VOTE 74215

DEVELOPMENT OF FAULT DETECTION, DIAGNOSIS AND CONTROL SYSTEM IDENTIFICATION USING MULTIVARIATE STATISTICAL PROCESS CONTROL (MSPC)

(PEMBINAAN SISTEM PENGESANAN, DIAGNOSIS DAN PENGAWALAN SISTEM IDENTIFIKASI MENGGUNAKAN PROSES KAWALAN MULTI PEMBOLEHUBAH STATISTIK (MSPC))

KAMARUL ASRI BIN IBRAHIM
ARSHAD AHMAD
MOHAMED WIJAYANUDDIN ALI
NOORLISA HARUN
NORAZANA IBRAHIM
MAK WENG YEE

RESEARCH VOTE NO:
74215

Jabatan Kejuruteraan Kimia
Fakulti Kejuruteraan Kimia Dan Kejuruteraan Sumber Asli
Universiti Teknologi Malaysia

2006

ABSTRACT

Processes exhibit complex behavior in chemical industries which makes the development of reliable theoretical models a very difficult and time consuming task. The resulting models are also often complex which poses additional problem for robust on-line process fault detection, diagnosis and control of these processes. Efficient process fault detection and diagnosis in processes is important to reduce the cost of producing products with undesired specifications. Multivariate Statistical Process Control (MSPC) uses historical data of processes to develop useful process fault detection, diagnosis and control tools. Thus, the availability of theoretical models is not an important factor in the implementation of MSPC on processes. The present fault detection and diagnosis (FDD) method based on MSPC uses statistical control charts and contribution plots. These charts are efficient in fault detection but ambiguous in diagnosis of fault cause of detected faults due to the absence of control limits in the contribution plots. In this research work, an FDD algorithm is developed using MSPC and correlation coefficients between process variables. Normal Correlation (NC), Modified Principal Component Analysis (PCA) and Partial Correlation Analysis (PCorrA) are used to develop the correlation coefficients between selected key process variables and quality variables of interest. Shewhart Control Chart (SCC) and Range Control Chart (RCC) are used with the developed correlation coefficients for FDD. The developed FDD algorithm was implemented on a simulated distillation column which is a single equipment process. Results showed that the developed FDD algorithm successfully detect and diagnosed the pre-designed faults. The implementation of the developed FDD algorithm on a chemical plant can reduce the operational cost due to early detection and diagnosis of faults in the process and improving the performance of the plant.

ACKNOWLEDGEMENT

In preparing this research, many people, researchers, academicians, and practitioners were in contact with. First and foremost, the authors would like to thank Government of Malaysia and Universiti Teknologi Malaysia (through IRPA Grant Vot 74215) for sponsoring this research.

The authors would also like to thank the following people who have helped in completing this research:

- Mr.Wong Teck Siang for his help in providing all the data for the process.
- Miss Heng Han Yann for her help.

Our sincere appreciation also extends to others who have provided us in direct or non-direct assistance.

ABSTRAK

Proses dalam industri kimia mempamerkan kelakuan yang rumit. Ini menyebabkan pembangunan model yang berkesan amat menyusahkan dan mengambil masa yang panjang. Model-model yang dibangunkan sering kali rencam dan mengemukakan banyak masalah tambahan dalam pengesanan kecacatan, diagnosis dan kawalan bagi proses-proses ini. Pengesanan dan diagnosis kecacatan yang cekap adalah penting dalam suatu proses bagi mengurangkan kos penghasilan produk dengan spesifikasi yang tidak diingini. Kawalan Proses Multipembolehubah Statistik (MSPC) menggunakan data proses masa lampau dalam pembangunan kaedah pengesanan kecacatan, diagnosis dan kawalan proses. Oleh sebab itu, kehadiran model-model teori adalah faktor yang tidak penting dalam pelaksanaan MSPC di dalam proses. Pengesanan dan diagnosis kecacatan (FDD) berdasarkan MSPC pada masa kini menggunakan carta-carta kawalan statistik dan carta-carta sumbangan. Carta-carta ini adalah cekap dalam pengesanan kecacatan tetapi tidak meyakinkan dalam diagnosis punca kecacatan disebabkan ketiadaan had kawalan di dalam carta-carta sumbangan. Dalam kerja penyelidikan ini, algoritma FDD dibangunkan menggunakan MSPC dan pekali-pekali korelasi antara pembolehubah-pembolehubah proses. Korelasi Normal (NC), Komponen Analisis Prinsipal yang diubah suai (PCA) dan Korelasi Analisis Separa (PCorrA) diguna untuk menerbitkan pekali-pekali korelasi antara pembolehubah-pembolehubah kunci proses dengan pembolehubah-pembolehubah kualiti yang dikaji. Carta Kawalan Shewhart (SCC) dan Carta Kawalan Julat (RCC) diguna bersama dengan pekali-pekali korelasi yang telah diterbitkan untuk tujuan FDD. Algoritma FDD yang telah dibangunkan dilaksanakan menggunakan sebuah alat iaitu turus penyulingan tersimulasi. Keputusan menunjukkan bahawa algoritma FDD ini berjaya mengesan dan mendiagnosis kecacatan-kecacatan yang dimasukkan ke dalam proses. Pelaksanaan algoritma FDD ini dalam sebuah loji kimia boleh mengurangkan kos operasi kerana pengesanan dan diagnosis kecacatan yang awal dan meningkatkan prestatsi loji.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

A_B	Bottom column area
$A_i\ B_i\ C_i$	Antoine constants for component-*i*
A^*,B^*,C^*,D^*	Constants for liquid heat capacity
A_P	Pumparound drum area
A_R	Reflux drum area
A_t	Tray active area
a_{ij}	Wilson constant
a	Number of factors in the PLS regression
B	Bottom flow rate
B_{ii}	Virial coefficient for component-*i*
B_{ij}	Virial coefficient for component-*i* and component-*j*
b_i	*i*-th regression coefficient
C_{ik}	Correlation coefficient
D	Distillate flow rate
E	Residual matrix of **X**
F	Residual matrix of **Y**
h	Matrix and vector index
h_b	Enthalpy for bottom stream
H_N	Vapor enthalpy for stage-*N*
h_N	Liquid enthalpy for stage-*N*
h_{ow}	Over weir height
K_i	Vapor-liquid distribution ratio for component-*i*
L	Liquid stream flow rate from reflux drum
L_f	Liquid feed flow rate

$L_{H,B}$	Bottom column liquid level
$L_{H,N}$	Liquid level at stage-N
$L_{H,P}$	Pumparound drum liquid level
$L_{H,R}$	Reflux drum liquid level
L_N	Liquid flow rate leaving stage-N
M	Number of components
M_B	Bottom column molar holdup
M_P	Pumparound drum molar holdup
M_R	Reflux drum molar holdup
N	Number of stages
P	Pumparound flow rate
P_{tot}	Total pressure of the system
P_i^{sat}	Vapor pressure of component-i
p_i	Loading vectors of the i-th PLS factor
Q_R	Reboiler duty
q_i	Loading vectors of the i-th PLS factor
R	Universal gas constant
Re	Reflux flow rate
r_i	i-th prediction error
r_{ij}	Correlation between variable i and j
S	Sidedraw flow rate
T	Temperature
T_b	Normal boiling point
T_{ci}	Critical temperature for component-i
T_{cij}	Critical temperature for component-i and component-j
T	Transpose vector
t_i	Latent score vectors of the i-th PLS factor
u_i	Latent score vectors of the i-th PLS factor
V_N	Vapor flow rate leaving stage- N
V_{ci}	Critical volume for component-i
V_{cij}	Critical volume for component-i and component-j

