DEVELOPMENT OF DISSIMILARITY-BASED MSPM SYSTEM

NURULSAIDATULHANIZA BINTI ZAHARI

BACHELOR OF CHEMICAL ENGINEERING

UNIVERSITI MALAYSIA PAHANG

Thesis Access Form

No____________________________Location______________________________

Author : ………………………………

Title : ………………………………

Status of access OPEN / RESTRICTED / CONFIDENTIAL

Moratorium period:_____________ years, ending__________/_______200____

Conditions of access proved by (CAPITALS): DR MOHD YUSRI BIN MOHD YUNUS

Supervisor (Signature)………………………………………………

Faculty: ………………………………

Author's Declaration: *I agree the following conditions:*

OPEN access work shall be made available (in the University and externally) and reproduced as necessary at the discretion of the University Librarian or Head of Department. It may also be copied by the British Library in microfilm or other form for supply to requesting libraries or individuals, subject to an indication of intended use for non-publishing purposes in the following form, placed on the copy and on any covering document or label.

*The statement itself shall apply to **ALL** copies:*

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Restricted/confidential work: All access and any photocopying shall be strictly subject to written permission from the University Head of Department and any external sponsor, if any.

Author's signature……………………………………….Date: ………………………………

users declaration: for signature during any Moratorium period (Not Open work):
I undertake to uphold the above conditions:

Date	Name (CAPITALS)	Signature	Address

DEVELOPMENT OF DISSIMILARITY-BASED MSPM SYSTEM

NURULSAIDATULHANIZA BINTI ZAHARI

Thesis submitted in partial fulfilment of the requirements

for the award of the degree of

Bachelor of Chemical Engineering

Faculty of Chemical & Natural Resources Engineering

UNIVERSITI MALAYSIA PAHANG

JANUARY 2014

SUPERVISOR'S DECLARATION

We hereby declare that we have checked this thesis and in our 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 main supervisor	: DR. MOHD YUSRI BIN MOHD YUNUS
Position	: LECTURER
Date	:

STUDENT'S DECLARATION

I hereby declare that the work in this thesis is my own except for 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 : NURULSAIDATULHANIZA BINTI ZAHARI

ID Number : KA10034

Date :

Dedicated to my beloved parents and myself for work hard to finish up this thesis

ACKNOWLEDGEMENT

In the name of Allah, the Most Gracious and Most Merciful.

Alhamdulillah, I would like to thank all those people who had made this thesis completed and possible.

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

Furthermore, I would like to express my 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 family during the honour years 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 this thesis.

ABSTRACT

This research is about development of dissimilarity matrix based on Multivariate Statistical Process Monitoring (MSPM) system. MSPM is an observation system to validate whether the process is happening according to its desired target. Nowadays, the chemical process industry is highly based on the non-linear relationships between measured variables. However, the conventional Principal Component Analysis (PCA) which applied based on MSPM system is less effective because it only valid for the linear relationships between measured variables. In order to solve this problem, the technique of dissimilarity matrix is used in multivariate statistical process monitoring as alternative technique which models the non-linear process which simultaneously can improve the process monitoring performance. The procedures in MSPM system consists of two main phases basically for model development and fault detection. This research focused on converting dissimilarity matrix to minor product moment before proceeding to PCA process which runs by using Matlab software. The monitoring performance in both techniques were compared and analysed to achieve the aims of this research. 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, it can be the other alternative method 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 pembentukkan perbezaan matrik berasaskan sistem proses pemantauan multivariat statistik (MSPM). MSPM adalah sistem pemerhatian untuk mengesahkan sama ada proses yang berlaku mengikut sasaran yang dikehendaki. Pada masa kini, industri proses kimia adalah berdasarkan hubungan bukan linear antara pembolehubah yang diukur. Walau bagaimanapun, system konvensional Proses Analisis Komponen (PCA) yang dijalankan mengikut sistem MSPM kurang berkesan kerana ia hanya sah untuk hubungan linear antara pembolehubah yang diukur. Dalam usaha untuk menyelesaikan masalah ini , teknik perbezaan matrik digunakan dalam proses pemantauan multivariat statistik sebagai teknik alternatif yang berasaskan proses bukan linear yang pada masa yang sama boleh meningkatkan prestasi proses pemantauan. Pada dasarnya, prosedur di dalam sistem MSPM terdiri daripada dua fasa utama iaitu untuk pembentukkan model dan pengesanan masalah. Kajian ini memberi tumpuan kepada penukaran perbezaan matrik menjadi masa produk kecil sebelum bersambung ke proses PCA yang dibentuk menggunakan perisian Matlab. Prestasi pemantauan dalam kedua-dua teknik dibandingkan dan dianalisis untuk mencapai matlamat kajian ini. Hasil kajian ini digambarkan dalam bentuk pemantauan statistik Hotelling T^2 dan Squared Ramalan Kesilapan (SPE) untuk dianalisis. Kesimpulannya , sistem perbezaan matrik adalah setanding dengan kaedah konvensional . Oleh itu, ia boleh menjadi kaedah alternatif dalam melaksanakan proses pemantauan. Akhirnya, ia adalah disyorkan untuk menggunakan data dari sistem pemprosesan kimia yang lain untuk memberi justifikasi yang lebih padat berkenaan teknik baru ini..

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

X	Normal operating data
X^T	Normal operating data transpose
$\breve{X}$	Standardised data
C_{mxm}	Variance-covariance matrix
λ	Eigen values
V	Eigenvectors
P	PC scores
E	Residual matrix
N	Samples
m	Variables
i	Row
j	Column
B	Scalar product matrix
q_i	Loading vector of PCA
x	Data
$\overline{x}$	Data means
σ	Standard deviation
ϕ(x)	Nonlinear transformation
k	Principal component
A	Number of PCS retained in PCA model
n	Number of nominal process measurements per variable
$P_{i,j}$	i^{th} score for Principal Component *j*
λ_j	Eiganvalue corresponds to Principal Component *j*
z_α	Standard normal deviate corresponding to the upper (1- α) percentile
X_z	Standardized matrix of original matrix, X
I	Identity matrix
J	Centring matrix
e_i	i^{th} row in residual matrix
Q_i	SPE statistics
$\{\delta_{rs}\}$	Dissimilarity
Λ	Diagonal matrix
V^T	Normalized orthogonal matrix
α	Level of control limit

LIST OF ABBREVIATIONS

PBR	Packed bed reactor
PFR	Plug flow reactor
CA	Canonical correlation analysis
CMDS	Classical multidimensional scaling
CVA	Canonical variate analysis
FA	Factor analysis
F1	Fault 1
F2	Fault 2
F3	Fault 3
ICA	Independent component analysis
IT-net	Input-training neural network
KPCA	Kernel PCA
MDS	Multidimensional scaling
MPCA	Multi-way PCA
MSPCA	Multi-scale PCA
MSPC	Multivariate statistical process control
MSPM	Multivariate statistical process monitoring
NOC	Normal operating data
PARAFAC	Parallel factors analysis
PC	Principal component
PCA	Principal component analysis
P.D.F	Probability density function
PLS	Partial least square
SD	Singular decomposition
SVD	Singular value decomposition
SPC	Statistical process control
SPE	Squared prediction errors

CHAPTER 1

INTRODUCTION

1.1 RESEARCH BACKGROUND

The ultimate aim of any production system is to produce the maximum amount of high quality products as per requested and specified by the customers. This is regarded as highly challenging due to the nature of the processes that always change over time and are also affected by various factors such as variations of raw materials as well as operating conditions, the presence of disturbances and also modification in the process technologies. In any of the situations, one of the main critical problems is to promptly detect the occurrence of faulty or abnormal operating conditions in the routine process operation and subsequently remove them. Such issues can be addressed quite effectively by the use of process monitoring techniques. 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).

SPC techniques involve univariate methods, that is, observing and analysing a single variable at a time. Industrial quality problems are multivariate in nature, since they involve measurements on a number of characteristics, rather than one single characteristic. The conventional SPC charts such as Shewhart chart and CUSUM chart have been widely used for monitoring univariate processes, however they do not function well for multivariable processes with highly correlated variables. Most of the limitations of univariate SPC can be addressed through the application of Multivariate Statistical Process Control (MVSPC) techniques, which consider all the variables of interest simultaneously and can extract information on the behaviour of each variable or characteristic relative to the others. Thus, multivariate statistical process monitoring (MSPM) can be considered as the most practical method for monitoring complicated and large scale industrial processes (Manabu et al., 2000).

