IMPLEMENTING PCA-BASED FAULT DETECTION SYSTEM BASED ON SELECTED IMPORTED VARIABLES FOR CONTINUOUS-BASED PROCESS

SITI NUR LIYANA BINTI AHAMD

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

Faculty of Chemical Engineering
UNIVERSITI MALAYSIA PAHANG

FEBRUARY 2013

ABSTRACT

Nowadays, the production based on chemical process was rapidly expanding either domestically or internationally. To produce the maximum amount of consistently high quality products as per requested and specified by the customers, the whole process must be considering included fault detection. This is to ensure that product quality is achieved and at the same time to ensure that the quality variables are operated under the normal operation. There were several methods that commonly used to detect the fault in process monitoring such as using SPC or MSPC. However because of the MSPC can operated with multivariable continuous processes with collinearities among process variables, this technique was used widely in industry. In MSPC have a few methods that were proposed to improve the fault detection such as PCA, PARAFAC, multidimensional scaling technique, partial least squares, KPCA, NLPCA, MPCA and others. Here, in this thesis was to proposed new technique which was by implementing PCA-based fault detection system based on selected imported variables for continuous-based process. This technique was selected depends on the highest number of magnitude of correlation of variables using Matlab Software. The result in this thesis was the fault can be detected using only selected important variables in the process.

ABSTRAK

Kini, pengeluaran berdasarkan proses kimia telah berkembang pesat sama ada domestik atau antarabangsa. Untuk menghasilkan jumlah maksimum produk yang berkualiti tinggi secara konsisten seperti diminta dan yang dinyatakan oleh pelanggan, keseluruhan proses mesti mempertimbangkan termasuk pengesanan kesalahan. Ini adalah untuk memastikan bahawa kualiti produk dapat dicapai dan pada masa yang sama untuk memastikan bahawa pembolehubah kualiti dikendalikan di bawah operasi normal. Terdapat beberapa kaedah yang biasa digunakan untuk mengesan kesalahan dalam pemantauan proses seperti menggunakan SPC atau MSPC. Walau bagaimanapun, kerana MSPC boleh dikendalikan dengan proses pembolehubah berterusan dengan collinearities di kalangan pembolehubah proses, teknik ini telah digunakan secara meluas dalam industri. Dalam MSPC mempunyai beberapa kaedah yang telah dicadangkan untuk meningkatkan pengesanan kerosakan seperti PCA, PARAFAC, multidimensi teknik bersisik, separa kuasa dua terkecil, KPCA, NLPCA, MPCA dan lain-lain. Di sini, dalam tesis ini terdapat teknik baru yang dicadangkan iaitu dengan melaksanakan sistem pengesanan PCA berasaskan kesalahan berdasarkan pembolehubah terpilih yang terpenting untuk proses yang berterusan berasaskan. Teknik ini telah dipilih bergantung kepada bilangan tertinggi magnitud korelasi pembolehubah menggunakan Perisian Matlab. Hasil dalam tesis ini adalah kesalahan boleh dikesan menggunakan hanya dipilih pembolehubah penting dalam proses.

TABLE OF CONTENTS

LIST OF FIGURE

LIST OF TABLES

LIST OF SIMBOLS

A	Number of PCs retained in the PCA model.
C	Covariance matrix.
i	Row
j	Column
k	Principal component
n	Number of nominal process measurements per variable.
ns	Number of samples.
P	PCA model.
R	Correlation matrix.
V	Eigenvectors.
X	NOC data.
SPE_i	SPE statistics.
$p_{i,j}$	i^{th} score for Principal Component j.
λ	Eigenvalue.
λ_j	Eigenvalue corresponds to Principal Component j.
e_i	i^{th} row in residual matrix.
z_α	Standard normal deviate corresponding to the upper (1-α) percentile.
σ	Standard deviation
Λ	Diagonal matrix.
V^T	Normalized orthogonal matrix.
$\bar{x}$	Data means.
$\check{X}$	Standardized data

LIST OF ABBREVIATIONS

KPCA	Kernel principal component analysis.
MPCA	Moving principal component analysis
MSPC	Multivariate statistical process control.
NLPCA	Nonlinear principal component analysis.
NOC	Normal operation condition.
ND	Not Detected
PCA	Principal component analysis.
PCs	Principal components.
SPC	Statistical process control
SPE	Squared prediction errors.
T^2	Hotelling's T^2.

