et al., 2007). This chapter is divided into five sections which are introduction, fundamental of MSPM, process monitoring issues and extension, dissimilarity in the MSPM framework and summary.

2.2 Fundamental of MSPC

A monitoring system is an observation system for the process to validate whether the process are happening according to planning and achieve their desired target. The system must supply the process with continuous flow of information throughout the time to make it possible to take the right decisions. This means, monitoring can be defined as a frequent observation and record of parameter taking place in a process and to check on how process are in progress. The report enables the collected information to be used in making the correct decisions for improving the process performance. The purposes of monitoring are to analyse the condition in the process, to determine whether the inputs in the process are well utilized, to identify the problems occur in the process and to determine whether the way the process was planned is the most appropriate way of solving the problem (Bartle, 2007).

In general, there are four main steps in MSPM in the field of the process monitoring performance and fault diagnosis. The four main steps consist of the fault detection, fault identification, fault diagnosis and process recovery. Graphically, the steps can be viewed in an arranged manner by referring to the following flow chart in Figure 2.1.

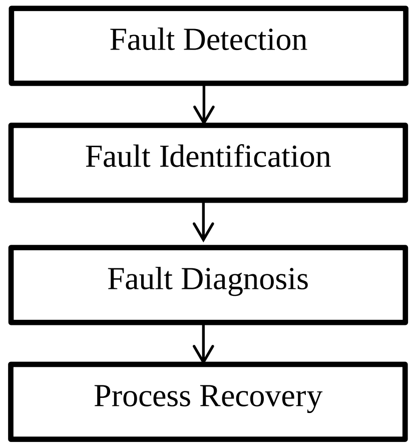


Figure 2.1 Main steps in MSPM system

Firstly, the fault detection is actually to indicate the departure of the observed sample of an acceptable range by using a set of parameters. Meanwhile for fault identification, it is to identify the observed process variables that are most relevant to the fault or malfunction which is usually identified by using the contribution of plot technique. Then, fault diagnosis is describes to determine the specific type of fault that significantly and also needs to be confirmed contributes to the signal. Finally, the process recovery is explains to remove the root of causes that contribute to the detected fault.

Based on the study by World, et al. (1987); Mardia, et al. (1989); Jackson (1991), recently, MSPC which applies not only product quality data (Y), but also all of the process variable data (X) can be obtained are based on multivariate statistical projection methods which is Principal Component Analysis (PCA). PCA is a statistical method for dimensionality reduction of the quality variable space (as cited in MacGregor and Kourti, 1995). This statement is quite similar to definition given by Neto, Jackson and Somers (2005), PCA which is one of the usual procedures used to give a condensed description and explain pattern of variation in multivariate data sets. According to Romagnoli and Palazoglu (2006), PCA is one of the multivariate statistical techniques which are basically classified as dimensionality reduction methods. The definitions of PCA from all researchers are quite similar to each other.

The first method in dimensionality reduction of PCA is collecting the normal operating data (NOC) which is X. Then, the data are then standardized to zero mean with respect to each of the variables, $\check{X}$. This is because PCA results depend on the data scales. Next, the calculation of a variance-covariance matrix, C_{mxm} by using this formula, $C = \frac{1}{n-1} X \check{X}$ is used to develop PCA model for the NOC data. From the calculation variance-covariance matrix, the eigen values, λ, and eigen vectors, V can be obtained. Finally, the Principal Component (PC) scores, P can be simply develop by using this formula, $P = \check{X}V$. Based on the study by MacGregor, et al., (1995), their covariance matrix almost singular when the number of the variables measured quality (q) which is large one often finds that they are highly correlated with one another. The first PC of y mean that linear combination $\lambda_1 = V_1^T y$ that has maximum variance subject to $|p_1| = 1$. The second PC which has the greatest variance subject to $|p_2| = 1$, that can be defined linear combination $\lambda_2 = V_2^T y$ and subject to the condition which means that it is not correlated with the first PC on in other word it is orthogonal. The PC loading vector V_i are the eigen vectors of the covariance matrix of Y and the subject of λ_i are the variances of the PC's. The PC scores are well defined as value of the PC that has been observed for each of the n observation vectors.

2.3 Process Monitoring Issues and Extension

There are various extensions have been proposed by other researchers. The process monitoring issues and extension can be divided into two categories which are process monitoring extension based on PCA and process monitoring extension based on multivariate technique which not based on PCA.

2.3.1 Process Monitoring Extension based on PCA

Furthermore, there are many extension proposed by other researchers based on PCA which are Non-Linear PCA, Kernel PCA, Multi-Way PCA, Dynamic PCA, Multi-Scale PCA and others. In this research, only three process monitoring extensions based on PCA will be described more details, which includes Non-Linear PCA, Multi-Scale PCA and Multi-Way PCA.

According study by Tan and Mavrovouniotis (1995), Non-Linear PCA is one of the process monitoring extensions based on linear technique of PCA. A data set X that consist m variables can be expressed in terms of non-linear PCA as follows;

$$X = F(T) + E \tag{2.1}$$

where T is the matrix of non-linear PC scores, F(.) the non-linear PC loading function and E the residual matrix. The concept of Input-Training neural network (the IT-net) is based on non-linear methods. Each input pattern was irregular but is adjusted with the internal network parameters to generate the same output pattern based on the steepest gradient descent network optimization rule (as cited in Jia, Martin, and Morris, 1998). The model of Non-Linear PCA is based on the Input-Training network has been developed. There are three basis steps to form the work. Firstly, the Linear PCA is used to perform the linear transformation in which the observation is rotated to a new set of uncorrelated ordinates permitting the main linear information to be extracted and condensed at the same time still maintaining sufficient data variance in the transformed data, so that the non-linear correlations is not excluded from the model. Next, the linear PC scores are rescaled to unit variance to enable the recovery of the non-linear structure in the new ordinates space of the transformed data. Finally, network optimization is improved through the use of Levenberg-Marquardt algorithm to interpret the non-linear structure in the transformed data. The Non-Linear PCA intended to address with the non-linearity among the variables and it can dominate more variance in a smaller dimension compared to the Linear PCA method (Wang and Romagnoli, 2005). However, the information of data tends to uniformly distributed among the PC, thus, disappearance the inherent orthogonality features of linear PCA (Maulud, Wang and Romagnoli, 2006).

Other extensions of PCA are Multi-Scale PCA (MSPCA) which is the nature of MSPCA makes it appropriate to work with the data is usually not fixed and represent the cumulative impact of many underlying process phenomena which each operating at different scale. MSPCA which retain the correlation and the maximum variance through the measurements from the benefit of PCA and the merit of the wavelets gives the correlation within the measurements. The signal trend and correlation are combined with MSPCA to extract maximum information from multivariate measurements. The methodology of decomposing each variable on the selected group of wavelets is determined. Then, the PCA model is identified independently for the coefficient signals at each scale. The yield of multi-scale model was developed from the model at significant scales are then combined in an efficient scale-recursive manner. The area of normal operation is determined at each scale from data representing normal operation for multivariate process monitoring by MSPCA (Wang, et al., 2005). MSPCA procedure is to determined different PCA models at each scale to identify the scales where important events occur. However, the weaknesses of this method is that, it requires basic understanding of the physical and chemical principles control the process operation to help in clustering the highly correlated variables together before constructing the PCA model (Bakshi, 1998).

Based on the study by Chen and Liu (2000), Multi-Way PCA which is also called MPCA is a technique that relate with all measured batch process which consist of noise and high correlation. MPCA subjects the information onto reduce dimensionality subspaces that include all the relevant information about the batches. The formation of the three-dimensional array as shown in Figure 2.1 is constructed based on the batch experimental data. A data matrix X with $(I \times J \times K)$ compress a series of runs of a usual batch process where j=1,2,…,J variables measured at times k=1,2,…,K intervals throughout one batch. Similar data will be run at several numbers of batches i=1,2,…,I. This technique carry out the data into time-ordered block and each of it represent one batch run. The blocks can be represented with multi-way matrices.

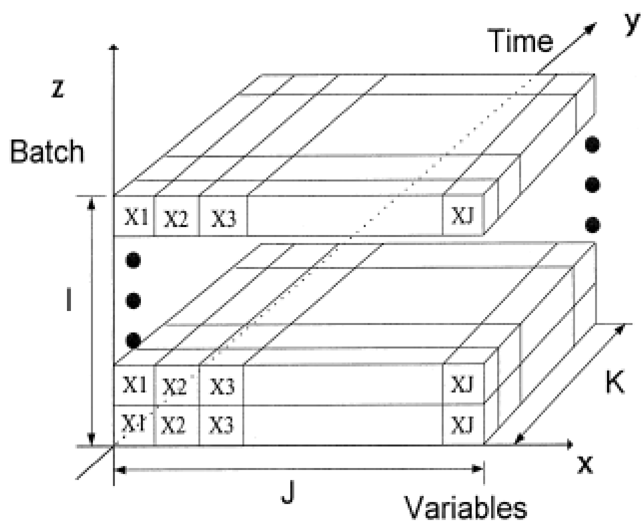


Figure 2.2 Three-dimensional data array of the batch experiments

(**Source**: Nomikos and MacGregor, 1994)

Three dimensional array data (I, batch samples x J, process variables x K, time) is decomposed to two dimensional array (I x JK) data for easier analysis (Nomikos and MacGregor, 1994).

2.3.2 Process Monitoring Extension based on Multivariate Technique

In this literature review will explain more detail only three process extension based on multivariate technique. There are Partial Least Square (PLS), Independent Component Analysis (ICA) and Canonical Variate Analysis (CVA). Actually, there are many others of extensions based on multivariate technique includes Parallel Factors Analysis (PARAFAC), Canonical Correlation Analysis (CA) and Factor Analysis (FA) which not discusses in this literature.

Gunther, Conner and Seborg (2009) states that “Partial Least Square (PLS) is a modeling technique that relates two data matrices (X and Y) by a multivariate linear model. PLS is used to predict quality variable measurements, Y from process variable data, X”. The PLS model is established from a set of calibration data which produced during normal operating data. If there are only a small number of batches, it can be difficult to select the NOC. If there is non-batches were selected, thus it can be cut down the entire monitoring effort. According to MacGregor, et al., (1995), the scores are linear combination of X-variables that maximizes the covariance between it and the Y space, whereby, the scores are orthogonal with the associated loading vectors. In term of concept, PLS is similar to PCA except that it simultaneously lowers the dimensions of the X and Y spaces.

Generally, Independent Component Analysis (ICA) is statistical technique for expose the secret factor that underlying a set of random variables, measurements or signals. The aims of ICA was proposed to settle the blind source separation problems which includes recovering independent source signals for example different voice or music after they have been combined by unknown matrix, A (Lee, Yoo and Lee, 2004). According to the study by Jutten and Herault, (1991); Girolami (1999), they state that “ICA is a signal processing technique for transforming observed multivariate data into statistically independent components, which are expressed as linear combinations of observed variables” (as cited in Kano, Hasebe, Hashimoto and Ohno, 2004). The definition of ICA by Kano, et al., (2004) is quite similar to the definition of Lee, et al., (2004).

Another extension of process monitoring based on multivariate technique is Canonical Variate Analysis (CVA). Based on the study by Simoglou, Martin and Morris (2002), the concept of PLS is quite similar to CVA which is in the method of linear combine calculation of past values of the system input or output that are most highly correlated with linear combine of the future of the outputs process. CVA give an advantage compared to other technique which is in terms of model stability and parsimony for example, CVA only required fewer identified parameter in the final models. CVA can provide more rapid detection when comparing CVA with PLS T^2 based on process monitoring schemes.

2.4 Dissimilarity in the MSPC Framework

According to the study by Cox (2001), "multidimensional scaling (MDS) is to find a configuration of points in a space which every point reflects one of the object and the distance d_{rs} between r and s 'matches', as well as possible, the original dissimilarity δ_{rs} for all pairs". The dissimilarities are expressed as Euclidean distance which is one of the techniques in classical scaling (Cox and Cox, 1994).

Based on the study Yunus and Zhang (2010), classical multidimensional scaling (CMDS) is another technique which used compressing multivariate data by using dissimilarity measures for process monitoring. This technique actually is same used in this research. In this work, the dissimilarity measure have been particularly constructed based on two different scales, city block and mahalanobis distances, which are shown respectively by Equation (2.2) and (2.3) (Cox et al., 1994). But the study by Yunus et al., (2010) used based on two different scales, Euclidean and city block distances.

City block distance: $\delta_{rs} = \sum_i |x_{ri} - x_{si}|$ (2.2)

Mahalanobis distance: $\delta_{rs} = \left\{(x_r - x_s)^T \sum^{-1} (x_r - x_s)\right\}^{1/2}$ (2.3)

Furthermore, this research were model by PCA in other word this research actually integrating dissimilarity with PCA while other researcher stated earlier were model by CMDS.

The algorithm for finding the dissimilarity can be summarized as (Borg and Groenen, 2005):

$$\mathbf{A} = [\delta_{rs}^2] \tag{2.4}$$

$$\mathbf{B} = -\frac{1}{2}\,\mathbf{J}\,\mathbf{A}\,\mathbf{J} \tag{2.5}$$

$$\mathbf{B} = \mathbf{V\Lambda V^T} \tag{2.6}$$

Matrix **A** contains the squared dissimilarities. Then **A** is doubly centred using the centring matrix $\mathbf{J} = \mathbf{I} - \frac{\mathbf{11'}}{\boldsymbol{n}}$ and multiplied by -1/2 to form matrix **B**. Then **B** is expressed in terms of its spectral decomposition, $\mathbf{V\Lambda V^T}$, where $\mathbf{\Lambda}$ is the diagonal matrix of ordered eigenvalues of **B**, **V** the matrix of corresponding eigenvectors.

