DEVELOPMENT OF PCA-BASED FAULT DETECTION SYSTEM BASED ON VARIOUS MODES OF NOC MODELS FOR CONTINUOUS-BASED PROCESS

NURUL FADHILAH BINTI ROSLAN

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

Faculty of Chemical Engineering and Natural Resources
UNIVERSITI MALAYSIA PAHANG

JANUARY 2013

SUPERVISOR'S DECLARATION

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

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

STUDENT'S DECLARATION

I declare that this thesis entitled "Development of PCA-based Fault Detection System Based on Various of NOC Models for Continuous-based Process" is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree"

Signature: …………………………………………......

Name: Nurul Fadhilah Binti Roslan

ID Number: KA09157

Date:

To my father and beloved mother

ACKNOWLEDGEMENT

In the name of Allah S.W.T the Most Beneficent and the Most Merciful. The deepest sense of gratitude to the Almighty for the strength and ability to complete this project. Infinite thanks I brace upon Him.

The special thank also goes to my helpful supervisor, Dr. Mohd Yusri Bin Mohd Yunus. The supervision and support that he gave truly help the progression and smoothness of the project. The co-operation for encouragement, guidance, morale support,and critics in bringing this project fruition is much indeed appreciated. Without him outstanding support and interest, this thesis would not been at the best it would right now.

Not forget, great appreciation goes to the rest of parents and family that helps us from time to time during the project. The whole program really brought us together to appreciate the true value of friendship and respect of each other.

My grateful thanks also go to all my friends. A big contribution and hard worked from all of you during this semester is very great indeed. All projects and assignment during this semester would be nothing without the teamwork and hard work every one.Besides, this project makes we realized the value of working together as a team and as a new experience, which challenges us every minute.

Great deals appreciated go to the contribution of my faculty - Faculty of Chemical Engineering (FKKSA). I am also would like to thankful the Coordinator of FKKSA – Mrs. Noraziah that patient in helping us complete this project.

Last but not least I would like to thank my friends especially those who work has involved neither directly nor indirectly in succession of this thesis. Indeed I could never adequately express my indebtedness to all of them.

ABSTRACT

Multivariate statistical techniques are used to develop detection methodology for abnormal process behavior and diagnosis of disturbance which causing poor process performance (Raich and Cinar, 2004). Hence, this study is about the development of principal component analysis (PCA) -based fault detection system based on various modes of normal operating condition (NOC) models for continuous-based process. Detecting out-of-control status and diagnosing disturbances leading to the abnormal process operation early are crucial in minimizing product quality variations (Raich and Cinar,2004). The scope of the proposed study is to run traditionally multivariate statistical process monitoring (MSPM) by defining mode difference in variance for continuous-based process. The methodology use to identify and detection of fault which undergo two phase which phase I is off-line monitoring while phase II is on-line monitoring. As a result, it will be analyze and compared of the implementing traditional PCA of Single NOC modes and Multiple NOC modes. Particularly, this study is critically concerned more on the performance during the fault detection operations comprising both off-line and on-line applications, hence it will analyze until fault detection and comparing between two modes of NOC data.

ABSTRAK

Multivariat teknik statistik yang digunakan untuk membangunkan kaedah pengesanan proses untuk tingkah laku yang tidak normal dan diagnosis gangguan yang menyebabkan prestasi proses miskin (Raich dan Cinar, 2004). Oleh itu, kajian ini adalah mengenai pembangunan analisis komponen utama (PCA) berasaskan kesalahan sistem pengesanan berdasarkan pelbagai mod keadaan operasi normal (NOC) model untuk proses yang berterusan berasaskan. Mengesan status out-of-kawalan dan mendiagnosis gangguan yang membawa kepada operasi proses abnormal awal adalah penting dalam mengurangkan variasi kualiti produk (Raich dan Cinar, 2004). Skop kajian yang dicadangkan adalah untuk menjalankan pemantauan tradisional multivariat proses berstatistik (MSPM) dengan menentukan perbezaan mod dalam varians proses yang berterusan berasaskan. Metodologi yang digunakan untuk mengenal pasti dan pengesanan kesalahan yang menjalani dua fasa fasa yang saya off-line pemantauan manakala fasa II adalah on-line pemantauan. Hasilnya, ia akan menganalisis dan berbanding PCA pelaksana tradisional mod Single NOC dan Pelbagai mod NOC. Terutama sekali, kajian ini secara kritikal berkenaan lanjut mengenai prestasi semasa operasi pengesanan kesalahan yang terdiri daripada kedua-dua aplikasi off-line dan on-line, maka ia akan menganalisis sehingga pengesanan kerosakan dan membandingkan antara dua mod data NOC.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDIX

CHAPTER 1

INTRODUCTION

1.1 Background of Proposed Study

Statistical process control (SPC) is the basic performance of monitor and detection of abnormal process (Zhao et al., 2004). According to MacGregor and Kourti (1995) the main objective of SPC is to monitor the process performance over time in order to verify the status of the process whether it is remaining in a “state of statistical control” or not. However, most SPC methods are based on charting only a small number of variables and examining them one at time (MacGregor and Kourti, 1995). As a result, multivariate statistical process control (MSPC) has been proposed especially to monitor multivariable process (Kumar and Madhusree, 2001; Kano et al., 2002; Zhao et at., 2004; MacGregor et al., 1995; Maestri et al. 1995). According to Kourti et al. 1995, multivariate method can treat and extract information simultaneously on the directionality of the process variation. Jackson and Mudholkar (1979) investigated principal component analysis (PCA) as a tool of MSPC and introduce a residual analysis. Typically, the Shewhart-type control chart is applied, for depicting the progression of two different types of monitoring statistics, namely as T2 and Q statistic. The T^2 statistics is a measure of the variation within PCA model while Q statistic is a measure of the amount of variation not capture by the PCA modes. When PC's is being scaling by the reciprocal of its variance, it will compute same role as T^2 irrespective of the amount of variance it's explain in the Y matrix, which Y is matrix of mean centered and scaled measurements. T^2 is not sufficient for first PC because it only detect whether the variation in the quality variables in the plane or not. Kresta et al., (1991) say new event can be detected by computing the squared prediction error (SPE) or also known as Q statistics. According to Jackson, (1991) and Nomikos and MacGregor (1995) Q statistics represents the square perpendicular distance of a new multivariate observation from the plane. Q statistics also represent unstructured fluctuation that cannot be accounted for by the model when the process is "in control". Hence it will be more effective multivariate control chart when T^2 chart on dominant orthogonal PC's plus a SPE chart.

1.2 Problem Statement

In order to ensure the successfulness of any operation, it is important to detect process upsets, equipment malfunctions or other special events as early as possible and then to diagnose and remove the factors that cause those events. However, Zhao et al., (2004) mentioned that a process which is having multiple operating modes tends trigger continuous warning signal even when the process itself is operating under another steady-state. In other word, the comprehensive mode is to sensitive as it will show the false alarm although the process are normal. Hence, MSPC is the only method, of which, the data is treated simultaneously into a single monitoring by way of reducing the dimensionality of the data observed without losing any of important information.

1.3 Research Objectives

The main purpose of this research is to study the impact of applying various modes of normal operating condition (NOC) in terms of the number of samples and variable variations on the process monitoring performance for continuous-based process. Therefore, the main objectives of this research are:

i. To develop the conventional MSPM method based on a single NOC
ii. To implement the conventional MSPM method based on different modes of NOC.
iii. To analyze the monitoring performance between system (i) and (ii).

1.4 Research Question

i. What is the main impact of reducing the number of samples as well as variations on the monitoring performance?
ii. What are the criteria should be used in selecting the NOC model?

1.5 Scopes of Study

Scope of propose study are on the development of PCA-based fault detection system based on various modes of NOC models for continuous-based process. There are three main scope will be investigated using MATLAB.

i. The conventional MSPM method will be develop based on single NOC mode. The linear PCA algorithm is used for reducing the multivariate data dimensions.

ii. The MSPM will be run traditionally by implementing different mode, which in this research is on two modes. According to Zhao et al. (2004),in spite of the success of applying PCA based MSPM tools to process data for detecting abnormal situations, when these tools are applied to a process with multiple operating modes, many missing and false alarms appear even when the process itself under other steady-state nominal operating conditions.

iii. As all data have been obtained, it will be analyze further with two multivariate control charts namely Hotelling’s T2 and Squared Prediction Errors (SPE) statistic for the fault detection operation.

1.6 Contributions

i. A new set of criteria is proposed for selecting the optimized NOC data for monitoring.

ii. As a result of (i), the monitoring performance can be enhanced in terms of missing and false alarm.

1.7 Organization of This Report

The new monitoring algorithm has been proposed in this study by developing PCA-based fault detection system based on various modes of NOC models for continuous-based process. Hence, this report is divided into five main chapters. The first chapter discusses the background of the works which includes the problem statement, objectives, scopes and contributions**. Chapter II** which is literature review describes the fundamental of MSPC and justification of applying PCA in MSPM frameworks. **Chapter III** explains the research methodology of this study. **Chapter IV** presents some of the preliminary results. Conclusions and further research works are given in **Chapter V**.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

According to Venkatasubramaniam, Rengaswamy, Kavuri and Yin (2003) MSPM tools are data driven technique that generally reduce the dimension of process data and extract key features and trends that are of interest to plant personnel. MSPM tools used to reduces dimensions of process data, like PCA and subsequent refinements, which have show great success. In chapter 2, we will discuss on the fundamental or theory of process monitoring on MSPM using PCA tools, process monitoring issues and extension and justification of applying PCA in MSPM frameworks. Lastly, a summary is given at the end of this chapter.

2.2 Fundamentals / Theory of Process Monitoring on MSPM Using PCA Tools

Reformation and upgrading of conventional Statistical Process Control (SPC) method has produce MSPC. MSPC tools such as principal component analysis (PCA) were used to reduce the explaining dimension of the process data. Maestri et al. say this method has show great success and particularly suited to data set comprising correlated and collinear variables. Ge and Song (2008) define process data as different group based, for instance, on variation in the operating capacity, seasonal variations or changes in the feedstock characteristics and also on modifications in the operation strategies. From a geometric point of view, whenever such as a change occurs, the process data tend to group into a new cluster in a different location in the high dimensional space containing the process normal operating region. However when the data is considered belong to a unique normal operating region, the volume of this region becomes incorrectly large. Zhao et al, (2006) say this region will lead to an increasing number of missing and false alarm. According to Zhao et al, (2004) when PCA based MSPC tools applied to a process with multiple operating modes, many missing and false alarm can appear even when the process itself is operating under other steady-state nominal operating conditions. Particularly this technique is for reducing the number of dimensions used from the original data as well as projected them into a number of uncorrelated variables, by means of forming the appropriate linear combinations of the original variables. Hence, MSPC is the only method where the data is treated simultaneously by way of reducing the dimensionality of the data observed without losing any of important information. In addition, this method can reduce the burden of constructing a large amount of single-variable control charts and enable detecting events that are impossible or difficult to detect from the single-variable control charts (Phatak, 1999).

According to Venkatasubramaniam et al, (2003) multivariate statistical techniques are powerful tool that capable to compressing data and reducing its dimensionality. Hence the essential information is retained and easy to analyze than the original huge data set. Moreover, it is able to handle noise and correlation to extract true information effectively. Initially, PCA method is proposed by Pearson (1901) later, it been develop by Hotelling (1947). This is a standard multivariate technique which has been including in many textbooks (Jackson, 1991; Anderson, 1984) and research paper (Wold, Esbensen and Geladi, 1987; Wold, 1978). Venkatasubramaniam et al, (2003) say PCA is based on orthogonal decomposition of the covariance matrix of the process variables along directions that explain the maximum variation of the data. Yu and Zhang say this method involved a mathematical procedure that transforms a number of correlated variables into a smaller number of uncorrelated variables, which are called principal component.

2.3 Extensions of Principal Component Analysis

There are many extension of Principle Component Analysis (PCA) which is some of these is Kernel of PCA, Multiway-PCA, , Three Modes PCA and many more.

2.3.1 Kernel of PCA

Some extension of PCA is nonlinear principle components (NLPCA) or also Kernel PCA (KPCA). According to Vidal, Ma, and Sastry, (2005) KPCA is method of identifying a nonlinear manifold from sample points. NLPCA is a standard solution based on embedding the first data into a higher space, then applying PCA. As a result it will give large dimension space, so the eigen value is being decomposition or also known as kernel matrix.

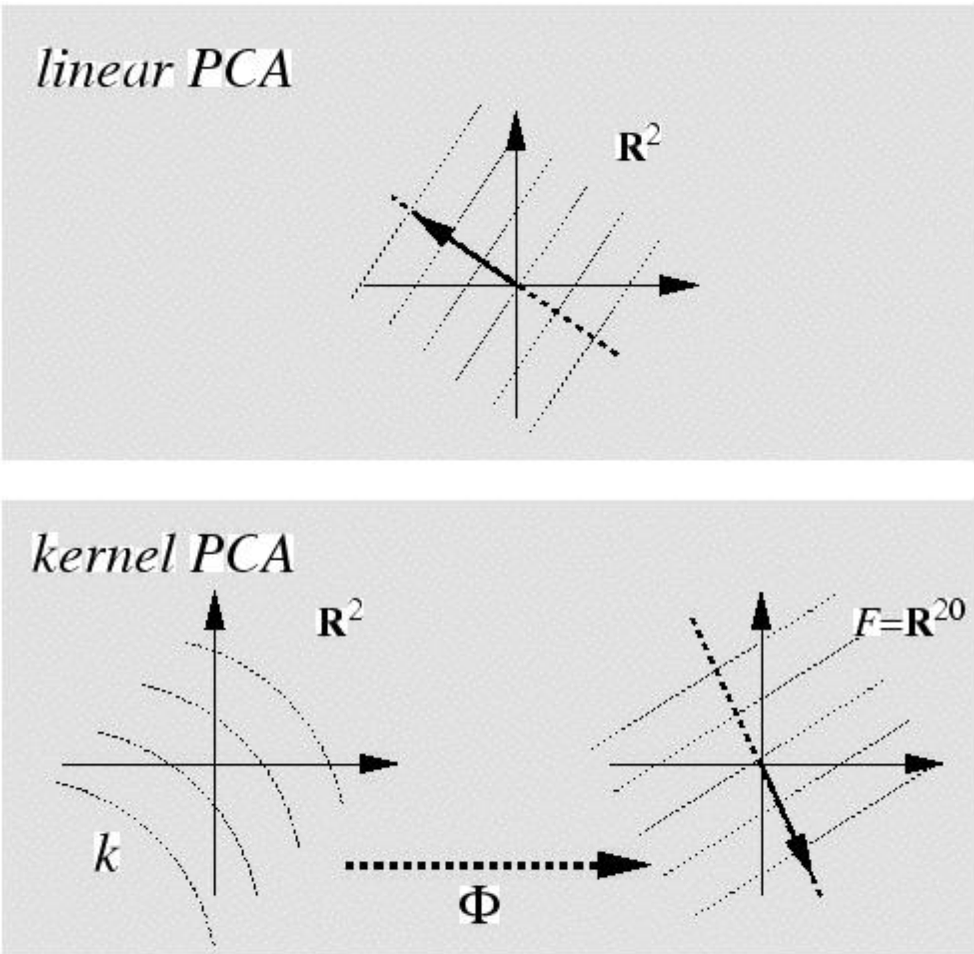


Figure 2.1 Linear PCA and Kernel PCA

From Figure 2.1 above, it show the basic idea of kernel PCA. By using a nonlinear function *k* instead of the standard d dot product, we implicitly perform PCA in a possibly high dimensional space F which is nonlinearly related to input space. The dotted lines are contour lines of constant feature value. Suppose that the number of observations *m* exceeds the input dimensionality *n*. In linear PCA, most samples are nonzero eigen values (Welling, nd). While for Kernel PCA variable will be nonzero eigen values. Thus, this is not necessarily a dimensionality reduction (Scholkopf, Smola and Muller, 2001). Furthermore, it may not be possible to find an exact preimage in input space of a reconstructed pattern based on a few of the eigenvectors. One of the disadvantages of KPCA is that, in practice, it is difficult to determine which kernel function to use because the choice of the kernel naturally depends on the nonlinear structure of the manifold to be identified (Vidal, Ma, and Sastry, 2005). In fact, learning kernels is an active topic of research in machine learning.