V_i	Liquid molar volume of component-*i*
W_L	Weir length
w	Weight used in PLS regression
$\mathbf{X}$	Cause data matrix
x_i	*i*-th input variable
$x_{f,\,i}$	Liquid feed mole fraction for component-*i*
x_i	Liquid phase mole fraction for component-*i*
$x_{D,\,i}$	Distillate mole fraction for component-*i*
$x_{P,\,i}$	Pumparound mole fraction for component-*i*
$x_{R,\,i}$	Reflux mole fraction for component-*i*
$x_{S,\,i}$	Sidedraw mole fraction for component-*i*
$\mathbf{Y}$	Effect data matrix
y_i	*i*-th output variable
y_i	Vapor phase mole fraction for component-*i*
Z_{ci}	Critical compressibility factor for component-*i*
Z_{cij}	Critical compressibility factor for component-*i* and component-*j*
χ^2	Chi square
ρ	Liquid density
ω_i	Accentric factor for component-*i*
ω_i	Accentric factor for component-*i* and component-*j*
Λ_{ij}	Wilson binary interaction parameters for component-*i* and *j*
γ_i	Liquid phase activity coefficient for component-*i*
ϕ_k	Vapor fugacity coefficient for component-*i*
ΔH_{vap}	Heat of vaporization
$\Delta H_{vap,\,n}$	Heat of vaporization at normal boiling point

LIST OF ABBREVIATIONS

Abbreviation	Meaning
CSTR	Continuous Stirred Tank Reactor
DPLS	Dynamic Partial Least Squares
DPCA	Dynamic Principal Component Analysis
ERD	Empirical Reference Distribution
ICA	Independent Component Analysis
MSPC	Multivariate Statistical Process Control
ms-PCA	Multi-scale Principal Component Analysis
NC	Normal Correlation
PLS	Partial Least Squares
PCorrA	Partial Correlation Analysis
PCA	Principal Component Analysis
PCR	Principal Component Regression
QTA	Qualitative Trend Analysis
RLS	Recursive Least Squares
RPLS	Recursive Partial Least Squares
SPC	Statistical Process Control
SPE	Square Prediction Error
MESH	Mass, Equilibrium, Summation and Heat

MSPC	Multivariate Statistical Process Control
MW	Molecular Weight
NOC	Nominal Operation Condition

LIST OF APPENDICES

CHAPTER I

INTRODUCTION

1.1 Introduction

Chemical industries are facing a lot of challenges. The industries have to keep sustainable production and within the quality specifications for the products. The whole production process has to operate at the minimum production of waste, minimum consumption of utilities, minimum cost of re-work and re-processing. In order to achieve these targets, modern chemical plants need to operate as fault free as possible because faults that present in a chemical process increase the operating cost due to the increase in waste generation and products with undesired specifications. Therefore, an efficient fault detection and diagnosis algorithm need to be developed to detect faults that are present in a process and pinpoint the cause of these detected faults.

This research is aimed to formulate a fault detection and diagnosis algorithm based on Multivariate Statistical Process Control (MSPC). The objectives of this algorithm are to ensure safe operation, better understanding of the process behaviour and to prevent continuously producing off-specification products.

1.2 Research Background

Multivariate Statistical Process Control (MSPC) is an extension of univariate Statistical Process Control (SPC). This extension enables MSPC to become applicable in chemical industries which are multivariable in nature. MSPC monitoring method consists of collecting nominal operation condition process data, building process models by using multivariate projection methods and comparing the incoming process measurements against the developed process models.

The present MSPC method has several weaknesses in detecting and diagnosing faults. According to Yoon and MacGregor (2000), MSPC is a very powerful tool for fault detection but its main limitation lies in the ability to isolate or diagnose the actual causes of the detected fault. Although contribution plots are use to diagnose the faults, they tend to be noisy and ambiguous. The contribution plots also do not have confidence limit, making it difficult to determine whether a situation is normal or abnormal.

From the previous paragraph, the major weakness of MSPC lies in its ability to diagnose the actual causes of the detected faults. Therefore, this research is trying to solve this problem by introducing new elements into the fault detection and diagnosis method in MSPC. The new elements are:

a) A new fault detection procedure based on correlation coefficient between the quality variables of interest and the selected key process variables.
b) Fault diagnosis using statistical control charts with control limits showing clearly the status of a situation.
c) Formulation of the correlation coefficient based on Normal Correlation (NC), Partial Correlation Analysis (PCorrA) and Principal Component Analysis (PCA).

1.3 Objectives and Scopes of the Research

a) To develop fault detection, diagnosis and control system identification using MSPC.

b) To develop a program package this contains several analysis strategies and multiple types of monitoring charts for detecting, diagnosing and controlling faults in a process.

Scopes of the research consist of:

- A distillation column from plant simulated data (**Appendix B**) is used as the case study. The dynamic models for the column are developed. The distillation column models will be used to describe the column behavior.

- A dynamic simulation algorithm is formulated based on the developed distillation column dynamics models. Later, the dynamic simulation algorithm is developed using Matlab software.

- The performance of the developed dynamic simulation program is assessed. The Matlab simulation results are compared to the simulation results from the plant simulated data (**Appendix B**).

- Controllers tuned and installed for stable operation of the column program.

- Selection of quality variables of interest and key process variables
    - Linoleic Acid composition (x_8) and Oleic Acid composition (x_9) in the bottom stream are chosen as the quality variables of interest.
    - Key process variables selected are process variables that are highly correlated with the two selected quality variables of interest. Process variables that have a Normal Correlation (NC) of 0.1 or more with the two quality variables of interest are selected as key process variables. The selected key process variables are feed flow rate (L_f), feed

temperature (T_f), reflux flow rate (Re), pumparound flow rate (P), reboiler duty (Q_r) and bottom temperature (T_{bot}).

- Determination of **Process Sampling Time, T_{MSPC}**
  - An autocorrelation test based on Wetherill and Brown (1991) was used to determine the suitable **Process Sampling Time, T_{MSPC}** of the process. The T_{MSPC} is determined at a value of 4.6 hours. In this research, T_{MSPC} refers to the time used to sample a data from the process into the data set used for calculation of correlation coefficients.

- Generation of Data
  - Data (values of the selected key process variables and quality variables of interest) are sampled from the process using the determined T_{MSPC}. The collected data are mean-centered and variance-scaled. This data are checked of its average, standard deviation, kurtosis and skewness to establish its normal distribution properties. Once the data follow the normal distribution, it is further checked to determine whether it is the desired Nominal Operation Condition Data (NOC).
  - Nominal Operation Condition (NOC) data are a set of data in which, the selected quality variables and key process variables have values within the statistical control limits of their statistical control charts. The statistical control charts used in this research are Shewhart Control Chart and Range Control Chart. For NOC, the statistical control limits are ± 3σ for the quality variables and ± 3σ/C_{ik} for selected key process variables (C_{ik} is the correlation coefficients between the selected key process variables with the quality variables of interest).
  - Fault Data (OC) are a set of data in which, the selected quality variables and key process variables have values outside the statistical control limits of their statistical control charts in certain times. Fault Data are also sampled from the process using the determined T_{MSPC}.

- Formulating fault detection and diagnosis (FDD) algorithm based on Normal Correlation (NC), Principal Component Analysis (PCA) and Partial Correlation Analysis (PCorrA). The procedures in formulating the algorithm are shown as follow:

    a) Develop the correlation coefficients using NC, PCA and PCorrA.
    b) Develop the fault detection tools.
    c) Develop the fault diagnosis tools.

- The developed FDD algorithm is used with Shewhart Control Charts (SCC) and Range Control Charts (RCC) for fault detection and diagnosis on the generated set of Fault Data.