Moreover, a search was also carry out for investigating the correlation between PCA and dissimilarity. This relationship is viewed from the close fundamental algorithms between conventional PCA and dismilarity procedures. Cox et al. (1994) had described the relationship between minor product moment and dissimilarity matrix by using algorithm manipulations approach. They started the procedure by defining the scalar product matrix, **B**, $\mathbf{B} = \mathbf{X}\mathbf{X}^T$, in which **X** is a data matrix, n x p, and has to be mean corrected. By applying the Singular Decomposition (SD) operation on **B**, the following are obtained:

$$\mathbf{B}\mathbf{u}_i = \lambda_i \mathbf{u}_i \tag{2.7}$$

$$\mathbf{X}\mathbf{X}^T\mathbf{u}_i = \lambda_i \mathbf{u}_i \tag{2.8}$$

Multiplying both side with $\mathbf{X}^T$

$$\mathbf{X}^T[\mathbf{X}\mathbf{X}^T\mathbf{u}_i] = \mathbf{X}^T[\lambda_i \mathbf{u}_i] \tag{2.9}$$

By which,

$\mathbf{C} = \mathbf{X}^T\mathbf{X}$; **C** represent the minor product moment

$\mathbf{q}_i = \mathbf{X}^T\mathbf{u}_i$; $\mathbf{q}_i$ represent loading vector of PCA

So,

$$\mathbf{C}\mathbf{q}_i = \lambda_i \mathbf{q}_i \tag{2.10}$$

By embedding the algorithm of the conventional PCA through dissimilarity, it may provide variety of results in terms of configuration plots for process monitoring. This is because the result can figure out both linear and non-linear relationships measured variables.

2.5 Summary

As a conclusion, there are four main steps in MSPM in the field of the process monitoring performance and fault diagnosis which are fault detection, fault identification, fault diagnosis and process recovery. This research focuses more to the fault detection. The conventional PCA is the one of the basic technique in MSPM. The definition of PCA is a statistical method for dimensionality reduction of the quality variable space. Besides that, there two types of process monitoring issues and extension which are process monitoring extension based on PCA and process monitoring extension based on multivariate technique. Extension based on PCA includes Non-Linear PCA, Multi-Scale PCA and Multi-Way PCA, while, extension based on multivariate technique are Partial Least Square (PLS), Independent Component Analysis (ICA) and Canonical Variate Analysis (CVA). It may provide variety of results in terms of configuration plots for process monitoring by embedding the algorithm of the conventional PCA through dissimilarity.

CHAPTER III

METHODOLOGY

3.1 Introduction

This chapter discusses a methodology on dissimilarity-based MSPM. Generally, there are varieties of techniques in multidimensional scaling (MDS). The technique includes classical scaling, non-metric scaling, procrustes analysis, biplot and general dissimilarity. This chapter can be divided into three sections which are introduction, methodology of this research and summary.

3.2 Methodology on Dissimilarity-based MSPM

In this research, the main focuses of the methodology is fault detection in MSPM system. According to the study of Yunus and Zhang (2010), the complete procedures of fault detection consists of two main phases namely as off-line modelling and monitoring (Phase I) and on-line monitoring (Phase II):

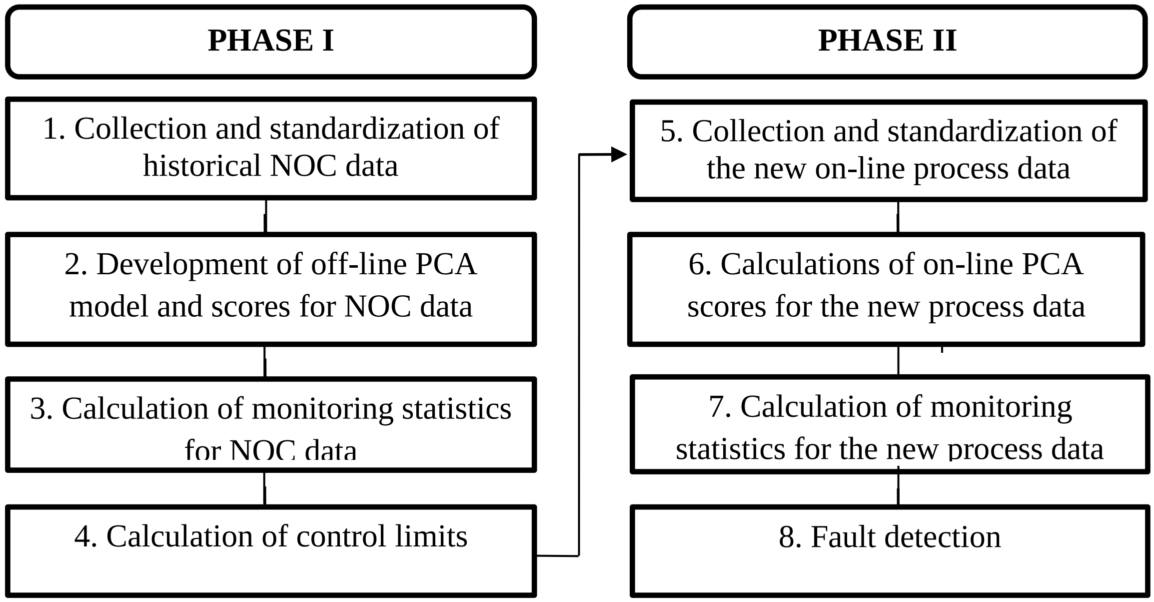


Figure 3.1 Procedures of fault detection

Before proceeding to the explanation about two main phases which are off-line modelling and monitoring (Phase I) and on-line monitoring (Phase II), actually this research deeply focuses on converting the dissimilarity matrix to minor product moment before proceeding to using conventional PCA processes. Clearly, this dissimilarity matrix technique is between step 1 and step 2 in the Phase I which is for off-line modeling monitoring based on the figure above. Similarly, the dissimilarity matrix technique for Phase II which is on-line monitoring is between step 5 and step 6. This is done based on the method proposed by Cox and Cox (1994). The dissimilarity matrix technique can illustrate as in the Figure 3.2.

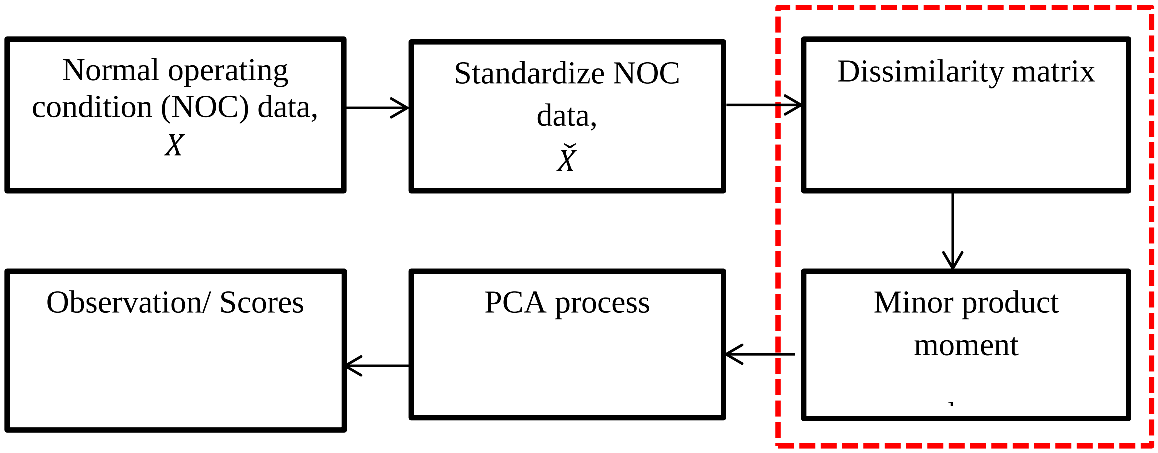


Figure 3.2 Main focuses for integration of dissimilarity matrix and PCA

Based on the study by Yunus and Zhang (2010), the conceptual framework of fault detection in MSPM system consists of two main phases which are off-line modelling and monitoring (Phase I) and on-line monitoring (Phase II). Basically, Phase I is for model development which is to gain understanding of the process and to form a statistical benchmark for the future process outcomes by using NOC data to develop the model. Meanwhile, Phase II is for fault detection operation. It describes to observe the process in actual time by comparing the new process data with the pre-specified model that is formed during the first phase. From the comparing data, the result of the process may be normal or abnormal. If the process is normal it means no changes in the process whereas if the process is abnormal, it means there are fundamental changes in the process which requires intervention.

Firstly, a set of normal operation condition (NOC) data, $X_{n\times m}$(n: samples, m: variables), are identified off-line based on the historical process data archive.

$$\mathbf{X} = \begin{bmatrix} X_{1,1} & X_{1,2} \cdots & X_{1,m} \\ X_{2,1} & X_{2,2} \dots & X_{2,m} \\ \vdots & \vdots \;\; \vdots & \vdots \\ X_{n,1} & X_{n,2} \dots & X_{n,m} \end{bmatrix} \tag{3.1}$$

The NOC data was a process operating at the desired target and produce a satisfactory product that meets the qualitative and quantitative standard stated (Martin et al., 1996). Next, 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.

$$\breve{X}_{\mathrm{j,i}} = \frac{(\mathrm{x_{j,i}} - \bar{\mathrm{x}}_{\mathrm{i}})}{\sigma_{\mathrm{i}}} \tag{3.2}$$

Now, there are three steps are added to the PCA algorithms. The starting point of an MDS analysis is to find the set of dissimilarity $\{\delta_{rs}\}$ between pairs of objects. There is variety of dissimilarity measures available for quantitative data but in this thesis only two dissimilarity measures are used which are city block and mahalanobis distance as shown by Equation (2.2) and (2.3)(Cox et al., 1994).

Then, matrix **A** contains the squared dissimilarities. Then **A** is doubly centred using the centring matrix $\mathbf{J} = \mathbf{I} - \frac{\mathbf{11'}}{\boldsymbol{n}}$ and multiplied by -1/2 to form matrix **B**. Then **B** is expressed in terms of its spectral decomposition, $\mathbf{V\Lambda V^T}$, where $\mathbf{\Lambda}$ is the diagonal matrix of ordered eigenvalues of **B**, **V** the matrix of corresponding eigenvectors. The algorithm for finding the dissimilarity can be summarized as shown from Equation (2.4) until

Equation (2.6) (Borg and Groenen, 2005). After that, the next step applies for the conversion of dissimilarity matrix to minor product moment which shown from Equation (2.7) until Equation (2.10).

Here, the step is continuing with the PCA algorithm. Finally, the PCA model of can be simply developed by:

$$\mathbf{P} = \check{\mathbf{X}}\,\mathbf{V} \tag{3.5}$$

Where,

$$\mathbf{P} = [\mathbf{p_1} \quad \cdots \quad \mathbf{p_m}]$$

$$= \begin{bmatrix} \check{x}_{1,1} \quad v_{1,1} + \cdots + \tilde{x}_{1,m} \quad v_{m,1} & \cdots & \check{x}_{1,1} \quad v_{1,m} + \cdots + \check{x}_{1,m} \quad v_{m,m} \\ \vdots & \cdots & \vdots \\ \check{x}_{n,1} \quad v_{1,1} + \cdots + \check{x}_{n,m} \quad v_{m,1} & \cdots & \check{x}_{n,1} \quad v_{1,m} + \cdots + \check{x}_{n,m} \quad v_{m,m} \end{bmatrix}$$

The following equation presents a measure of data variations captured by the first principal components (Jolliffe, 2002).

$$k = \frac{\lambda_1 + \lambda_2 + \cdots + \lambda_a}{\lambda_1 + \lambda_2 + \cdots + \lambda_a + \cdots + \lambda_m} \tag{3.6}$$

The third step basically involves calculation of the Hotelling's T^2 and Squared Prediction Errors (SPE) monitoring statistics. Finally, step four in the Phase I deal with developing the control limits for both of the statistics. The Hotelling's T^2 statistics come along with time can be used to establish the Hotelling's T^2 control chart. This chart presents the information in terms of how well the current process data fit in the range of the normal process operation data. Furthermore, SPE statistic is a measure to distinguish the process variations that are not captured by the process models, which forms the 'Residual Subspace' of the conventional PCA. Both control charts have 95% confidence limit to serve as the warning alarm while 99% confidence limit provides the action or control limit signal. The Hotelling's T^2 statistic, SPE statistic and their confidence limits are determined from the following formulas:

$$\text{Hotelling's } T^2 \text{ statistic, } T_i^{\,2} = \sum_{j=1}^{A} \frac{p_{i,j}^2}{\lambda_j^2} \tag{3.7}$$

$$\text{Control limits } = \frac{A(n-1)}{(n-A)} F_{A,n-A,\alpha} \tag{3.8}$$

α typically takes the value of 0.05 or 0.01 for the warning and action limits respectively. An out-of-control signal is identified if

$$T_i^2 > \frac{A(n-1)}{(n-A)} F_{A,n-A,0.01} \tag{3.9}$$

Where,

A = number of PCs retained in the PCA model

n = number of nominal process measurements per variable

$p_{i,j}$ = i^{th} score for Principal Component j

λ_j = *eigenvalue* corresponds to Principal Component j

$$\text{Residual Matrix, } \mathbf{E} = \mathbf{X_Z}(\mathbf{I} - \mathbf{V}_A \mathbf{V}^T{}_A) \tag{3.10}$$

$$\text{SPE statistic,} \quad Q_i = \mathbf{e}_\mathrm{i}\mathbf{e}_\mathrm{i}^\mathrm{T} \tag{3.11}$$

$$\text{Confidence limit, } Q_\alpha = \theta_1\left(\frac{z_\alpha\sqrt{2\theta_2 h_0^2}}{\theta_1} + \frac{\theta_1 h_0(h_0-1)}{\theta_1^2} + 1\right)^{\frac{1}{h_0}} \tag{3.12}$$

$$\theta_1 = \sum_{i=A+1}^{N} \lambda_i \tag{3.13}$$

$$\theta_2 = \sum_{i=A+1}^{N} \lambda_i^2 \tag{3.14}$$

$$\theta_3 = \sum_{i=A+1}^{N} \lambda_i^3 \tag{3.15}$$

$$h_o = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2} \tag{3.16}$$

Where,

z_α= standard normal deviate corresponding to the upper (1-α) percentile

$\mathbf{X_Z}$ = standardized matrix of original matrix, **X**

E = residual matrix (*n* x *m*)

I = identity matrix

$\mathbf{V}_{\boldsymbol{A}}$ = *eigenvector* matrix contains up to *A* eigenvectors

e_i = i^{th} row in residual matrix

Phase II: On-line Monitoring

On the other hand, the fifth to seventh steps follow procedures of the first to the third step in the Phase I. With regards to the last of eight steps describes earlier, there is one main operations which are fault detection. The fault detection can be traced by comparing the new process data with the developed model in the first phase. All the steps stated above are run by using Matlab software platform version 7 as a tool to achieve the main goal.