2.3.2 Multi-way-PCA

A monitoring approach using a multivariate statistical modelling technique namely multi-way principle component analysis is a method that overcome the assumption that the system is at steady state and it's provide a real time monitoring approach for continuous processes (Chen and McAvoy,1998). Recently MacGregor and Nomikos (1992) and Nomikos and MacGregor (1994) employed multiway PCA (MPCA) to extend multivariate SPC methods to batch processes. This multi-way PCA model can detect fault in advance compare to other monitoring approaches as it will analyzing a historical reference distribution of the measurement trajectories from past successful batches (Nomikos and MacGregor, 1995). Besides Nomikos et al. also say that the latent-vector space is reducing as the variation in the trajectories is characterized.

This make multi-way PCA is a useful procedure because each dynamic response signature is highly auto-correlated. Gallagher, Wise and Stewart (1996) say the correlation at different times within each signature, hence there is a high degree of correlation between signatures. Wold et al.(1987) has discuss that multi-way PCA will allows the multivariate data to be described in far fewer components than original variables. The multi-way PCA procedure can be described as follows. The data from a historical database of batch runs are organized in a three-way array X ($I \times J \times K$). The batch runs (I) are organized along the vertical axis, the measurement variables (J) along the horizontal axis, and their time evolution (K) occupies the third dimension. Usually, the minimum duration of the batch process defines the time length of a batch (K) and the data are synchronized based on a trigger variable whose change indicates the beginning of the batch. Nomikos et al. (1996) say multi-way PCA will give a great result as more information related with analysis is provided such as quantities from mass or energy balances, properties related to quality, and degradation rates. Hence, X is decomposed into scores vectors t and loadings vectors p using traditional principal components analysis (PCA) (Jackson and Mudholkar, 1979, Wold, 1987).

The p-loading matrices, which define the reduced space upon the actual data are projected and summarize the time variation of the measurement variables around the average trajectories. The elements are the weights applied to the observations of a particular batch to give the t-scores for this batch which each element of a t-vector corresponds to a single batch and represent the projection of this batch onto the reduced space. Finally, the sum of squared residuals for a given batch represents the squared distance of this batch perpendicular to the reduced space. A small number (R) of principal components usually 3 to 5 can express most of the variability in the batch data since the measurement variable are highly cross-correlated with one another and highly auto-correlated over time (Nomikos et al.,1996).

A process abnormality will result in poor quality product, hence multi-way PCA will help to detect and classify the cases. This is because multi-way PCA is an easily interpret tool which characterized batches based on their process operation. Then it is up to the engineers to remove the root cause and eliminate any future appearances of this fault. In some cases, MPCA might detect an abnormal behavior which may not have an immediate impact on quality, but may constitute an alarm for an incipient equipment failure such as an agitator or sensor deterioration. In these cases, one will have the opportunity to correct such process deteriorations which otherwise could lead to permanent malfunctions (Nomikos et al., 1996; Gallagher et al., 1996;Chen et al. 1998).

2.3.3 Three-Mode PCA

Tucker (1963) was first formulated the three-mode model principal component analysis or also known as Tucker3 model and it subsequently extended in articles by Tucker (1964, 1966) and Levin (1963). Kroonenberg and Leeuw say the articles review on the mathematical description and programming aspects of the model. In term of multidimensional scaling references to the mode l occur frequently (Harshman, 1970; Jennrich, 1972; Carroll & Chang, 1972; Takane , Young & de Leeuw, 1977), hence the Tucker3 mode l is the general mode l comprising various individual differences models. Tucker (1972), Carroll & Wish (1974), and Takane , Young & de Leeuw (1977) has discuss more on the relationships between multidimensional scaling and three-mode PCA. In article by Tucker (1966) remarks that the procedures "do not produce a least squares approximation to the data. Investigations of the mathematics of a least squares fit for three-mode factor analysis indicates a need for an involved series of successive approximations. "The procedures described in the sequel are designed to provide least squares estimates of the parameters in the three -mode model. The alternating least squares approach used can also be extended to accommodate other levels of measurement, as has been recently demonstrated by Sands & Young (I980) for a more restricted model.

Three-way data are data that can be classified in three ways. For an example is scores of a number of subjects on different variables measured on different occasions. Three-mode principal components analysis (Tucker, 1966) is a method for summarizing three-way data, and is a generalization of standard two-way principal components analysis (PCA). In two-way PCA the data are decomposed into two matrices, namely the component scores matrix and the component loading matrix. In three-mode PCA, the three-way data are decomposed into three component matrices, where the numbers of components to be used are not necessarily equal for each component matrix. When the numbers of components are not suggested by the nature of the data, a method is needed to indicate these numbers. In order to choose the numbers of components, Tucker (1966) proposed the application of a method ordinarily used in two-way PCA. However, it is not clear that this method is suitable for use in three-way problems. Therefore, a new method is proposed for indicating the numbers of components in three-mode PCA, and this method is compared to two methods ordinarily used in two-way PCA by means of a simulation study.

Timmerman and Kiers (2000) three-mode PCA model is usually fitted to the data by Tuckals3 which is an alternating least squares algorithm. Unfortunately, this kind of algorithm may end in a local optimum. At the cost of computational effort, the possibility of missing the global optimum can be reduced by using multiple 'starts' for a single three-mode PCA model. Since the new method of determining the numbers of components requires a large number of three-mode PCAs, it is useful to examine the necessity of using multiple starts. In several applications of three-mode principal component analysis to sets of correlation matrices, results turned out to be very similar to results obtained via perfect congruence analysis for weights (Louwerse and Smilde, 2000). Three-mode PCA is meant for the analysis of possibly preprocessed three-way data xijk that give the score of individual i on variable j at measurement occasion k, i=1,...,I, j=1,...,J, k=1,...,K. In 3MPCA, as in PCA, matrices A and B are found that summarize the individuals and the variables, respectively, but in addition, a matrix C is found that summarizes the occasions. Usually, in three-mode PCA these matrices are all referred to by the general term "component matrices" and a distinction between component scores and loadings is not made (Kiers and Mechelen, 2001).

2.4 Extension of Multivariate Statistical Process Monitoring

Besides PCA, there also have several more extension of MSPM such as Projection to Latent Structures (PLS), Independent Component Analysis (ICA), Subspace Identification and many more.

2.4.1 Projection to Latent Structures (PLS)

Projection to latent structures or partial least squares (PLS) is a multivariable statistical regression method based on projecting or viewing the information in a high dimensional data space down onto a low dimensional one defined by some latent variables (Zhao et al., 2006). Abdi (2010) say PLS is a recent technique that combines features and generalizes PCA and multiple linear regressions. Zhao et al. (2000) support Abdi statement as PLS is one of the most powerful linear regression techniques to deal with noisy and highly correlated data. Its goal is to predict a set of dependent variables from a set of independent variables or predictors. This prediction is achieved by extracting from the predictors a set of orthogonal factors called latent variables which have the best predictive power (Abdi, 2010).

PLS already has been successfully applied in diverse fields including process monitoring and quality control and identification of process dynamics & control with a limited number of observations available (Lee et al, 2006). When dealing with nonlinear systems, this approach assumes that the underlying nonlinear relationship between predictor data and response data can be approximated by quadratic PLS (QPLS) or neural network based PLS (NNPLS) while retaining the outer mapping framework of linear PLS algorithm and matrices were auto-scaled before they were processed by PLS algorithm (Wold, 2005). PLS model consists of outer relations which data are expressed in terms of their respective scores and inner relations that link the data to the data in the latent subspace. PLS finds the latent variables from the measured data by capturing the largest variance in the data and achieves the maximum correlation between the predictor variables and response variables.

A tutorial description along with some examples on the PLS model was provided by Geladi and Kowalaski (1986). PLS reduces the dimensionality of the measured data, finds the latent variables from the measured data by capturing the largest variance in the data and achieves the maximum correlation between the predictor X variables and response Y variables. In PLS based process dynamics, the inner relationship between variance and scores. The process dynamics in latent subspace could not be well identified by linear or quadratic relationships. For multivariable processes, the Partial least squares (PLS) controllers offer the opportunity to be designed as a series of SISO controllers (Qin and McAvoy (1992, 1993). Because of the diagonal structure of the dynamic part of the PLS model, input-output pairings are automatic. Series of SISO controllers designed on the basis of the dynamic models identified into latent subspaces and embedded in the PLS framework are used to control the process. Till date there is no reference on NNPLS controllers in the open literature though PLS & NNPLS based process identification, PLS controllers are well documented. The quality of the prediction obtained from a PLS regression model is evaluated with cross-validation techniques such as the bootstrap and jackknife. There are two main variants of PLS regression which is the most common one separates the roles of dependent and independent variables and the second one is used mostly to analyze brain imaging data that gives the same roles to dependent and independent variables.

2.4.2 Independent Component Analysis (ICA)

Hyvarinen (n.d) identified independent component analysis (ICA) is a statistical and computational technique for revealing hidden factors that underlie sets of random variables, measurements, or signals. It is a generative models for the observed multivariate data, which is typically given as a large database of samples. In the model, the data variables are assumed to be linear mixtures of some unknown latent variables, and the mixing system is also unknown (Lee et al., 2004). The latent variables are assumed non Gaussian and mutually independent and they are called the independent components of the observed data (Lee et al., 2003). ICA seeks to extract these independent components as well as the mixing matrix of coefficients.

Although ICA can be looked upon a useful extension of PCA, its objective differs from that of PCA. Bakshi (1998) say PCA is a dimensionality reduction technique that reduces the data dimension by projecting the correlated variables onto a smaller set of new variables that are uncorrelated and retain most of the original variance. However, its objective is only to correlate variables, not to make them independent. PCA can only impose independence up to second order statistics information which is mean and variance (Kano et al., 2004). While constraining the direction vectors to be orthogonal, whereas ICA has no orthogonality constraint and involves higher-order statistics, for an example it not only correlates the data for second order statistics but also reduces higher order statistical dependencies (Yoo et al., 2004). Hence, ICs reveal more useful information from observed data than principal components (PCs).

The data analyzed by ICA could originate from many different kinds of application fields, including digital images, document databases, economic indicators and psychometric measurements (Kano et al., 2004). In many cases, the measurements are given as a set of parallel signals or time series; the term blind source separation is used to characterize this problem. Typical examples are mixtures of simultaneous speech signals that have been picked up by several microphones, brain waves recorded by multiple sensors, interfering radio signals arriving at a mobile phone, or parallel time series obtained from some industrial process (Lee at al., 2004). A number of applications of ICA have been reported in speech processing, biomedical signal processing, machine vibrationanalysis, nuclear magnetic resonance spectroscopy, infrared optical source separation, radio-communications, and so on (Girolami, 1999). Kano et al. (2004) say the key idea or motivation of using ICA is that the monitoring performance can be improved by focusing on essential variables that drive a process.

2.4.3 Subspace Identification

Subspace methods or also known as 4SID or Subspace State Space System Identifcation are used to find linear state space model from experimental data as a relatively new alternative to widely used regression methods such as ARX, ARMAX and many more (Trnka and Havlena, 2005). Subspace identification algorithms are widely known and appreciated for their quick and reliable estimation of linear models based on available input/output measurements (Goethals et al. 2004). Much of the reliability of subspace identification algorithms is attributed to the fact that a model is obtained, solely by using numerically reliable matrix and vector manipulations such as projections and singular value decompositions. This is different from the classical predictor error methods, which usually involve with the minimization of a non-convex cost-function. There will be no guarantee that the obtained local minimum yields a good model. Treasure et al. (2003) say subspace identification can determine a set of state variables for describing process dynamics, produce a reduced set of variables to monitor process performance and offer contribution charts to diagnose anomalous behavior. This is demonstrated by an application study to a realistic simulation of a chemical process (Treasure et al., 2003). This point of view on subspace methods will show suitable fields for their application, where we can take advantage of their good properties like numerical robustness implemented by QR and SVD factorization, implicit rank reduction, non-iterative algorithm and few user parameters (Trenka et al., n.d.).

Goethals et al. (2004) say although subspace identification algorithms are fast and robust, a major drawback is that their use is mostly restricted to the class of linear systems. Some attempts to extend the use of subspace identification algorithms to nonlinear systems have been made in the past for general nonlinear systems, or more restricted model structures such as bilinear models, Wiener models, and Hammerstein models. Overschee and Moor (1996) say that Subspace identification algorithms are based on concepts from system theory, (numerical) linear algebra and statistics. The subspace identification approach does not suffer from any of these inconveniences. The only parameter to be user-specified is the order of the model, which can be determined by inspection of certain singular values.

Ljung et al. (1993) say when implemented correctly, subspace identification algorithms are fast, despite the fact that they are using QR and singular value decompositions. As a matter of fact, they are faster than the “classical” identification methods, such as Prediction Error Methods, because they are not. Hence there are also no convergence problems. Moreover, numerical robustness is guaranteed precisely because of these well-understood algorithms from numerical linear algebra. As a consequence, the user will never be confronted with hard-to-deal-with-problems such as lack of convergence, slow convergence or numerical instability (Chiuso et al., 2004).

2.5 Summary

In conclusion, this chapter review and discuss on fundamental of MSPM using PCA tools, and some of extension of PCA and MSPM. Thus it have show that PCA and MSPM are really important with many advantages. Beside that, the purpose of MSPM and PCA really give many benefit to the industrial process as many industrial are using multiple operating mode. Moreover Chapter three will illustrates the basic of methodology and techniques applied to achieve the objectives.

CHAPTER 3

METHODOLOGY

3.1 Introduction

Generally MSPC procedure could be categorized into two phases. For phases 1, it involves the off-line monitoring operation, while for phases 2 it involves on-line monitoring operation. However, for this study before undergo this two phase, the normal operating condition sample data is structured based on the assumption that has been made. The Figure 3.1 below show the procedures involves in MSPC to detect, identified, diagnosis and recover the fault detection.

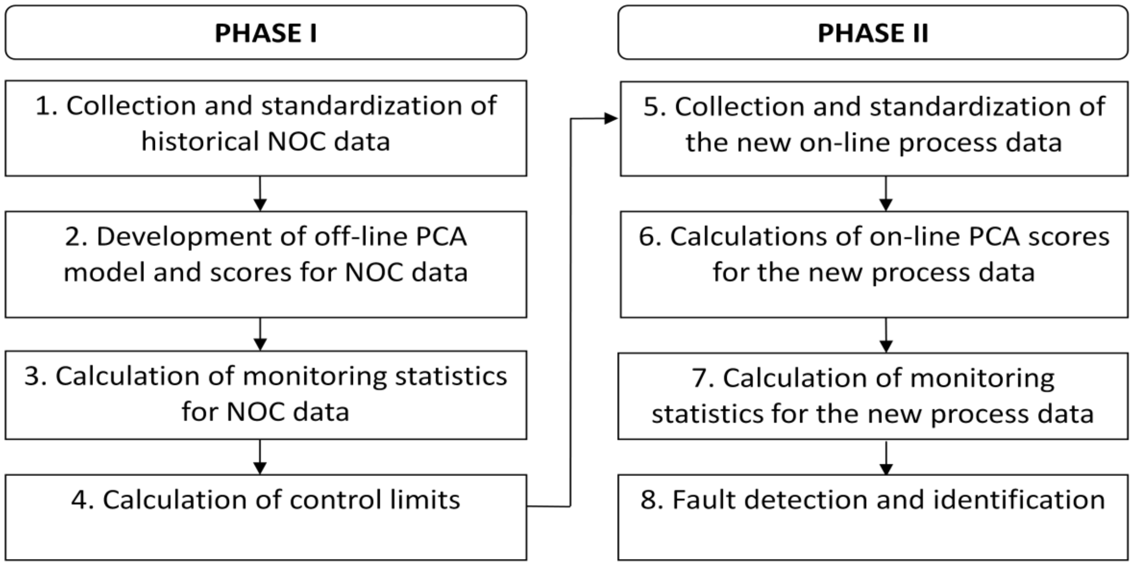


Figure 3.1 MSPC procedure (Adapted from Yusri and Zhang, 2010)

3.2 Phase I Procedures

There are four step that been proposed in this phases. After normal operating condition samples data have been structure based on the assumption have be made. Then a set of normal operation condition (NOC) data, $X_{n \times m}$ which n is samples and m is variables, are identified off-line based on the historical process data archive.