According to Yunus and Zhang (2010), MSPM has been shown to be a very effective process monitoring tool. The framework which has been originated from the method of statistical process control (SPC) is aimed to maintain consistent productivity by way of anticipating early warning of possible process malfunctions in the multivariate process. MSPM methods are basically algorithms that can be used for extracting important information from large multivariable data sets such as plant data. Its performance depends on how well the model describes relationships between the variables. Therefore, the key feature of such methods is the possibility to handle highly correlated, highly dimensional and noisy data. MSPM methods describe original data by the reduced set of variables which in turn makes analysis of the data much easier (Sliskovic et al., 2012).

1.2 MOTIVATION AND PROBLEM OF STATEMENT

Over last decade, many chemical process industries used MSPM as an alternative method in process monitoring performances and fault diagnosis for their plants. One of the tools in multivariable statistical techniques is Principal Component Analysis (PCA). Lindsay (2002) has defined PCA as a way to identify patterns in data and express the data in such a way to highlight their similarities and differences. PCA is a powerful tool for analysing data since patterns in data can be hard to find in data of high dimension. The other main advantage of PCA is once the patterns are found the data can be compress by reducing the number of dimensions without loss much of information.

Research done by Faezah and Athena (n.d) proved that PCA provide a roadmap to shrink a complex data set to lower dimension and it can analyse the basis of variation present in multi-dimensional data set. However, Choi, Morris and Lee (2008) said that conventional PCA based on MSPM is only valid for the non-auto correlated data with linear relationships between measured variables. Often, inefficient and unreliable process monitoring scheme can materialize as a consequence of the underlying assumption of PCA-based MSPM being violate. Recently, the chemical process industry is highly based on the non-linear relationships between measured variables. Thus, the conventional PCA based on 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 good quality control expectation as the goal to produce the maximum amount of highly quality product that requested and specified by the customer. In react to this issue, dissimilarity method based on MSPM is expected to solve the current problem which models the non-linear process. Dissimilarity method 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 done to study and explore about the dissimilarity and perhaps can introduce it as another alternative in process monitoring.

1.3 RESEARCH 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 MSPC. 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

i. What are the types of scales which can be used by the new system in achieving consistent process monitoring performance?
ii. How effective and efficient the new system may improve the process monitoring performance as compared to the conventional MSPM?
iii. 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. Using Shewhart chart to monitor the process performance.
vi. Using Tennessee Eastman process as a case study.

1.6 SIGNIFICANCE OF STUDY

This study produces a new idea on how to reduce the complexity of monitoring performance by using dissimilarity matrix method in modelling all the variables involved. The method is expected to improve the monitoring progressions especially in terms of fault detection sensitiveness.

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 significance of this research. The literature review is presented in chapter 2, where it describes the fundamental of MSPM, process monitoring issues and extension and multidimensional scaling in the MSPM framework. Chapter 3 explains the methodology for both conventional PCA and dissimilarity matrix methods. Chapter 4 is discussing on the result and discussion of the research and finally, conclusion and recommendations have been discussed in chapter 5.

CHAPTER 2

LITERATURE REVIEW

2.1 INTRODUCTION

Quality and safety are the two important aspects of any production process. Identification and control of chemical process is a challenging task because of their multivariate, highly correlated and non-linear nature. As mentioned in the first chapter MSPM is the effective tool in process monitoring. 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. 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 MSPM

Statistical performance monitoring of a process detects process faults or abnormal situations, hidden danger in the process followed by the diagnosis of the fault. The diagnosis of abnormal plant operation can be greatly facilitated if periods of similar plant performance can be located in the historical database (Yingwei and Yang, 2010). In general, there are four main steps of 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:

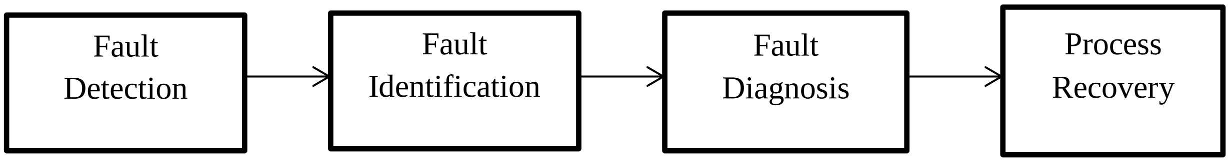

Figure 2.1: Main Steps in MSPM

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.

MSPM is based on the chemo metric techniques such as principal component analysis (PCA) and partial least squares (PLS). In previous work by Sliskovic et al. (2012), PCA was described as tool for data compression and information extraction which finds linear combination of variables that describes major trends in a data set. By using PCA, control limits are set for two kinds of statistics, T^2 and Q after a PCA model is developed. Q is the sum of squared errors, and it is a measure of the amount of variation not captured by the first few principal components. A measure of the variation within the PCA model is given by Hotelling's T^2 statistic. T^2 statistic is the sum of normalized squared scores, and it is a measure of the distance from the multivariate mean to the projection of the operating point on the subspace formed by the PCA model. PCA is also a linear transformation that is easy to be implemented for applications in which huge amount of data is to be analysed. In other words it is a numerical procedure for analyse the basis of variation present in a multi-dimensional data set (Faezah & Athena, n.d). Zhou (2010) also had described PCA is widely used in data compression and pattern matching by expressing the data in a way to highlight the similarities and differences without much loss of information. According to Spring (2010), PCA is one of techniques for taking high-dimensional data, and using the dependencies between the variables to represent it in a more tractable, lower-dimensional form, without losing too much information. The definitions of PCA from all researchers are quite similar to each other.

Based on study by Yusri (2012), first method in dimensionality reduction of PCA is a set of normal operating condition (NOC) data, X are identified off-line based on the historical process data. Then, the data are standardized to zero mean and unit variance with respect to each of the variables by using Equation (2.1) because PCA results depend on data scales.

$$\breve{\mathrm{X}}_{\mathrm{j,i}} = \frac{(\mathrm{X}_{\mathrm{j,i}} - \bar{\mathrm{X}}_{\mathrm{i}})}{\sigma_{\mathrm{i}}} \tag{2.1}$$

Where, $\breve{\mathrm{X}}_{\mathrm{j,i}} = \text{standardized data for variiable } 'i' \text{at sample } 'j'$

$\mathrm{X}_{\mathrm{j,i}}$ = original measurement for variable ‘i’ at sample ‘j’

$\bar{\mathrm{X}}_{\mathrm{i}} = \text{mean for variable } 'i'$

$\sigma_{\mathrm{i}} = \text{standard deviaton for variabe } 'i'$

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 eigenvalues, λ, and eigen vectors, V can be obtained. Finally, the Principal Component (PC) scores, P can be simply develop by using this formula, $P = \check{X}V$. 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 EXTENSIONS

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

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

Nikolov (2010) proposed that Non-Linear PCA is one of the process monitoring extensions based on linear technique of PCA. There several approaches to dealing with nonlinear datasets within the framework of PCA. One possibility is to model the data with a mixture of principal component analysers that trace out the nonlinear distribution using multiple linear principal subspaces. Assuming a Gaussian distribution for each subspace, the probability of a given data point is then defined by the probability each subspace assigns to the point and the probabilities that the point belongs to each subspace.

In Non-linear PCA, the Input-Training network has been developed to reduce the network complexity (Tan & Mavrovouniotis, 1995). 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 while 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.

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. The MSPCA methodology consists of decomposing each variable on a selected family of wavelets. The PCA model is then determined independently for the coefficients at each scale. The models at important scales are then combined in an efficient scale-recursive manner to yield the model for all scales together. For multivariate statistical process monitoring by MSPCA, the region of normal operation is determined at each scale from data representing normal operation. For new data, the important scales are determined as those where the current coefficient violates the detection limits. The actual state of the process is confirmed by checking whether the signal reconstructed from the selected coefficients violates the detection limits of the PCA model for the significant scales (Bakshi, 1998). Study done by Vijaykumar et al. (2012) shown that the multi-scale principal component generalizes the usual PCA of a multivariate signal seen as a matrix by performing simultaneously a PCA on the matrices of details of different levels. In addition, a PCA is performed also on the coarser approximation coefficients matrix in the wavelet domain as well as on the final reconstructed matrix. By selecting conveniently the numbers of retained principal components, interesting simplified signals can be reconstructed.