3.3 Summary

As a conclusion, it is hope that by using dissimilarity matrix techniques a model of non-linear process which is highly used by chemical industry can be developed. Simultaneously, a new technique which can improve the process of monitoring performance by using MSPM procedures can be developed.

CHAPTER IV

RESULTS AND DISCUSSION

4.1 Introduction

The results and discussion of research have been presented in this chapter and stressed on the integration of process monitoring algorithms based on the conventional PCA-based MSPM and dissimilarity-based MSPM. Firstly, this chapter describes about the case study use in this research. Next, result of the first phase which is discussed about normal operating condition both on the conventional PCA-based MSPM and dissimilarity-based MSPM. Then, second phase of result which is described about the fault by using both algorithms. Finally, the summary is briefly explained.

4.2 Case Study

A simulated continuous stirred tank reactor with recycle (CSTRwR) shown in Figure 4.1 was used as the case study (Zhang, 2006).

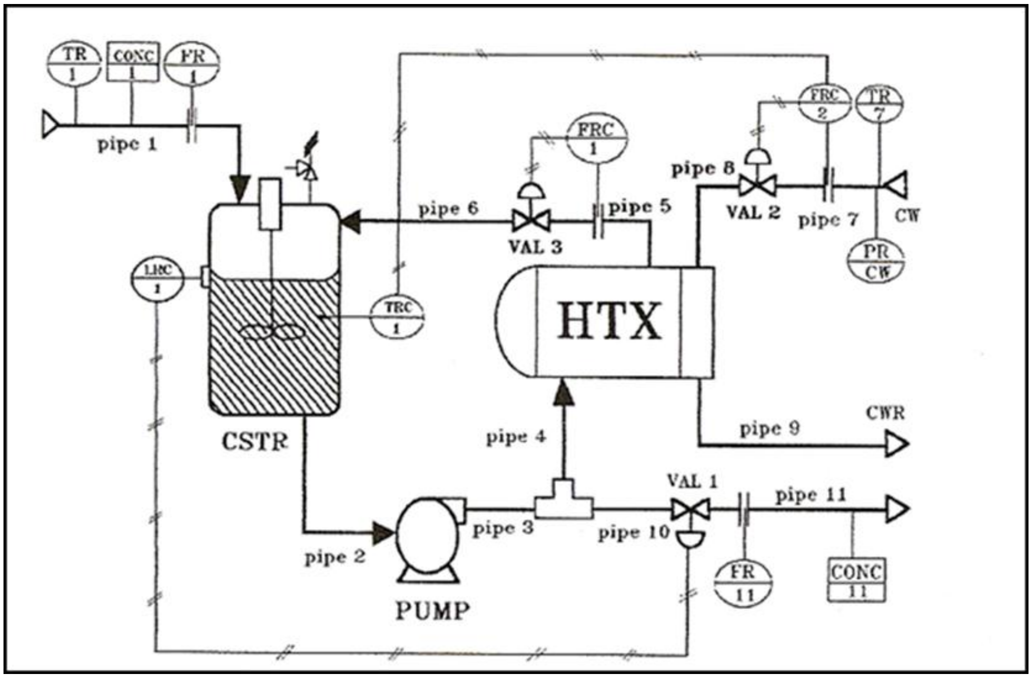


Figure 4.1 CSTRwR system

This system conducts an irreversible heterogeneous catalytic exothermic reaction from reactant A to form product B in the reactor vessel. The process is installed with three separate control loops, which consists of tank temperature, tank level and recycling flow variables, in order to maintain the product concentration at a desired level. In particular, the cold water flow is adjusted through a cascade system corresponding to the changes in the reactor temperature. The reactor level, on the other hand, is maintained by controlling the flow rate of the product. Lastly, the product composition in the reactor is indirectly controlled by manipulating 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 for monitoring

No.	Variables	Variable Names
1.	V1	Tank temperature
2.	V2	Tank level
3.	V3	Flow rate feed
4.	V4	Flow rate inlet
5.	V5	Flow rate cooling
6.	V6	Flow rate outlet
7.	V7	Flow rate recycle
8.	V8	Product concentration
9.	V9	Feed concentration
10.	V10	Tank pressure
11.	V11	Controller 1
12.	V12	Controller 2
13.	V13	Controller 3

4.3 Overall Monitoring Performance

4.3.1 .First Phase *(Off-line Modelling and Monitoring)*

A set of NOC data containing 100 samples was obtained from simulation. In order to evaluate the robustness of the monitoring limits, another set of NOC data which is the second set of NOC data containing 50 samples were also collected. The second set of NOC data is also called as NOC test data.

Firstly, the standardized NOC data is analysed through the conventional PCA-based MSPM algorithm. The analysis is to identify the number of PCs that required in the process which is to reduce the dimensions of multivariate data as shown in Figure 4.2.

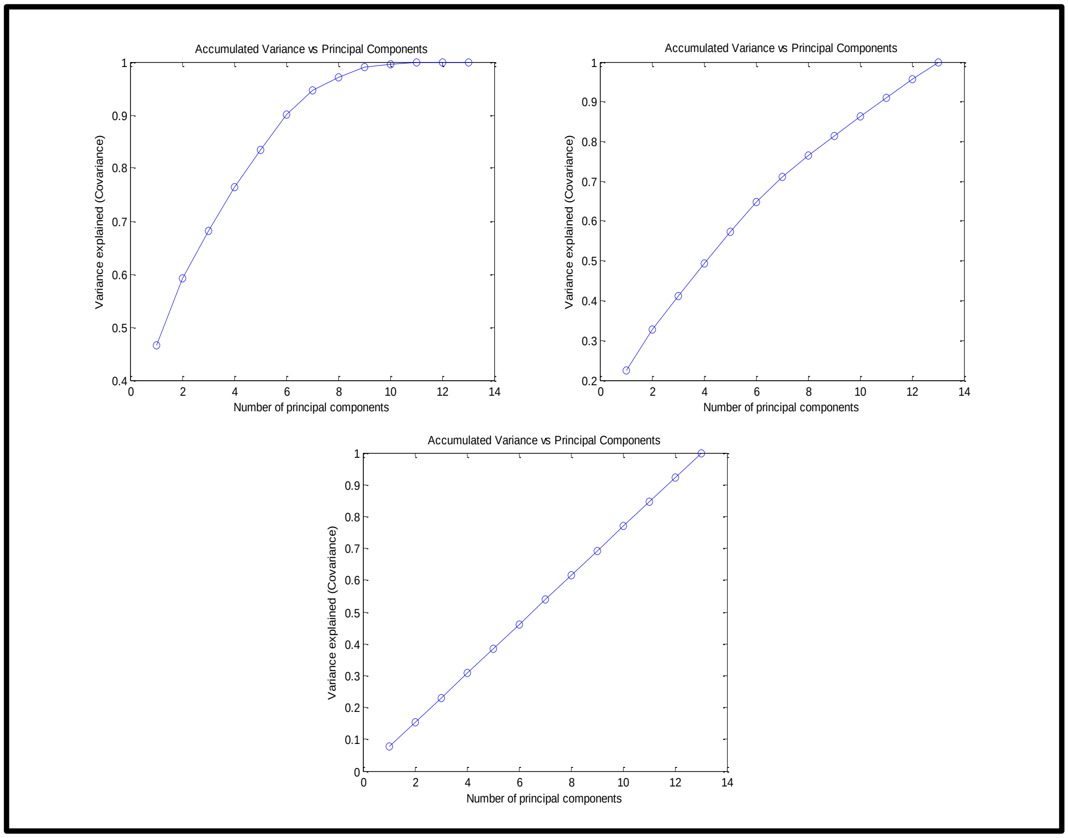


Figure 4.2 Accumulated data variance explained by different PCs for conventional PCA-based MSPM (left), dissimilarity-based MSPM of city block distance (right) and dissimilarity-based MSPM of mahalanobis distance (bottom)

From Figure 4.2 (left), it shows that to explain over 90% of the total NOC data variances at least six PCs are required, meanwhile, to represent over 70% of the total variances only three PCs are needed. Therefore, both PCs are using in the PCA model for the calculation of NOC scores for this case study in this research.

Secondly, the standardized NOC data is analysed through the new algorithm which are dissimilarity-based MSPM. Generally, dissimilarity-based MSPM algorithms are based on the city block distance and mahalanobis distance. The different between them are method to find the set of dissimilarities between pairs of objects. Figure 4.2 (right) and 4.2 (bottom) show the analysis which is to identify the number of PCs for dissimilarity-based MSPM based on city block distance and mahalanobis distance respectively. Based on the findings, the new system required more PCs to represent 70% and 90%. However, in this thesis used three and six PCs for the new algorithm to be analysed.

Based on Figure 4.2 (right), it shows that only three PCs are required to explain over 45% of the total NOC data variances. However, 70% of the total variances is required only six PCs to represent it compared to conventional PCA-based MSPM, 90% of the total NOC data variances required over six PCs to explained. From this, it can be seen new algorithm based on city block can identify the same number of PCs with conventional PCA-based MSPM but the percentage of the total variances transformed is less. Therefore, it can be conclude that new algorithm based on city block is less efficient to identify number of PCs.

From Figure 4.2 (bottom), it illustrates that to explain over 50% of the total NOC data variances are required at least six PCs. Besides that, only three PCs are needed to represent over 30% of the total variances. Based on this finding it can be summarized that dissimilarity-based MSPM algorithms based on mahalanobis is less effective to find the number of PCs that required in the process due to the less percentage of the total variances transformed compared to algorithms stated earlier. Then, the steps like conventional PCA-based MSPM will be performed for this case study.

4.3.1.1. Monitoring Outcomes Based on Three PCs

The calculation of the confidence region of the scores basically involves of the Hotelling's T^2 and Squared Prediction Errors (SPE) monitoring statistics. Both control charts have 95% confidence limit to show as the warning alarm while 99% confidence limit provides the action or control limit signal. In Figure 4.3 shows both of charts by using process monitoring algorithms based on the conventional PCA.

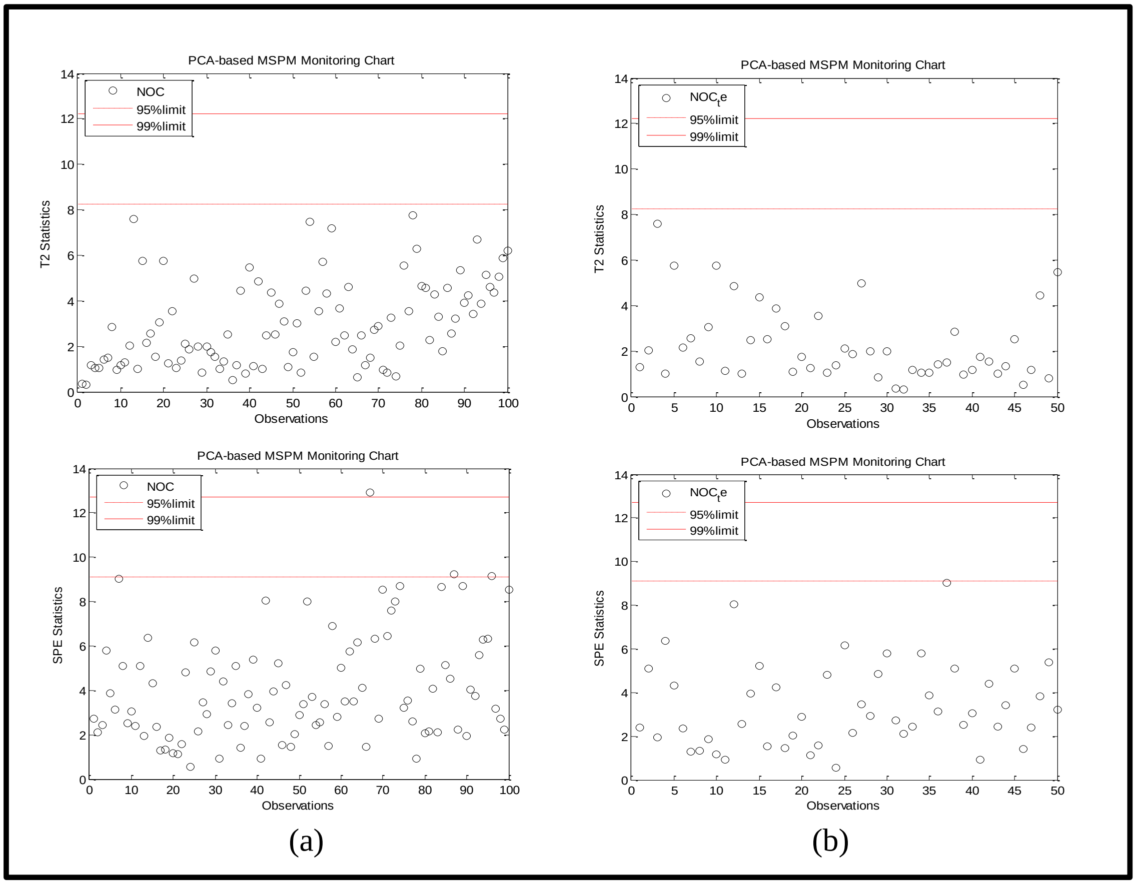


Figure 4.3 Hotelling's T^2 and Squared Prediction Errors (SPE) monitoring statistics chart plotted together with the 95% and 99% confidence limits of conventional PCA: (a) NOC data (b) NOC test data

Based on Figure 4.3(a), there is no observation that is out of the control limit based on the Hotelling's T^2 monitoring statistics chart, meanwhile for the SPE monitoring statistics chart, there are only one observation out of the 99% confidence limit. The observations are still within the control limit. For Figure 4.3(b), both Hotelling's T^2 and SPE monitoring statistics chart based on the NOC test data, there is no observation that is out of control limit, thus it can be concluded that the process is in a normal condition.