CHAPTER 1

INTRODUCTION

1.1 Research Background

Process monitoring can be define as the observation of chemical process variables by means of pressure, temperature, flow, and other types of indicators usually occurs in a central control room (The McGraw-Hill Companies, 2003). This process is very important because the fault detection can be determine early in the process that means the high quality of the products that requested and specified by the customers can be achieved.

In the process monitoring, there are two types of monitoring schemes that are commonly used in chemical-based industry which are individual-based monitoring or known as statistical process control (SPC) and multivariate-based monitoring also called multivariate statistical process control (MSPC). Studies by Yu, (2010), traditional SPC have been widely used, where the usual practice has been to maintain a separate chart like Shewhart (univariate control chart) for each process variable. However, most chemical process operations are multivariable continuous processes

with collinearities among process variables that are why SPC always make a false result in the process. SPC also could result in many fault alarms when the process variables are highly correlated in processes (Yu, 2010). According to Raich, & Cinar, (1995), SPC methods do not makes the best use of data available and often make an error during the multivariate chemical process. It is true where in univariate system, the process shows that the system is in control process but in multivariate actually the system is out of control process.

Because of the weakness from the SPC, the advance SPC is use which is multivariate SPC (MSPC) has been introduced. MSPC techniques have been successfully used to detect and identify departure from normal operation within industrial processes (Yu, 2010). Besides, followed by MacGregor, & Kourti (1995), MSPC methods can separate confirming information from observations on extract confirming information from observations on many variables and can reduce the noise levels through averaging.

1.2 Problem Statement and Motivation

Basically, in process monitoring the Principal Component Analysis (PCA) is one of the branches found in the MSPC has been used widely and the method is very popular in non-linear continuous application process. The basic strategy of PCA that have been reported by Jeng, (2010) is to discard the noise and collinearity between process variables, while preserving the most important information of the original data set. However, a PCA technique is used to develop a model describing variation under normal operating conditions (NOC). It is utilized to detect outliers from NOC,

as excessive variation from normal target and as unusual patterns of variation (Raich et al 1996). In addition, PCA have been used for detection of out-of-control status, identification of the process variables with significant variations and diagnosis of source causes (Raich et al., 1995).

PCA it is actually appropriate for linear process but in industry nowadays used highly nonlinear process which can give a lot of problems. Based on the Yu, (2010), the nonlinear and multimodal characteristics in some processes have posed difficulties to the conventional approaches, because a fundamental assumption is often that the operating data is unimodal and Gaussian distributed. The typical nonlinearity or multimodality makes modelling of PCA -based T^2 and SPE charts difficult, which results in the decreasing of their monitoring performance, significantly (Yu, 2010). Therefore, from that problem, the alternative method or techniques was proposed to solve the current problem in the process monitoring to achieved high quality product as per requested and specified by the customers. The main concept in this new method was used less number of variables due to the less number of Principle Components (PCs) used. Based on this number of PCs, number of important variables selected based on the high number of magnitude of correlation. From that, the fault detection time can detected more faster compared with conventional PCA.

1.3 Research Objectives

1.3.1 To conduct the conventional PCA-based MSPC system.

1.3.2 To determine whether by choose some of variables can detect the fault for Continuous-based process.

1.3.3 To compare and analyse the monitoring performance between the conventional PCA and New PCA.

1.4 Research Questions

1.4.1 Does the fault can be detect based on the selected important variables for Continuous-based process?

1.4.2 Does the high number of magnitude of correlation can detect the fault?

1.5 Scope of Proposed Study

1.5.1 This research is mainly focus to select only the important variable to determine the fault detection in the process.

1.5.2 Furthermore, these researches only focus to the high magnitude of correlation of the variable in order to simplify and easier the work.

1.5.3 The number of the magnitude of correlation is choosing above 70%, 80%, and 90% to identify if this value can contribute to the determining of the fault detection.

1.5.4 This research is used the Continuous-Stirred Tank Reactor (CSTR).

1.5.5 Besides, the natures of the fault that are wanted to be focus are fault cases number 9 and 10. Fault number 9 is about Pipe 4, 5, or 6 is blocked or control valve 3 fails low and fault number 10 is control valve 3 fails high.

1.5.6 Next, this research also assume to used operating mode 2 which is by select one sample at 13 number of variables (NOC data) and another is same number of sample but reduce the number of variable.

1.5.7 To determine the result, the Shewhart chart is introduced to shows and detect the of the out-of control status of the variables.