Besides that, Kernel PCA (KPCA) has been proposed by Kruger, Zhang & Zie (n.d) as one of PCA extensions. In construct the kernel matrix, a nonlinear transformation $\phi(x)$ from the original D-dimensional feature space to an M-dimensional feature space, where usually M > D. Then each data point x_n is projected to a point $\phi(x_n)$. Traditional PCA can be performs in the new feature space, but this might be extremely costly. Thus kernel methods are used to simplify the computation (Wang, 2012). The main benefit is that the original nonlinear behaviour can be mapped into the feature space and then analysed through linear correlation (through a specified means of kernel function), and as a result, linear PCA can be effectively executed for monitoring.

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 types 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.

Yusri (2012) stated that Partial least square (PLS) is the main competitor of PCA with regard to its popularity in the area of MSPM application. Among others, the original works have been proposed by Nomikos and MacGregor, (1995), as well as Kourti et al., (1995), for batch process monitoring using multi-way PLS, whereas Kourti and MacGregor, (1995) proposed using PLS for both continuous and batch processes. PLS regression is a recent technique that generalizes and combines features from principal component analysis and multiple regressions. It is particularly useful to predict a set of dependent variables from a very large set of independent variables. The goal of PLS regression is to predict Y from X and to describe their common structure. When Y is a vector and X is full rank, this goal could be accomplished using ordinary multiple regression. When the number of predictors is large compared to the number of observations, X is likely to be singular and the regression approach is no longer feasible (Abdi, n.d). In such cases, although there are many factors, there may be only a few underlying or latent factors that account for most of the variation in the response. The general idea of PLS is to try to extract these latent factors, accounting for as much of the manifest factor variation as possible while modelling the responses well.

Generally, Independent Component Analysis (ICA) is statistical technique for expose the secret factor that underlying a set of random variables, measurements or signals. ICA identifies non-Gaussian components which are modelled as a linear combination of the biological features. These components are statistically independent such as there is no overlapping information between the components. ICA therefore involves high order statistics, while PCA constrains the components to be mutually orthogonal, which involves second order statistics. As a result, PCA and ICA often choose different subspaces where the data are projected. As ICA is a blind source signal separation, it is used to reduce the effects of noise or artefacts of the signal since usually noise is generated from independent sources (Yao**,** Coquery and Kim, 2012)**.** According to the study by Matei (n.d), there are two distinct approaches towards computing the ICA. One employs high order cumulant and is found mainly in the statistical signal processing literature and the other uses the gradient-descent of non-linear activation functions in neuron-like devices and is mainly developed in the neural networks community. Each of the above approaches has advantages and shortcomings: the computation of high order cumulants is very sensitive to outliers and lack of sufficient support in the data especially for signals having a long-tailed probability density function (p.d.f.), while the neural-networks algorithms may become unstable, converge slowly and most often require some extra knowledge about the p.d.f. of the source signals in order to choose the non-linearities in the neurons.

Another extension of process monitoring based on multivariate technique is Canonical Variate Analysis (CVA). According to 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 based on process monitoring schemes.

2.4 DISSIMILARITY IN THE MSPM FRAMEWORK

In the present work, in order to improve the performance of process monitoring, a new statistical process monitoring method is proposed. The proposed method is based on the idea that a change of operating condition can be detected by monitoring a distribution of time-series data, which reflects the corresponding operating condition. In order to quantitatively evaluate the difference between two data sets, a new index representing dissimilarity is defined. According to Manabu et al. (2000), concept of dissimilarity is used for classifying a set of data for example, the degree of dissimilarity between two classes is measured by the distance between barycentre of the data and two classes with the smallest degree of dissimilarity are combined for generating a new class.

Based on the study of 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 measures 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):

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

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

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

$$\mathbf{A} = [\delta_{rs}^2] \qquad (2.4)$$

$$\mathbf{B} = -\frac{1}{2}\,\mathbf{J}\,\mathbf{A}\,\mathbf{J} \qquad (2.5)$$

$$\mathbf{B} = \mathbf{V\Lambda V^T} \qquad (2.6)$$

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 dissimilarity 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**, $\mathbf{B} = \mathbf{XX^T}$, in which **X** is standardized NOC data. By applying the Singular Decomposition (SD) operation on **B**, the following are obtained:

$$\mathbf{Bu_i} = \boldsymbol{\lambda}_i\mathbf{u_i} \qquad (2.7)$$

$$\mathbf{XX^Tu_i} = \lambda_i\mathbf{u_i} \qquad (2.8)$$

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

$$\mathbf{X^T}\,[\mathbf{XX^Tu_i}] = \mathbf{X^T}\,[\lambda_i\mathbf{u_i}] \qquad (2.9)$$

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,

$$\mathbf{Cq_i} = \boldsymbol{\lambda_i}\mathbf{q_i} \tag{2.10}$$

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 Kernel 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 3

METHODOLOGY

3.1 INTRODUCTION

This chapter will illustrate procedures on MSPM through development of PCA and dissimilarity matrix methods. Generally, there are varieties of technique in multidimensional scaling (MDS). It includes classical scaling, non-metric scaling, procrustes analysis, biplot and general dissimilarity. This chapter can be divided into three sections which are introduction, methodology 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 Mason and Young (2002), 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):

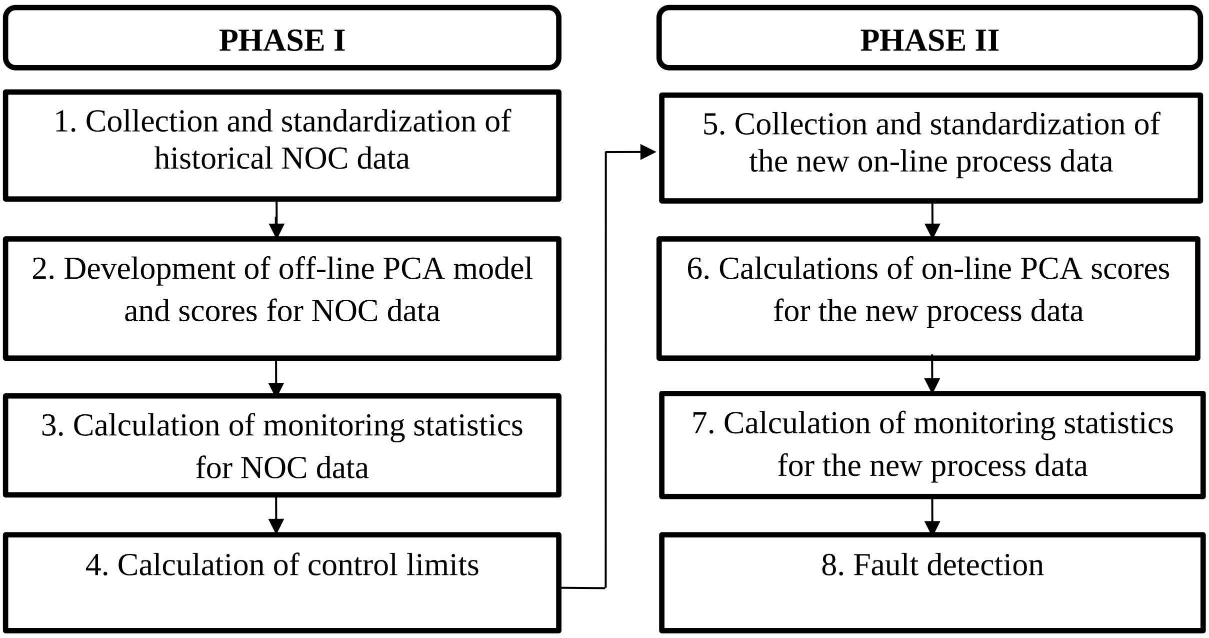

Figure 3.1: Procedures of fault detection

Source: Mason and Young 2002

Clearly, dissimilarity matrix technique is between step 1 and step 2 in the Phase I which is for off-line modelling 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 et. al.(1994). The dissimilarity matrix technique can illustrate as in the figure below:

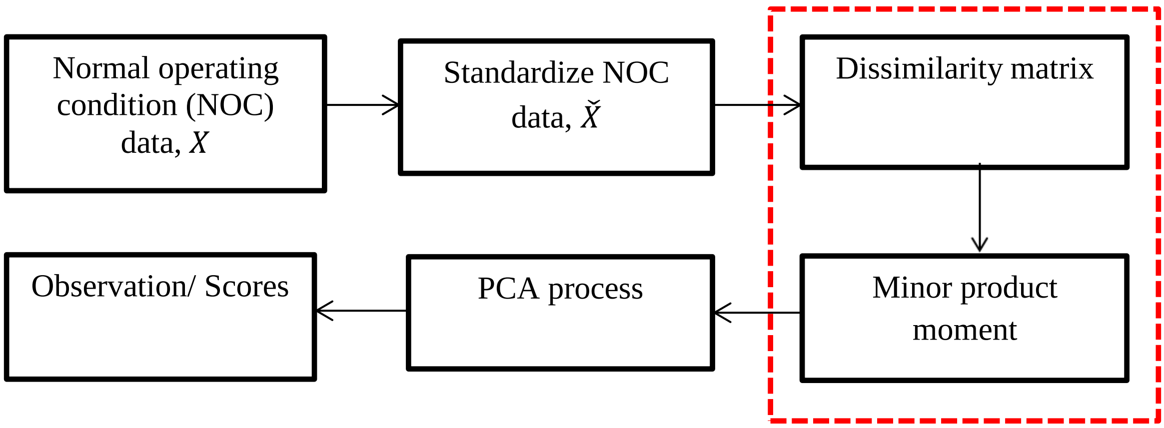

Figure 3.2: Main focuses for integration of dissimilarity matrix and PCA

Basically referred to Figure 3.1, 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 (Raich and Cinar, 1996).

3.2.1 Phase I: Off-line Modelling and Monitoring

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.

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

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.

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

Where, $\breve{X}_{j,i} =$ standardized data for variiable $'i'$at sample $'j'$

$X_{j,i}$ = original measurement for variable ‘i’ at sample ‘j’

$\bar{X}_i =$ mean for variable $'i'$

$\sigma_i =$ standard deviaton for variabe $'i'$

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 (Cox et al., 1994):

City block distance: $$\delta_{rs} = \sum_i |x_{ri} - x_{si}| \tag{3.3}$$

Mahalanobis distance: $$\delta_{rs} = \left\{(x_r - x_s)^T \Sigma^{-1} (x_r - x_s)\right\}^{1/2} \tag{3.4}$$

Then, matrix **A** is developed which contains the squared dissimilarities. Matrix **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 (Borg and Groenen, 2005).

After that, the next step applies for the conversion of dissimilarity matrix to minor product moment which is shown in Chapter 2 from equation (2.7) until equation (2.10). Here, the step is continuing with the PCA algorithm. Finally, the PCA model can be simply developed by:

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

Where,

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

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

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. 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

$$\text{Residual Matrix,}\quad \tilde{\mathbf{E}} = \breve{\mathbf{X}} - \hat{\mathbf{X}}$$
$$= \breve{\mathbf{X}} - \mathbf{P}_a\mathbf{V}_a^{\mathrm{T}}$$
$$= \breve{\mathbf{X}} - \breve{\mathbf{X}}\mathbf{V}_a\mathbf{V}_a^{\mathbf{T}}$$
$$= \breve{\mathbf{X}}\left(\mathbf{I} - \mathbf{V}_a\mathbf{V}_a^{\mathbf{T}}\right) \qquad (3.10)$$

$$\text{SPE statistic,}\qquad Q_i = \mathbf{e}_i\mathbf{e}_i^{\mathrm{T}} \qquad (3.11)$$

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

$$\theta_1 = \sum_{i=A+1}^{N}\lambda_i \qquad (3.13)$$

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

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

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

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

3.2.2 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 4

RESULTS AND DISCUSSIONS

4.1 INTRODUCTION

This chapter will presents the results and discussions of the research which focuses on integration of process monitoring algorithms through development of conventional PCA-based MSPM and dissimilarity-based MSPM. Firstly, this chapter will describe about the case study in this research. Then, results of first phase which based on normal operating condition are displayed for both methods. Meanwhile, results for second phase is described about the fault detection for both PCA and dissimilarity techniques. Finally, the summary of all results is briefly explained.

4.2 CASE STUDY

The simulator of the Tennessee Eastman process (TEP) as shown in Figure 4.1 was used as the case study which were originally presented in Down and Vogel (1993). This particular plant has been proposed in many works as benchmark in evaluating the effectiveness of various schemes of controls, optimization techniques and monitoring applications.

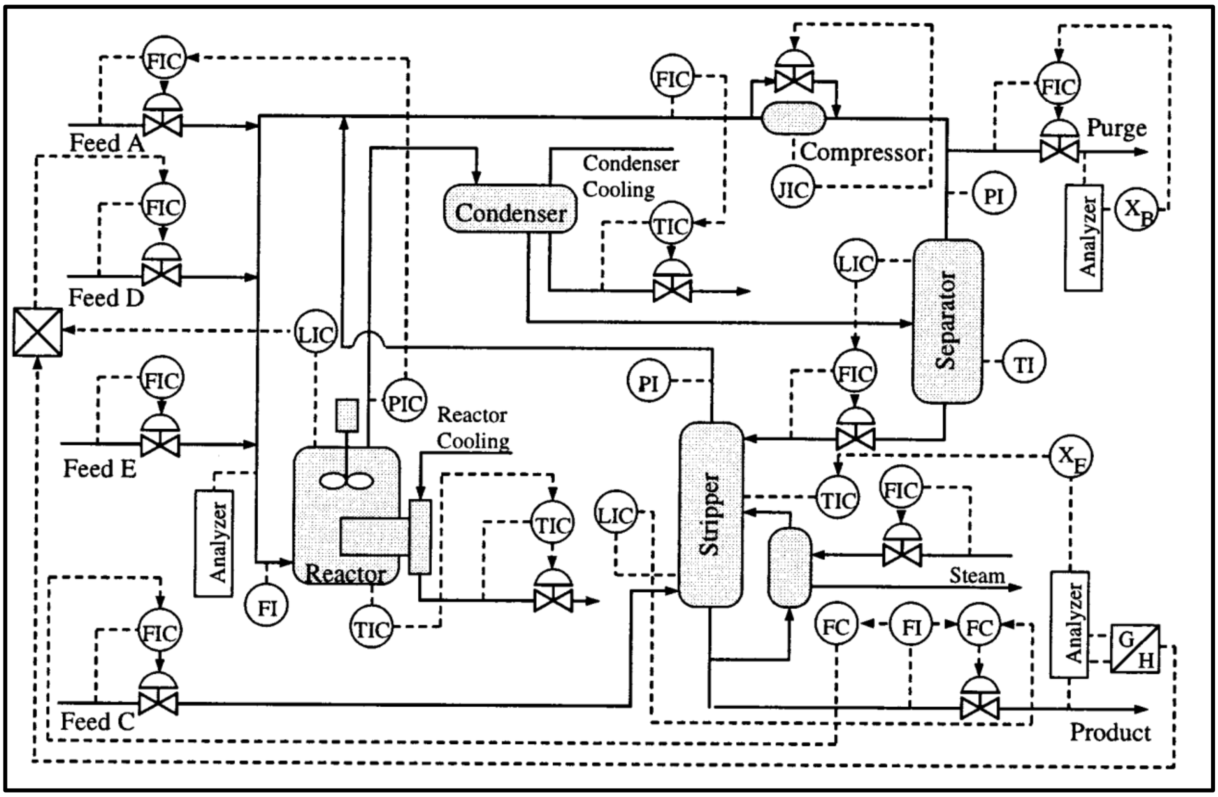

Figure 4.1: Decentralized control system of the Tennessee Eastman process

Source: Down and Vogel 1993

This process is well suited for a wide variety of studies including both plant-wide control and multivariable control problems. It consists of a reactor/separator/recycle arrangement involving two simultaneous gas-liquid exothermic reactions of the following form:

$$A(g) + C(g) + D(g) \longrightarrow G(liq) \qquad \text{(Product 1)}$$

$$A(g) + C(g) + E(g) \longrightarrow H(liq) \qquad \text{(Product 2)}$$

In the process, the gaseous reactants are fed to the reactor where they react to form liquid products. The gas phase reactions are catalysed by a non-volatile catalyst dissolved in the liquid phase. The reactor has an internal cooling bundle for removing the heat of reaction. The products leave the reactor as vapors along with the unreacted feeds while catalyst is remains in the reactor. Then, reactor product stream passes through a cooler for condensing the products and from there to a vapour-liquid separator. Non-condensed components are recycle back through a centrifugal compressor to the reactor feed. Condensed components will move to a product stripping column to remove remaining reactants. Product G and H exit the stripper base and separated in a downstream refining section meanwhile inert and by-product are purged from the system as a vapour from the vapour-liquid separator.