Figure 4.4 illustrates Hotelling's T^2 and Squared Prediction Errors (SPE) Monitoring Statistics charts by using dissimilarity-based MSPM process monitoring algorithms based on the city block distance.

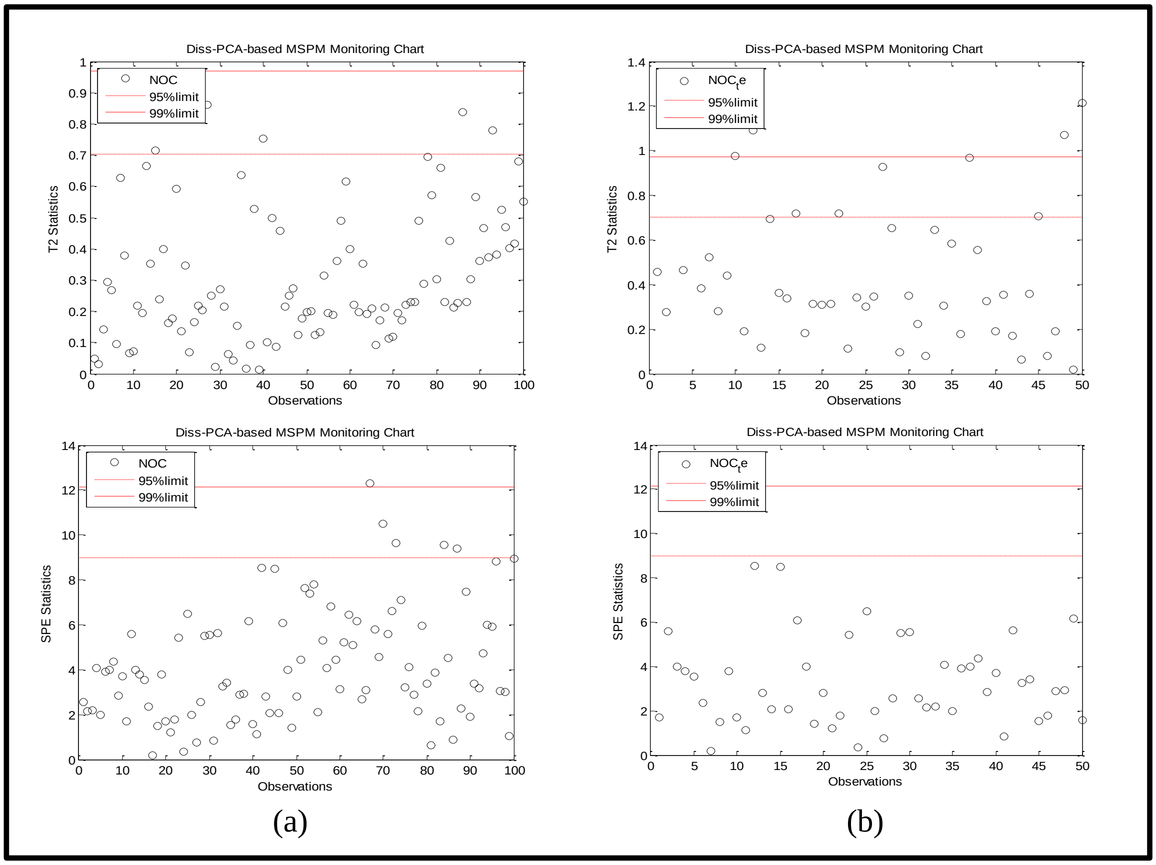


Figure 4.4 Hotelling's T^2 and Squared Prediction Errors (SPE) monitoring statistics chart plotted together with the 95% and 99% confidence limits of dissimilarity based on city block distance: (a) NOC data (b) NOC test data

It can be seen from Figure 4.4(a) based on the Hotelling's T^2 monitoring statistics chart, all the observation is in the control limit but for SPE monitoring statistics chart, only one observation is out of control limit signal. Another figure which is Figure 4.4(b) can be described that there is no observation is out of control limit in SPE monitoring statistics chart based on the NOC test data. Otherwise, there are about five observations that out of the 99% confidence limit in Hotelling's T^2 monitoring statistics chart. However, both figures of the process can be summarized in the typical condition. Figure 4.5 illustrates Hotelling's T^2 and Squared Prediction Errors (SPE) Monitoring Statistics charts by using dissimilarity-based MSPM process monitoring algorithms based on mahalanobis distance.

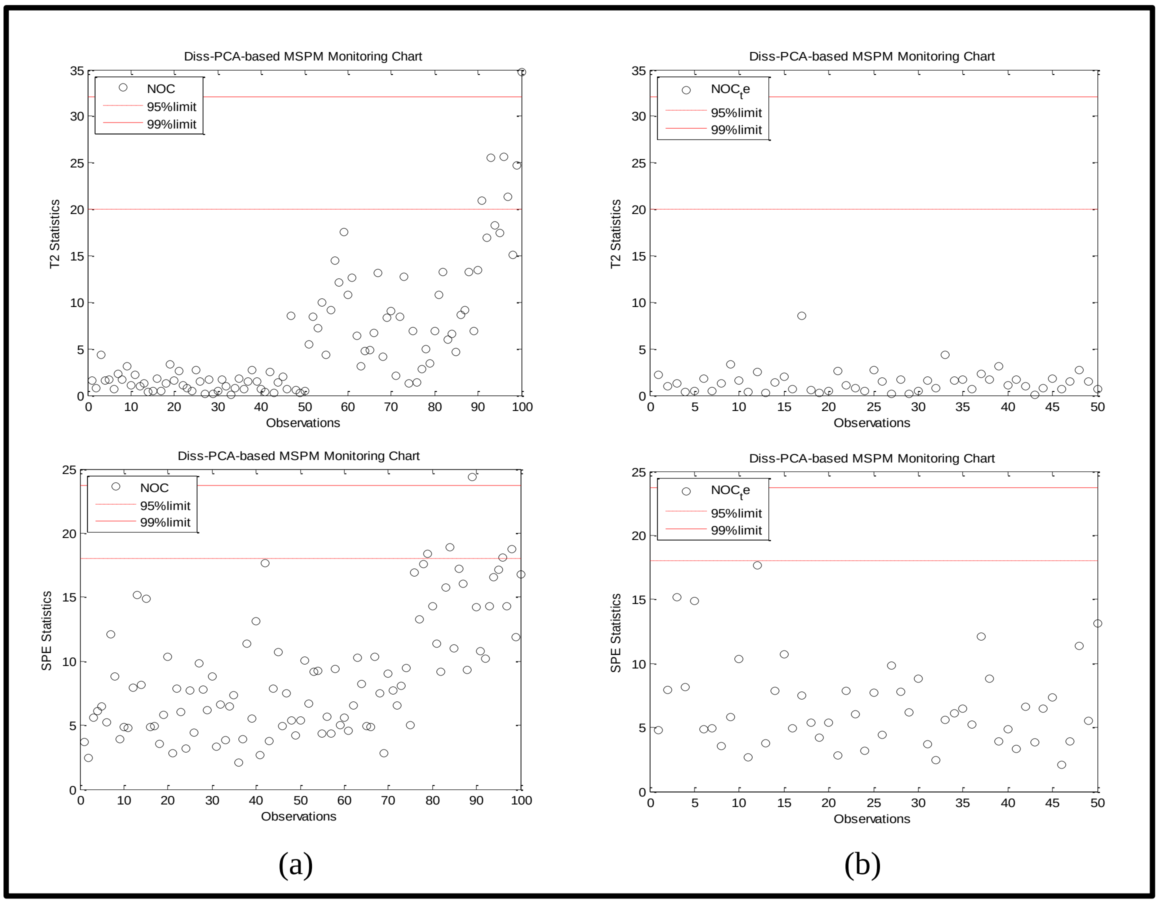


Figure 4.5 Hotelling's T^2 and Squared Prediction Errors (SPE) monitoring statistics chart plotted together with the 95% and 99% confidence limits of dissimilarity based on mahalanobis distance: (a) NOC data (b) NOC test data

Besides that, from the Figure 4.5 above, all the observation in both Hotelling's T^2 and SPE monitoring statistics charts are still within the control limit accept for Hotelling's T^2 and SPE monitoring statistics charts based on the NOC data, there is one out of the 99% confidence limit. Nevertheless, the process is still in the ordinary condition.

4.3.1.2. Monitoring Outcomes Based on Six PCs

The calculation of the confidence region of the scores basically involves of the Hotelling's T^2 and Squared Prediction Errors (SPE) monitoring statistics. Both control charts have 95% confidence limit to show as the warning alarm while 99% confidence limit provides the action or control limit signal as shown in Figure 4.6 based on the conventional PCA.

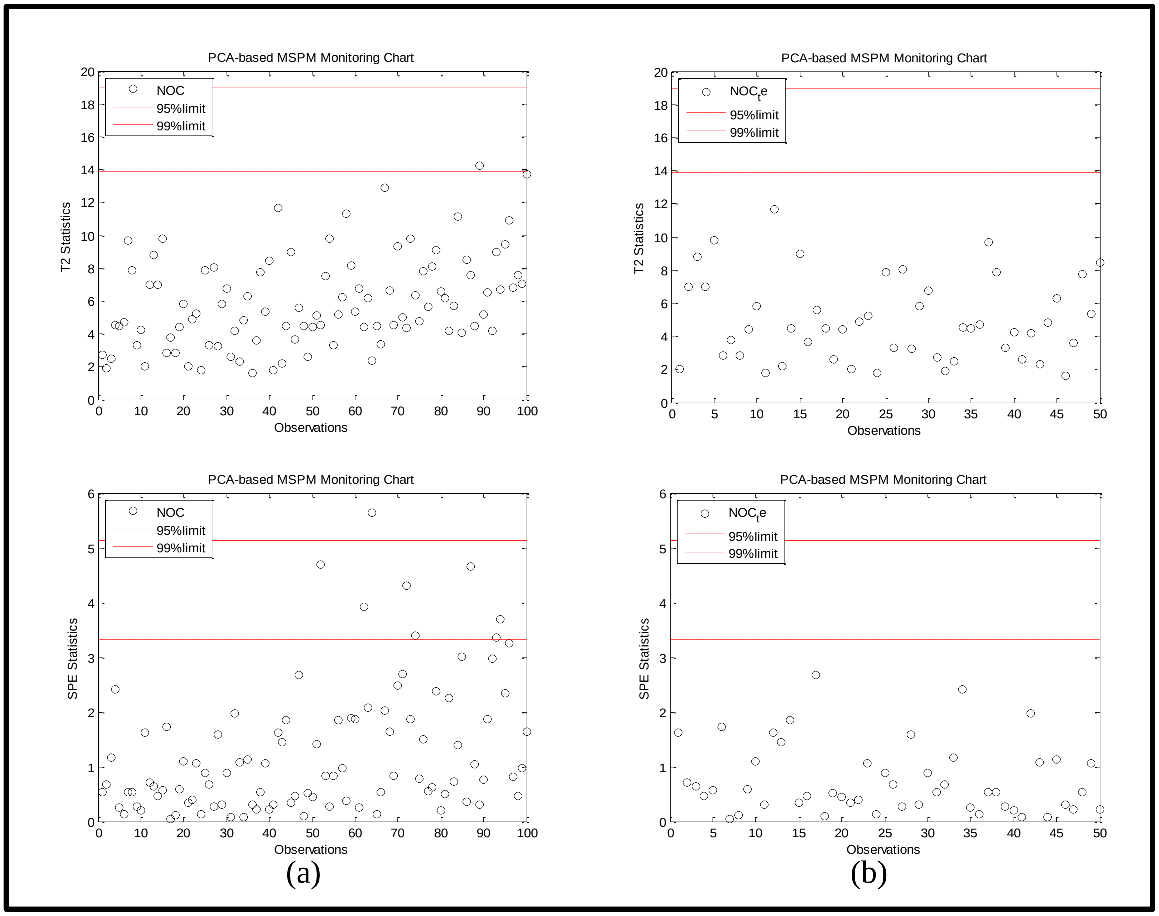


Figure 4.6 Hotelling's T^2 and Squared Prediction Errors (SPE) monitoring statistics chart plotted together with the 95% and 99% confidence limits of conventional PCA: (a) NOC data (b) NOC test data

Based on Figure 4.6(a), there is no observation that is out of the control limit based on the Hotelling's T^2 monitoring statistics chart, meanwhile for the SPE monitoring statistics chart, there are only one observation out of the 99% confidence limit. If compare figure 4.6(a) with Figure 4.3(a) of SPE monitoring static chart, there is different between them which are the some of the observation in figure 4.6(a) is out of 95% confidence limit but in figure 4.3(a) still within the control limit. Although this happen, the observations of figure 4.6(a) are still within the control limit because not more than one observation is out 99% confidence control limit. For Figure 4.6(b), both

Hotelling's T^2 and SPE monitoring statistics chart based on the NOC test data, there is no observation that is out of control limit, thus it can be concluded that the process is in a commonplace condition.

Figure 4.7 illustrates Hotelling's T^2 and Squared Prediction Errors (SPE) Monitoring Statistics charts by using dissimilarity-based MSPM process monitoring algorithms based on the city block distance.