- The performance of the FDD algorithm is evaluated. The results for fault detection and diagnosis are discussed in depth.

1.4 Contributions of the Research

The contributions of this research can be summarized as follows:

a) The introduction of the correlation coefficient between quality variables of interest and the selected key process variables in formulating the FDD algorithm.

b) The derivation of the correlation coefficient based on Normal Correlation (NC), Principal Component Analysis (PCA) and Partial Correlation Analysis (PCorrA).

c) The application of Partial Correlation Analysis (PCorrA) as an important analysis tool in MSPC.

1.5 Summary of Report

In general the report was organized as follows:

Chapter II elaborates the literature review concept of Multivariate Statistical Process Control (MSPC), Principal Component Analysis (PCA), Partial Correlation Analysis (PCorrA) and the development of MSPC.

Chapter III presents the dynamic modeling of a distillation column as the case study, formulation and establishment of the dynamic simulation program, the tuning of controllers in the column and the evaluation of the performance of the developed simulation program.

Chapter IV mainly consists of the procedures in formulating the fault detection and diagnosis (FDD) algorithm based on NC, PCA and PCorrA. The introduction of the correlation coefficient between the quality variables of interest and the selected key process variables were also presented in this chapter.

Chapter V presents the results obtained from the developed FDD algorithm and the discussion of these results.

Chapter VI gives the conclusions that can be made from the results obtained and also recommendations for future work.

Papers publishing the results of this research are given in **Appendix A**.

CHAPTER II

LITERATURE REVIEW

2.1 Introduction

This chapter consists of six parts: Introduction, Techniques in Fault Detection and Diagnosis, Principal Component Analysis (PCA), Partial Correlation Analysis (PCorrA), The Development in Multivariate Statistical Process Control (MSPC) and Summary. This research focuses on the development of fault detection and diagnosis tool via correlation coefficients using PCA and PCorrA. Hence, the techniques used in fault detection and diagnosis are discussed in the next section of this chapter.

Various techniques are used in MSPC for fault detection and diagnosis such as Principal Component Analysis (PCA) (Kresta *et al*., 1991), Partial Least Squares (PLS) (Nomikos and MacGregor, 1994), Principal Component Regression (PCR) (Randall, 1997) and Independent Component Analysis (ICA) (Manabu *et al*., 2002). The applications of these techniques in MSPC are to obtain a correlation between the quality variables (usually the specifications of a product) with the process variables (measurements in the process such as temperature and pressure). From this correlation, the methods of process monitoring for fault and maintaining the process in nominal operating conditions are formulated. Techniques such as PCA, PLS and PCR are of great importance in determining the desired correlation.

The details and procedures of PCA and PCorrA are presented in the next two sections of this chapter.

The fifth section of this chapter will be discussing the development in MSPC. Finally, the summary of this chapter will be presented in the last section of this chapter.

2.2 Techniques in Fault Detection and Diagnosis

This section will give a brief explanation on the need for formulation of an efficient fault detection and diagnosis method, the definition of fault, noise and also the various techniques of fault detection and diagnosis.

2.2.1 The Need for Fault Detection and Diagnosis

Currently, chemical plants have large scale operation. The final products have strict requirements on its quality. Chemical plants need to operate at the minimum risk level to ensure the safety of equipments in the plant and the lives of the workers in the plants. Modern chemical plants must be also friendly to the environmental be able to handle a wide range of varying feedstock and also facing the constant upgrading of product quality in the market.

Due to these numerous challenges, there is a great need of formulation of an efficient fault detection and diagnosis algorithm in the industry. This algorithm will not only helps to maintain the process in the desired operating conditions but also reduce the risk level in a chemical plant. This algorithm should have the capability to detect and diagnose any fault that is present in the process. This will enables the operators of the plant to take the appropriate actions before any unwanted event take place. The following sections will discuss briefly about key terms used in fault

detection and diagnosis such as fault definition, types of fault, noise definition and unstructured uncertainties.

2.2.2 Fault Definition

According to Himmelblau (1978), the term fault is generally defined as a departure from an acceptable range of an observed variable or a calculated parameter associated with a process. This defines a fault as a process abnormality or symptom, such as high temperature in a reactor or low product quality and so on. The underling cause(s) of this abnormality, such as a failed coolant pump or a controller, is (are) called the basic event(s) or the root cause(s). The basic event is also referred to as malfunction or failure (Venkata *et al*., 2003a).

2.2.3 Types of Fault

Generally, one has to deal with three classes of faults (Venkata *et al*., 2003a). The first type of fault, parameter failures, arises when there is a disturbance entering the process from the environment through one or more independent variables. An example of such fault is a change in the concentration of the reactant from its steady state value in reactor feed. Here, the concentration is an independent variable, a variable whose dynamics is not provided with that of the process.

The second type of fault, structural changes, refers to changes in the process itself. They occur due to hard failures in equipment. Structural faults result in a change in the information flow between various variables. An example of a structural fault would be failure of a controller. Other examples include sticking valve, a broken or leaking pipe.

The third type of fault is malfunctioning in sensors and actuators. Gross errors usually occur with actuators and sensors. These could be due to a fixed failure,

a constant bias (positive or negative) or an out of range failure. In a chemical process, some instruments provide feedback signals which are essential for the control of the process. A fault in one of the instruments could cause the plant state variables to deviate beyond acceptable limits unless the fault is detected promptly and corrective actions are accomplished in time.

2.2.4 Unstructured Uncertainties, Process Noise and Measurement Noise

Unstructured uncertainties are mainly faults that are not modeled a priori. Process noise refers to the mismatch between the actual process and the predictions from the model equations. Measurement noise refers to high frequency additive component in sensor measurements (Venkata *et al*., 2003c). In Multivariate Statistical Process Control (MSPC), process noise and measurement noise are classified as common cause variations of a chemical process. Common cause variations refer to Nominal Operating Condition (NOC), conditions where the process is operating at its desired range of operation. Faults are classified as causal cause variations. Causal cause variations refer to conditions where there are abnormalities in the process and cause(s) of the observed abnormalities can be found. The next section will discuss briefly about the various techniques in fault detection and diagnosis.

2.2.5 Fault Detection and Diagnosis Techniques

There are few desirable attributes of a diagnostic system according to Sourabh and Venkata (1999):

a) Early detection and diagnosis
b) Isolability
c) Robustness
d) Multiple fault identifiability

e) Explanation facility

f) Adaptability

g) Novelty identifiability

Diagnostic methods differ not only in the way the process knowledge is utilized but also in the form of knowledge required. A classification based on the form of process knowledge is shown in Figure 2.1.