In general, the system is consisted of five major unit operations including a reactor, a product condenser, a vapour-liquid separator, a recycle compressor and finally product stripper. In particular, there are 21 manipulated and 41 measurement variables are monitored in the process. For this study, there are 20 types of abnormal operations considered as the fault cases as shown in Table 4.1.

Table 4.1: List of fault in the TEP system for process monitoring

Fault Cases	Fault causes	Type
1	A/C feed ratio, B composition constant	Step
2	B composition, A/C ratio constant	Step
3	D feed temperature	Step
4	Reactor cooling water inlet temperature	Step
5	Condenser cooling water inlet temperature	Step
6	A feed loss	Step
7	C header pressure loss	Step
8	A, B, C feed compositions	Random variation
9	D feed temperature	Random variation
10	C feed temperature	Random variation
11	Reactor cooling water inlet temperature	Random variation
12	Condenser cooling water inlet temperature	Random variation
13	Reaction kinetics	Slow drift
14	Reactor cooling water valve	Sticking
15	Condenser cooling water valve	Sticking
16	Unknown	-
17	Unknown	-
18	Unknown	-
19	Unknown	-
20	Unknown	-

4.3 OVERALL MONITORING PERFORMANCES

4.3.1 First Phase (Off-line Modelling and Monitoring)

A set of NOC data containing 500 samples was obtained from simulation. Firstly, the standardized NOC data is analysed through both methods to identify the number of PC's that is required in the process to reduce the dimensions of multivariate data as shown in Figure 4.2.

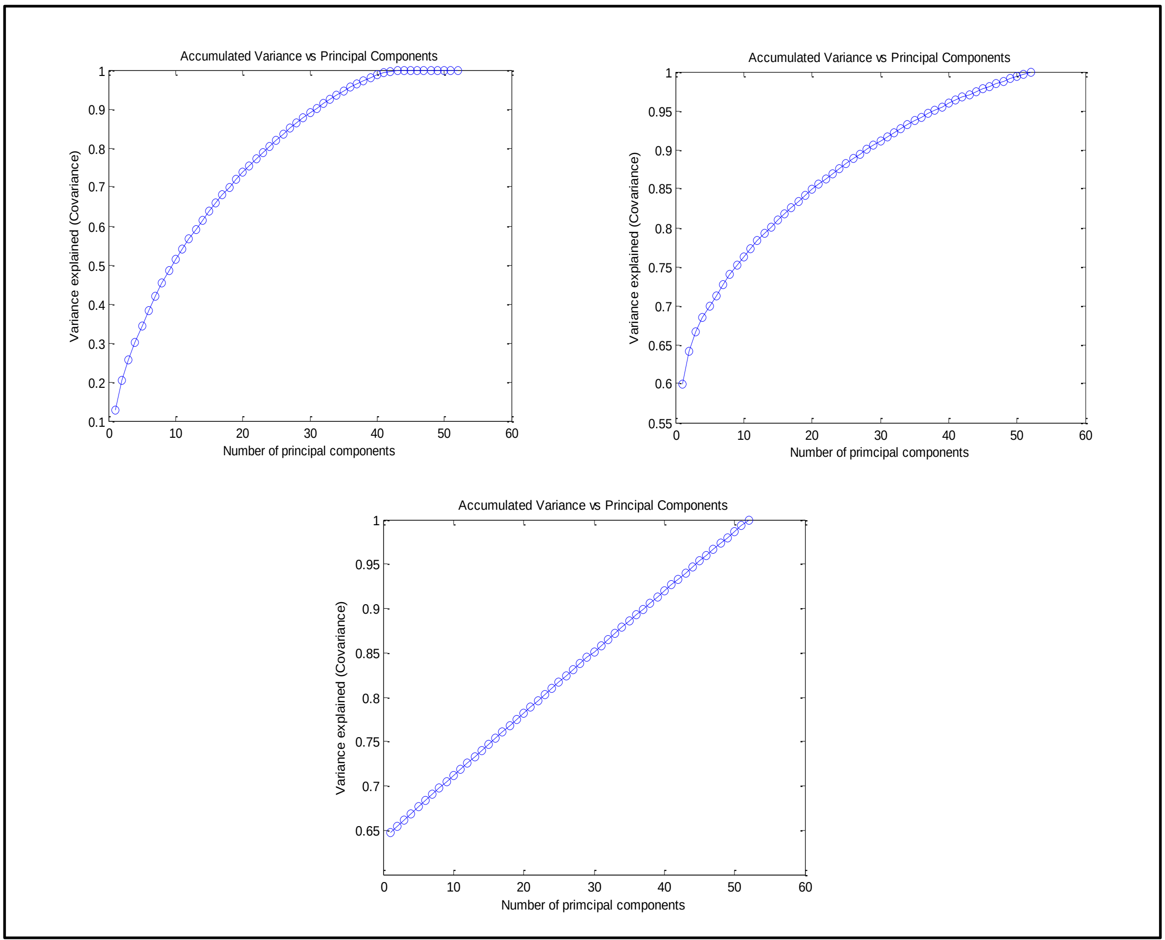

Figure 4.2: Accumulated data variance explained by different PCs for conventional PCA-based MSPM (left), dissimilarity-based MSPM of city block distance (right) and dissimilarity-based MSPM of mahalanobis distance (bottom)

When standardized NOC data is analysed through conventional PCA method, Figure 4.2 (left) shows that 24 PCs are required to explain 80% of total NOC data variances. Next, the standardized NOC data is analysed through new algorithm which is dissimilarity-based MSPM which is based on city block and mahalanobis distances. According to Figure 4.2 (right), for city block distances it shows that 14 PCs are required to explain 80% of total NOC data variances whereas for mahalanobis distance (bottom), it required 23 PCs which is more than city block distances. However, when compared to PCA-based MSPM method, new algorithm based on city block and mahalanobis distances are more efficient as it required less number of PCs to explained 80% of total data variances.

Next for the calculation of confidence region of scores, Hotelling's T^2 and Squared Prediction Errors (SPE) monitoring statistics are involved which displayed in Shewhart-type chart. Both control charts have 95% confidence limit for warning alarm and 99% confidence limit for action or control limit signal. Figure 4.3 will shows both of monitoring statistics charts for conventional PCA and dissimilarity methods.