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

According to Martin et al., (1996) NOC simply implies that the process is operated at the desired setting condition and produces satisfactory products that meet the qualitative as well as quantitative specified standard. Then, the data are standardized to zero mean and unit variance with respective to each of the variables because PCA results depend on data scales.

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

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

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

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

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

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

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

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

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

Jolliffe (2002) say the equation below presents a measure of data variations captured by the first a principal component.

$$k = \frac{\lambda_1 + \lambda_2 + \ldots + \lambda_a}{\lambda_1 + \lambda_2 + \ldots + \lambda_a + \ldots + \lambda_m} \tag{3.7}$$

The third step is to calculate monitoring statistics for NOC data. This calculation basically involves calculation of the Hotelling's T^2 and SPE monitoring statistics.

$$T_i^{\,2} = \sum_{j=1}^{a} \frac{p_{i,j}^2}{\lambda_j} \tag{3.8}$$

$$\begin{aligned} \tilde{\mathbf{E}} &= \breve{\mathbf{X}} - \hat{\mathbf{X}} \\ &= \breve{\mathbf{X}} - \mathbf{P}_a \mathbf{V}_{\mathrm{a}}^{\mathrm{T}} \\ &= \breve{\mathbf{X}} - \breve{\mathbf{X}} \mathbf{V}_a \mathbf{V}_a^{\mathbf{T}} \\ &= \breve{\mathbf{X}} \left(\mathbf{I} - \mathbf{V}_a \mathbf{V}_a^{\mathbf{T}} \right) \end{aligned} \tag{3.9}$$

$$SPE_i = \tilde{\mathbf{e}}_{\mathbf{i}} \tilde{\mathbf{e}}_{\mathbf{i}}^{T} \tag{3.10}$$

The final task in phase I which is 4th step deals with developing the control limits for both of the statistics.

$$T_{\alpha} = \frac{A(n-1)}{(n-A)} F_{A,n-A,\alpha} \tag{3.11}$$

$$SPE_{\alpha} = \theta_1 \left(\frac{z_{\alpha}\sqrt{2\theta_2 h_0^2}}{\theta_1} + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} + 1 \right)^{\frac{1}{h_0}} \tag{3.12}$$

3.3 Phase II Procedures

For the second phases, all the first phases of fault detection procedure will be developed for on-line application. However for the last step which is step 8 there are two main operations which have to be conducted separately which is fault detection and fault identification. Fault detection is a result of an occurrence of a special event that is not in conformance to the common cause nature. This fault detection will be declared if monitoring statistics exceeding its respective control limit for a pre-defined successive number of samples. While, for the fault identification is based on contribution plot. Typically, the circumstance of the on-line MSPC application always involves monitoring the real-time states of process condition. In other words, the dynamic behavior of the operation conditions should be analyzed in a real time manner to reflect the process status immediately. Even though, the procedures are still considering, more or less, the major steps taken during the off-line application development previously.

3.4 Summary

In conclusion, there are four main steps of MSPC system which is fault detection, fault identification, fault diagnosis and process recovery. This method will designated the departure of observed samples from an acceptable range using a set of parameters. For the new monitoring method that have been propose which is multiple operating mode, the NOC sample of 500 sample which is mode 1 are divide into two part where mode 2 consist of first 250 sampe of NOC while mode 3 is consist of the last 250 sample of NOC Moreover, it will identify the observed process variables that most relevant to the fault by using the contribution plot technique. This will specifically determine the type of fault which has been significantly contributed to the signal.

CHAPTER 4

RESULT AND DISCUSSION

4.1 Introduction

Initially some work has been done, that specially emphasize on the integration of process monitoring algorithms based on the conventional PCA. Hence, several works are represented in this chapter. Firstly, a description on the case study is briefly explained for the particular analysis. Next, is discussion on the result of NOC data and the tested fault data. The tested data are discussed for abrupt and incipient. Finally, is the summary of the chapter.

4.2 Case Study of an industrial chemical process in Tennessce Eastmant

This case study desribe a model of an industrial chemical process in Tennessce Eastmant as shown in figure 4.1 below (Downs and Vogel, 1993). This process is involving two simultaneously gas-liquid exothermic reactions which also produce two additional byproduct reactions. The reactions are:

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

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

$$A(g) + E(g) \longrightarrow F(liq) \quad \text{Byproduct}$$

$$3D(g) \longrightarrow 2F(liq) \quad \text{Byproduct}$$

The process has five major unit operations which is the reactor, the product condenser, a vapor-liquid separation, arecycle compressor and a product stripper. The gaseeous reactants are fed to the reactor where it reacts to form liquid products. The gas phase reactions are catalyzed by a nonvolatile 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 the catalyst remains in the reactor. The reactor product stream passes through a cooler for condensing the products and from there to a vapor-liquid separator. Noncondensed components recycle back through a centrifugal compressor to the reactor feed. Condensed components move to a product stripping column to remove remaining reactants by stripping with feed stream number 4. Products G and H exit the stripper base and are separated in a downstream refining section which is not included in this problem. The inert and byproduct are primarily purged from the system as a vapor from the vapor-liquid separator.

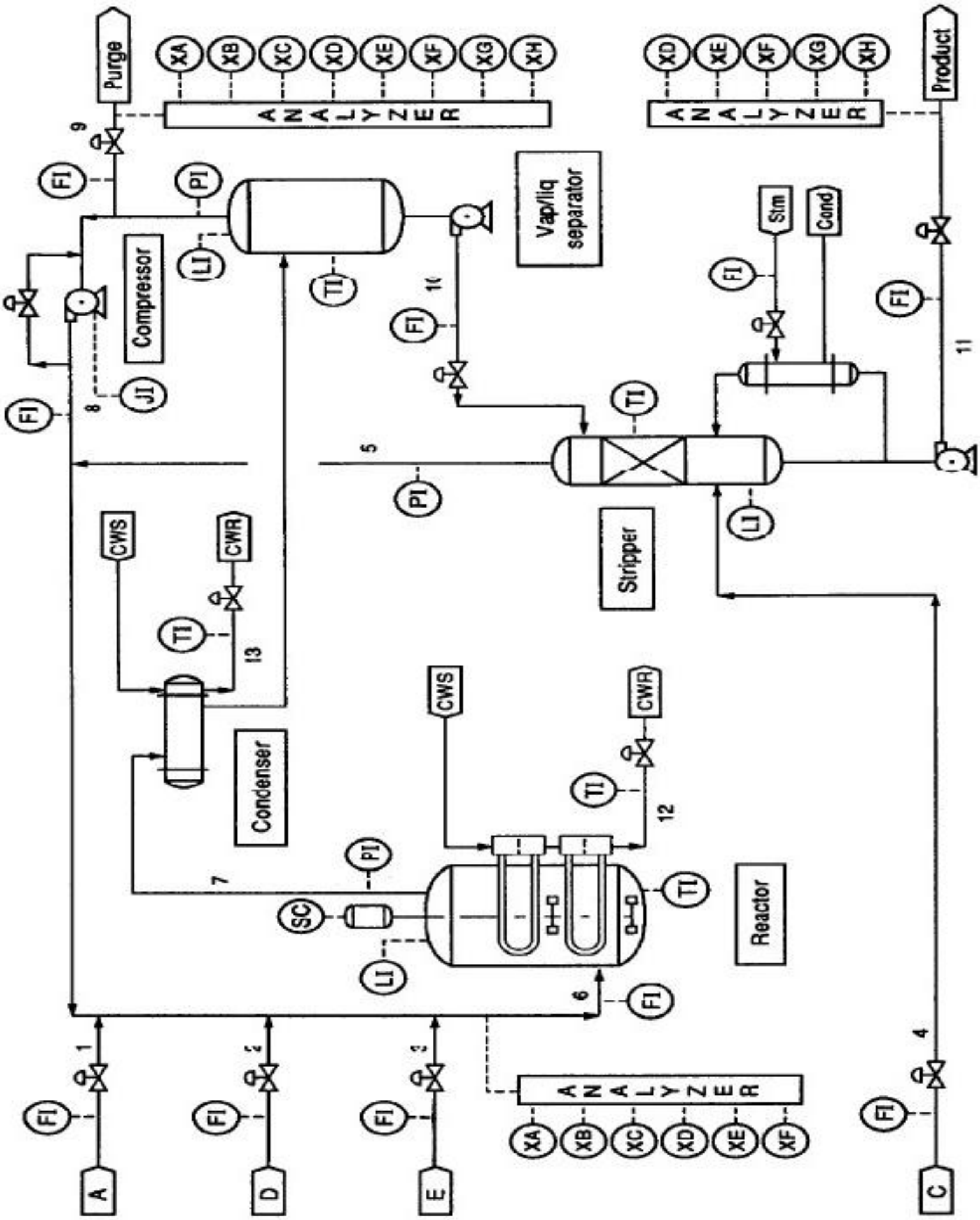


Figure 4.1 Tennessce Eastmant industrial chemical process

The process has 41 measurements and 12 manipulated variables as listed in Table 4.1 (a)-(c) below.

Table 4.1 (a) Process manipulated variables

Variable name	Variable number	Units
D feed flow (stream 2)	XMV (1)	kg h^{-1}
E feed flow (stream 3)	XMV (2)	kg h^{-1}
A feed flow (stream 1)	XMV (3)	kscmh
A and C feed flow (stream 4)	XMV (4)	kscmh
Compressor recycle valve	XMV (5)	%
Purge valve (stream 9)	XMV (6)	%
Seperator pot liquid flow (stream 10)	XMV (7)	m^3 h^{-1}
Stripper liquid product flow (stream 11)	XMV (8)	m^3 h^{-1}
Stripper steam valve	XMV (9)	%
Reactor cooling water flow	XMV (10)	m^3 h^{-1}
Condenser cooling water flow	XMV (11)	m^3 h^{-1}
Agitator speed	XMV (12)	rpm

Table 4.1 (b): Continuous process measurements

Variable name	Variable number	Units
A feed (stream 1)	XMEAS (1)	kscmh
D feed (stream 2)	XMEAS (2)	kg h^{-1}
E feed (stream 3)	XMEAS (3)	kg h^{-1}
A and C feed (stream 4)	XMEAS (4)	kscmh
Recycle flow(stream 8)	XMEAS (5)	kscmh
Reactor feed rate (stream 6)	XMEAS (6)	kscmh
Reactor pressure	XMEAS (7)	kPa gauge
Reactor level	XMEAS (8)	%
Reactor temperature	XMEAS (9)	˚C
Purge rate (stream 9)	XMEAS (10)	Kscmh
Product seperator temperature	XMEAS (11)	˚C
Product seperator level	XMEAS (12)	%
Product seperator pressure	XMEAS (13)	kPa gauge
Product seperator underflow (stream 10)	XMEAS (14)	m^3 h^{-1}
Stripper level	XMEAS (15)	%
Stripper pressure	XMEAS (16)	kPa gauge
Stripper underflow (stream 11)	XMEAS (17)	m^3 h^{-1}

Stripper temperature	XMEAS (18)	˚C
Stripper steam flow	XMEAS (19)	kg h^{-1}
Compressor work	XMEAS (20)	kW
Reactor cooling water outlet temperature	XMEAS (21)	˚C
Seperator cooling water outlet temperature	XMEAS (22)	˚C

Table 4.1 (c): Sample process measurement

Component	Variable number	Units
Reactor feed analysis (stream 6)		
A	XMEAS (23)	mol%
B	XMEAS (24)	mol%
C	XMEAS (25)	mol%
D	XMEAS (26)	mol%
E	XMEAS (27)	mol%
F	XMEAS (28)	mol%
Purge gas analysis (stream 9)		
A	XMEAS (29)	mol%
B	XMEAS (30)	mol%
C	XMEAS (31)	mol%
D	XMEAS (32)	mol%
E	XMEAS (33)	mol%
F	XMEAS (34)	mol%
G	XMEAS (35)	mol%
H	XMEAS (36)	mol%
Product analysis (stream 11)		
D	XMEAS (37)	mol%
E	XMEAS (38)	mol%
F	XMEAS (39)	mol%
G	XMEAS (40)	mol%
H	XMEAS (41)	mol%

4.3 Normal Operating Condition Data Collection

This simulation has identified and collects a set of normal operating condition. This set of NOC data containing 500 measurements of 53 variables. By using PCA algorithm, the standardized NOC data is analysed. This analysis is performing to identify the required number of PCs as to reduce the dimension of the multivariate data.

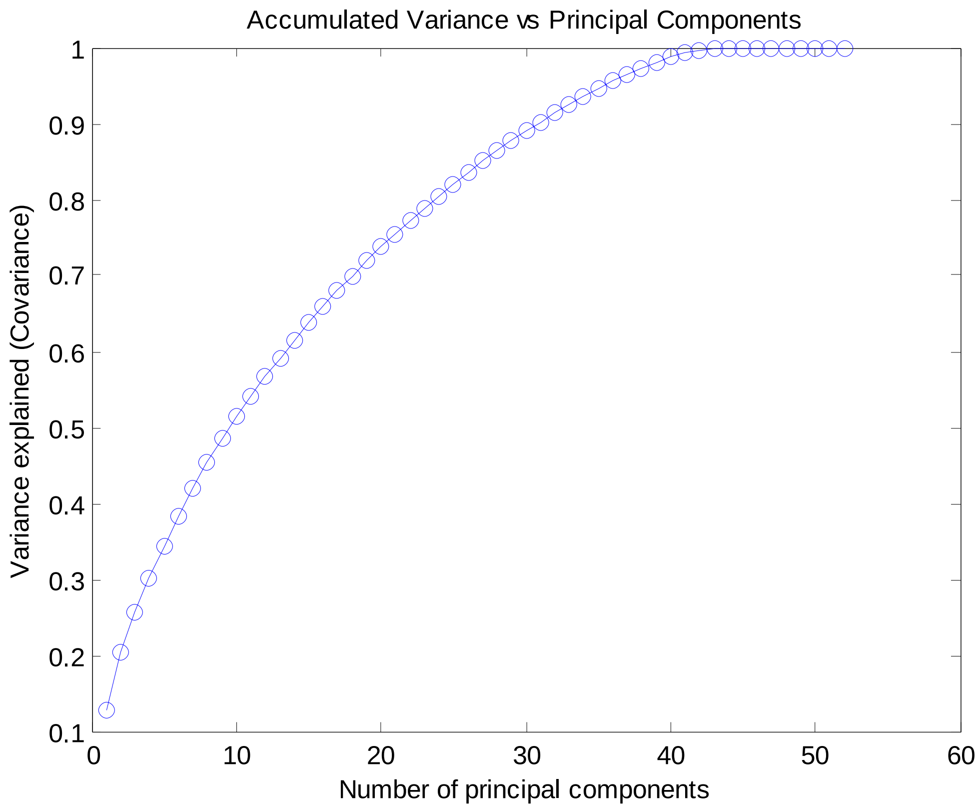
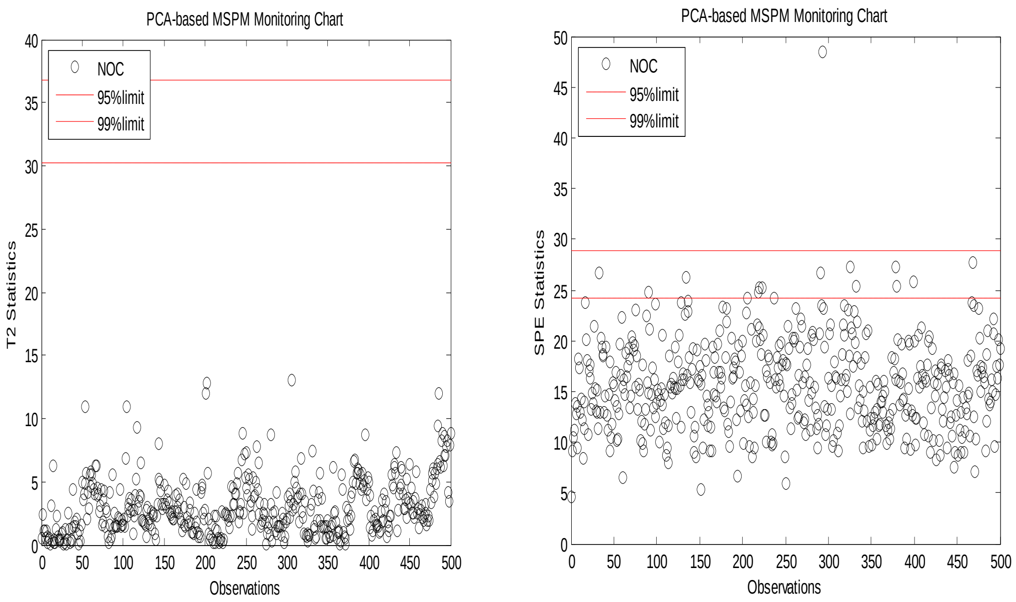