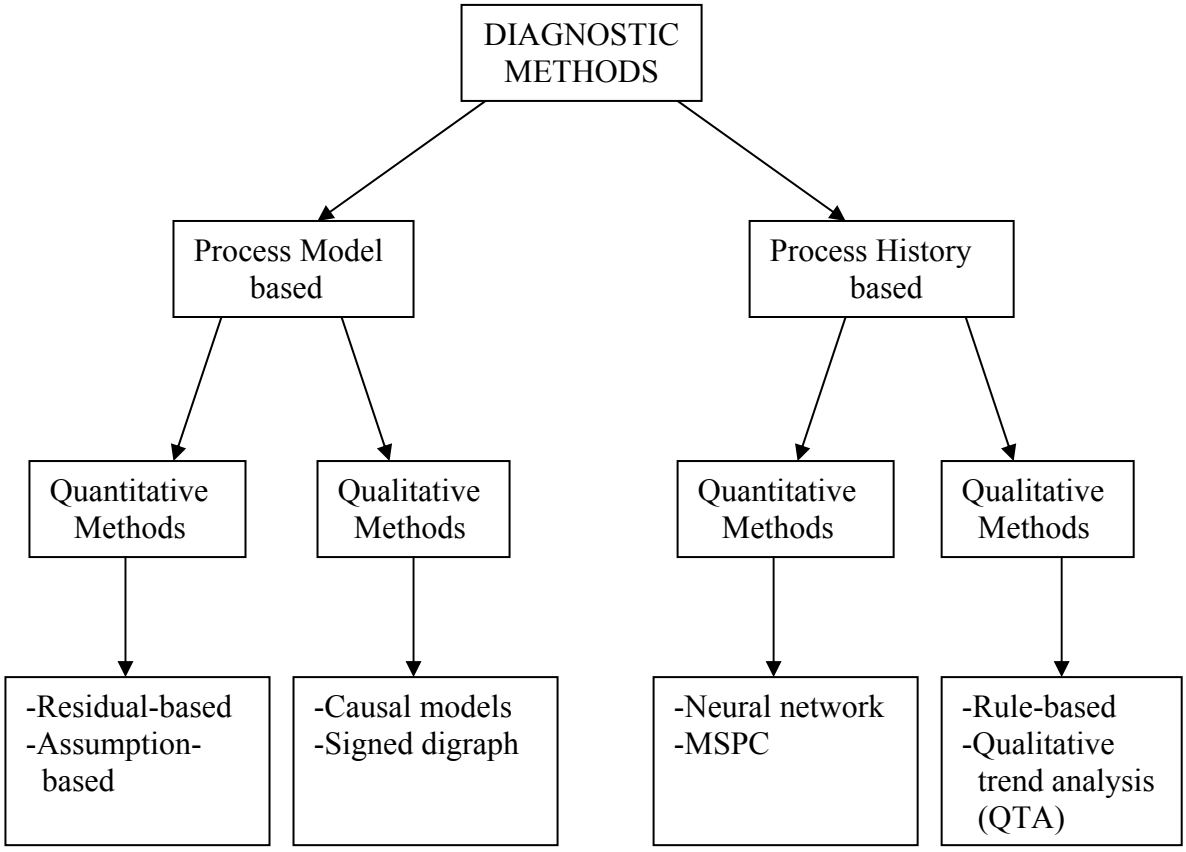


Figure 2.1: Classification of diagnostic methods

Fault detection and diagnosis (FDD) using model-based approach can be broadly classified as qualitative or quantitative. The model is usually developed based on some fundamental understanding of physics of the process. In quantitative models, this understanding is expressed in terms of mathematical functional relationships between the inputs and outputs of the system. In contrast, in qualitative model equations these relationships are expressed in terms of qualitative functions centered on different units in a process (Venkata *et al*., 2003a).

For process history based methods, only the availability of large amount of historical process data is assumed. There are different ways in which this data can be transformed and presented as a priori knowledge to a diagnostic system. This is known as the feature extraction process from the history data, and is done to facilitate later diagnosis. This extraction process can mainly proceed as either quantitative or qualitative feature extraction. In quantitative feature extraction, one can perform either a statistical or non-statistical feature extraction (Venkata *et al*., 2003b).

Among the various techniques shown in Figure 2.1, only MSPC will be discussed in detail since this research focuses on the application of MSPC in developing fault detection and diagnosis algorithm.

2.2.6 Multivariate Statistical Process Control (MSPC)

MSPC is an extension of Statistical Process Control (SPC). The univariate SPC method in process monitoring for fault ignores the influence or cross-correlations between the process variables. Thus, the conventional SPC method does not make the best use of data available (Raich and Cinar, 1996) and this often yield misleading results on multivariate chemical processes. Therefore, it is evidently shown that SPC is not suitable for multivariate chemical processes. Due to this, MSPC has been developed to overcome the weakness of SPC. According to Phatak (1999), the advantages of MSPC compared to SPC are:

1. Multivariate methods can reduce a lot of dimensions in the process data. This will reduce the burden of constructing too many univariate control charts at one time.
2. MSPC could derive a lot of information from process data than SPC since MSPC take account into all process data and not only a single variable as in SPC.

The main techniques in MSPC as mentioned before in the introduction of this chapter are PCA and PLS. In this research, two multivariate techniques are used: the

extensively used PCA and the relatively new PCorrA which is widely used in other fields but not in MSPC. Therefore, the details and procedures of PCA and PCorrA are discussed in the next two sections.

2.3 Principal Component Analysis (PCA)

PCA is a useful tool in MSPC for handling abundant of information from process measurements. This method may be used for regression or similarly to Partial Least Square (PLS), reduction of the effective dimensionality of data (Lennox *et al*., 2001). The variables that need to be analyzed are formed into a data matrix, **X**. These variables are consisting of standardized variables (mean-centered and variance-scaled). This is to avoid the various magnitude of variance of the data sets to influence the outcome of the PCA method. The presence of various measurement data such as composition, temperature, pressure, flow rate and so on were the cause of the various magnitude of variance in the data as stated as above. The PCA regression builds a linear model by decomposing matrices **X** into terms as shown in Equations 2.1.

$$\mathbf{X} = \sum_{i=1}^{a} t_i p_i^T + \mathbf{E} \qquad (2.1)$$

Where: **X** = data matrix in standardized from,

t_i = latent score vectors of the *i*-th PCA factor

p_i = loading vectors of the *i*-th PCA factor

E = residual matrices of the matrix **X**

a = number of factors in the PCA regression

T = transpose vector of the particular vector

PCA can be performed on a set of data using various approaches such as Non-Linear Iterative Partial Least Squares (NIPALS) (Geladi and Kowalski, 1986)

and eigenvector-eigenvalue approach using Singular Value Decomposition (SVD) (Lam and Kamarul, 2002). Both methods will produce the same score and loading vectors. Any discrepancies will be due to rounding off error during the calculation procedure in each method. In each method, a set of score and loading vectors will be extracted from the data matrix and this step will be repeated to extract the next set of vectors. Equation 2.2 shows how SVD decompose a data matrix into its eigenvalue matrix and eigenvector matrices.

$$\mathbf{X} = \mathbf{VSV}^{\mathbf{T}} \qquad (2.2)$$

Where: $\mathbf{X}$ = data matrix in standardized from

$\mathbf{V}$ = eigenvector matrix

$\mathbf{V}^{\mathbf{T}}$ = transpose eigenvector matrix

$\mathbf{S}$ = eigenvalue matrix

The number of factors to be used in building the PCA model is determined in various techniques. Among them are the cross-validation procedure (Geladi and Kowalski, 1986) and sometimes a threshold is chose for the residual **E** and once the magnitude of this residual is below the threshold chosen, the PCA algorithm is stopped. From the PCA procedure via SVD, it is possible to regress a set of selected process variables to a set of quality variables of a process. Although the PCA method is applied for multiple process variables and one quality variable in determining the desired correlation, this method can be extended to multiple process variables and multiple quality variables. First, the set of selected process variables and the first quality variable of interest are formed into a data matrix **X** as shown in Equation 2.1. Then, PCA is performed on this data matrix. The same procedure is use for the second quality variable with the column data of the first quality variable replaced with column data of the second quality variable and so on for the next quality variables. This method is then particularly useful for this research since the case study of this research is a distillation column which has a set of selected process variables and a set of quality variables of interest. The PCA approach in this research uses the SVD method and the main objective of PCA is to find the correlation between the process variables.