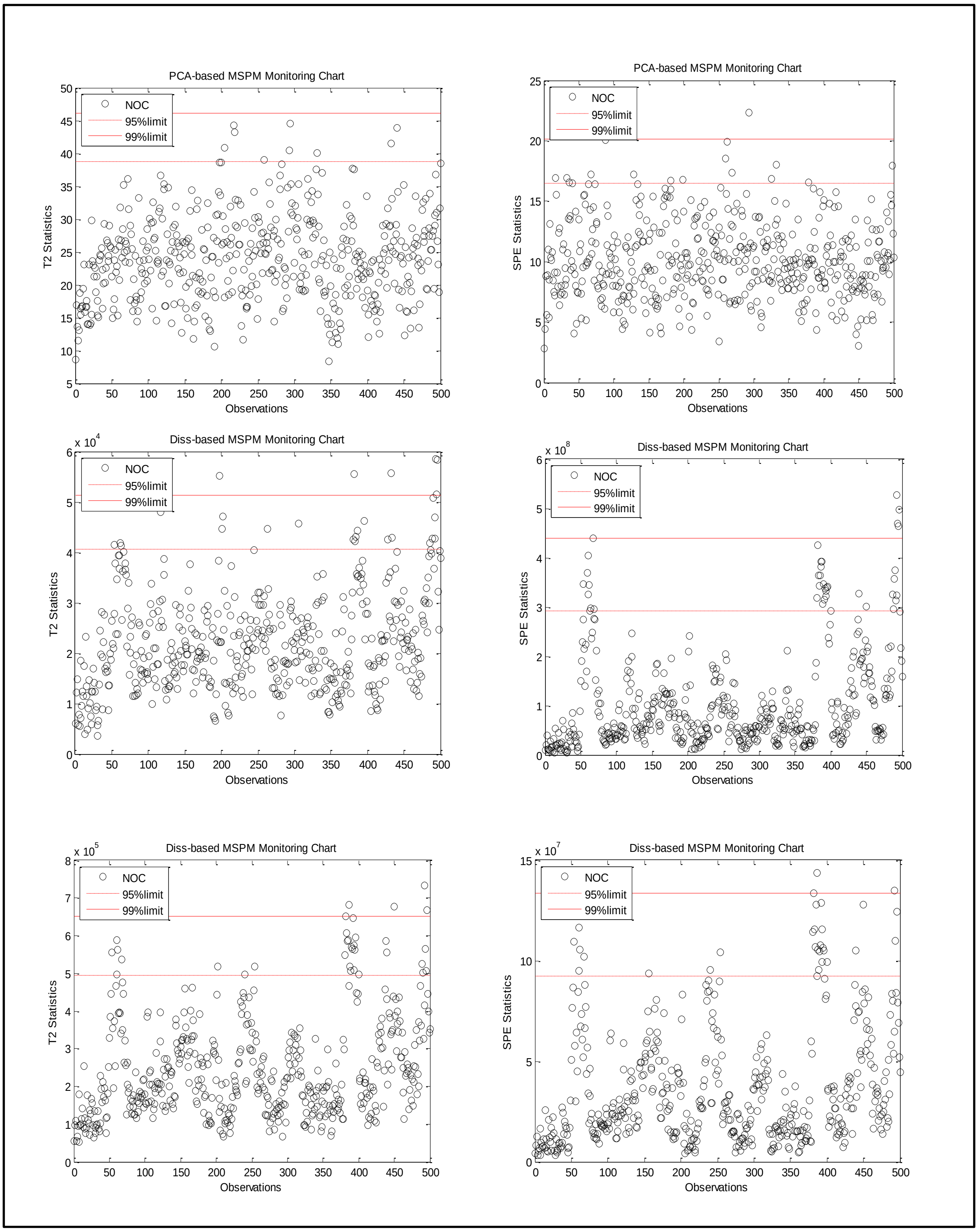

Figure 4.3: Hotelling's T^2 and SPE monitoring statistics chart with 95% and 99% confidence limits of NOC data : conventional PCA-based MSPM (top diagrams), dissimilarity-based MSPM of city block distance (middle diagrams) and dissimilarity-based MSPM of mahalanobis distance (bottom)

Based on Figure 4.3, for conventional PCA method there is only one observation that is out of 99% control limit in SPE monitoring statistics and all observations are in control limit for Hotelling's T^2 statistic. The observations for conventional PCA method are still within the control limit. For dissimilarity matrix based on city block distance, there are about five observations that are out of 99% confidence limit for both Hotelling's T^2 and SPE monitoring statistics charts. However, as the samples are not denoting 5 consecutives samples that out of control limit thus the observations are still within the limit. It is the same case for dissimilarity matrix based on mahalanobis distance where there are about five observations out of control limit in Hotelling's T^2 statistic charts meanwhile in SPE chart, there are only three observations that are out of 99% control limit. From the observations, it can conclude that both conventional PCA and dssimilarity matrix methods are capable and efficient in process monitoring performances.

4.3.2 Second Phase (On-line Monitoring)

A set of abnormal process contain 960 samples were applied to the conventional PCA and dissimilarity matrix. In each of fault case, the fault was introduced in sample 160 whereby the fault will be detected if 5 consecutive observations are located outside the 99% control limit. The set of abnormal data is reflects to only three faults appeared in TEP system according to it causes respectively which are:

i. Fault 1: A/C feed ratio with constant B composition
ii. Fault 2: B composition with constant A/C feed ratio
iii. Fault 3: Temperature of feed D

4.3.2.1 Monitoring Outcomes Based on Fault 1 (F1)

Table 4.2 shows the tabulated results obtained from process monitoring runs by using two different methods which are conventional PCA-based MSPM and dissimilarity-based MSPM algorithm. The result based on Fault 1 which occurred by the step change of A/C feed ratio with constant B composition.

Table 4.2: Fault detection time for F1

Fault	**Conventional PCA**			**Dissimilarity Matrix**					
				City Block			**Mahalanobis**		
	T^2	SPE	Final	T^2	SPE	Final	T^2	SPE	Final
F1	5	3	3	7	8	7	7	7	7

Based on tabulated results, it clearly seen that both conventional PCA and dissimilarity methods are able to detect all the fault appeared in the system. It shows that fault detection time for conventional PCA is faster than dissimilarity-based MSPM method where the first detection occurred at sample 163. Meanwhile, both city block and mahalanobis distances in dissimilarity algorithm have same fault detection time where first fault detected at sample 167. Then, monitoring performances for F1 are illustrated through T^2 and SPE monitoring statistics charts as shown in Figure 4.3.

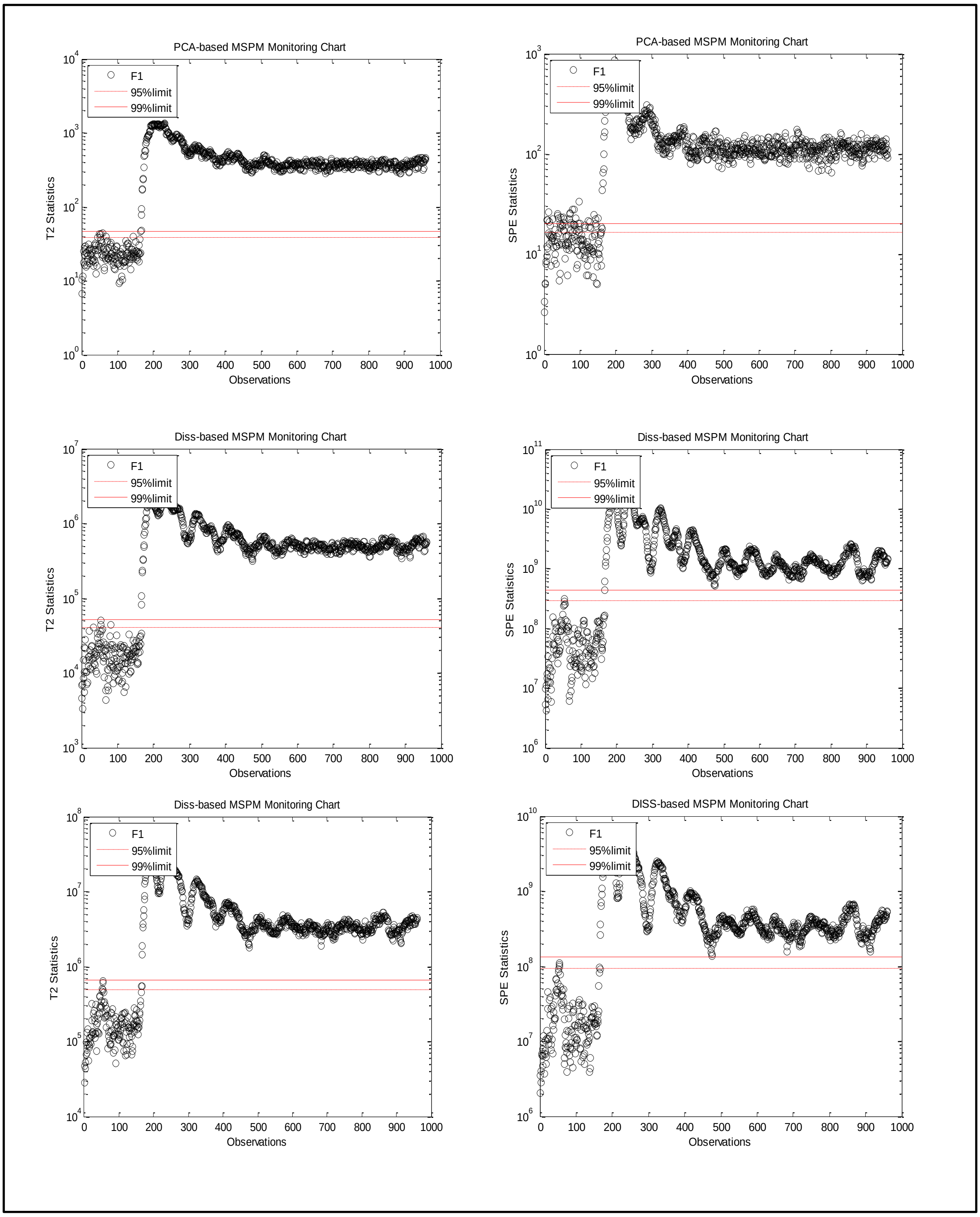

Figure 4.4: Hotelling's T^2 and SPE monitoring statistics chart plotted together with the 95% and 99% confidence limits of F1: conventional PCA-based MSPM (top diagrams), dissimilarity-based MSPM of city block distance (middle diagrams) and dissimilarity-based MSPM of mahalanobis distance (bottom)