Figure 4.2: Accumulated data variance explained by different PCs

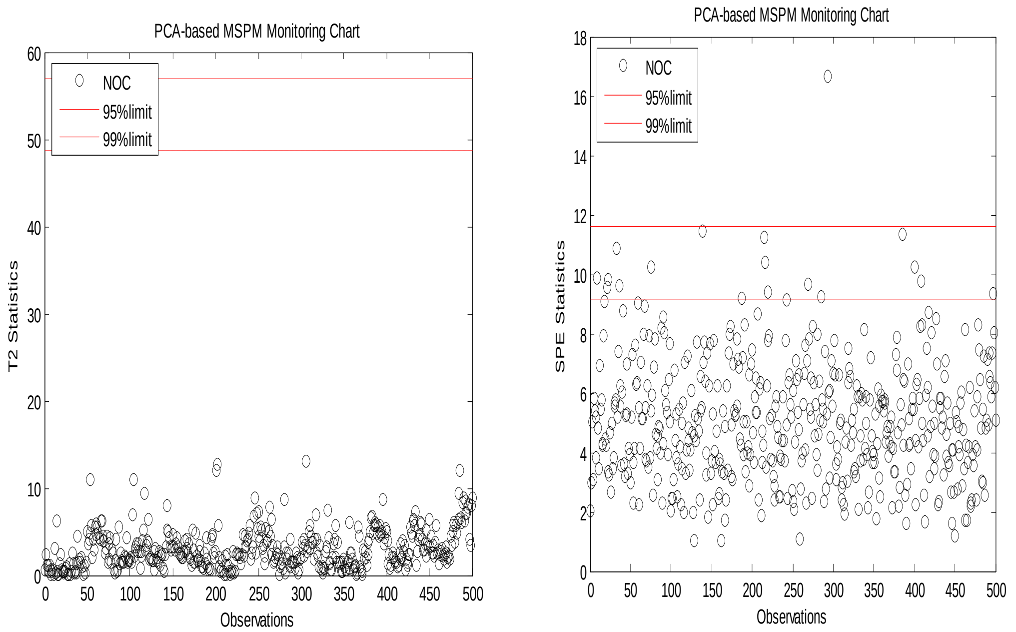
From figure 4.2 above, show the accumulated data variance explained by different PCs. Its shows that at leasts 31 PCs are needed to represent over 90% of the total NOC data variance and 18 PCs are required to explain over 70% of the total variance. Therefore, for this particular case study, 18 and 31 PCs are retained in the PCA model for the calculation of NOC scores. The Hoteling's T^2 statistic and SPE statistic was then to be calculated and plotted together with the 95% and 99% confidence limits.



(a) T^2 statistic for NOC data

(b) SPE statistic for NOC data

Figure 4.3 Mode I at 18PCs



(a) T^2 statistic for NOC data

(b) SPE statistic for NOC data

Figure 4.4 Mode I at 31PCs

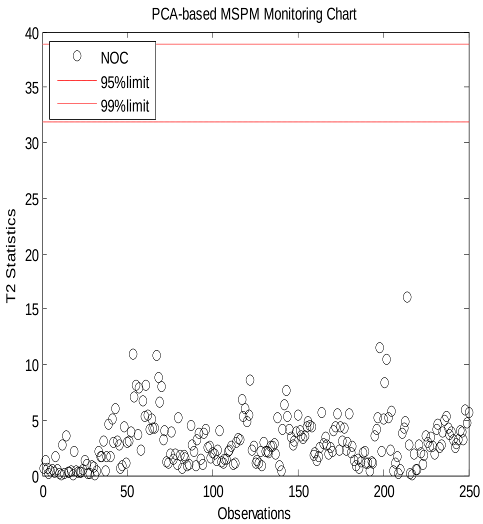
Figure 4.3 illustrate graph of NOC data of mode I which consist 500 sampel, for T^2 statistic and SPE statistic for at 18 PCs at 70% total NOC variance. It can be seen that the T^2 and SPE statistic for the NOC data is below the confident limits. The data is still normal as the fault is considered when 3 samples in series are out from the boundaries. Thus, T^2 statistic and SPE statistics illustrate normal operating condition.

While, figure 4.4 show graph of mode I of 500 sampel of NOC data for T^2 statistic and SPE statistic for 31 PCs at 90% total NOC variance. For the T^2 statistic for the NOC data show it below the confident limits however for SPE there are one sample outside the limit boundaries. The data is still normal as the fault is considered when 3 samples in series are out from the boundaries. Thus, T^2 statistic and SPE statistics illustrate normal operating condition.

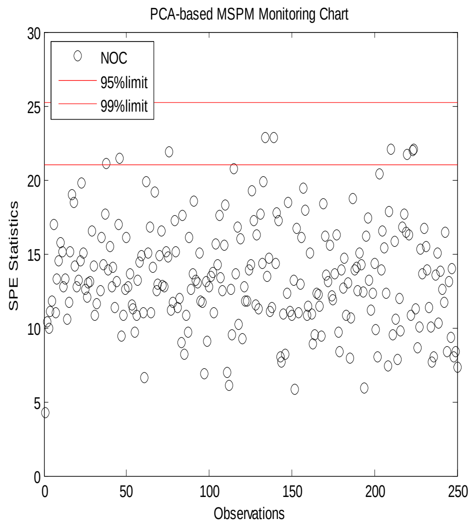
This NOC data has been test by reduce the sample using 250 measurements with 53 variables. By using PCA algorithm, the standardized NOC data test is analysed. This analysis objective is to identify the required number of PCs as to reduce the dimension of the multivariate data. Same with NOC data, the Hoteling's T^2 statistic and SPE statistic was calculated and plotted together with the 95% and 99% confidence limits.

Figure 4.5 and figure 4.6 below show statistic for NOC data which is T^2 statistics and SPE statistics for 18 PCs and 31 PCs respectively. Besides that, both figure also representing T^2 and SPE statistics of mode II and mode III. Moreover mode 1 is data of the first 250 sample of NOC from mode I, while mode II is consist the last 250 of NOC data from mode I.This figure show that all the statistics process are normal as all the statistics NOC data test is below the confident limit. Thus, T^2

statistic and SPE statistics illustrate normal operating condition for mode II and mode III at 18 PCs and 31 PCs.

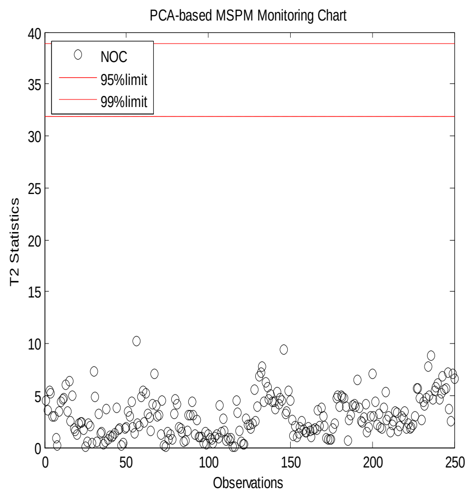


(a) T^2 statistic for NOC data

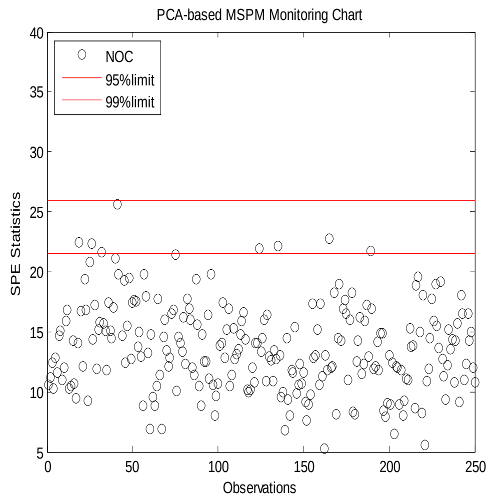


(b) SPE statistic for NOC data

Mode II First 250 Sample



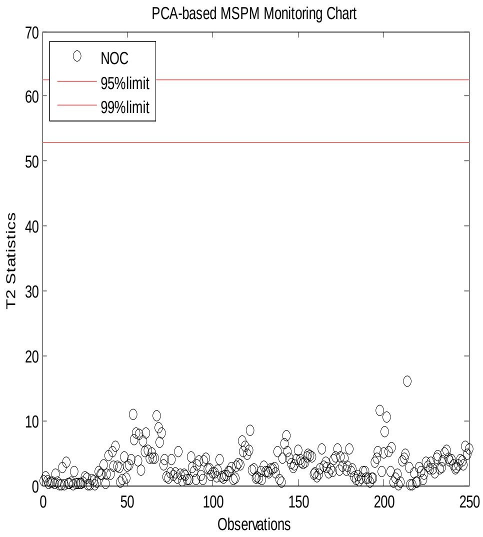
(a) T^2 statistic for NOC data



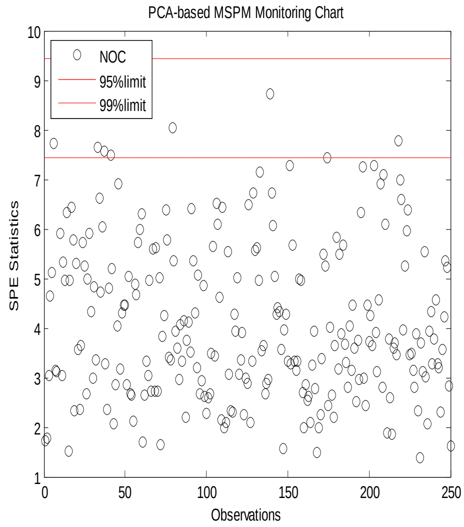
(b) SPE statistic for NOC data

Mode III Second 250 Sample

Figure 4.5 18 PCs at 70% total NOC variance

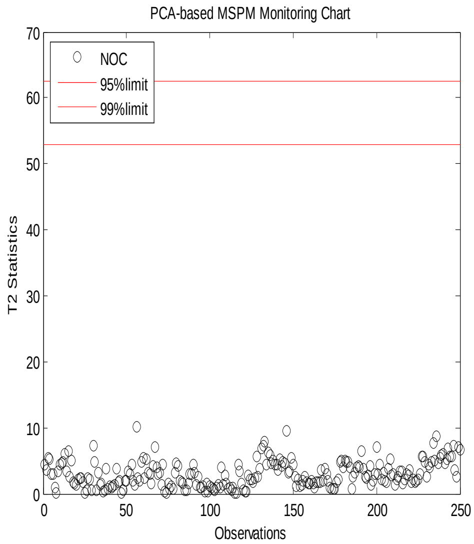


(a) T^2 statistic for NOC data

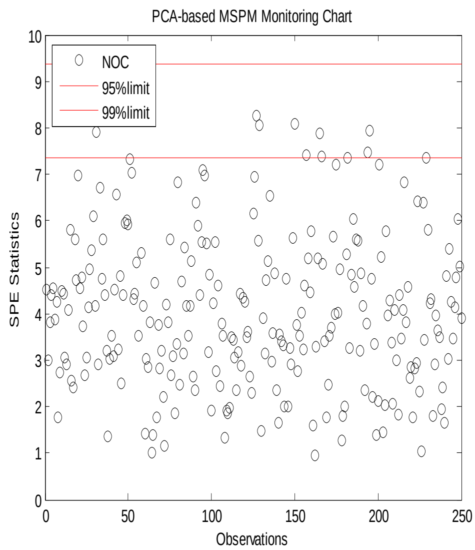


(b) SPE statistic for NOC data

Mode II First 250 Sample



(a) T^2 statistic for NOC data



(b) SPE statistic for NOC data

Mode III Second 250 sample

Figure 4.6 31 PCs at 90% total NOC variance

4.4 Fault data collection

The system also subjects to be affected from several malfunction conditions and detecting this kind of malfunctions should be easy for any multivariate monitoring system as the deviations are usually very obvious. By concentrated to fault 1,2,8 and 9, the study are divided into three modes which first mode for 250 sample, while for next second and third mode it analyzing on the first and last of 500 sample from mode 1 respectively. Then the contributed variable to the fault and the impact will be discussed by comparing mode II and mode III with mode I.

4.4.1 Fault Detection and The Comparison Between The Mode

Table 4.2 and Table 4.3 below show the result of different mode at 18 PC's and 31 PC's respectively. Moreover, the both table also show that the mode 2 and mode 3 have fast fault detection compare to the mode 1. Besides that, the low number of PC use the total variance of normal operating condition cover are also low as well, but however the fault detection are faster. Table 4.4 in appendix A is to show variance of the normal operating condition (NOC) of 500 sample and 53 fault at each mode of different PC.Beside to study the number of variance of each variable at every mode this figure also approve that the data that have been use is same.