2.4 Partial Correlation Analysis (PCorrA)

Partial Correlation Analysis (PCorrA) is used extensively when there are numerous independent variables and a dependent variable. However, PCorrA can also be applied to cases where there are multiple independent and dependent variables. PCorrA is used to determine whether the correlation obtained between an independent variable with the dependent variables is spurious. The meaning of spurious correlation is when two variables, *x* and *y*, have a non-zero (positive or negative) correlation, but neither variable directly causes changes in the other variable because some other variable, *z*, is simultaneously causing changes in both *x* and *y* thus creates a correlation between *x* and *y*. The third variable, *z* is called an extraneous variable (Cliff and Ord, 1973).

There is another case when *x* and *y* are correlated but neither variable have a direct effect upon the other because the presence of intervening variables. These intervening variables act as an intermediate in the correlation between *x* and *y*. Therefore, the correlation between two variables is affected by both extraneous variables and intervening variables (Cliff and Ord, 1973).

PCorrA might not be the complete solution for determining whether a correlation really exists between two variables or it is caused by other variables. But, it can be very helpful in determining the correlation between two variables while holding other variables at a constant value. In general, partial correlations between variables show the correlation that will be obtained when all the control variables (many variables can be statistically control at once) were held constant at their mean values. This aspect of PCorrA will be useful in this research since the correlation between the quality variable and the each input variable will be determined while the other variables will be held constant at the same time (Kamarul, 1997).

For a set of input variables (cause variables) like $\mathbf{X} = [x_1, x_2, \ldots, x_n]$ and a set of output variables (effect variables or quality variables) like $\mathbf{Y} = [y_i, y_2, \ldots, y_n]$, the correlation between the two set of data can be determined by the procedure as shown in Figure 2.2:

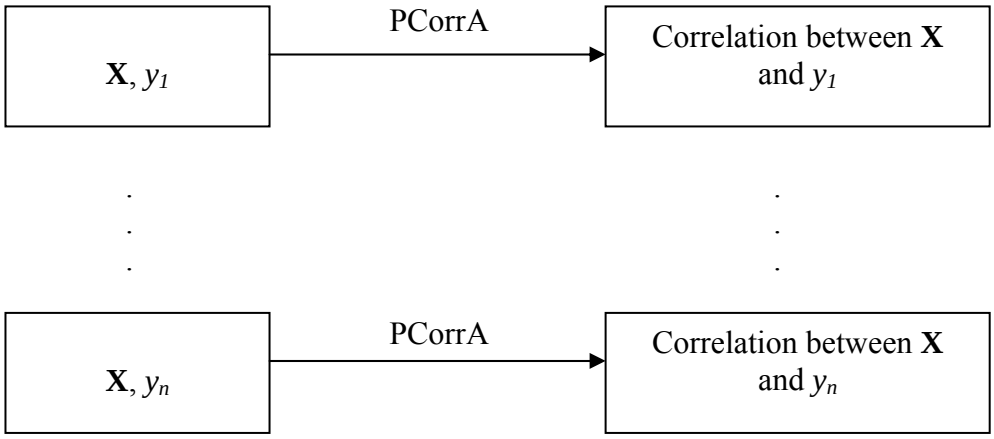


Figure 2.2: The procedure for using PCorrA to determine the correlation between **X** and **Y**.

From Figure 2.2, the method for using PCorrA to determine the correlation between **X** and **Y** is a variable from **Y** is selected, y_1. Then, PCorrA is used to determine the correlation between **X** and y_1. Then, another variable from **Y** is selected, y_2 and the similar procedure is done to obtain correlation between **X** and y_2. This method is repeated until all the variables in **Y** are selected.

For observations x_{1i}, x_{2i} (i = 1, 2, …, n) of the random variables x_1 and x_2, the correlation, r_{12}, between the two variables is given by Equation 2.3 (McNeese and Klein, 1991).

$$r_{12} = \frac{\sum_{i=1}^{n} x_{1i}x_{2i} - n\bar{x}_1\bar{x}_2}{(\sum_{i=1}^{n} x_{1i}^2 - n\bar{x}_1^2)^{1/2}(\sum_{i=1}^{n} x_{2i}^2 - n\bar{x}_2^2)^{1/2}} \qquad (2.3)$$

The partial correlation between x_1 and x_2, after allowing for the effect of another variable, x_3 is denoted by $r_{12.3}$ and is as given in Equation 2.4 (Cliff and Ord, 1973).

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{(1-r_{13}^2)^{1/2}(1-r_{23}^2)^{1/2}} \qquad (2.4)$$

Higher order partial correlations are defined in terms of partial correlation of the next lowest order. Higher order means allowing more effect of other variables while determining the partial correlation between two variables. For example, the partial correlation between x_1 and x_2, $r_{12.34}$, after allowing for the effect of x_3 and x_4, is given in Equation 2.5.

$$r_{12.34} = \frac{r_{12.4} - r_{13.4}r_{23.4}}{(1-r_{13.4}^2)^{1/2}(1-r_{23.4}^2)^{1/2}} \qquad (2.5)$$

Generally, the partial correlation between two variables, x_1 and x_2 after allowing the effects of j - 2 other variables, $r_{12.(3,4,\ldots,j)}$, is given in Equation 2.6.

$$r_{12.(3,4,\ldots,j)} = \frac{r_{12.(4,\ldots,j-2)} - r_{13.(4,\ldots,j-2)}r_{23.(4,\ldots,j-2)}}{(1-r_{13.(4,\ldots,j-2)}^2)^{1/2}(1-r_{23.(4,\ldots,j-2)}^2)^{1/2}} \qquad (2.6)$$

The major advantage of using partial correlation in comparison to the normal correlation is due to the fact that partial correlation has taken into account the effect of other variables while determining the partial correlation between two variables. This advantage will be significant in this research due to the presence of numerous variables in the case study of this research, a distillation column.

2.5 The Development in Multivariate Statistical Process Control (MSPC)

As mentioned earlier on in this chapter that various multivariate techniques had been used in MSPC. Among them are Principal Component Analysis (PCA) and

Partial Least Square (PLS). It can be seen that at the early development of MSPC, these techniques are mostly used on batch processes such as the application of PCA and PLS to a low density polyethylene batch reactor (MacGregor *et al*., 1992), application in a semi-batch reactor (Nomikos and MacGregor, 1994) and application in an industrial batch polymerization reactor (Nomikos and MacGregor, 1995).

In Nomikos and MacGregor (1995), batch systems are hard to monitor due to several reasons such as lack of on-line sensors for measuring product quality variables, the finite duration of batch processes, the presence of significant non-linearity, the absence of steady-state operation and the difficulties in developing a model that represents the process. Therefore, Nomikos and MacGregor (1995) suggested using historical data on measured process variables and final product quality measurements to obtain the needed information about a process. The use of multivariate projection techniques such as PLS and PCA will be able to extract the information needed in developing a good monitoring framework for the process. Process monitoring are done using Hotelling's T^2 chart and Square Prediction Error chart of the process variables. Any detected fault will be diagnosed using contribution plots of the process variables. Two major assumptions were made in this working paper: the method is valid as long as the reference database is representative of the process operation and the requirement that event which one wishes to detect must be observable from the measurements collected. Therefore, there are a few drawbacks of this method. Continuous processes are present in the chemical industry and the behavior of the process continues to change with time. The method proposed above will not work well with process exhibiting constant behavior change. The numerous faults in a process might not always be observable especially those variables that are not within the control loops (Chen *et al*., 2001). Therefore, the method will not detect all the faults in a process.