According to Figure 4.4, most of the observations in the Hotelling's T^2 and SPE statistics for F1 by using both conventional PCA and dissimilarity matris methods are located far beyond the 99% control limits after sample 160. In comparing both charts, SPE statistics has demonstrated effective and consistent performance against T^2 as most of samples located beyond the 99% control limit starting at sample 163 faster than T^2 statistic particularly for conventional PCA method. However, dissimilarity matrix based on city block method shown that T^2 statistic is slightly faster than SPE statistic meanwhile monitoring performances using dissimilarity matrix based on mahalanobis distance obtained same fault detection time for both monitoring statistic which is at sample 167. For overall monitoring performances fo F1, eventhough dissimilarity matrix method slightly slower in fault detection time compared to convetional PCA, both methods are comparable as both methods can detect the fault efficiently.

4.3.2.2 Monitoring Outcomes Based on Fault 2 (F2)

In TEP system, Fault 2 is occurred because of change in B composition with constant A/C feed ratio. Conventional PCA and dissimilarity algorithm are used to detect the fault in the process which based on MSPM system. Results according to fault detection time for both methods are shown in Table 4.3.

Table 4.3: Fault detection time for F2

Fault	**Conventional PCA**			**Dissimilarity Matrix**					
				City Block			**Mahalanobis**		
	T^2	SPE	Final	T^2	SPE	Final	T^2	SPE	Final
F2	17	11	11	15	23	15	15	23	15

According to Table 4.3, fault detection time of conventional PCA is at sample 171 which is faster than fault detection time performed by dissimilarity matrix based on city block and mahalanobis distances respectively. In dissimilarity matrix method, both city block and mahalanobis distances are able to detect fault at the same time which is at sample 175. Even though both methods in dissimilarity matrix have found a number of samples located outside 99% control limit much earlier but none of the samples denoting at least 5 consecutives fault sample thus it not considered as fault detection time.

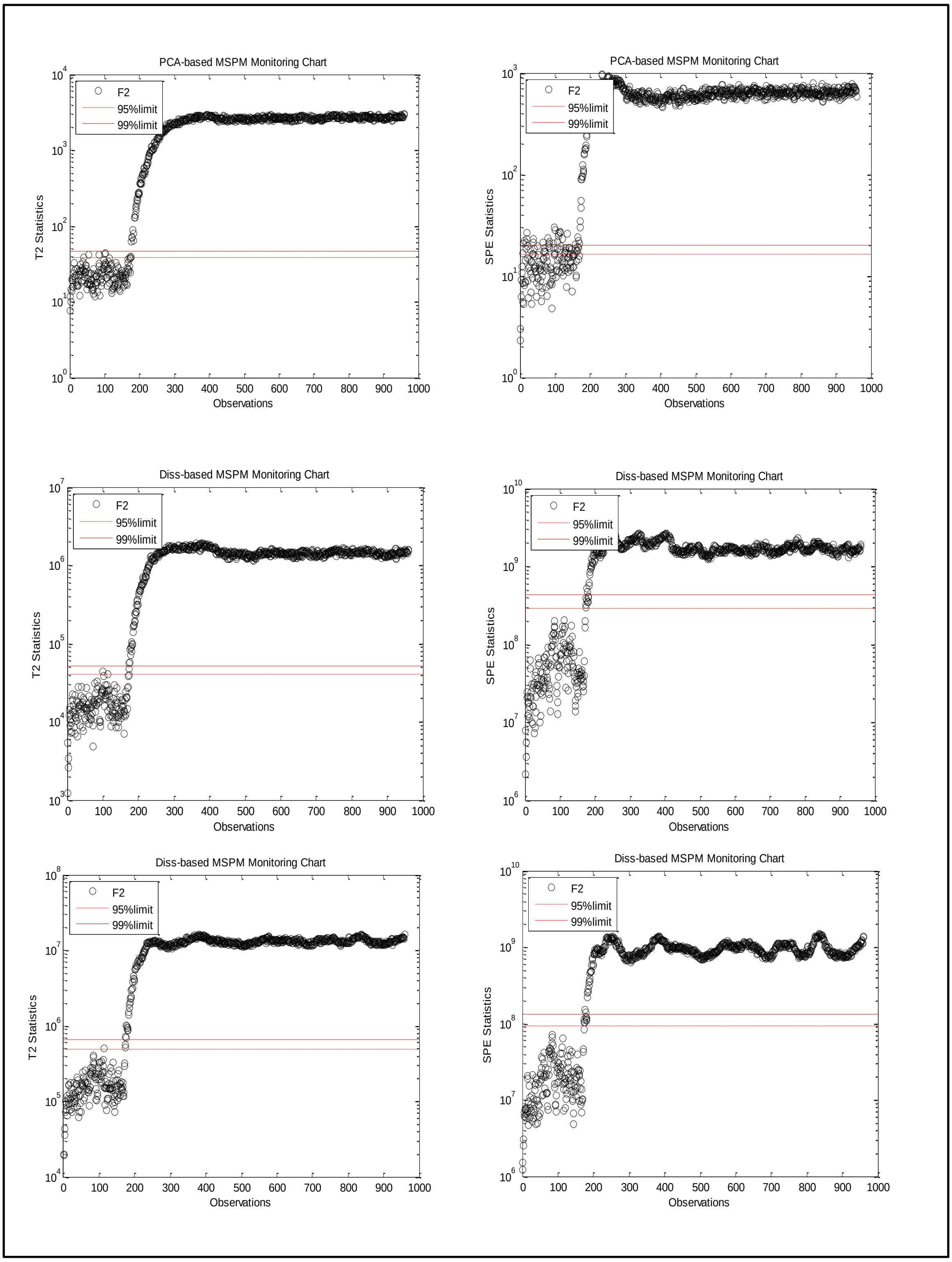

Figure 4.5: Hotelling's T^2 and SPE monitoring statistics chart plotted together with the 95% and 99% confidence limits of F2: conventional PCA-based MSPM (top diagrams), dissimilarity-based MSPM of city block distance (middle diagrams) and dissimilarity-based MSPM of mahalanobis distance (bottom)

Figure 4.5 illustrated the monitoring performances for both methods applied through the Hotelling's T^2 and SPE monitoring statistic charts. By comparing both charts, T^2 monitoring statistic is more effective in fault detection compared to SPE monitoring statistic when applied with dissimilarity-based MSPM method. Through T^2 statistic, the first detection can be noticed at sample 175 faster than SPE statistic which noticed at sample 183. However, monitoring performance using conventional PCA shown that SPE monitoring statistic is more efficient against T^2 statistic. Thus, it can be concludes that both monitoring statistic are comparable as both statistics are efficient to detect the fault in the system.

4.3.2.3 Monitoring Outcomes Based on Fault 3 (F3)

Table 4.4 shows the tabulated results obtained from process monitoring runs by using two different methods which are conventional PCA-based MSPM and dissimilarity-based MSPM algorithm according to the fault detection time. The result based on Fault 3 which happened due to temperature of feed D.

Table 4.4: Fault detection time for F3

Fault	**Conventional PCA**			**Dissimilarity Matrix**					
				City Block			**Mahalanobis**		
	T^2	SPE	Final	T^2	SPE	Final	T^2	SPE	Final
F3	0	51	51	43	12	12	15	7	7

Based on tabulated results, it clearly seen that both conventional PCA and dissimilarity methods are able to detect all the fault appeared in the system. It shows that fault detection time for dissimilarity matrix based on mahalanobis distance is faster than dissimilarity matrix based on city block and conventional PCA where the first detection can be noticed at sample 167. Meanwhile, dissimilarity matrix based on city block and conventional PCA are able to detect fault at sample 172 and 211 respectively. Then, monitoring performances for F3 is illustrated in Figure 4.5 for both T^2 and SPE monitoring statistics charts.