Table 4.2: Result of fault detection for 18 PC's of 70% total variance

FAULT	Mode 1		Final	Mode 2		Final	Mode 3		Final
	T^2	SPE		T^2	SPE		T^2	SPE	
1	55	13	13	7	9	7	7	9	9
2	11	12	11	17	11	11	17	12	12
8	81	25	25	10	6	6	25	13	13
9	45	9	9	12	6	6	12	7	7

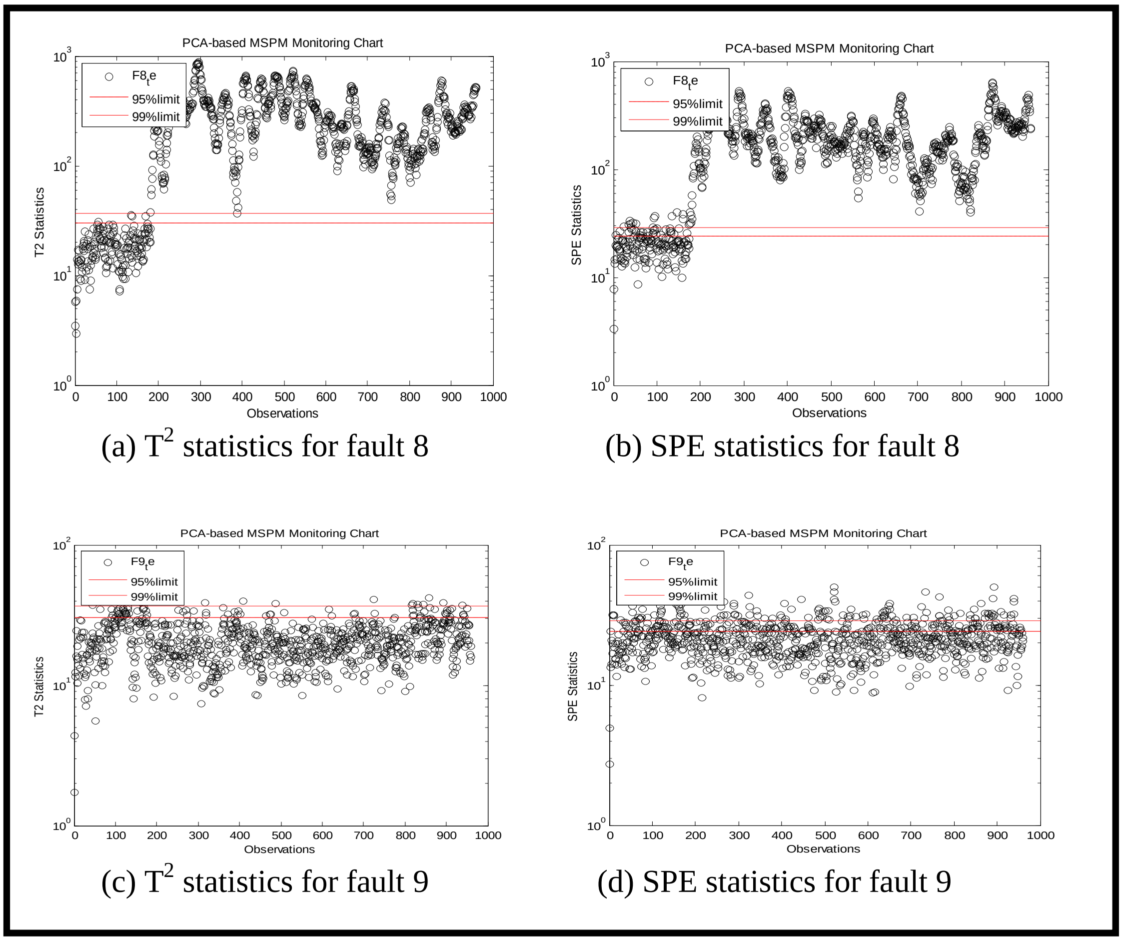
Table 4.3: Result of fault detection for 31 PC's of 90% total variance

FAULT	Mode 1		Final	Mode 2		Final	Mode 3		Final
	T^2	SPE		T^2	SPE		T^2	SPE	
1	65	51	51	15	9	9	50	9	9
2	33	11	11	11	12	11	33	11	11
8	43	14	14	16	6	6	16	6	6
9	42	11	11	42	15	15	45	8	8

This both tabel approve that, mode 2 and mode 3 which have low number of sampel use the higher the sensitivenes due to decreasing the variance. Besides that the mode 1 which use 500 sampel of NOC have high variances, thus the sensitivenes of fault detection are decreasing as the process in steady state , it will detect fault. Thus mode 2 and mode 3 analyse all the sampel better than mode 1 because the variance is low.

4.4.2 Mode I

Figure 4.7 below show result of T^2 statistics and SPE statistics for fault 8 and 9 for the 500 sample of normal operating condition (NOC). This figure have show that for all fault which is fault 8 and 9 the result of both T^2 statistic and SPE statistic for 18 pc of 70% total variance is outside of the boundaries limit which show abnormal process thus occurs in suddenly. This figure show, T^2 statistics for fault 8 and 9 the fault detection are detect at sampel 81 and 45 respectively. While for SPE statistics the fault are detect at sampel 25 and 9 for fault 8 and 9 respectively. Thus the final fault detection is 25 and 9 for fault fault 8 and 9 respectively.



(a) T^2 statistics for fault 8

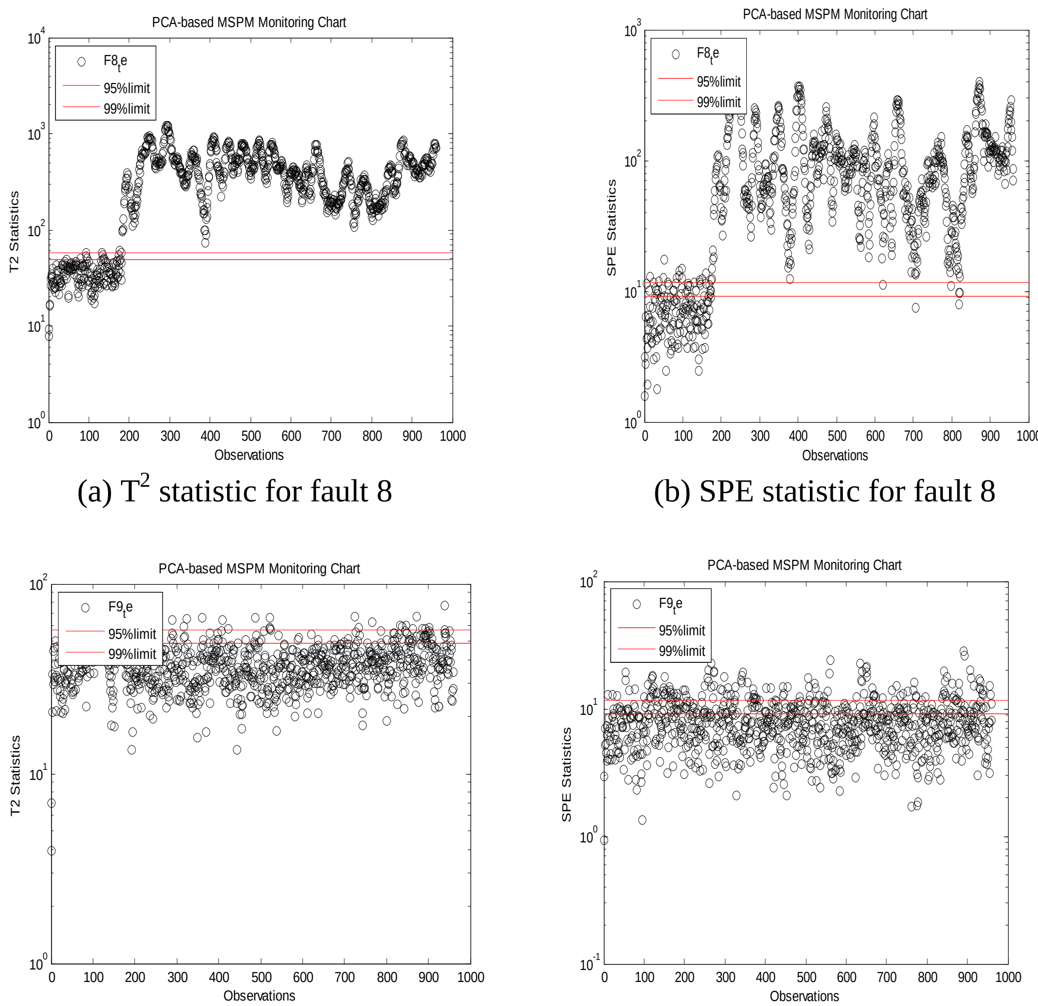
(b) SPE statistics for fault 8

(c) T^2 statistics for fault 9

(d) SPE statistics for fault 9

Figure 4.7 Mode I: T^2 statistics and SPE statistics for fault 8and9 for 18 pc of 70% total variance

While for Figure 4.8 below the result of T^2 statistics and SPE statistics for fault 8 and 9 for the 500 sample of normal operating condition (NOC) at 31 pc of 90% total variance are show. This figure show that all fault is outside of the boundaries limit which show abnormal process thus occurs in suddenly. Beside that for fault 8 the fault detection are detect at sampel 43 and 14 for T^2 statistics and SPE statistics respectively. While, for the fault 9 the T^2 statistics and SPE statistics are detect at sampel 42 and 11 respectively. Hence, the final fault detection for fault 8 and 9 are at sampel 14 and 11 respectively.



(a) T^2 statistic for fault 8

(b) SPE statistic for fault 8

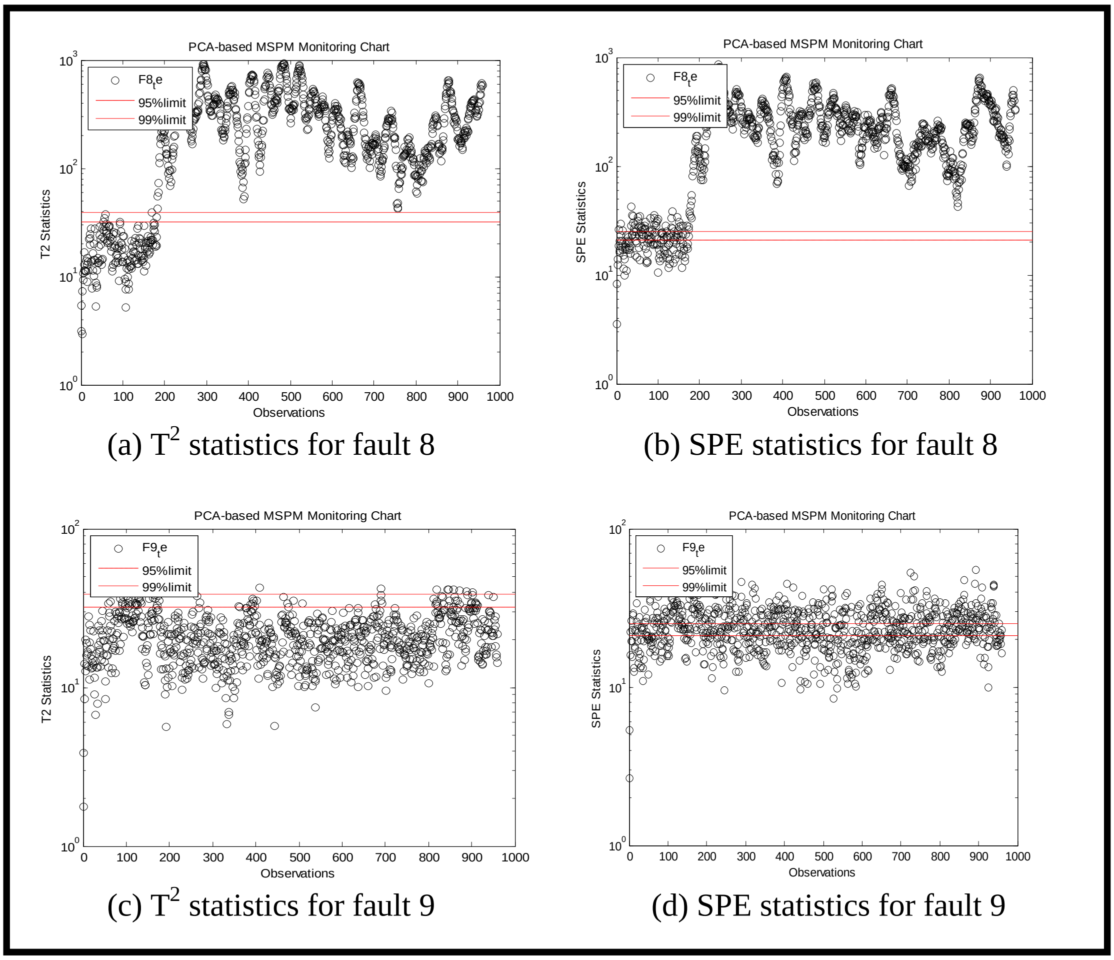
(c) T^2 statistic for fault 9

(d) SPE statistic for fault 9

Figure 4.8 Mode I: T^2 statistics and SPE statistics for fault 8 and 9 for 31 pc of 90% total variance

4.4.3 Mode II

This mode II are base on first 250 sample of NOC from the first mode. This figure 4.9 show the result of T^2 statistics and SPE statistics for fault 8 and 9 for 18 pc of 70% total variance have show that for all the fault is outside of the boundaries limit which this show abnormal process thus occurs in sudden. Beside that for fault 8 the fault detection are detect at sampel 10 and 6 for T^2 statistics and SPE statistics respectively. While, for the fault 9 the T^2 statistics and SPE statistics are detect at sampel 12 and 6 respectively. Thus the final fault detection is 6 and 6 for fault fault 8 and 9 respectively



(a) T^2 statistics for fault 8

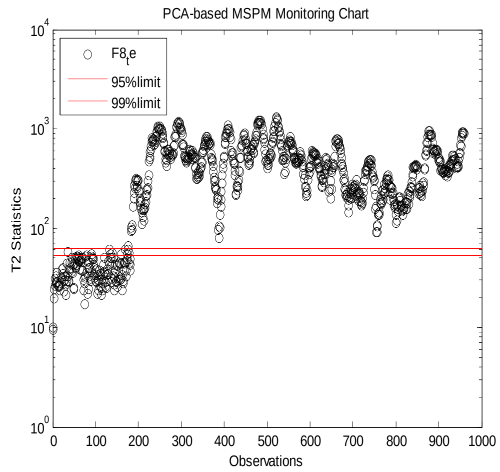
(b) SPE statistics for fault 8

(c) T^2 statistics for fault 9

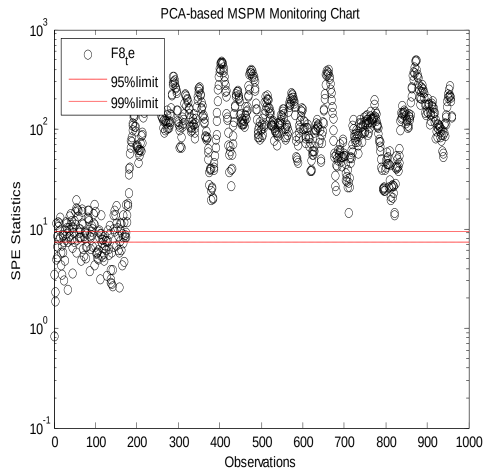
(d) SPE statistics for fault 9

Figure 4.9 Mode II:T^2 statistics and SPE statistics for fault 8 and 9 for 18 pc of 70% total variance

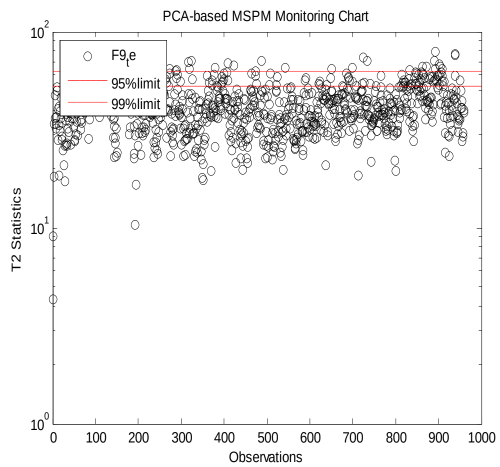
This figure 4.10 show the result of T^2 statistics and SPE statistics for fault 8 and 9 for 31 pc of 90% total variance have show that for all the fault is outside of the boundaries limit which this show abnormal process thus occurs in sudden. This figure show, T^2 statistics for fault 8 and 9 the fault detection are detect at sampel 16 and 42 respectively. While for SPE statistics the fault are detect at sampel 6 and 15 for fault 8 and 9 respectively. Hence, the final fault detection for fault 8 and 9 are at sampel 6 and 15 respectively.