Chen and Kun (2002) have proposed using Dynamic Principal Component Analysis (DPCA) and Dynamic Partial Least Square (DPLS) to address the problem in Nomikos and Macgregor (1995). The proposed method will incorporate new data into the past data collection as an updating procedure. This data matrix will be subjected to analysis using PCA and PLS to form the monitoring algorithm of the process. An auto-regressive model structure for each batch was developed using a

window length of the process. This will construct a time lagged data window of the whole measurements for each batch run. The proposed method will be able to handle the auto-correlation effect in the series of each batch. In modeling dynamic batch processes, one is concerned with both the correlation among batches and also the auto-correlation of the process variables during one batch run. The proposed method was found to be superior to the conventional static PCA and PLS approach.

Although the DPCA and DPLS methods proposed by Chen and Kun (2002) show improvement over conventional static PCA and PLS approach, there is a major disadvantage of these two dynamic methods. For a highly complex process, there will be excessively large number of variables to be incorporated into the updating data matrix. This will cause major computational load on the process monitoring system. Treasure *et al*. (2004) has proposed using Error in Variable (EIV) identification integrated into Subspace Model Identification (SMI) using PCA to overcome the previous stated weakness of the DPCA and DPLS methods. The proposed method was applied on a simple simulated continuous stirred tank reactor (CSTR) model. Results show encouraging process monitoring of the proposed method over DPCA method. Two advantages of the proposed method over DPCA are lower number of process variables used in constructing the control limits of the monitoring charts and more concise model developed compared to DPCA. However, the fault diagnosis of the proposed method is far from satisfactory and there were cases where more fault causes diagnosed than pre-designed.

MSPC has also been applied successfully to continuous processes (Manish *et al*., 2002). However, in most of the applications, the techniques are implemented on data collected at a single scale. In other words, they present a convolved picture of events occurring at various time scales. Chemical processes are known to operate at different scales and have contributions from events occurring at different scales. Manish *et al*. (2002) proposed using a multi-scale PCA (ms-PCA) to address this problem. In the work, wavelets and PCA were combined to capture the correlation between sensors and also the correlation within a sensor. The procedure is based on data classification prior to further analysis. The process variables that are highly correlated go to a regular PCA module for analysis while the variables that are not highly correlated are fed directly into a sensor validation module. This procedure

uses the conventional Hotelling's T^2 control chart and Square Prediction Error (SPE) chart for fault detection. Fault diagnosis is performed using contribution plots of the variables. This method is applied to an industrial reboiler and a tubular reactor process. Results shown that proposed method managed to detect abnormal events faster than the conventional PCA or PLS approach. Although there are improvements in detection of fault, the diagnosis procedure is still based on conventional contribution plots. The numerous possible suspects in a contribution plot still make the isolation of fault difficult to be performed.

Gertler *et al*. (1999) have combine parity relations and PCA into a fault isolation enhanced PCA method for process monitoring. The method uses incidence matrix developed from PCA and parity relations for fault diagnosis and also decoupling of disturbances. Representation sub-space is used to represent true plant variables and residual sub-space used to represent faults. The residuals were threshold tested for process monitoring. This method was applied to the Tennessee Eastman model (only the reactor model). Fault diagnosis using the developed incidence matrix shows encouraging results. The threshold value in the residuals removes any ambiguity in fault detection diagnosis. For case of multiple faults, the incidence matrix will be really complex and this will restrict the application of this method in fault isolation. The observed fault codes isolated in the case study are those identified earlier on or in other words, the residuals for these faults were designed prior to fault detection and diagnosis. The major limitation of this method when applied to a real plant is that plant data containing faults and fault free data are both needed. This is impractical since the size of faults in a chemical plant is not necessarily known.

Vedam and Venkata (1999) use a combination of PCA and Signed Digraph (SDG) for enhancing the fault diagnosis ability of conventional process monitoring based on PCA. The method, PCA-SDG, overcome the weaknesses of SDG in previous works such as large number of threshold needed for process monitoring, difficulty in selection of these threshold values and lack of resolution in the diagnosis ability of SDG. This method uses PCA for fault detection while SDG is used for fault diagnosis. This PCA-SDG method only needs one threshold value for fault isolation. It was applied on the Amoco Model IV: Fluidized Catalytic Cracking Unit

(FCCU). Both single and multiple faults can be successfully detected and diagnosed using the proposed PCA-SDG method. The main advantages of this method lie in its abilities to isolate the faults faster, can diagnosed both single and multiple faults and more resolute diagnosis of faults. The only apparent drawback of this method is the selection of threshold value is a trade-off between robustness and the presence of false alarms. The value of the threshold value can made the monitoring system become too sensitive even to small process changes (presence of false alarms). Therefore, the selection procedure for the threshold value needs to be studied to improve the ability of this method.

Lee *et al*. (2004b) try to improve the performance of process monitoring using Independent Component Analysis (ICA) in the place of the conventional PCA. The idea of ICA is to extract latent variables (independent variables in process that are not directly measurable) from the process data and these independent components (IC) are assumed to be non-Gaussian and mutually independent. This work proposed the combination of ICA and the conventional process monitoring techniques in developing a new process fault detection and diagnosis method. The developed method was applied to a benchmark of biological wastewater treatment process. The method based on ICA was superior in fault detection of the pre-designed faults compared to the conventional method based on PCA. The limitation of this method is the computational load when establishing the normal operation condition data of the study process. The process faults diagnosis using contribution plots (without control limit) is ambiguous since there is no limit to differentiate the situations of fault or normal operation.

According to Kruger *et al*. (2001), conventional PLS-based fault detection and diagnosis algorithm does not detect every kind of abnormal process behavior and also insensitive to process changes. The usage of too many charts will be required if the retained latent variables in the projection method is greater than two or three. These two weaknesses of the previous works have suggested the need of a new PLS based fault detection algorithm. An extended PLS approach has been proposed by Kruger *et al*. (2001). In this work, the determination of two new PLS scores which are called generalized scores has yield better fault detection performance. These scores are calculated by restructuring of the data matrix that resulted from the

standard PLS decomposition using both the predictor and response variables. The conventional T^2 and SPE values are then calculated using these generalized scores. The control limits are determined using empirical studies. This method is implemented on two case studies: fluid catalytic cracking unit by McFarlane *et al*. (1993) and a real industry process simulation from ICI Plant Polymer Chemicals. The proposed method is able to detect all the generated faults including those involving variables which are not incorporated in the control loops. The conventional PLS approach fails to detect these faults due to their nature of not present in the control loops. The usage of only two control charts in this method is far simpler than the numerous control charts in other common PLS implementation for process monitoring.

PLS based method has shown satisfactory results for monitoring of simple processes. For complex processes, results obtained from PLS based method are often hard to interpret and understand. Choi and Lee (2005) proposed using a multi-block PLS (MBPLS) system for complex processes. The method proposed uses super score deflation method for process fault detection and diagnosis. This method uses four kinds of monitoring statistics: Hotelling's T^2 and SPE statistics for block and overall process for fault detection and contribution plots using relative contribution (contribution value of variables divided by the control limit of contribution plots) for fault diagnosis. The proposed method was applied to a wastewater treatment plant with multi-unit operations. The method was able to detect the pre-designed faults in the study process. The main contribution of this work is the introduction of relative contribution in contribution plots with limit providing a more effective and confident fault diagnosis method compared to the conventional contribution plots which have no limit. The limitation of this method is the limit of the contribution plots is selected arbitrarily and not by any statistical method, thus affecting the validity of the limit selected.