1.5.8 All the steps that want to be focus only to the fault detection is used Matlab software.

1.6 Significant of Proposed Study

This research have proposed new method that can give some advantages where by only selected a few number magnitude of correlation of variables, the fault can be detect an earlier and can save times to simplified works. This study will examine if by select only the high number of magnitude of correlation, the fault time detection will be faster compared with conventional PCA.

1.7 Report Organization

As a conclusion, process monitoring is important in order to detect the fault earlier in the process to ensure the quality of the product can be achieve, safety of the process is under control and reduce the cost of the material use. There are two types of the monitoring schemes which are individual-based monitoring that used statistical process control (SPC) and multivariate-based monitoring that used multivariate SPC (MSPC). Basically, the MSPC is better than SPC because SPC can only monitor one variable at one time but MSPC can monitor many variables at one time. This thesis is divided into five main chapters.

Chapter 1 introduces the background of the research which includes the problem statement and motivation, objectives, scopes and significant of proposed study. In Chapter 2, which is literature review it will discuss about the fundamental or theory of PCA and more about process monitoring issues of PCA. In Chapter 3, it will discuss the procedure of the fault detection and identification which consist of two main phases namely as off-line modelling and monitoring (Phase I) and on-line monitoring (Phase II). Besides, in Chapter 4 will discuss details about the results that obtain from this thesis which is the overall monitoring performance for first phase and second phase. Lastly, for the Chapter 5 will discuss about the conclusion and recommendation.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

In process monitoring there were several methods that used in detecting the fault in the process. Nowadays, the chemical process exists in multivariable process and the MSPC was used widely to solve this problems. From that many researchers trying to improve this MSPC by proposed a new technique. This chapter divided in three sections which were Principal Component Analyses (PCA), Principal Component Analysis (PCA) extension and conclusion.

2.2 Principal Component Analysis (PCA).

Principal component analysis (PCA) was a technique that uses to detect out the deviation such as the excessive variation from the normal operating data (NOC) patterns of variation. When the out of control data was detected the observation was

compared to PCA models for known the disturbance or fault. PCA has been successfully applied to the monitoring of industrial processes, including chemical and microelectronics manufacturing processes (Li, Yue, Cervantes, & Qin, 2000). This situation was also supported by Bhavik (1998) in his journal where PCA was used to solve several tasks including, data rectification (Kramer & Mah, 1994), gross error detection (Tong and Crowe, 1995), disturbance detection and isolation (Ku et al., 1995), statistical process monitoring (Kresta et al., 1991; Wise et al., 1990), and fault diagnosis (MacGregor et al., 1994; Dunia et al., 1996).

However this PCA which was one of the MSPC techniques only suitable for linear operation but nowadays most chemical process in industry were multivariable continuous process with collinearities among process variables. In the study by Raich & Cinar (1996), the collinearities generate strong departures from the assumptions utilized in developing this approach and limit its usefulness in multivariable processes. For an example shows in figure below:

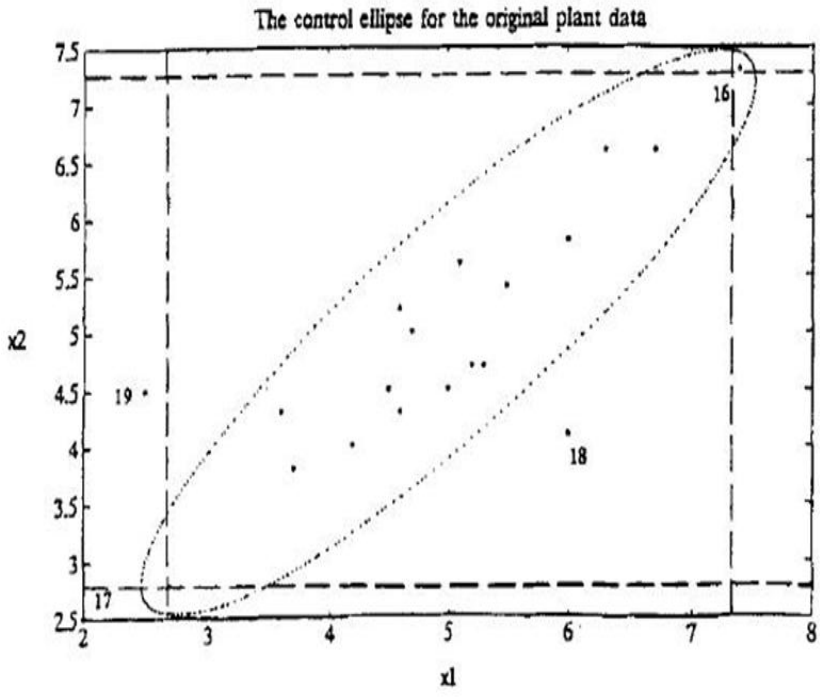

Figure 2.1: Multivariate control ellipse based on Hotelling T2 for x1 and x2.

