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

NUR AFIFAH BINTI HASSAN

BACHELOR OF CHEMICAL ENGINEERING
UNIVERSITI MALAYSIA PAHANG

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

NUR AFIFAH BINTI HASSAN

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

Faculty of Chemical Engineering and Natural Resources
UNIVERSITI MALAYSIA PAHANG

FEBRUARY 2013

ABSTRACT

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

ABSTRAK

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

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

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

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

LIST OF ABBREVIATIONS

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

SPE	Squared prediction errors

CHAPTER I

INTRODUCTION

1.1 Research Background

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

Traditionally, SPC performs a toolkit for managing process malfunction by way of providing early warning through fault detection (Montgomery, 1985; Grant and Leavenworth, 1988; Wetherill and Brown, 1991). The control chart is one of the key tools which are used to monitor the processes that are in control by using mean and range. According to Cinar, Palazoglu and Kayihan (2007), the generic purpose of statistical process control (SPC) is to detect the nature of faults in the process that lead to

disastrous deviation from the desired goal. Among others, the main procedures should include data collection, control chart development, and followed by control chart progression analysis. The next step involves process diagnosis, which is to find the root cause of the changes as well as execute corrective actions corresponding to the nature of the faults. Thus, on-line monitoring and diagnosis are important to ensure that high quality product can be maintained over the period of operations (MacGregor, 1994).

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

1.2 Problem Statement and Motivation

Over the last decade, the field of the process monitoring performance and fault diagnosis in chemical process industry has used MSPM as an alternative method based on the existing knowledge. One of the tools multivariable statistical techniques is Principal Component Analysis (PCA) and its extension which can indicate the strong correlations of the data set through a set of empirical orthogonal function (Cinar, et al., 2007). However, "conventional PCA-based MSPM is only valid for the non-autocorrelated data with linear relationships between measured variables. Often, inefficient and unreliable process performance monitoring schemes can materialize as a consequence of the underlying assumptions of PCA-based MSPM being violated" (Choi, Morris and Lee,

2008). Furthermore, based on the study of Choi, Martin and Morris (2005), a large amount of the principal components are retained to clarify a large proportion of the sample variance when dealing with the non-linear relationship between measured variable. Simultaneously, this leads to an increase in the probability of false alarms which happen especially in the T^2 statistic and result of the decrease in order of the components that only explain minimum level of variability.

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

1.3 Research Aims and Objectives

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

i. To run the conventional PCA-based MSPM system.

ii. To develop the dissimilarity-based MSPM system.

iii. To compare and analyse the monitoring performance between the conventional PCA and dissimilarity techniques.

1.4 Research Questions

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

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

1.4.3 Do the outcomes support the research aim?

1.5 Research Scopes

The research scopes of this research are listed as follow:

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

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

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

iv. Focusing on the fault detection scheme only.

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

vi. Using Shewhart chart to monitor the process performance.

vii. Using CSTRwR system as a case study.

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

1.6 Expected Research Contributions

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

1.7 Chapter Organizations

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

CHAPTER II

LITERATURE REVIEW

2.1 Introduction

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

2.2 Fundamental of MSPC

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

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

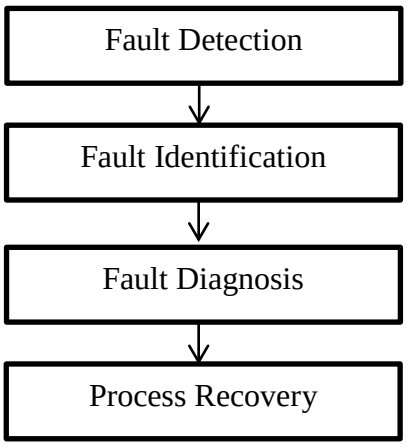

Figure 2.1 Main steps in MSPM system

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

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

description and explain pattern of variation in multivariate data sets. According to Romagnoli and Palazoglu (2006), PCA is one of the multivariate statistical techniques which are basically classified as dimensionality reduction methods. The definitions of PCA from all researchers are quite similar to each other.

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

2.3 Process Monitoring Issues and Extension

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

2.3.1 Process Monitoring Extension based on PCA

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

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

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

where T is the matrix of non-linear PC scores, F(.) the non-linear PC loading function and E the residual matrix. The concept of Input-Training neural network (the IT-net) is based on non-linear methods. Each input pattern was irregular but is adjusted with the internal network parameters to generate the same output pattern based on the steepest