The application of PCA for large-scale processes has been done by Kruger *et al*. (2004). In this work, the auto-correlation in process variables from a large-scale process has been proven to affect the performance of process monitoring using PCA because of the increase in the number of false alarms. Kruger *et al*. (2004) proposed using an auto-regressive moving average (ARMA) filter to filter data processed by

PCA earlier on in order to reduce the auto-correlation of the developed PCA variables. The proposed method was applied to the Tennessee Eastman (TE) simulation process. The method was able to reduce the number of false alarms in the monitoring method based on PCA thus increasing the performance of the monitoring method. The limitation of the proposed method is this method is only applicable for fault detection for the time being. The selection of the ARMA filters also affect the performance of the developed process monitoring method, causing the selection of unsuitable filters will yield unsatisfactory results in process monitoring.

For highly non-linear processes, the application of conventional PCA method and its improvement like multi-way PCA, DPCA and so on will yield poor performance due to the non-linear behavior of the process. Lee *et al*. (2004a) has proposed using a novel Kernel PCA (KPCA) approach to model the non-linear behavior of certain highly non-linear processes. The proposed method wish to overcome the weakness of the previous KPCA: direct application of KPCA to process monitoring problematic because the conventional SPE chart can not be generated using the previous KPCA method. The proposed method addresses this problem by introduction of a feature space and the generation of Hotelling's T^2 and SPE statistics from this feature space as process monitoring tools. The method was applied on a simple multivariate model and a simulated wastewater treatment process.

The performance of the new KPCA method was compared to conventional linear PCA method. In the first case study, KPCA shows faster detection of abnormal process behavior and good detection of insignificant faults which was failed to be detected by the PCA method. For the second case study, KPCA captures the non-linear properties of the process and distinguish well between normal operating condition and out of control condition. The PCA method could not differentiate between the two conditions due to non-linearity that present in the process behavior. Fault diagnosis using the proposed KPCA method in non-linear situations is very difficult because it is very hard to find an inverse mapping function from the feature space to the original data space. Therefore, the proposed KPCA method can be only applied for fault detection for the time being.

Choi and Lee (2004) improve the application of KPCA for non-linear processes by considering the dynamic properties of the process data in developing the process monitoring method based on KPCA. The method, dynamic KPCA (DKPCA) uses KPCA with Gaussian function as the kernel function and also combines the conventional Hotelling's T^2 and SPE statistics into a unified index, D_f, for process monitoring. The proposed method was applied to a wastewater treatment process. DKPCA shows good fault detection performance of the pre-designed faults (includes bias and drift types of faults in sensors). This method also improves the simplicity of process monitoring by the application of the unified index. The limitation of this method is that the dimension of the feature space in DKPCA depends on the type of kernel function and lag time used. The width of the Gaussian function will affect the robustness and sensitivity of the developed process monitoring method. Fault diagnosis using this method will be very difficult since the re-mapping from the feature space into the process variables space (input space) is very complicated.

Cho *et al*. (2005) also applied KPCA with Hotelling's T^2 and SPE statistics for process monitoring (both fault detection and diagnosis) on a non-isothermal continuous stirred tank reactor (CSTR) process. The new element in the work of Cho *et al*. (2005) is the derivation of an analytical solution using the gradient of kernel function of contribution for kernel PCA. This is different from the direct application of PCA on the feature space by Lee *et al*. (2004a). Results show that the KPCA method successfully detect and diagnosed the pre-designed faults in the study process. The main advantages here are there is no data reconstruction or approximation needed for fault diagnosis, thus reducing the loss of information. However, the fault diagnosis charts derived from KPCA is still ambiguous and it takes certain time to diagnose the fault causes.

In the area of on-line process monitoring, constant upgrading or recursive updating of process model with new process data are needed for maintaining a good process monitoring algorithm. Recursive least squares (RLS) is the most commonly used method for recursive on-line estimation of model parameters. RLS has several weaknesses such as estimation difficulties when process variables are highly correlated and the presence of auto-correlation and cross-correlation in the process

variables. Dayal and MacGregor (1997) have proposed a recursive PLS algorithm (RPLS) to overcome the previous weaknesses of the RLS technique. The usage of an improved kernel algorithm and a variable forgetting factor for updating the covariance structure of the model has yield better results than the RLS-based updating method. A variable forgetting factor is chosen rather than a constant forgetting factor because in the case of no new variation in new data, the old data that represent the process well are being discounted for a constant forgetting factor. This will make the covariance matrix lose the essential process information and become extremely ill-conditioned and the precision of the resulting parameters will be poor (Dayal and MacGregor, 1997).

Most of the past works on fault detection and diagnosis in MSPC uses the same method in determining the control limits for the control charts used. The Hotelling's T^2 chart is assumed to follow an F-distribution while the SPE chart is assumed to follow a χ^2-distribution respectively. This method has limitation since the assumption that the data obtained are serially independent is made. When process data are collected from a dynamic system, serial correlation will generally be present. This will lead to resulting in too many false alarms in the fault detection algorithm based on this assumption (Vasilopoulos and Stanoboulis, 1978).

Simoglou *et al*. (2002) proposed the usage of Empirical Reference Distribution (ERD) in determining the control limit for all the control charts. This was being carried out on a simulated continuous stirred tank reactor (CSTR) and the results are compared to the method using the assumption that both the charts follow the above mentioned distribution. The results clearly show that the control charts with the limit based on ERD performed better than the control charts based on the assumption that the data are serially independent. This is due to the fact that ERD takes into account the serial correlation in the measurements. However, this method was applied on the assumption that the process exhibit linear behavior. For non-linear process behavior, future studies need to be carry out.

MSPC has certainly shown its ability in detection of fault in a process. According to Yoon and MacGregor (2001), the fault diagnosis ability of MSPC is still far from satisfying. The usages of contribution plots in isolating faults are

limited. This is because the faults that are able to be isolated using normal contribution plots are those resulting from simple faults like actuator or sensor fault. Complex faults like fundamental change in the process and also faults that has effect propagated into other variables are hard to be isolated using these contribution charts. Therefore, Yoon and MacGregor (2001) have introduced the usage of fault signatures in enhancing the isolation ability of MSPC. Faults from process data are collected and fault signatures are then developed using PLS or PCA based model built from common-cause data. Fault signatures consist of directions of movement of the process in both the model space and in the orthogonal residue space during the period immediately after the fault has been detected. Any new fault will exhibit certain fault signature and this signature is then compared to the signatures from the fault signatures databank. The usage of joint angle plots of measurement vectors of fault has enabled the isolation of all faults including the complex faults. Although the proposed isolation method has shown good results, there is still need to incorporate dynamic data into the development of fault signatures (Yoon and MacGregor, 2001).

From the previous paragraphs, the technique used was still based on PCA and PLS. Although PLS has shown great usage when both quality variables and process variables are available, this technique still has its shortcoming. The inability to hold the other process variables constant while determining the correlation between the quality variable and one process variable has made the correlation determined questionable. This is due to the fact that the correlation produced between the two variables might be caused by the presence of other variables. Therefore, in this research, a new method of analyzing data in MSPC is introduced. This technique is the Partial Correlation Analysis (PCorrA).

PCorrA has been applied in many applications such as:

- Partial correlation analysis of fuzzy sets: Ding and Nancy (2000)
- Analysis of mineral content: Quemerais et al. (1998)
- Analysis of metals concentration: Wang and Chen (2000)

Although these applications are not in the field of MSPC, but they show that PCorrA has been widely used as analytical tool. In Kamarul (1997), PCorrA was used in deriving a correlation coefficient between the input variables and the control variable for a chemical process. This work is one of the few that has used PCorrA in the field of fault detection and diagnosis. The application of PCorrA in this research will be based on the work by Kamarul (1997).

The PCorrA technique focuses on determining the correlation between two variables while setting the other identified variables that might affect the correlation between the two study variables at a constant value (Ding and Nancy, 2000). Therefore, this technique will be able to overcome the shortcoming of PLS stated in the previous paragraph. Aside from the novel idea of using PCorrA, this research also utilizes a new technique in constructing the control limit of the control charts used for fault detection and diagnosis.