Figure above shows two-variable process (x1 and x2) was monitored by using the control ellipse based on Hotelling T^2. Actually, based on the Shewhart charts, the process is out of control at samples 16, 17, and 19 but using multivariate analysis, sample 16 is inside the control ellipse and the process is in control. Furthermore, sample 18, which indicates that the process is in control based on the Shewhart charts, is outside the control ellipse and the process is out of control. On the other hand, Raich & Cinar (1996) also stated that Hotelling T^2 can detect the out-of-control status consistently, but offers no assistance in identifying the variables responsible for this status.

Next, in the geometry of principal components models the new coordinate system was introduce which was Principal Components (PCs) to detect the fault. PCs were a new set of axes in which the data points are randomly scattered, usually with fewer dimensions than the original measurements like shows in Figure 2.2. Furthermore, in the study by Raich & Cinar (1996) data set that is well described by two PCs which the data can be displayed in a plane. The scatter is enclosed in an ellipse whose axes are the PC loadings. However, for higher than two or three dimensions, graphical displays do not provided a clear picture.

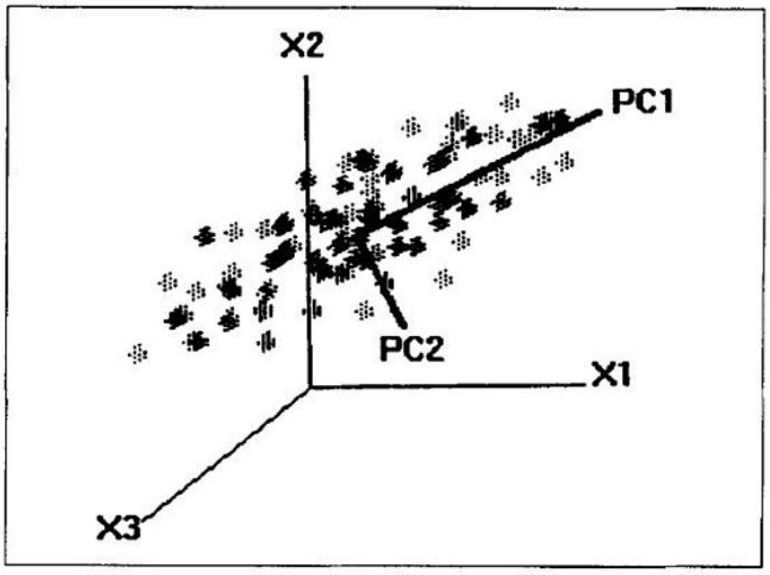

Figure 2.2: Principal components of multivariate data.

Actually, before this graph was plotted, the number of PC in the fault detection method using PCA is considered. As proposed by Tamura & Tsujita (2007), there are known criteria for selecting the number of PCs, such as the scree plot, eigenvalue limit, cumulative contribution limit, cross validation, variance of reconstruction error (VRE) criterion and so on. Moreover, Kano et al. (2002) also mentioned that fault detection ability depends on the number of PCs retained in the PCA model (Tamura & Tsujita, 2007). Because of the PCA only for linear correlation and have some difficult in nonlinear a linear PCA mapping results in substantial loss of information or large numbers of linear components are required to obtain the required accuracy. Therefore, there are several methods such as PCA extension has been introduced and will be discussed in the next subtopic.

In MSPC method have several steps for process monitoring performance and fault diagnosis. Figure 2.3 shows that there were four main steps which consist of fault detection, fault identification, fault diagnosis and process recovery.

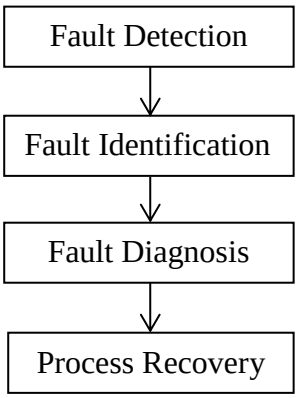

Figure 2.3: Overall Framework

Firstly, the fault detection was to purpose the deviation of observed samples from a normal range using a set of parameters. Secondly fault identification was the identifying the observed process variables that are most related to the fault which is normally identified by using the contribution plot technique. Thirdly, was fault diagnosis which was to specifically establish the type of fault which has been possibly (and should be also validated) contributed to the signal. Lastly was process recovery where to eliminate the causes that contribute to the detected fault (Yunus, 2012).