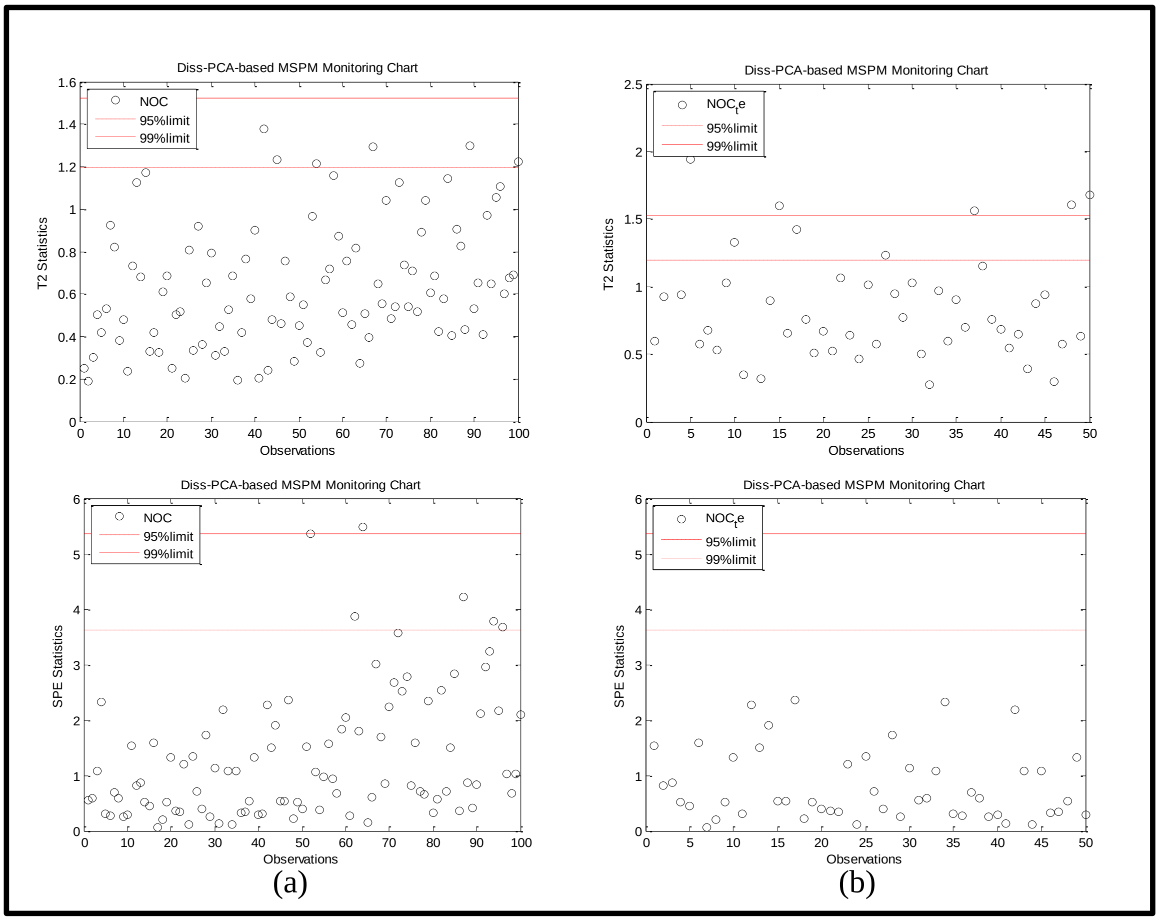


Figure 4.7 Hotelling's T^2 and Squared Prediction Errors (SPE) monitoring statistics chart plotted together with the 95% and 99% confidence limits of dissimilarity based on city block distance: (a) NOC data (b) NOC test data

From Figure 4.7(a) based on the Hotelling's T^2 monitoring statistics chart, all the observation is in the control limit but for SPE monitoring statistics chart, only one observation is out of control limit signal. Besides that, based on Figure 4.7(b) can be clarified that there is no observation is out of control limit in SPE monitoring statistics chart based on the NOC test data. For Hotelling's T^2 monitoring statistics chart, there are about seven observations that out of the 99% confidence limit. There are some different that can be seen when compared figure 4.7(b) and 4.4(b) of Hotelling's T^2 monitoring statistics chart which the number observations that out of control based on six PCs is more than three PCs. However, both figures of the process can be summarized in the normal condition. This is because there is no trend of the observations that out of control limit tend to be abnormal condition. Actually, abnormal condition can be happen when three of observations are out of control limit repeatedly.

Figure 4.8 illustrates Hotelling's T^2 and Squared Prediction Errors (SPE) Monitoring Statistics charts by using dissimilarity-based MSPM process monitoring algorithms based on mahalanobis distance.

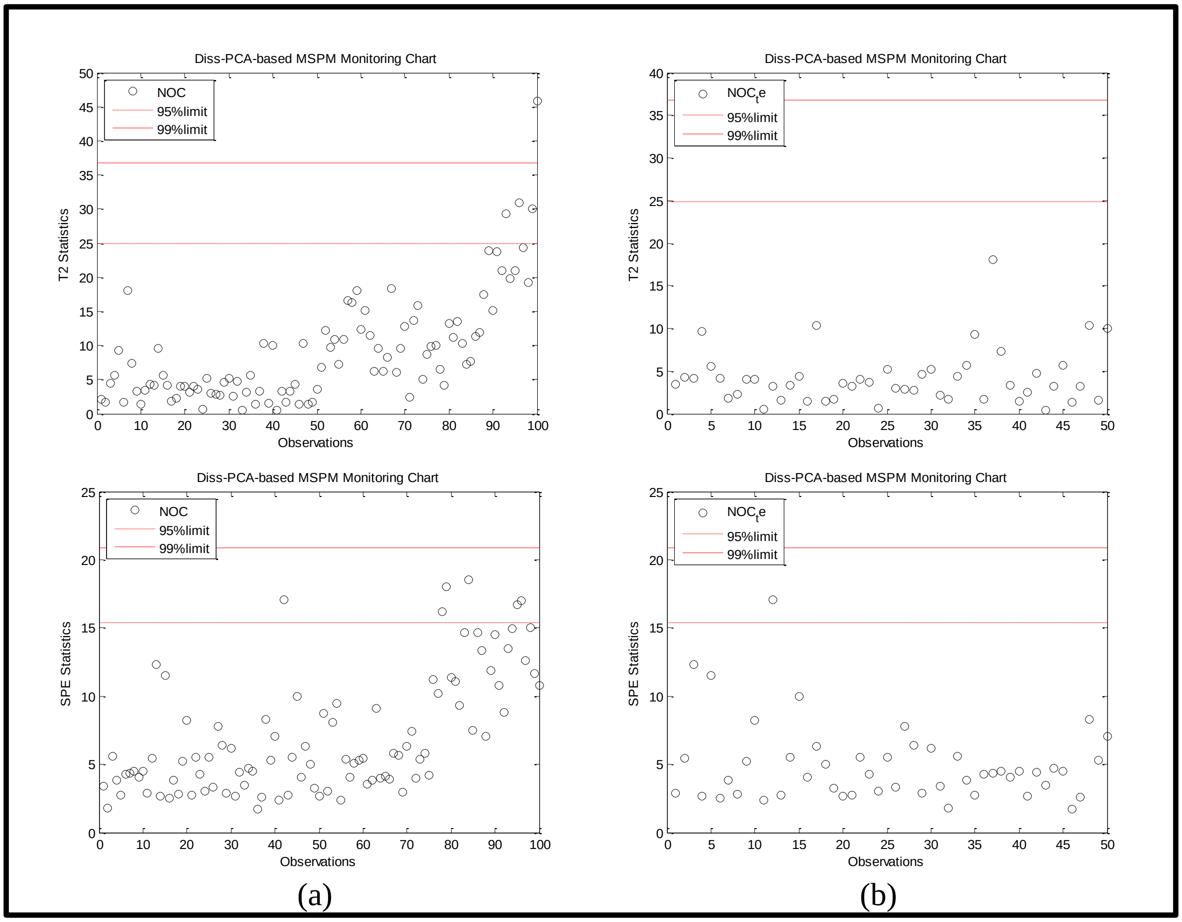


Figure 4.8 Hotelling's T^2 and Squared Prediction Errors (SPE) monitoring statistics chart plotted together with the 95% and 99% confidence limits of dissimilarity based on mahalanobis distance: (a) NOC data (b) NOC test data

Moreover, all the observation in both Hotelling's T^2 and SPE monitoring statistics charts are still within the control limit accept for Hotelling's T^2 monitoring statistics charts based on the NOC data, there is one out of the 99% confidence limit as can see from Figure 4.8. Although that happen, the process is still in the usual condition.

4.3.2 Second Phase *(On-line Monitoring)*

The system of the case study also subjects to be affected from several malfunction conditions which are pipe 1 blockage and external feed-reactant flow rate is too high. For each fault stated earlier, both abrupt and incipient faults are considered. An abrupt fault expresses 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 reflects a kind of fault that gradually deviates from the normal target. Thus, the monitoring system typically takes a while to detect these particular abnormal behaviours. In particular, all the faults were introduced at sample 2 and the sampling time was fixed at 4 seconds.

A set of abnormal process contain 20 samples were then applied to the conventional PCA and dissimilarity algorithm. This set of abnormal data reflects to the faults which are pipe 1 blockage and external feed-reactant flow rate is too high. The fault happen in two conditions which are abrupt and incipient that has been explained before.

4.3.2.1. Monitoring Outcomes Based on Three PCs

The Table 4.2 shows the tabulated results obtained from analysing the monitoring statistics for overall runs by using two different methods which are the conventional PCA-based MSPM and dissimilarity-based MSPM algorithm. The results also based on two faults stated earlier which involved two types of faults which are abrupt and incipient faults respectively.

Based on the tabulated results, it clearly seen that both conventional PCA-based MSPM and dissimilarity-based MSPM are able to detect all the fault directly at the same fault detection time accept for dissimilarity-based MSPM based on mahalanobis distance is detect the fault slightly slower compared to the conventional method. Now, fault one (F1) which is pipe 1 blockage include abrupt and incipient faults are discussed more deeply. The result of T^2 and SPE statistics for fault two (F2) can be viewed in appendix A. Figure 4.9 shows the T^2 and SPE statistics for the abrupt fault.

Table 4.2 Fault detection time for abrupt and incipient faults based on three PCs

Fault	Conventional PCA			Dissimilarity Matrix					
				City Block			Mahalanobis		
	T^2	SPE	Final	T^2	SPE	Final	T^2	SPE	Final
F1a	1	1	1	1	1	1	1	1	1
F1i	4	2	2	2	2	2	6	2	2
F2a	1	1	1	1	1	1	1	1	1
F2i	6	3	3	3	4	3	14	4	4

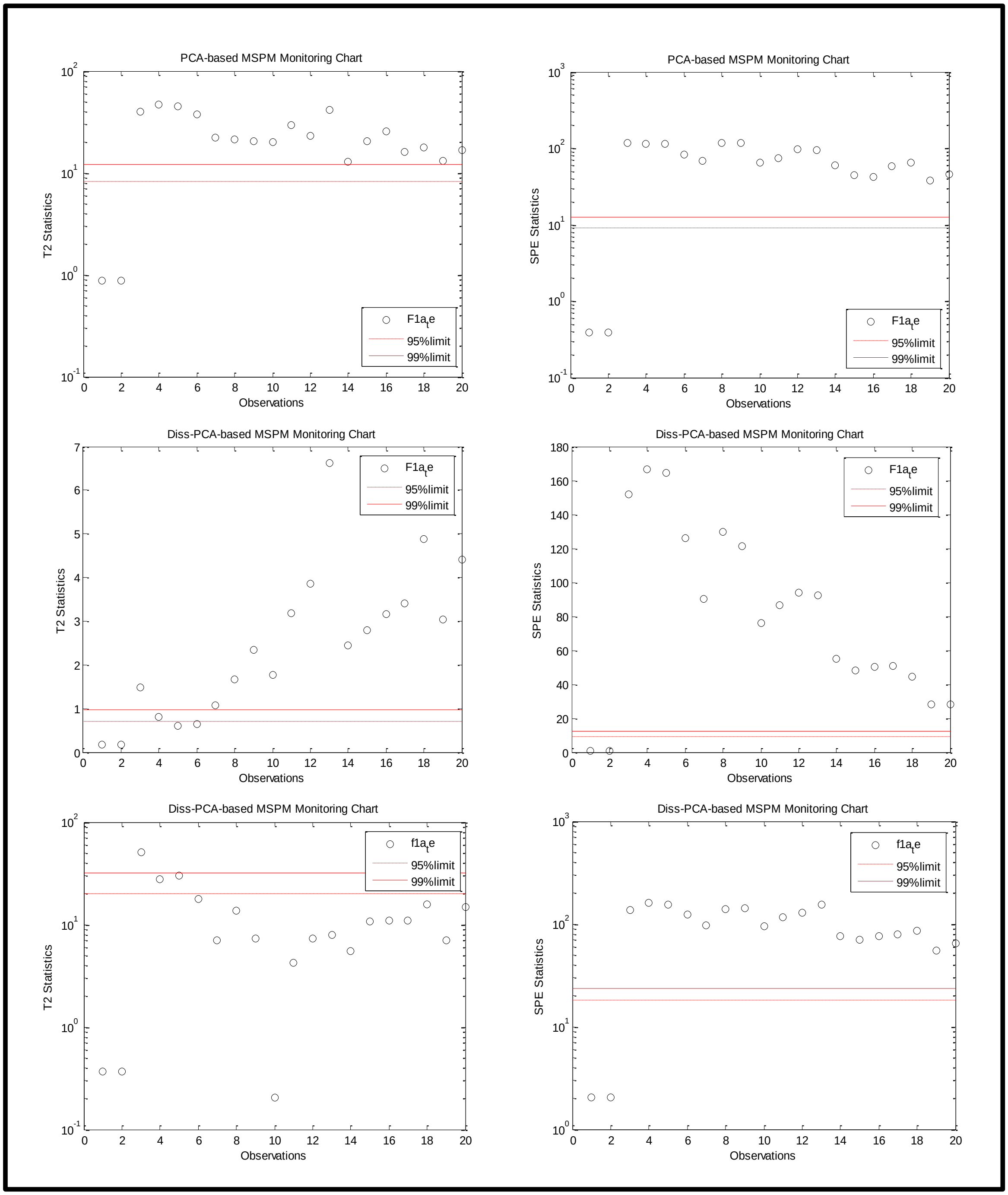


Figure 4.9 Hotelling's T^2 and SPE monitoring statistics chart plotted together with the 95% and 99% confidence limits of F1 for abrupt fault data: conventional PCA-based MSPM (top diagrams), dissimilarity-based MSPM of city block distance (middle diagrams) and dissimilarity-based MSPM of mahalanobis distance (bottom diagrams)

From Figure 4.9, it shows that when the fault introduced at sample 2, most of observations after sample 2 in the Hotelling's T^2 and SPE statistics for the abrupt fault data by using different method are located far beyond the 99% confidence limits. But for Hotelling's T^2 monitoring statistics charts of dissimilarity-based MSPM (mahalanobis distance), only one observation that is out of the 99% confidence limits. However, this figure can be stated as abnormal condition due to the observations trend after sample 2. Based on the finding, SPE more efficient than Hotelling's T^2 monitoring statistics charts which are able to detect fault with consistent of observations that out of control limit after sample 2 for this case. Besides that, based on the monitoring statistics charts for the conventional PCA-based MSPM and dissimilarity-based MSPM respectively, both methods are able to detect the fault directly after sample 2. It is obviously from the figure that the fault detection time for all methods is equal to 1. Therefore, it can be conclude that all methods can effectively detect the fault at the same time. Figure 4.10 shows the T^2 and SPE statistics for the incipient fault.