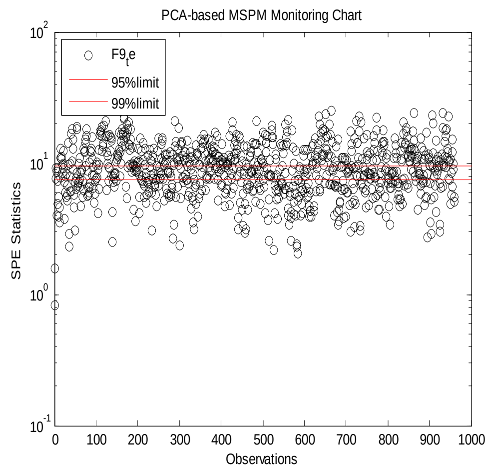
(a) T^2 statistics for fault 8



(b) SPE statistics for fault 8



(c) T^2 statistics for fault 9

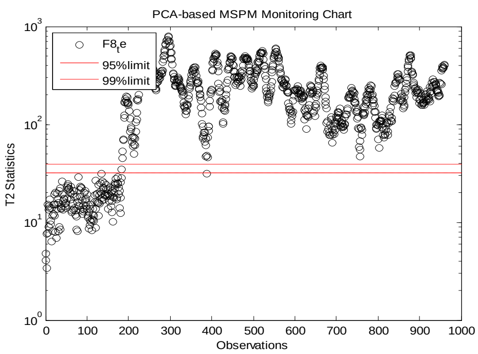


(d) SPE statistics for fault 9

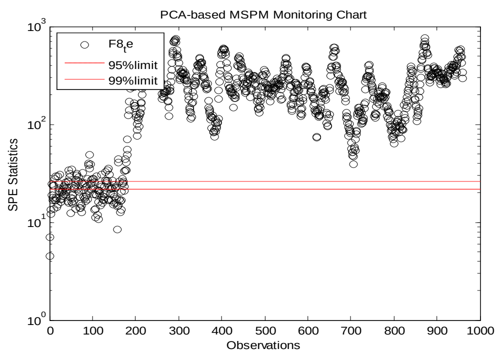
Figure 4.10 Mode II: T^2 statistics and SPE statistics for fault 8 and 9 for 31 pc of 90% total variance

4.4.4 Mode III

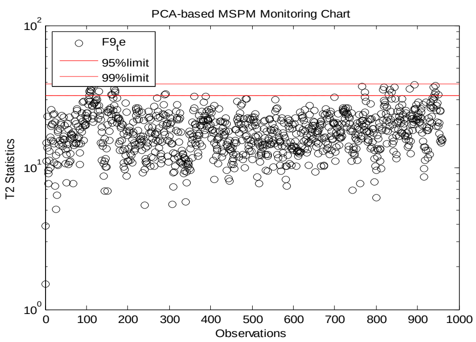
Figure 4.11 below show result of T^2 statistics and SPE statistics for fault 8 and 9 for the 500 sample of normal operating condition (NOC) for 18 pc of 70% total variance. This figure have show that for all fault which is fault 8 and 9 the result of both T^2 statistic and SPE statistic is outside of the boundaries limit which show abnormal process thus occurs in suddenly. Beside that for fault 8 the fault detection are detect at sampel 25 and 13 for T^2 statistics and SPE statistics respectively. While, for the fault 9 the T^2 statistics and SPE statistics are detect at sampel 12 and 7 respectively. Hence, the final fault detection for fault 8 and 9 are at sampel 13 and 7 respectively.



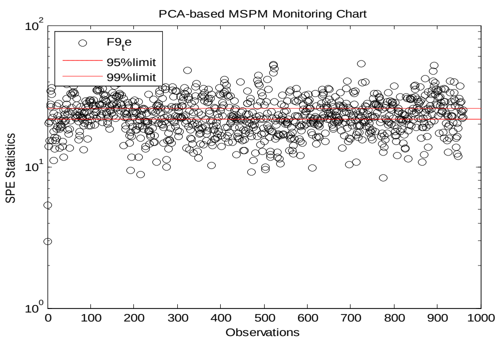
(a) T^2 statistics for fault 8



(b) SPE statistics for fault 8



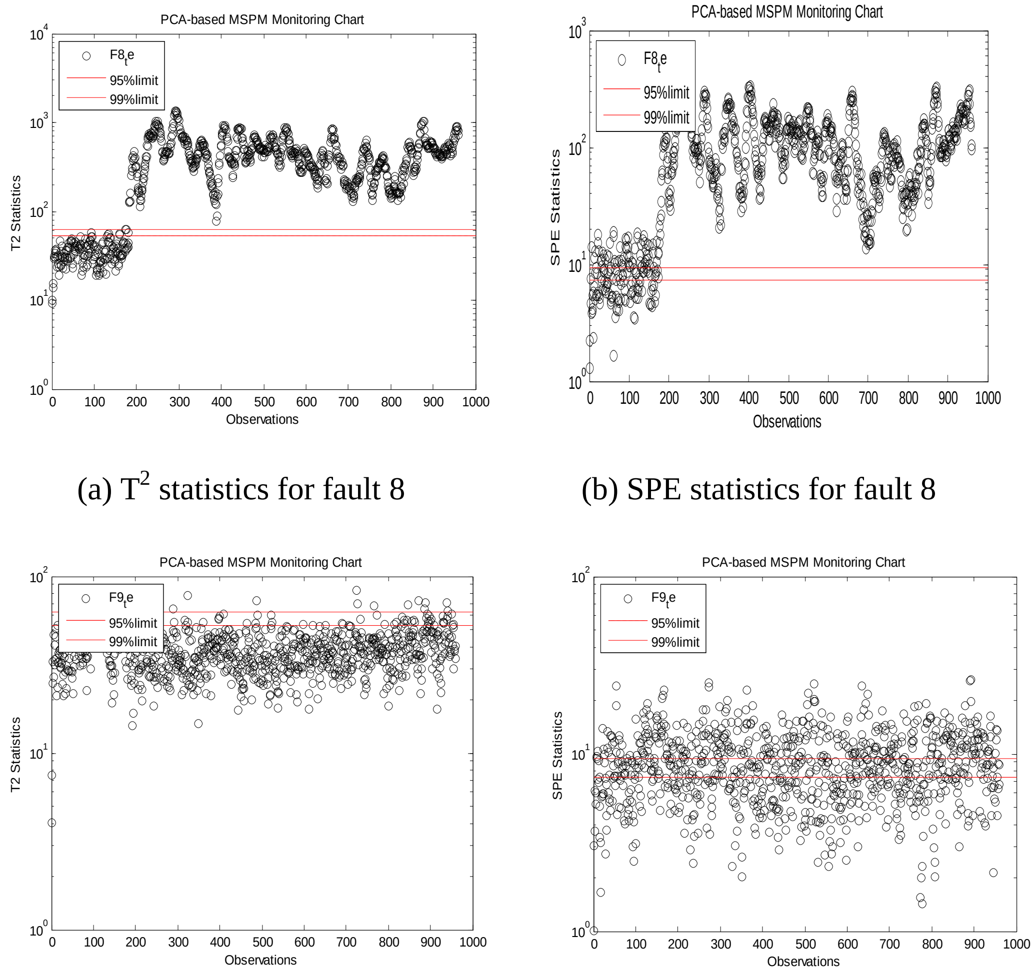
(c) T^2 statistics for fault 9



(d) SPE statistics for fault 9

Figure 4.11 Mode III: T^2statistics and SPE statistics for fault 8 and 9 for 18 pc Of 70% total variance

Figure 4.12 below show result of T^2 statistics and SPE statistics for fault 8 and 9 for the 500 sample of normal operating condition (NOC) for 31 pc of 90% total variance. This figure have show that for all fault which is fault 8 and 9 the result of both T^2 statistic and SPE statistic is outside of the boundaries limit which show abnormal process thus occurs in suddenly. This figure show, T^2 statistics for fault 8 and 9 the fault detection are detect at sampel 16 and 45 respectively. While for SPE statistics the fault are detect at sampel 6 and 8 for fault 8 and 9 respectively. Thus the final fault detection is 6 and 8 for fault fault 8 and 9 respectively



(a) T^2 statistics for fault 8

(b) SPE statistics for fault 8

(c) T^2 statistics for fault 9

(d) SPE statistics for fault 9

Figure 4.12 Mode III T^2statistics and SPE statistics for fault 8 and 9 for 31 pc of 90% total variance

4.5 Summary

A simulation of an industrial chemical process in Tennessce Eastmant is applying the conventional PCA to monitor the process. The main conventional PCA results have been discussed initially, which includes both of the NOC and fault data. Then the fault detection are discussing and each mode are then comparing the NOC variance of each mode and the percentages error. Hence, it show that the main conventional PCA use in mode I has slow fault detection compared to mode II and mode III which applying the different operating modes. For an example, in table 4.2 at fault 9, for the mode I the fault detection are detect at sampel 9, while for the mode II and mode III the fault are detect at sampel 6 and 7 respectively. Thus it approve that, the mode II and mode III that been apply different operating mode are having fast fault detection compare to mode I.

CHAPTER 5

CONCLUSION

5.1 Conclusions

In this research, MSPM using PCA tools is introduced. Some of the extension of MSPM and PCA is be review and the basic methodology to approach the proposed has been illustrated. The core technique to formulate the multivariate dimensional data reduction has been developing in order to approach the objectives using conventional PCA technique. The main goal in carrying out this study is to implement the conventional MSPM method based on different modes of NOC and analyze it with the conventional PCA technique on single NOC data. Based on the review on literature review there are many more method and technique to formulated multivariate data reduction. Every method has its own advantages and disadvantages. This research has proposed to run the traditional PCA by analyzing it with single NOC data and different modes of NOC data. Based on the result get it has shown the technique has affected the fault detection solutions. However, it does not mean that it is an excellent method for non-linear process monitoring. Therefore, more analyses are required.

5.2 Recommendation

The result and analysis for this research only valid to this indusrial process only whic is Tennessce Eastmant industrial chemical process. Hence, it is recommended for future research by using other data from the others chemical processing system. Beside that, more fault should be tested to come up with strongly conclusion that approve the reseach objectives.

REFERENCES

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

Bruker Daltonis, (n.d.) Principal Component Analysis [PowerPoint Slide], Retrieved From www.bdal.com.ppt (22/8/2012)

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

Downs, J.J. and Vogels, E.F., (1993), A Plant-Wide Industrial Process Problem Control,Pcrgamon Press Ltd, United State of America

Ge, Z., and Song, Z., (2008) Online Monitoring of nonlinear multiple mode processes Based on adaptive local model approach. *Control Engineering Practice*, 16, 1427-1437.

Kano, M., Hasebe, S., Hashimoto, I., and Ohno, H., (2001). A New Multivariate Statistical Process Monitoring Method Using Principal Component Analysis. *Computers and Chemical Engineering,* 25, 1103-1113.

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 and Chemical Engineering*, 28, 1157-1166.

Kano, M., Nagao, K., Hasebe, S., Hashimoto, I., Ohno, H., Strauss R, Bakshi, B.R., (2000). Comparison of Multivariate Statistical Process Monitoring Methods: Application To The Eastman Challenge Problem. *Computers and Chemical Engineering,* 24, 175-181.

Kano, M., Nagao, K., Hasebe, S., Hashimoto, I., Ohno, H., Strauss R, Bakshi, B.R., (2002).Comparison of Multivariate Statistical Process Monitoring Methods With Applications To The Eastman Challenge Problem. *Computers and Chemical Engineering,* 26, 161-174.

Kano, M., Tanaka, S., Ohno, H., Hasebe, S., and Hashimoto, I., (2002) The Use of Independent Component Analysis For Multivariate Statistical Process Control. *Control of Industrial Process. 423-428*

Kasonovich, K. A., Piovoso, M. J., and Hoo, K. A., (1996). *Improvements In the Development of Statistical Models for Process Monitoring & Detection*. ASA Conference. Chicago, USA.

Kourti, T, (2003). Abnormal Situation Detection, Three-way Data and Projection Methods; Robust Data Archiving and Modelling for Industrial Applications. Annual Reviews in Control, 27, 131-139.

Kourti, T., Lee, J., and MacGregor, J,F., (1996). Experiences with Industrial Applications of Projection Methods. *Computers and Chemical Engineering*, 20, S745-S750.

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

Kourti, T., Nomikos, P. and MacGregor, J.F., (1995) Analysis, Monitoring and Fault Diagnosis of Batch Processes Using Multi-block and Multi-way PLS. *Journal Process Control.* 5 (4), 277-284.

Kresta, J., MacGregor, J. F., and Martin, T. E., (1991). Multivariate Statistical Monitoring of Process Operation Performance. *Canadian Journal of Chemical Engineering,* 69, 34 – 47.

Kroonenberg, P.M., and Berge, J.M.F.T., (1989) Three-mode principal component Analysis and perfect congruence analysis for sets of covariance matrices, *British Journal of Mathematical ana Statistical Psychology*. 42**.** 63-80.

Kruger, U., Antory, D., Hahn, J., Irwin, G.W., McCullough, G., (2005). Introduction of A Nonlinearity Measure for Principal Component Models. *Computers and Chemical Engineering,* (29), 2355-2362.

Kruger, U., Zhou, Y., and Irwin, G.W., (2004). Improved Principal Component Monitoring of Large-scale Processes. *Journal of Process Control*, 14, 879-888.

Kruskal, J.B., and Wish, M., (1978). Multidimensional Scaling. California, USA: SAGE Publications.

Kumar, D.S., and Madhusree, K., (2011) Monitoring of Bioreactor using Statistical Techniques. *Research Journal of Chemical Sciences,* 1,3.

Ku, W., Storer, R. H., and Georgakis, C., (1995). Disturbance Detection and Isolation By Dynamic Principal Component Analysis. *Chemometrices and Intelligent Laboratory Systems*, 30, 179 – 196.

Lane, S., Martin, E. B., Koijman, R. and Morris, A. J. (1999). The Use of Generic Sub Space Models for the Monitoring of a Semi – Batch Process. IFAC World Congress, Beijing.

Lane, S., Martin, E. B., Koijmans, R. and Morris, A. J. (2001). Performance Monitoring of a Multi-Product Semi-Batch Process. *Journal of Process Control,* 11, 1-11.

Lee, D.S., Lee, M.W., Woo, S.H., Kim, Y.J., and Park, J.M., (2006). Nonlinear Dynamic Partial Least Squares Modelling of a Full-scale Biologiical Wastewater Treatment Plant. *Process Biochemistry,* 41, 2050-2057.

MacGregor, J. F., and Kourti, T., (1995). Statistical Process Control of Multivariate Processes. *Control Engineering Practice,* 3, 403 – 414.

Madsen, R.E., Hansen, L.K., and Winther, O., (2004) Singular Value Decomposition And Principal Component Analysis

Maestri, M., Farall, A., Groisman, P., Cassanello, M., and Horowitz, G., (n.d) A Robust Clustering For Detection Of Abnormal Situations In A Process With Multiple Steady-State Operating Modes, Argentina.

Mei, M., (2009) Principal Component Analysis

Mitchum, G.T., (n.d) Principal Component Analysis: Basic Methods and Extensions, *Department of Oceanography, University of Hawai, Manoa.*

Nomikos, P. (1996). Detection and Diagnosis of Abnormal Batch Operations Based On Multiway Principal Component Analysis. *ISA Transactions,* 35, 259 – 266.

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

Nomikos, P., and MacGregor, J.F., (1995). Multi-way Partial Least Squares in Monitoring Batch Processes. *Chemometrics and Intelligent Laboratory Systems,* 30, 97-108.

Overschee, P.V., and Moor B.D., (1996) Subspace Identification for Linear Systems. *Kluwer Academic Publishers*

Penha, R. M. L. and Hines, J. W. (2001). Using Principal Component Analysis Modeling to Monitor Temperature Sensors in a Nuclear Research Reactor. Maintenance and Reliability Conference (Marcon), Knoxville, TN.

Phatak, A., (1999). Whither Multivariate Statistical Process Monitoring. *NAmICS Newsletter.* No.19**.** April.