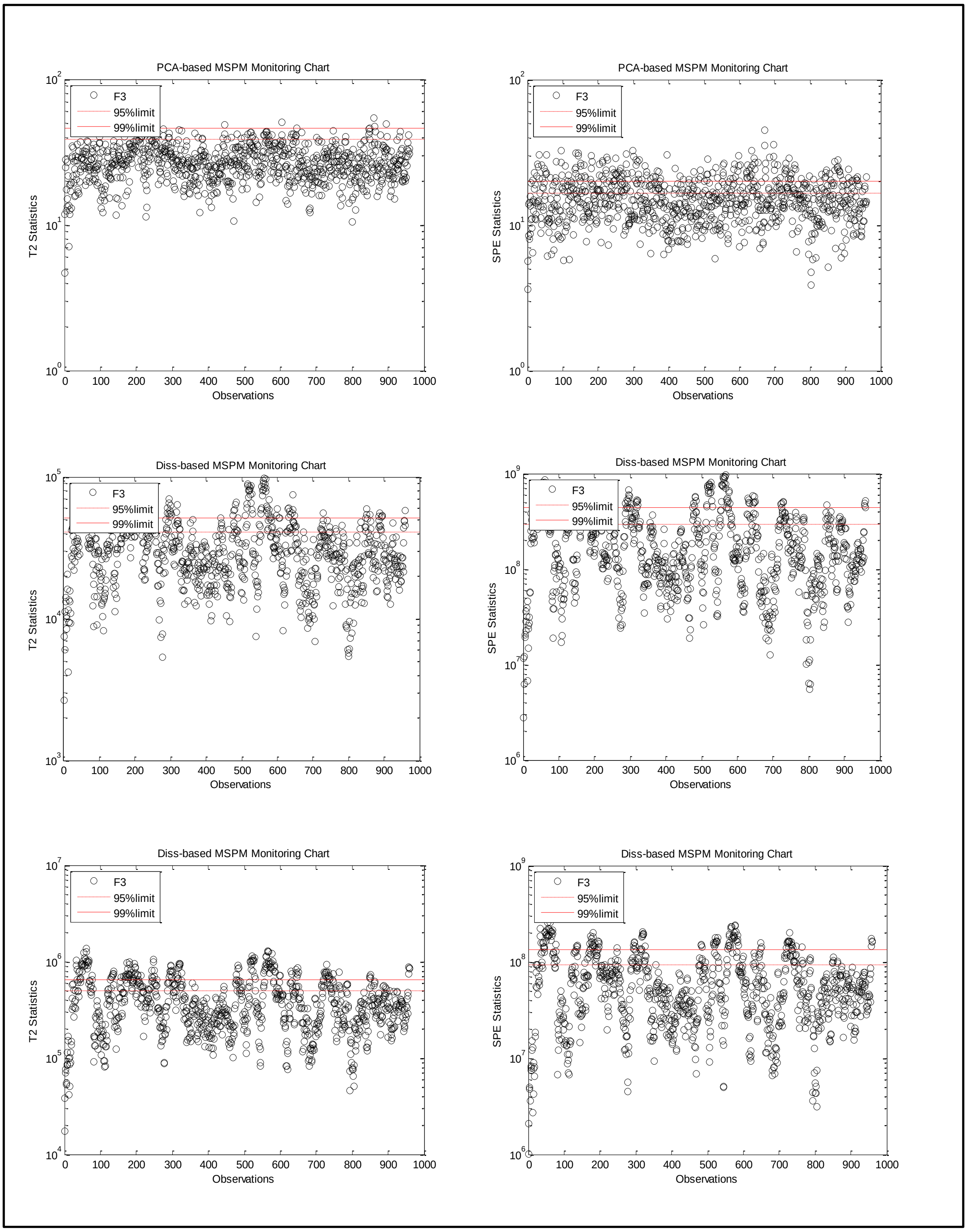

Figure 4.6: Hotelling's T^2 and SPE monitoring statistics chart plotted together with the 95% and 99% confidence limits of F3: conventional PCA-based MSPM (top diagrams), dissimilarity-based MSPM of city block distance (middle diagrams) and dissimilarity-based MSPM of mahalanobis distance (bottom)

According to Figure 4.6, Hotelling's T^2 and SPE monitoring statistics are used to display monitoring performances for both conventional PCA and dissimilarity algorithm methods for Fault 3. Based on overall performances of monitoring statistics, SPE statistics is more effective and consistent in fault detection against Hotelling's T^2 statistic especially by using dissimilarity matrix based on mahalanobis distance method. Hotelling's T^2 statistic has very slow performance in fault detection for both conventional PCA and dissimilarity matrix methods. When referred to Hotelling's T^2 chart for conventional PCA majority of samples are below 99% control limit. Even though there is sample that out of limit noticed at sample 161, but it is less than 5 consecutives samples thus there is no fault detection in the performance for Hotelling's T^2 statistic. From the result, it shown that dissimilarity matrix method is more efficient compare to conventional PCA method in fault detection for process monitoring.

4.4 SUMMARY

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

CHAPTER 5

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 is able to achieve its aim through all objectives stated 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 is able to detect the fault as efficient as the conventional technique. Moreover, dissimilarity-based MSPM system is more effective as it is able to find small number of PCs with high percentage of total variances transformed compared to the conventional PCA system especially for dissimilarity matrix based on city block distance. Simultaneously, it 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 Tennessee Eastman process. 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 three 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 of the system can be enhanced through the recommendations stated earlier.

REFERENCES

Abdi, H, (n.d). Partial Least Squares (PLS) Regression. The University of Texas at Dallas.

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.

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

Chiang, L.H., Russel and E.L. Braatz, (2001). Fault Detection and Diagnosis in Industrial Systems. Springer Verlag, Great Britain.

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.

Cox, et. al. (1994). Multidimensional Scaling. London, Great Britain: Chapman & Hall.

Downs, J.J. and E.F. Vogel, (1993). A Plant-wide industrial process control problem. Comput Chem. Engin., 17: 245-255.

Faezah, S. and Athena, A. (n.d). Principal Component Analysis (PCA).

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

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.

Kruger, U., Zhang, J., and Xie, L. (n.d). Developments and Applications of Nonlinear Principal Component Analysis.

Kourti, T., Nomikos, P. and MacGregor, J. F. (1995). Analysis, monitoring and fault diagnosis of batch processes using multiblock and multiway PLS. *Journal Of Process Control*, 5 277-284.

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

Manabu, K. et al (2000). Dissimilarity Of Process Data For Statistical Process Monitoring. Proceeding of IFAC Symposium on Advanced Control of Chemical Process, 1, p231-236.

Manabu, K., Shinji, H., & Iori, H. (2000). Contribution Plots for Fault Identification Based on the Dissimilarity of Process Data, 12-17.

Mason, R.L. and Young, J.C. (2002). Multivariate Statistical process Control with Industrial Applications. ASA_SIAM., USA.

Nomikos, P., and MacGregor, J. F., (1995). Monitoring of Batch Processes Using Multi- Way Principal Component Analysis. AIChE Journal, 40 (8), 1361 – 1375.

Raich, A. and Cinar, A. (1996). Statistical process monitoring and disturbances diagnosis in multivariate continuous processes. AIChE J., 42: 995-1009.

Sheshu, K. D. (2011). Multivariate Statistical Process Monotoring and Control. Chemical Engineering Department National Institute of Technology Rourkela.

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.

Sliskovic, D., Grbic, R. and Hocenski, Z.(2012). Multivariate Statistical Process Monitoring, 19, p33-41.

Spring (2010). Principal Components Analysis, 36-490.

Tan, S. and Mavrovouniotis, M. L. (1995). Reducing Data Dimensionality through Optimizing Neural Network Inputs. Am Inst Chemical Eng J, 41, pp 1471-1480.

Yingwei, Z. and Yang, Z. (2010). Process Monitoring, Fault Diagnosis and Quality Prediction Methods Based on Multivariate Statistical Techniques.

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

Yunus, M.Y.M, and Zhang, J. (2010c). Multivariate Statistical Process Monitoring using Multidimensional Scalling. 19th International Congress of chemical and Process Engineering , Prague, Checkolosvakia.

Yunus, M.Y.M. (2012). Multivariate Statistical Process Monitoring using Classical Multidimensional Scaling, 1-27.

Zhou, Y. (2010). Principal component analysis based image fusion routine with application to stamping split detection (Doctoral dissertation). Graduate School of Clemson University.