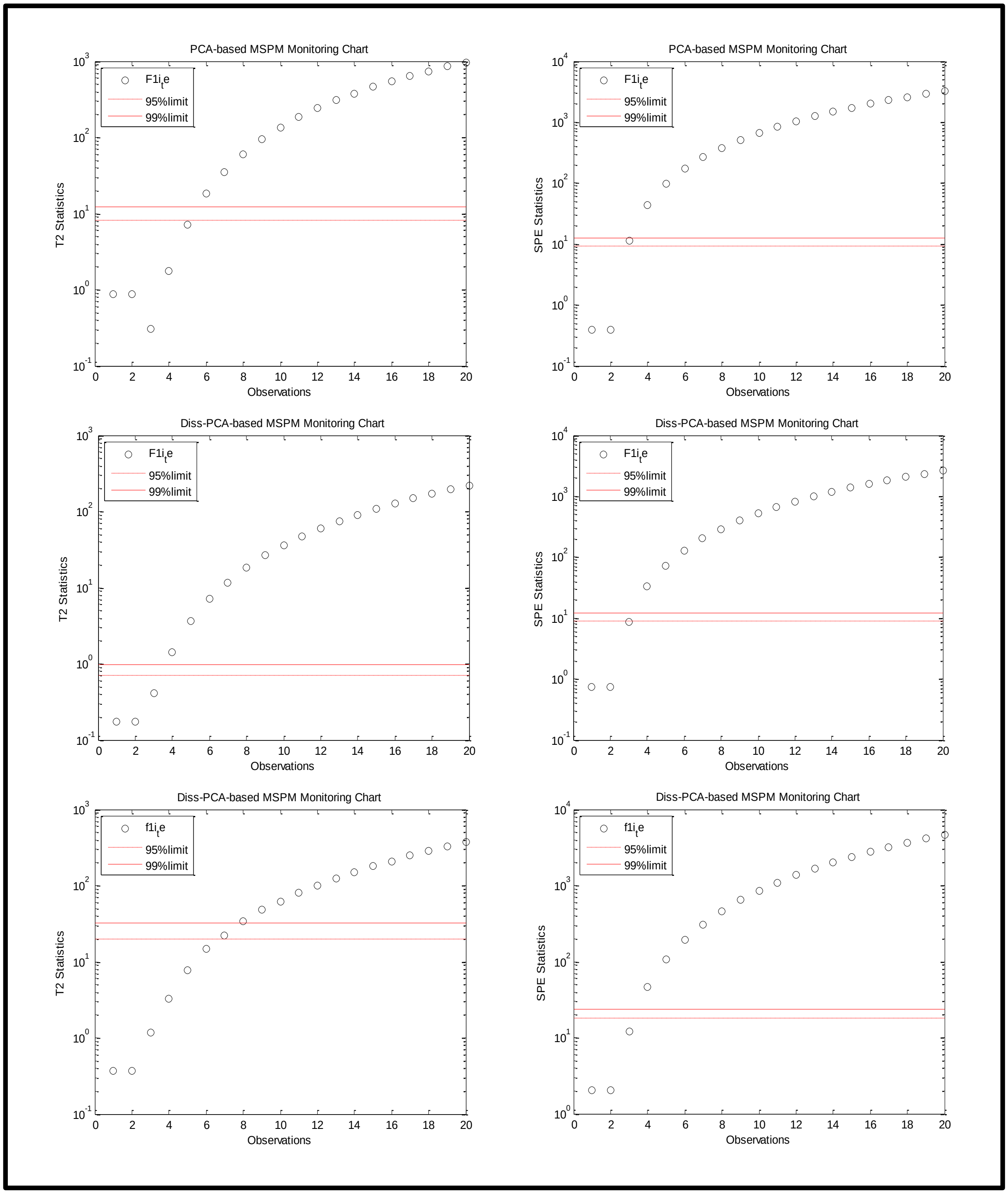


Figure 4.10 Hotelling's T^2 and SPE monitoring statistics chart plotted together with the 95% and 99% confidence limits of F1 for incipient fault data: conventional PCA-based MSPM (top diagrams), dissimilarity-based MSPM of city block distance (middle diagrams) and dissimilarity-based MSPM of mahalanobis distance (bottom diagrams)

Based on Figure 4.10, it clearly seen that when the fault introduced at sample 2, all of observations after sample 2 in the Hotelling's T^2 and SPE statistics for the incipient fault data by using different method are gradually increase far beyond the 99% confidence limits. Furthermore, there are different between Hotelling's T^2 and SPE statistics charts of conventional PCA-based MSPM and dissimilarity-based MSPM (mahalanobis distance). The Hotelling's T^2 is less effective than SPE statistics charts which are slightly slower to detect the observations that out of control limit after sample 2. Besides that, based on the monitoring statistics charts for the conventional PCA-based MSPM and dissimilarity-based MSPM respectively, both methods are able to detect the fault directly after sample 2. It is noticeable from the figure of SPE monitoring statistics charts that the fault detection time for all methods is equal to 2. Therefore, it can be conclude that all methods can effectively detect the fault at the same time.

4.3.2.2. Monitoring Outcomes Based on Six PCs

The Table 4.3 illustrates the finding acquired from analysing the monitoring statistics for overall runs by using two different methods which are the conventional PCA-based MSPM and dissimilarity-based MSPM algorithm. The results also based on two faults stated earlier which involved two types of faults which are abrupt and incipient faults respectively which based on six PCs.

Refer to tabulated results, it shows that both conventional PCA-based MSPM and dissimilarity-based MSPM are able to detect the entire fault. The conventional PCA-based MSPM and dissimilarity-based MSPM based on mahalanobis distance are detecting the fault slightly slower compared to the dissimilarity-based MSPM based on city block distance. Currently, fault two (F2) which is external feed-reactant flow rate is too high include abrupt and incipient faults are discussed more details. The result of T^2 and SPE statistics for fault one (F1) can be seen in appendix A. Figure 4.11 shows the T^2 and SPE statistics for the abrupt fault.

Table 4.3 Fault detection time for abrupt and incipient faults based on six PCs

Fault	Conventional PCA			Dissimilarity Matrix					
				City Block			Mahalanobis		
	T^2	SPE	Final	T^2	SPE	Final	T^2	SPE	Final
F1a	1	1	1	1	1	1	1	1	1
F1i	2	6	2	1	4	1	3	2	2
F2a	1	6	1	1	6	1	1	1	1
F2i	4	18	4	3	11	3	6	4	4

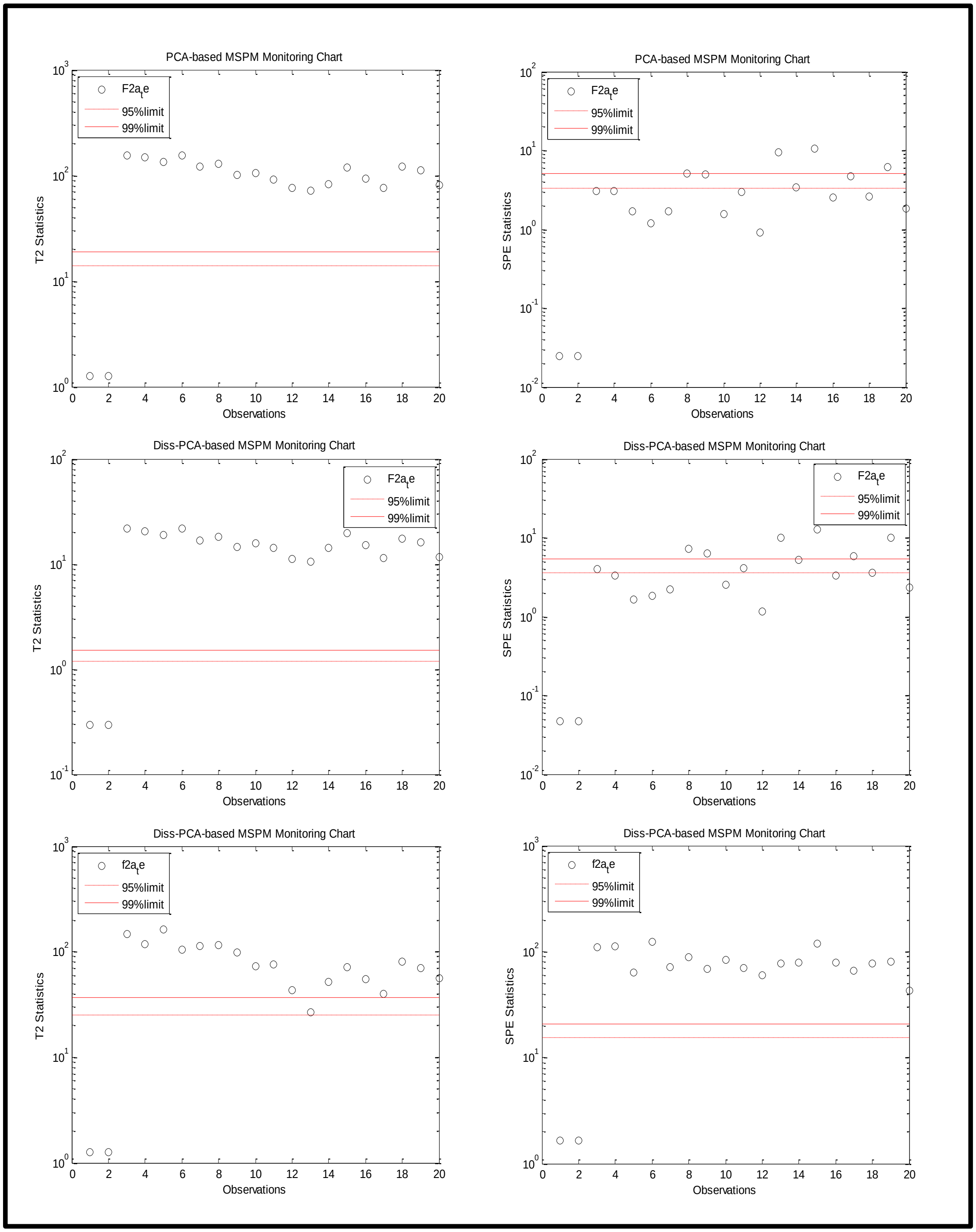


Figure 4.11 Hotelling's T^2 and SPE monitoring statistics chart plotted together with the 95% and 99% confidence limits of F2 for abrupt fault data: conventional PCA-based MSPM (top diagrams), dissimilarity-based MSPM of city block distance (middle diagrams) and dissimilarity-based MSPM of mahalanobis distance (bottom diagrams)

From Figure 4.11, it shows that when the fault introduced at sample 2, most of observations after sample 2 in the Hotelling's T^2 and SPE statistics for the abrupt fault data by using different method are located far beyond the 99% confidence limits. Furthermore, the SPE monitoring statistics charts of conventional PCA-based MSPM and dissimilarity-based MSPM based on city block distance, most of the observation are well below of 99% confidence limits after introduced fault at sample 2. However, this figure can be stated as abnormal condition due to the observations trend after sample 2. Based on the result above, Hotelling's T^2 is more efficient than SPE monitoring statistics charts which are able to detect fault with consistent of observations that out of control limit after sample 2 for conventional PCA-based MSPM and dissimilarity-based MSPM based on city block distance. But for dissimilarity-based MSPM based on mahalanobis distance, the observation of SPE monitoring statistics charts is more consistent compared to the other one. Besides that, based on the monitoring statistics charts for the conventional PCA-based MSPM and dissimilarity-based MSPM respectively, both methods are able to detect the fault directly after sample 2. It is clearly from the figure that the fault detection time for all methods is equal to 1. Therefore, it can be conclude that all methods can effectively detect the fault at the same time. Figure 4.12 shows the T^2 and SPE statistics for the incipient fault.