Raich, A. C., and Cinar, A., (1994). Multivariate Statistical Methods for Monitoring Continuous Processes: Assessment of Discrimination Power of Disturbance Models and Diagnosis of Multiple Disturbances. First International Chemometrics InterNet Conference (InCINC 1994).

Shiens, J., (2009) A tutorial on Principal Component Analysis. *New York City.*

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

Vidal, R., Ma, Y., and Sastry, S., (2005) Generalized Principal Component Analysis (GPCA), *IEEE Transactions On Pattern Analysis And Machine Intelligence*, 27,12.

Venkatasubramaniam, V., Renggaswamy, R., Kavuri, S.N., and Yin, K., (2003) A Review of Process Fault Detection and Diagnosis, *Computer and Chemical Engineering.* 27,293-346.

Wold, S., Kettaneh, N., MacGregor, J., and Kourti, D., (1999). Multivariate Process Monitoring. *NAmICS Newsletter,* No. 19. April.

Yoon, S., and Macgregor, J.F., (2004). Principal Component Analysis of Multi-scale Data for Process Monitoring and Fault Diagnosis. *AIChE Journal,* 50 (11), 2891-2903.

Young, G., and Householder, A.S., (1938). Discussion of a Set of Points in terms of Their Mutual Distances. *Psychometrika,* 3 (1).

Yu, H., and Zhang, J.J., (n.d.) An Extension of Principal Component Analysis, *National Centre for Computer Animation, Bournemouth University. U.K.*

Yunus, M.Y.M., (2012) Multivariate Statistical Process Monitoring [PowerPoint slides], Process Monitoring. Universiti Malaysia Pahang

Zhang, J., Martin, E.B., Morris, A.J., (1997). Process Monitoring Using Non-linear Statistical Techniques. *Chemical Engineering Journal*, 67, 181-189.

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

Zhao, S.J., Zhang, J., Xu, Y.M., (2004). Monitoring of Processes with Multiple Operating Modes through Multiple Principle Component Analysis Models. *Ind. Eng. Chem. Res,* 43, 7025-7035.

APPENDIX

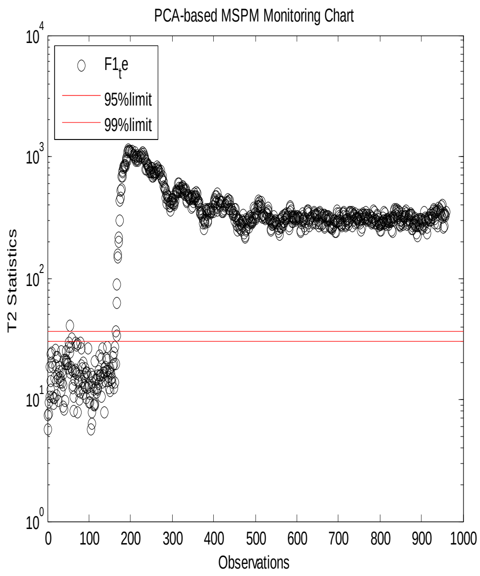
Appendix A

Table 4.4: Normal Operating Condition Variance of each mode

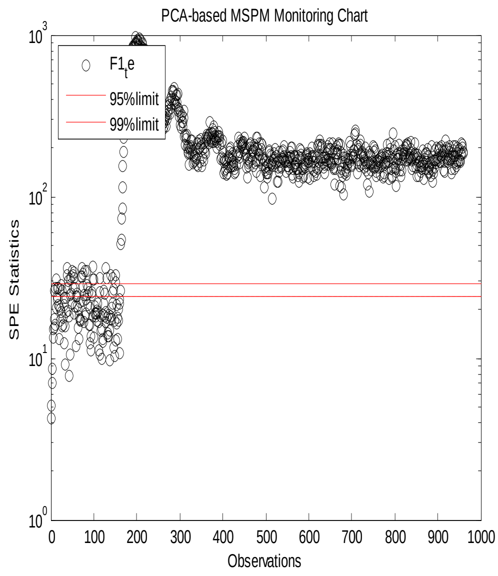
Fault	NPC 18			NPC 31		
	NOC	MODE 1	MODE 2	NOC	MODE 1	MODE 2
1	0.000815	0.000757	0.000876	0.000815	0.000757	0.000876
2	1026.023	896.1234	1158.955	1026.023	896.1234	1158.955
3	1006.447	1052.389	948.8497	1006.447	1052.389	948.8497
4	0.005854	0.005577	0.006152	0.005854	0.005577	0.006152
5	0.043541	0.042708	0.044546	0.043541	0.042708	0.044546
6	0.048638	0.047854	0.049955	0.04883	0.047854	0.049955
7	27.70322	19.41776	35.99501	27.70322	19.41776	35.99501
8	0.279002	0.288889	0.269493	0.279002	0.288889	0.269493
9	0.000348	0.000275	0.00042	0.000348	0.000275	0.00042
10	0.000137	0.00014	0.000134	0.000137	0.00014	0.000134
11	0.037726	0.031513	0.043972	0.037726	0.031513	0.043972
12	1.069353	1.101457	1.033217	1.069353	1.101457	1.033217
13	30.61652	21.758	39.50775	30.61652	21.758	39.50775
14	1.132037	1.071723	1.191732	1.132037	1.071723	1.191732
15	1.035486	0.977003	1.083558	1.035486	0.977003	1.083558
16	21.0923	15.95738	26.1791	21.0923	15.95738	26.1791
17	0.392885	0.376792	0.40568	0.392885	0.376792	0.40568
18	0.117225	0.136004	0.097734	0.117225	0.136004	0.097734
19	66.37931	80.84395	50.48489	66.37931	80.84395	50.48489
20	1.496126	1.322528	1.673472	1.496126	1.322528	1.673472
21	0.014232	0.01499	0.013506	0.014232	0.01499	0.013506
22	0.06176	0.062794	0.060966	0.06176	0.062794	0.060966
23	0.078979	0.074877	0.083261	0.078979	0.074877	0.083261
24	0.009808	0.009154	0.010113	0.009808	0.009154	0.010113
25	0.074212	0.079251	0.0689	0.074212	0.079251	0.0689
26	0.010496	0.009836	0.01119	0.010496	0.009836	0.01119
27	0.058737	0.061734	0.055823	0.058737	0.061734	0.055823
28	0.000636	0.000579	0.000693	0.000636	0.000579	0.000693
29	0.098922	0.082165	0.116076	0.098922	0.082165	0.116076
30	0.010375	0.010339	0.01044	0.010375	0.010339	0.01044
31	0.09865	0.084483	0.11132	0.09865	0.084483	0.11132
32	0.010819	0.01109	0.010416	0.010819	0.01109	0.010416

33	0.091703	0.086452	0.097124	0.091703	0.086452	0.097124
34	0.000689	0.000603	0.000778	0.000689	0.000603	0.000778
35	0.003347	0.002579	0.004104	0.003347	0.002579	0.004104
36	0.00275	0.002804	0.002689	0.00275	0.002804	0.002689
37	8.16E-05	7.41E-05	7.96E-05	8.16E-05	7.41E-05	7.96E-05
38	0.000207	0.000199	0.000214	0.000207	0.000199	0.000214
39	9.12E-05	9.84E-05	8.39E-05	9.12E-05	9.84E-05	8.39E-05
40	0.286971	0.273277	0.300976	0.286971	0.273277	0.300976
41	0.23923	0.17754	0.300645	0.23923	0.17754	0.300645
42	0.31155	0.310008	0.312823	0.31155	0.310008	0.312823
43	0.18297	0.184915	0.179847	0.18297	0.184915	0.179847
44	7.981287	7.39374	8.600689	7.981287	7.39374	8.600689
45	1.404937	1.523811	1.271541	1.404937	1.523811	1.271541
46	0.171138	0.14964	0.193299	0.171138	0.14964	0.193299
47	2.067926	2.11576	2.028227	2.067926	2.11576	2.028227
48	9.260773	9.538717	8.947894	9.260773	9.538717	8.947894
49	5.546086	5.232666	5.803699	5.546086	5.232666	5.803699
50	4.507303	5.434861	3.477856	4.507303	5.434861	3.477856
51	0.276211	0.242945	0.310578	0.276211	0.242945	0.310578
52	2.214309	2.11694	2.289996	2.214309	2.11694	2.289996

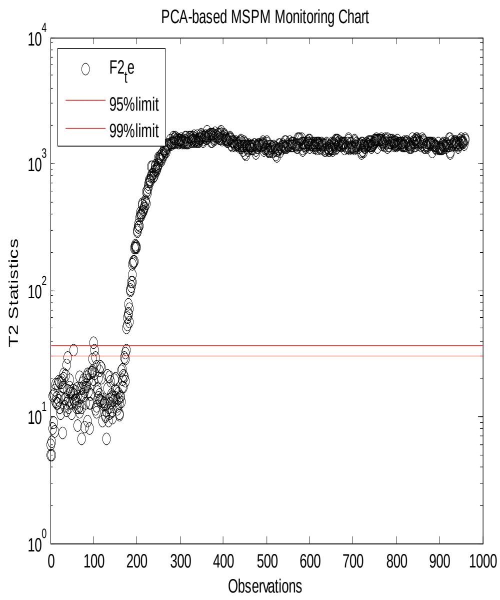
Appendix B



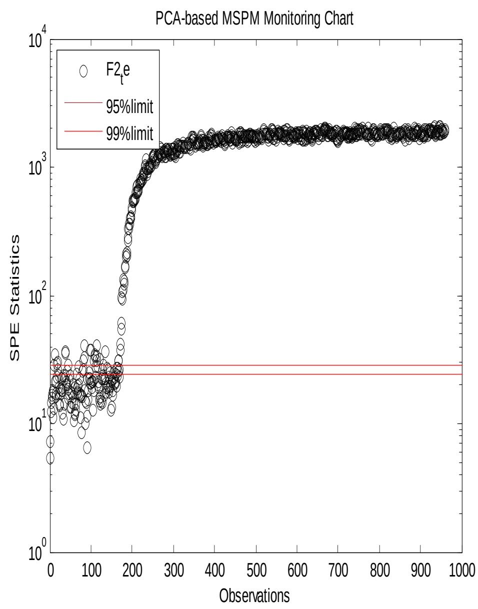
(a) T^2 statistic for fault 1



(b) SPE statistic for fault 1



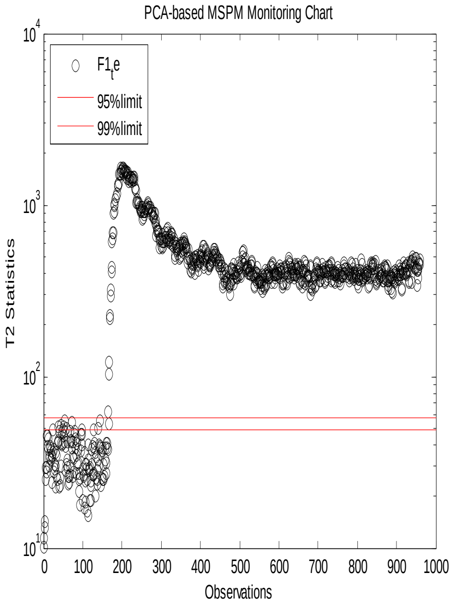
(c) T^2 statistics for fault 2



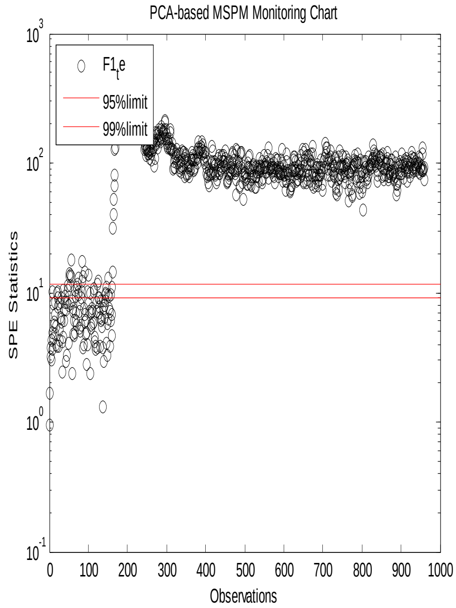
(d) SPE statistics for fault 2

Mode I: T^2 statistics and SPE statistics for fault 1and 2 for 18 pc of 70% total variance

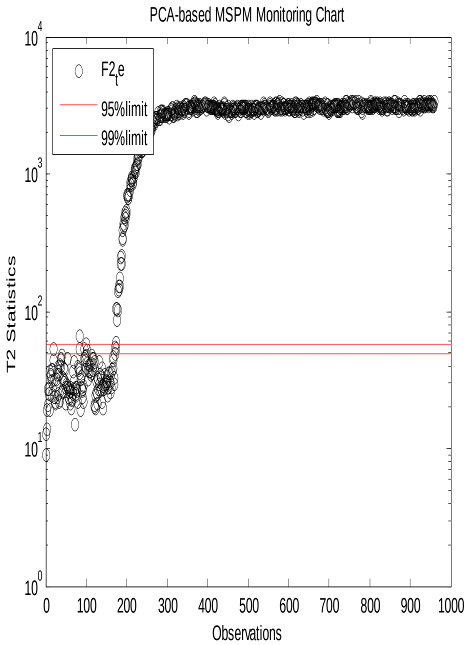
Appendix C



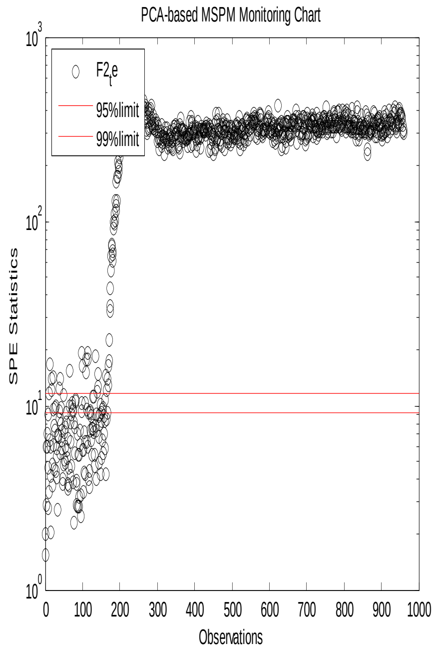
(a) T^2 statistic for fault 1



(b) SPE statistic for fault 1

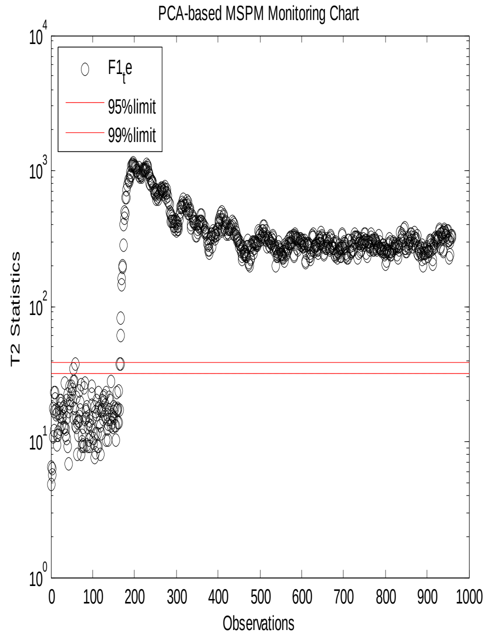


(c) T^2 statistic for fault 2

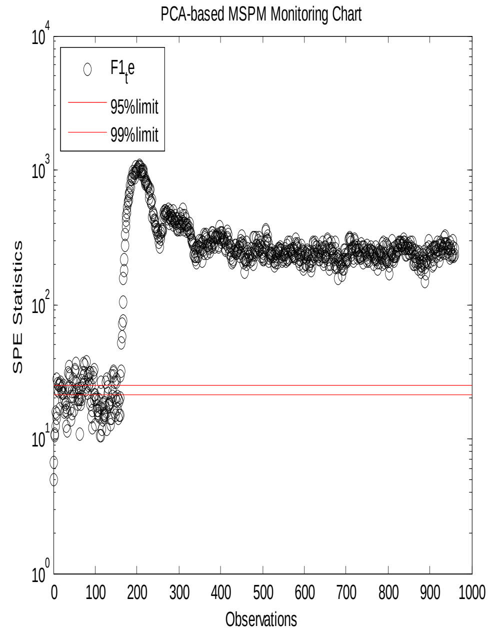


(d) SPE statistic for fault 2

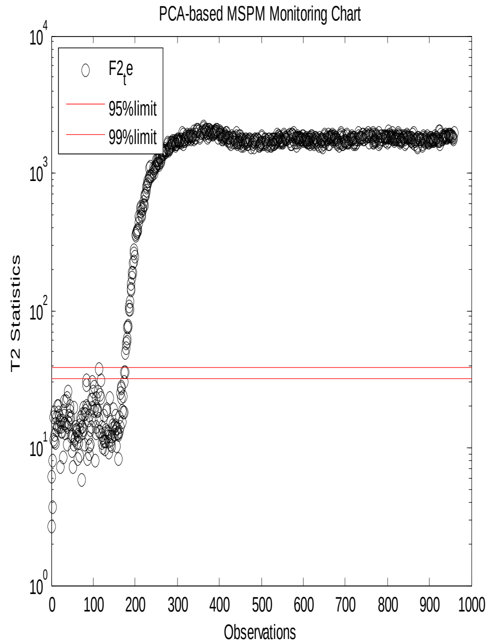
Mode I: T^2 statistics and SPE statistics for fault 1 and 2 for 31 pc of 90% total variance