2.3 Principal Component Analysis (PCA) extension.

Aside from PCA method, there were several methods used in the MSPC to detect the fault such as Kernel Principal Component Analysis (KPCA) model, Nonlinear Principal Component Analysis (NLPCA) methods and Moving Principal Component Analysis (MPCA). These three methods were proposed to address multivariate process performance monitoring and in particular fault diagnostics in nonlinear processes. Some of this method perhaps has own advantage or disadvantage in order to detect fault.

Firstly, Kernel Principal Component Analysis (KPCA) model which was proposed by Choi, Morris & Lee (2008) which not only captures nonlinear relationships between variables but also reduces the dimensionality of the data. In their thesis, the KPCA approach conceptually consists of two steps which were the extended nonlinear mapping of measurements in the original space into the extended feature space and the construction of PCA in the feature space. These two steps were completely carried out by using kernel functions without knowledge of the nonlinear functions and without the need to solve any optimization procedure. Besides, KPCA was accomplished to give a nonlinear multiscale model with the reconstructed signals in the time domain being separately transformed from all the scales in wavelet domain. Choi et al. (2008) also state that KPCA is a type of kernel-based learning machine (Schölkopf et al., 1998 & Müller et al., 2001). PCA finds principal components minimizing the loss of data information in the input space, whereas KPCA searches such components in the extended feature space. Furthermore, there were some advantage of KPCA which was no optimization procedure is involved unlike nonlinear PCA based on neural networks. Besides that Choi et al. (2008) said

that KPCA approach the straightforward eigenvalue problem was solved to compute the principal loading vectors in the feature space. From that, it has proven to be very powerful in the detection of faults in the nonlinear processes (Choi et al. 2005).

Secondly was Nonlinear Principal Component Analysis (NLPCA) method which was proposed by Maulud, Wang & Romagnoli (2006). This method has been proposed in the literature to improve the data extraction when the nonlinear correlations among the variables exist. Conventional NLPCA approaches were capable of projecting the multi-dimensional data to a lower dimensional data. Maulud et al. (2006) also mentioned that for nonlinearly correlated process data, a nonlinear PCA model based on auto associative neural network NLPCA model which employs a feed forward structure with a bottleneck layer to represent the nonlinear principal components (Kramer, 1991). NLPCA will be represented by the outputs of the mapping network. If the network training is properly conducted and reasonable approximation has been achieved, the data input features must be well represented. Although NLPCA one of the PCA extensions, but there was some shortcomings in this method where the data information tends to be evenly distributed among the principal components. In view to this drawback, a training algorithm for NLPCA needs to be incorporated in which the nonlinear scores produced are orthogonal at the end of training session (Maulud et al. 2006). Besides that, Maulud et al (2006) also stated in their research where the data information tends to be evenly distributed among the principal components or bottleneck layer thus, losing the inherent orthogonality characteristics of linear PCA.

Lastly was Moving Principal Component Analysis (MPCA) where in MPCA changes in the direction of each principal component or changes in the subspace spanned by several principal components were monitored (Kano, Hasebe, Hashimoto

& Ohno, 2001). In other words, changes in the correlation structure of process variables, instead of changes in the scores of predefined principal components were monitored. However, Kano et al. (2001) also reported that MPCA can detect a change of correlation among process variables, which is difficult to detect by conventional MSPC with T^2 and Q. Besides that, Kano et al. (2001) proposed MSPC method was based on the idea that a change of operating condition, represented by a change of correlation among process variables, can be detected by monitoring directions of principal components. On the other hand, MPCA changes in the direction of each PC or changes in the subspace spanned by several PCs are monitored on-line. MPCA can detect changes in the operating condition even when the deterministic changes in the monitored variables are not significant and the variances were not increased, because MPCA monitors the correlation among process variables. MPCA also have some disadvantage where since MPCA has a smoothing effect, which is caused by the use of a time-window, MPCA suffers from the delay in detecting a fault and also in detecting a return to the normal operating condition.

Based on these three methods it shows that every PCA extension has their own method to solve the fault detection. Here, the new technique by selected the important variable where just concentrate to the high number of magnitude correlation of the variable was introduce to fault detection. The selected important variables were selected based on the number of PCs. If less number of PCs used means less number of variables also used in this fault detection. Only three numbers of PCs was used and the graph of transform variance was obtained. The value of graph shows approach to 1 which was most suitable to detect the fault. This new