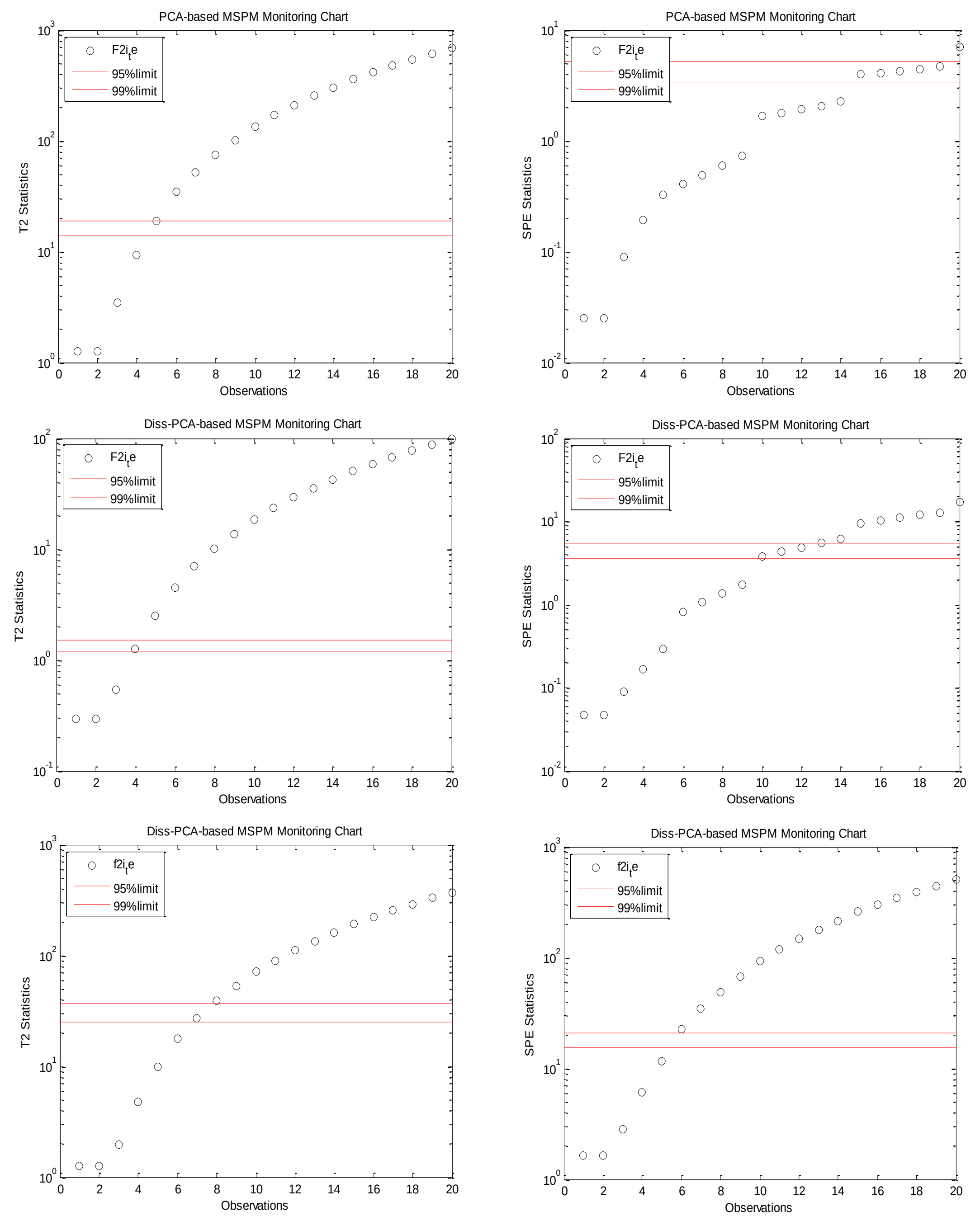


Figure 4.12 Hotelling's T^2 and SPE monitoring statistics chart plotted together with the 95% and 99% confidence limits of F2 for incipient fault data: conventional PCA-based MSPM (top diagrams), dissimilarity-based MSPM of city block distance (middle diagrams) and dissimilarity-based MSPM of mahalanobis distance (bottom diagrams)

Based on Figure 4.12, it seem that all of observations after sample 2 in the Hotelling's T^2 and SPE statistics for the incipient fault data by using different method are ascending by stage far exceed the 99% confidence limits when the fault introduced at sample 2. But there are different between Hotelling's T^2 and SPE statistics charts of conventional PCA-based MSPM and dissimilarity-based MSPM. The Hotelling's T^2 can introduce the fault earlier than SPE statistics charts for conventional PCA-based MSPM and dissimilarity-based MSPM (city block distance). Otherwise for dissimilarity-based MSPM (mahalanobis distance), the Hotelling's T^2 is less effective than SPE statistics charts the which are slightly slower to detect the observations that out of control limit after sample 2. Moreover, based on the finding for the conventional PCA-based MSPM and dissimilarity-based MSPM respectively, both methods are able to detect the fault directly after sample 2. It is appear clearly from the figure of SPE monitoring statistics charts that the fault detection time for conventional method and new method based on mahalanobis is equal to 4. Nevertheless, fault detection time for new method based on city block distance earlier than both method stated before which equal to 3. Therefore, it can be summarizing that new methods based on city block is more effective to detect the fault compared to other method used in this thesis.

4.4 Summary

The application on a simulated CSTRwR process is monitored by using the conventional PCA-based MSPM and dissimilarity-based MSPM. The conventional and new algorithm results have been discussed earlier, which include both of the NOC and two fault data. As a conclusion, it is proven that the new algorithms proposed are comparable to the conventional method. Thus can be the other alternative ways in the process monitoring performance.

CHAPTER V

CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

The main goal of this research is to introduce dissimilarity-based MSPM as new technique to detect fault in process monitoring performances. This study can achieved their aim through their objectives which the multivariate dimensional data reduction was developed by using dissimilarity methods instead of the conventional PCA technique. Then, both of techniques were run and the monitoring performance between the conventional PCA and dissimilarity techniques were compared as well as analysed.

From the finding, it is proven that the proposed system can able to detect the fault including both abrupt and incipient faults as efficient as the conventional technique. Although the proposed system is less effective to find the number of PCs that required in the process due to the less percentage of the total variances transformed compared to the conventional system. Besides that, the new system based on city block distance can achieved consistency of process monitoring performance compared to dissimilarity system based on mahalanobis distance.. Simultaneously, can support the reason why the new proposed system has potential in process monitoring performances.

5.2 Recommendations

Firstly, the finding from this research may valid only for the case study of CSTRwR system. Therefore, it is recommended for future research to use data from other chemical processing systems. Examples of other chemical processes are packed bed reactor (PBR), plug flow reactor (PFR) or other known chemical reactors. Fundamentally, dissimilarity method can cope with the input data in terms of both quantitative and qualitative measures while conventional method only used data in term of quantitative measures. For future research, it is suggested to use qualitative data to prove the fundamental stated earlier. Furthermore, in this research there are only two faults being considered to acquire the finding to be analysed. The result and strong justification which is to differentiate both conventional and new method can be improved by using more possible faults that can be predicted in the system as well as depend on the availability of the data itself. Finally, the performance of the system in process monitoring and effectiveness as well as their efficiency of the system can be enhanced through the recommended stated earlier.

REFERENCES

Bakshi, B. R., (1998). Multiscale PCA With Application To Multivariate Statistical Process Monitoring. *AIChE Journal*, 44, 1596 – 1610.

Bartle P. (2007). The Nature of Monitoring and Evaluation: Definition and Purpose. Retrieved from file:///D:/Document/PSM%201/What%20Monitoring%20Is%3B%20Definition %20and%20Purpose.htm

Borg, I., and Groenen, P. J. F. (2005). *Modern Multidimensional Scaling*. United States, America: Springer Science + Business Media.

Chen, J., and Liu, J. (2000). Post Analysis on Different Operating Time Processes using Orthonormal Function Approximation and Multiway Principal Component Analysis. *Journal of Process Control,* 10, 411-418.

Choi, S. W., Martin, E. B., and Morris, A. J. (2005). Fault Detection Based on a Maximum-Likelihoood Principal Component Analysis (PCA) Mixture. *Ind. Eng. Chem. Res.,* 44, 2316-2327.

Choi, S. W., Morris, J., and Lee, I. B. (2008). Nonlinear Multiscale Modelling for Fault Detection and Identification. *Chemical Engineering Science,* 63, 2252-2266.

Cinar, A., Palazoglu, A., and Kayihan, F. (2007). Chemical Process Performance Evaluation. *Univariate and Multivariate Statistical Monitoring Techniques.* Retrieved from file:///D:/Document/SEMESTER%206/BKC3853%20PROCESS%20MONITORING/ebooksclub.org__Chemical_Process_Performance_Evaluation.pdf.

Cox, T. F. (2001). Multidimensional Scaling used in Multivariate Statistical Process Control. *Journal of Applied Statistics,* 28, 365-378.

Cox, T.F., and Cox M.A.A. (1994). *Multidimensional Scaling.* London, Great Britain: Chapman & Hall.

Grant, E.L., and Leavenworth, R.S., (1988). *Statistical Quality Control*. USA: McGraw- Hill.

Gunther, J. C., Conner, J. S., and Seborg, D. E. (2009). Process Monitoring and Quality Variable Prediction Utilizing PLS in Industrial Fed-Batch Cell Culture. *Journal of Process Control*, 19, 914-921.

Jia, F., Martin, E.B., and Morris, A. J. (1998). Non-linear Principal Components Analysis for Process Fault Detection. *Computers Chem. Engng,* 22, S851-S854.

Kano, M., Hasebe, S., Hashimoto, I., and Ohno, H. (2004). Evolution of Multivariate Statistical Process Control: Application of Independent Component Analysis and External Analysis. *Computer & Chemical Engineering,* 28, 1157-1166.

Lee, J. M., Yoo, C., and Lee, I. (2004). Statistical Process Monitoring with Independent Component Analysis. *Journal of Process Control*, 14, 467-485.

MacGregor, J. F., (1994). Statistical Process Control of Multivariate Processes. Tutorial Session 3, Proc. IFAC Int. Symp., ADCHEM' 94, Kyoto, Japan.

MacGregor, J. F., and Kourti, T. (1995). Process Analysis, Monitoring and Diagnosis, using Multivariate Projection Methods. *Chemometrics and Intelligent Laboratory System,* 28, 3-21.

Maulud, A., Wang, D., and Romagnoli, J. A. (2006). A Multi-Scale Orthogonal Nonlinear Strategy for Multi-Variate Statistical Process Monitoring. *Journal of Process Control*, 16, 671-683.

Montgomery, D.C., (1985). *Statistical Quality Control.* Singapore: John Wiley & Sons.

Neto, P. R., Jackson, D. A., and Somers, K. M. (2005). How Many Principal Components? Stopping Rules for Determining the Number of Non-trivial Axes Revisited. *Computational Statistics & Data Analysis*, 49, 974-997.

Nomikos, P., and MacGregor, J. F., (1994). Monitoring of Batch Processes Using Multi-Way Principal Component Analysis. *AIChE Journal*, 40 (8), 1361 – 1375.

Romagnoli, J. A., and Palazoglu, A. (2006). *Introduction to Process Control*. Florida: CRE Press.

Simoglou, A., Martin, E. B., and Morris, A. J. (2002). Statistical Performance Monitoring of Dynamic Multivariate Processes using State Space Modelling. *Computer & Chemical Engineering*, 26, 909-920.

Wang, D., and Romagnoli, J. A. (2005). Robust Multi-Scale Principal Components Analysis with Applications to Process Monitoring. *Journal of Process Control*, 15, 869-882.

Wetherill, G.B., and Brown, D.W., (1991). *Statistical Process Control. Theory and Practice.*London, Great Britain: Chapman and Hall.

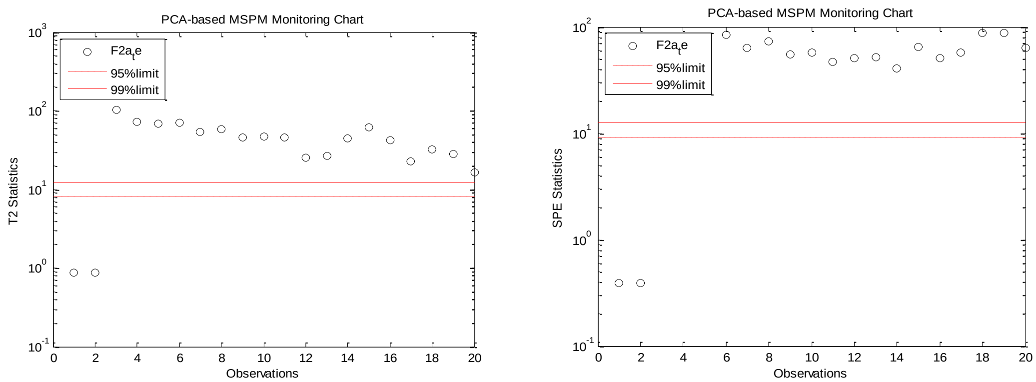
Yunus, M. Y. M., and Zhang, J. (2010). Multivariate Process Monitoring using Classical Multidimensional Scaling and Procrustes Analysis. *Dynamics and Control of Process Systems*, 151-156.

Zhang, J., (2006). Improved On-line Process Fault Diagnosis Through Information Fusion in Multiple Neural Networks. *Computers and Chemical Engineering*, 30, 558-571.