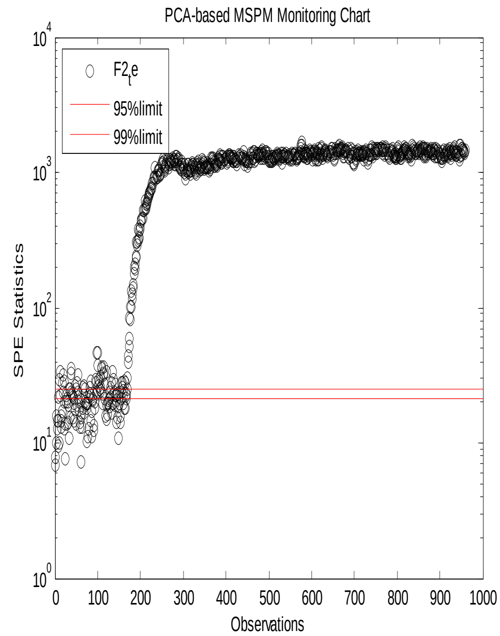
(a) T^2 statistics for fault 1



(b) SPE statistics for fault 1



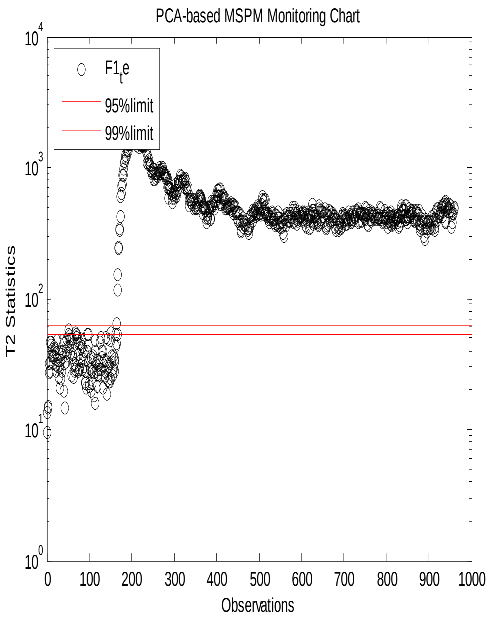
(c) T^2 statistics for fault 2



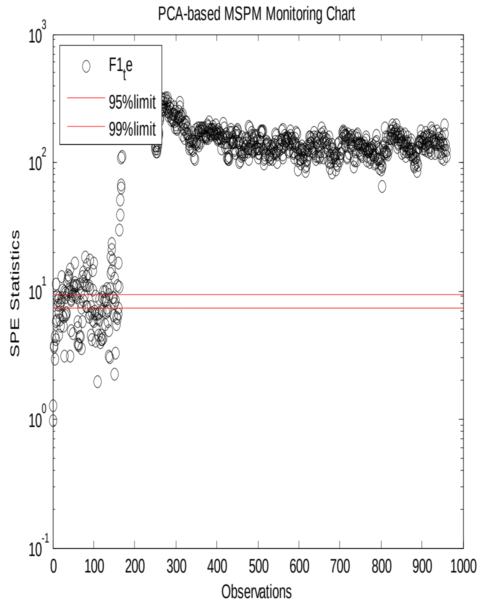
(d) SPE statistics for fault 2

Mode II:T^2 statistics and SPE statistics for fault 1 and 2 for 18 pc of 70% total variance

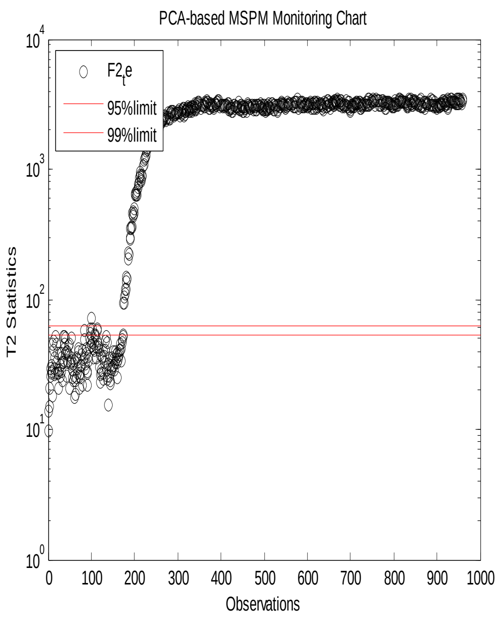
Appendix E



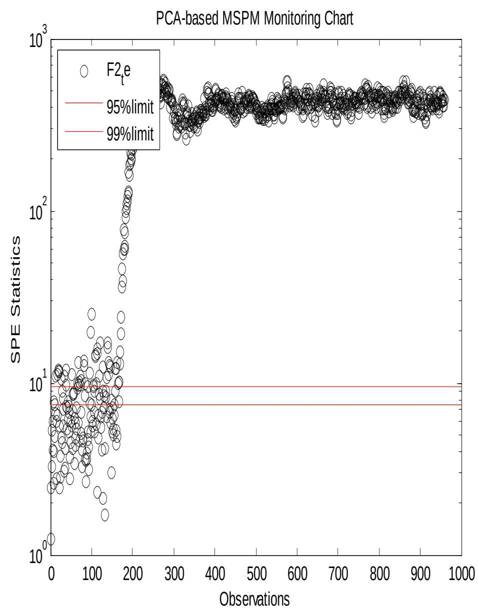
(a) T^2 statistics for fault 1



(b) SPE statistics for fault 1

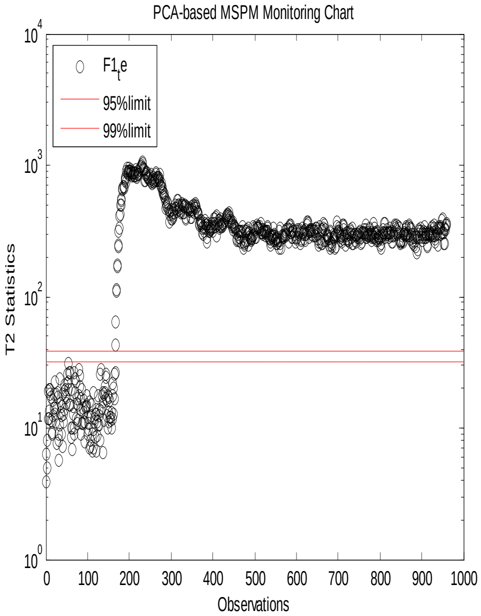


(c) T^2 statistics for fault 2

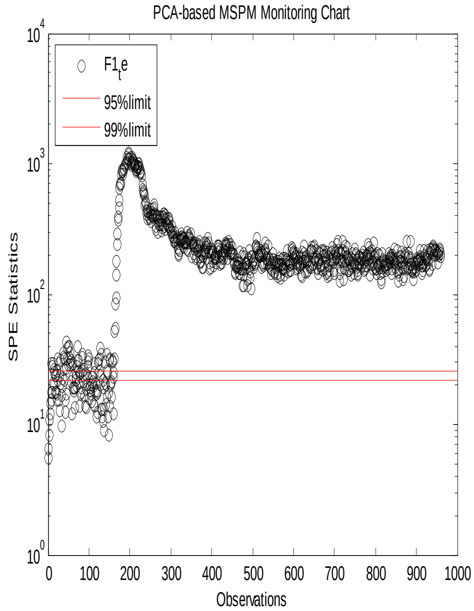


(d) SPE statistics for fault 2

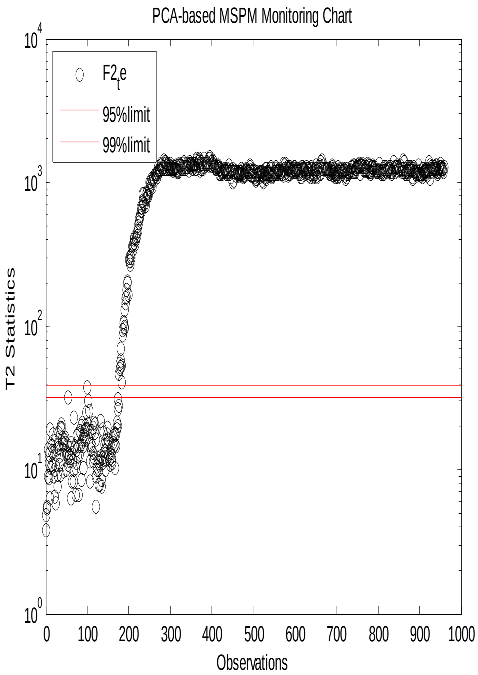
Mode II: T^2 statistics and SPE statistics for fault 1 and 2 for 31 pc of 90% total variance



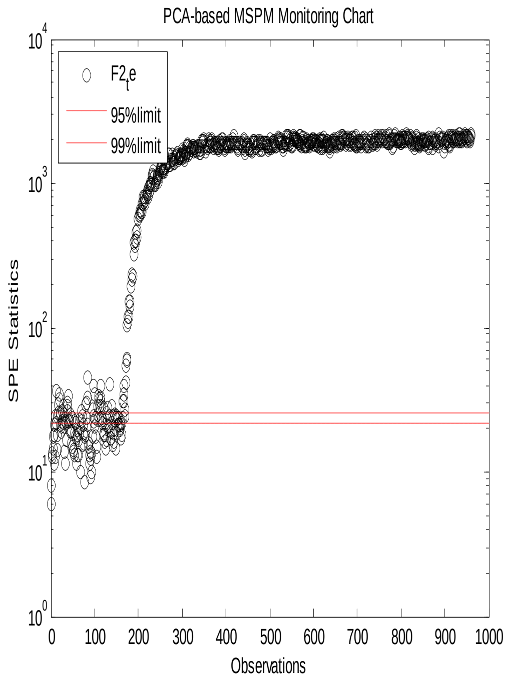
(a) T^2 statistics for fault 1



(b) SPE statistics for fault 1

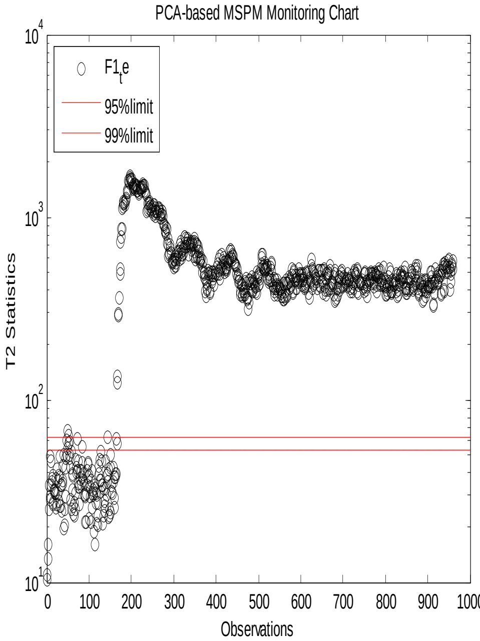


(c) T^2 statistics for fault 2

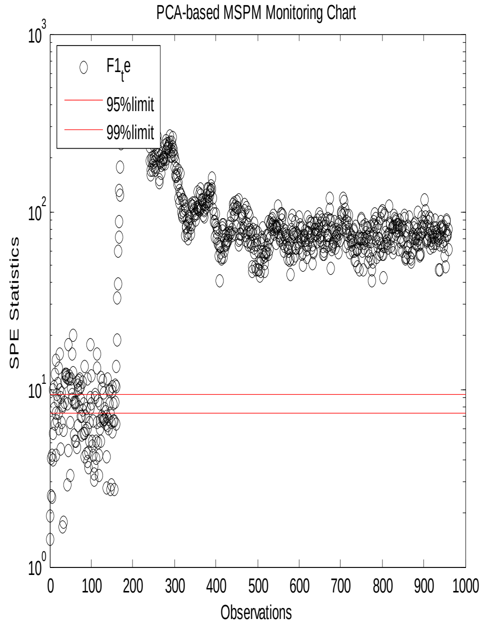


d) SPE statistics for fault 2

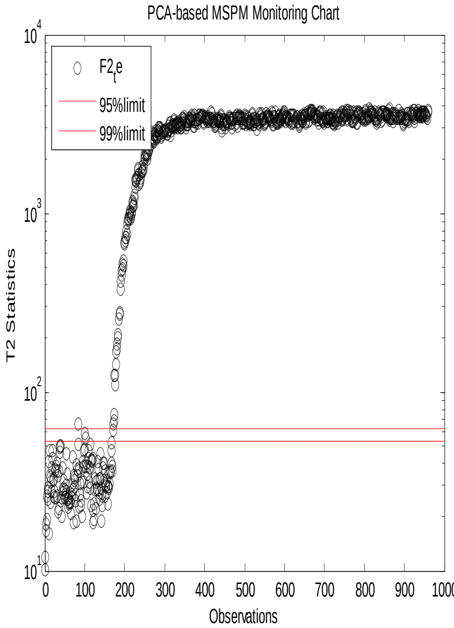
Mode III: T^2 statistics and SPE statistics for fault 1 and 2 for 31 pc of 90% total variance



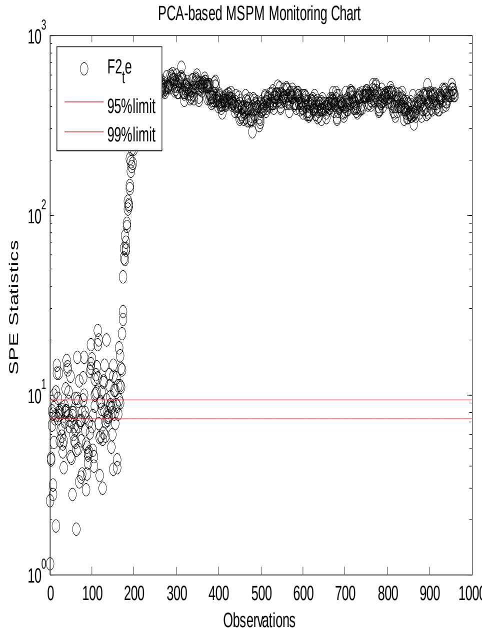
(a) T^2 statistics for fault 1



(b) SPE statistics for fault 1



(c) T^2 statistics for fault 2



(d) SPE statistics for fault 2

Mode III: T^2statistics and SPE statistics for fault 1 and 2 for 31 pc of 90% total variance