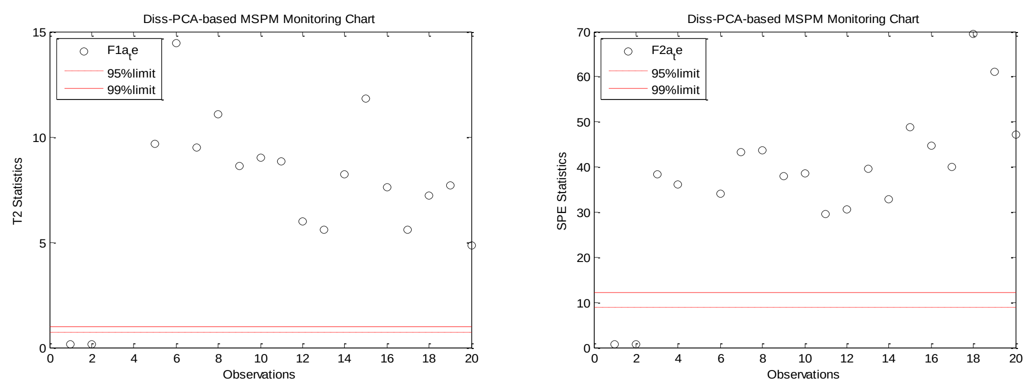
APPENDICES

MONITORING OUTCOMES A

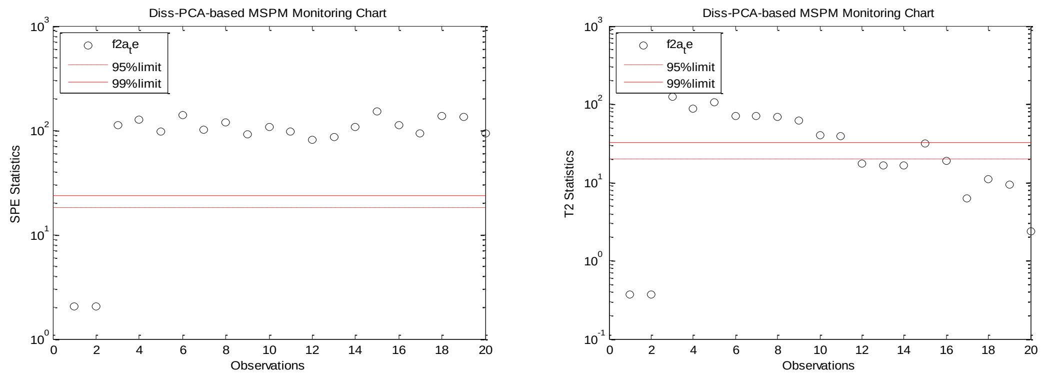
Monitoring Outcomes Based on Three PCs for Abrupt Fault 2



Conventional PCA-based MSPM (top diagrams)

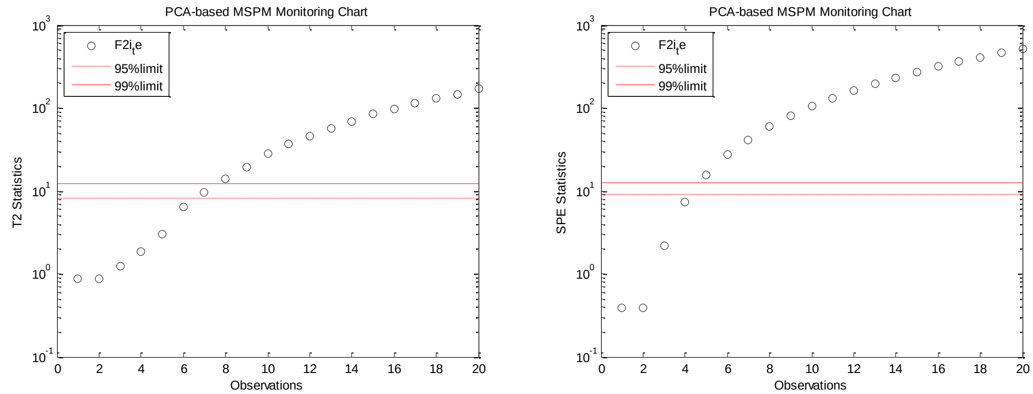


Dissimilarity-based MSPM of city block distance (middle diagrams)

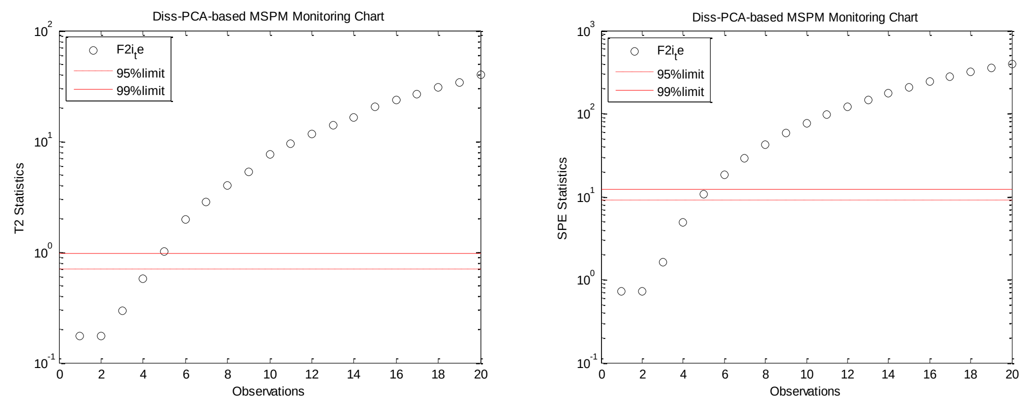


Dissimilarity-based MSPM of mahalanobis distance (bottom diagrams)

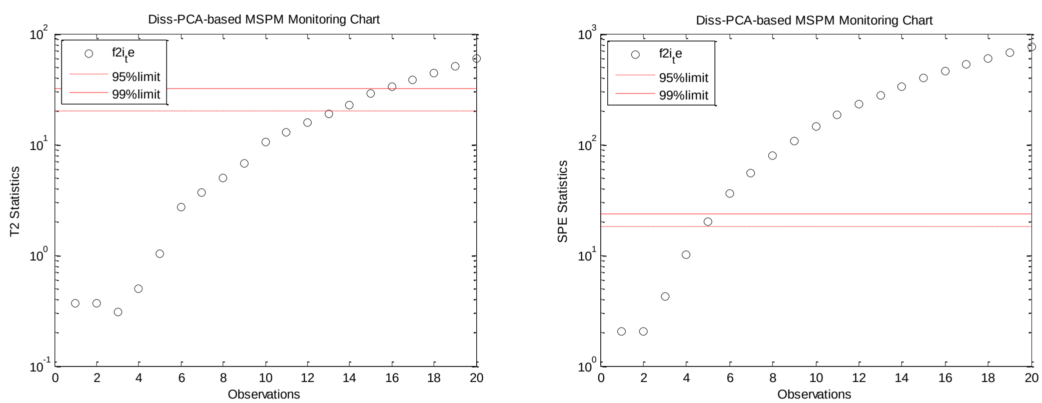
Monitoring Outcomes Based on Three PCs for Incipient Fault 2



Conventional PCA-based MSPM (top diagrams)

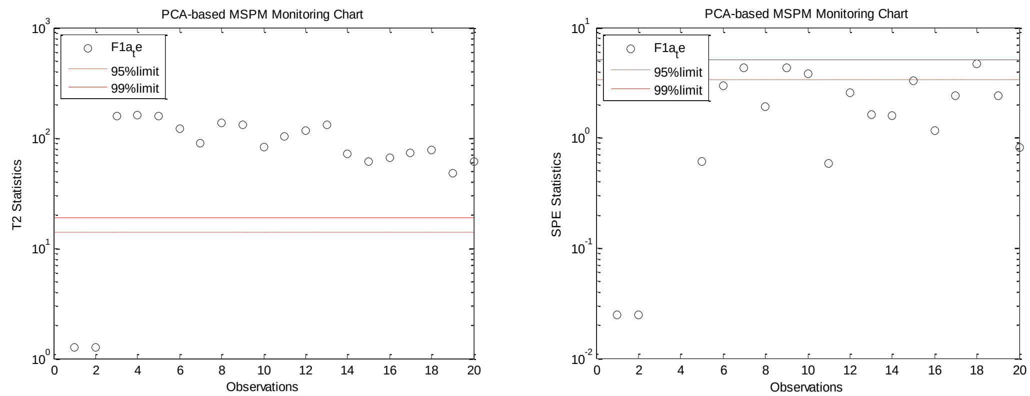


Dissimilarity-based MSPM of city block distance (middle diagrams)

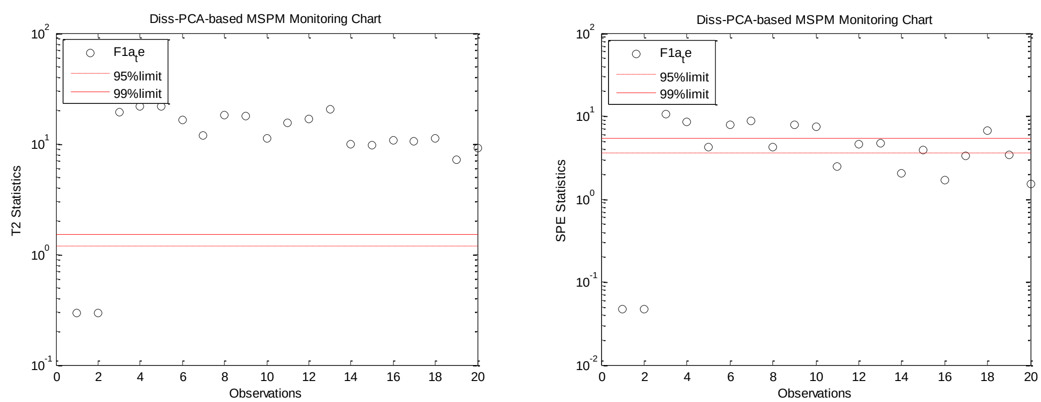


Dissimilarity-based MSPM of mahalanobis distance (bottom diagrams)

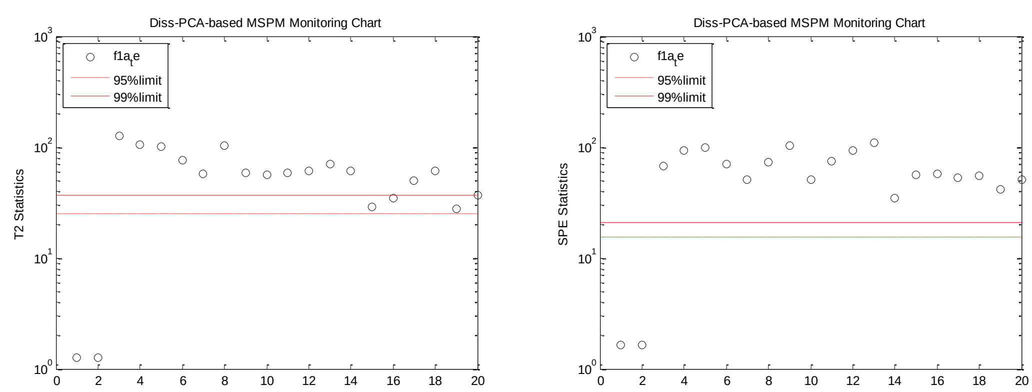
Monitoring Outcomes Based on Six PCs for Abrupt Fault 1



Conventional PCA-based MSPM (top diagrams)

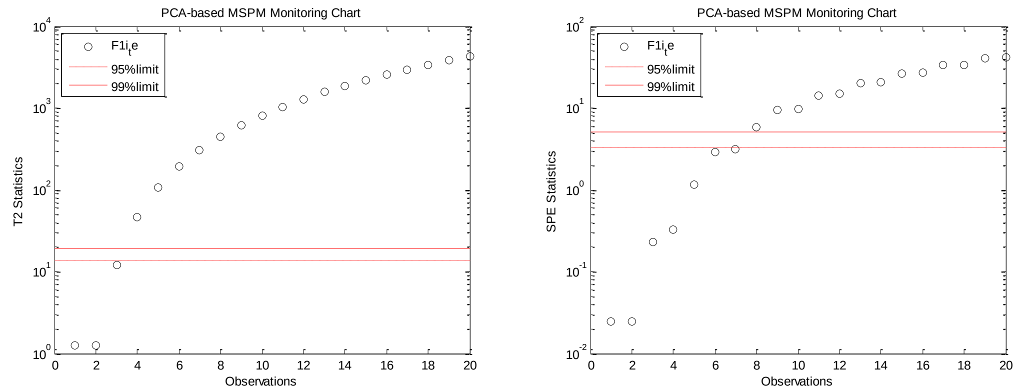


Dissimilarity-based MSPM of city block distance (middle diagrams)

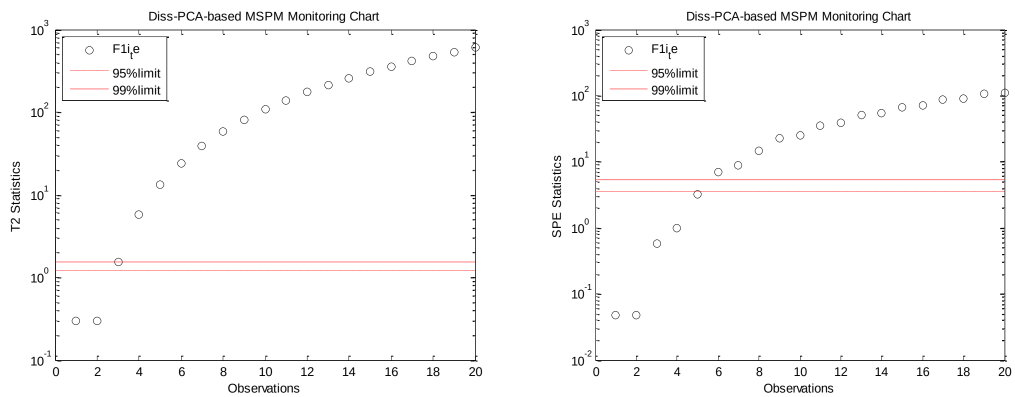


Dissimilarity-based MSPM of mahalanobis distance (bottom diagrams)

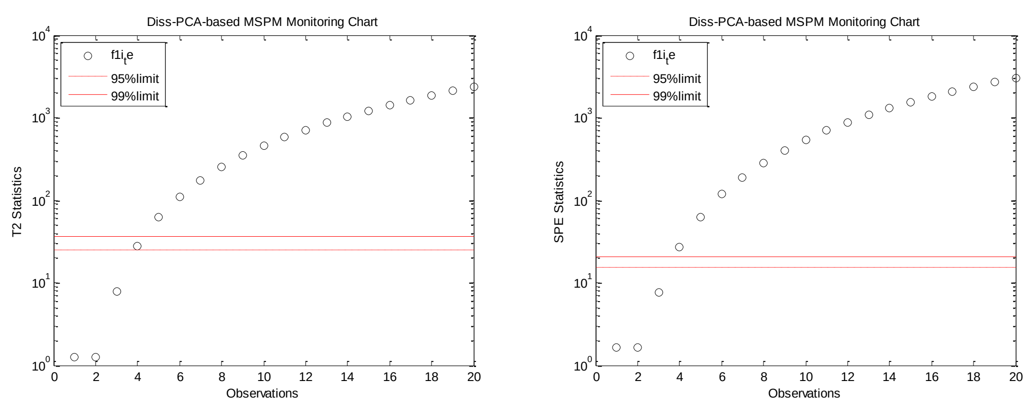
Monitoring Outcomes Based on Six PCs for Incipient Fault 1



Conventional PCA-based MSPM (top diagrams)



Dissimilarity-based MSPM of city block distance (middle diagrams)



Dissimilarity-based MSPM of mahalanobis distance (bottom diagrams)