From literature, the conventional control charts used in MSPC for the objective of fault detection and diagnosis are the Hotelling's T^2 Statistic control chart and the SPE Statistic control chart (Nomikos and MacGregor, 1994, Chen and MacAvoy, 1998 and Wachs and Lewin, 1999). The advantage of both the T^2 Statistic control chart and the SPE statistic control chart over the normal univariate control charts is the former could generate one common statistics from the values of many variables that can be plotted on a control chart (Ku *et al*., 1995). This will avoid the condition of data overload which is always the case for normal univariate chart such as Shewhart chart. In addition, normal univariate chart does not function well for multi-variable process with highly correlated variables (Manabu *et al*., 2000).

The new approach in this research is to use the normal univariate control chart with a slight modification that takes into account the correlation between the variables of the process. The correlation between the quality variables of interest and the selected key process variables obtained from using Normal Correlation (NC), PCA and PCorrA will be used to construct the control limits for the control charts of the selected key process variables. The major advantages of the proposed method are any faults that are present in the process is easily detected through the control

charts and the present of control limits in the control charts make fault diagnosis effective and non-ambiguous.

2.6 Summary

From the previous sections in this chapter, the development of MSPC and the various multivariate techniques used in MSPC were discussed. The techniques used vastly in MSPC are PLS, PCA and the modification of these two techniques such as Multiway PCA, Multiblock PCA, Multiway PLS, Multiblock PLS, DPLS, DPCA, KPCA, MPCA and so on. These techniques although improve the usage of MSPC to various fields and applications, they are still room for improvement. Thus, in this research, a technique that has not been extensively used in MSPC, PCorrA, will be used as the multivariate technique analysis for the data obtained together with the extensively used, PCA. A modified approach in determining the control limits of control charts for process variables will also be presented in this research. A complete explanation on the application of NR, PCA, PCorrA and the modified approach in setting up the control limit of control charts for selected key process variables will be presented in the methodology chapter.

CHAPTER III

DISTILLATION COLUMN MODELING AND SIMULATION

3.1 Introduction

This chapter consists of nine sections: chapter introduction, explanation on choosing distillation column as the study unit operation, the study column information, degree of freedom analysis prior to column modeling, distillation column models formulation, distillation column literature review, dynamic study of the study column, performance evaluation and chapter summary. The proposed fault detection and diagnosis algorithm can be applied to any chemical unit operation. However, a distillation column is chosen as the case study and explanations regarding this choice are given in section two. The distillation column model based on simulated plant data (**Appendix B**) is selected as the study column. The details and descriptions of the study column are presented in the third section.

A degree of freedom analysis is needed before the modeling of the column and this analysis is presented in the fourth section. Section five presents distillation column dynamic modeling, which involves using Mass balance equations (M-equations); Equilibrium equations (E-equations); Summation equations (S-equations) Heat balance equations (H-equations) and also hydraulic equations to describe the behavior of the column bottom, normal tray, feed tray, reflux point, pumparound point, sidedraw tray and the reflux tray.

The literature review on distillation column dynamic simulation is elaborated in details in the sixth section with the aim of selecting the suitable dynamic algorithm. The step-by-step formulation of the dynamic simulation algorithm for the case study is presented in section seven after the dynamic simulation algorithms in the literature have been studied. Later, a dynamic simulation program using Matlab is developed according to the formulated algorithm. The performance of the written simulation program is evaluated in section eight. The performance evaluation is conducted based on the ability of the program to produce results that are close to the data from the plant simulated data (**Appendix B**).

3.2 Choosing Distillation Column as the Case Study

Distillation columns are used in the chemical industries for fractionation, purification and separation. The separation process occurs due to the difference in boiling point or vapor pressure of the components that present in the liquid mixture. The component with the lowest boiling point or the component with the highest vapor pressure will be separated first.

In this research, distillation column is chosen as the case study because of its importance in the chemical industries. Distillation column is energy intensive equipment. According to some statistics, 40% of the energy in a chemical plant is consumed by distillation columns (Zhang *et al*., 1999). Successful maintenance of operation of distillation column enables consistent production and ensures sustainable product quality with minimum utility consumption. Any unwanted event happening in the column have to be identified as early as possible for preventing continuous production of out-of-specification products or cause serious problems in down stream processes. The assignable causes for the unusual events have to be investigated and prompt action has to be taken in order to rectify the situation.

On the other hand, distillation columns are multivariate, dynamic in nature and can operate either continuously or batch-wise. The process variables are correlated and interrelated. Interpreting either one of the column variables will not

give the state of the column behavior. In order to fully understand the column behavior and state, the key process variables like feed stream flow rate, feed stream temperature, distillate flow rate, reflux flow rate, pumparound flow rate, reboiler duty and other process variables have to be analyzed and monitored simultaneously. Therefore, distillation column is selected as the case study for developing a fault detection and diagnosis algorithm based on MSPC approach.

3.3 Information of the Case Study

The column information is based on the information obtained from the plant simulated data (**Appendix B**). From the plant simulated data, the original packed distillation column is approximated into a tray column in this research. The method of conversion from packed column to tray column is based on the method found in Geankoplis (1995) and additional column information from Wong (2003). The tray column model developed in Matlab is an approximation of the original packed column in the plant simulated data. The summaries of the column are presented in Table 3.1:

Table 3.1: Specifications of the case study

Variable	**Specification**		
Number of actual tray	28		
Actual feed position	Tray 14 from the top		
Plate spacing (cm)	50		
Plate diameter (cm)	120		
Type of condenser	No condenser in the system		
Type of reboiler	Total		
Reflux rate (kmole/hr)	10		
Pumparound rate (kmole/hr)	49.6		
Sidedraw rate (kmole/hr)	63.6		
Reboiler duty (kJ/hr)	2.16×10^6		
Reference Steady State			
	Composition (kmole/kmole)		
Component	Feed stream	Distillate stream	Bottom stream
Hexanoic Acid	0.00229	0.02132	0
Octanoic Acid	0.05106	0.52101	0
Decanoic Acid	0.04317	0.44283	0.00036
Dodecanoic Acid	0.52169	0.01484	0.57654
Tetradecanoic Acid	0.15689	0	0.17387
Hexadecanoic Acid	0.06793	0	0.07528
Stearic Acid	0.01456	0	0.01613
Oleic Acid	0.12191	0	0.13509
Linoleic Acid	0.02050	0	0.02272
Temperature (K)	483.15	344.15	512.05
Pressure (Bar)	2.5	0.11	0.1333
Flow rate (kmole/hr)	41.557	4.000	37.5
Plate type	Sieve tray		
Weir length (m)	0.912		
Weir height (m)	0.076		

*The compositions of the pumparound stream, reflux stream and the sidedraw stream are the same as the compositions of the distillate.

After obtaining the basic data of the column, the modeling of the column is performed. Prior to the modeling process, a degree of freedom analysis is initiated. The degrees of freedom analysis will be presented in the following section.

3.4 Degree of Freedom Analysis

Dynamic simulation involves setting an algorithm to solve sets of equations that describe the column behavior. These equations are Mass balance, Equilibrium, Summation and Heat balance (MESH) equations and hydraulic equations. In order to obtain unique solution to a set of independent equations, the number of independent equations must equal to the number of independent variables (Stephanopoulos, 1984). This rule must be followed so as to avoid either redundant equations or variable conditions because these conditions cause infinite number of solutions or no solution at all. The degree of freedom for a system is determined as in Equation 3.1:

$$\begin{aligned} \#\,df &= \#\,\text{Variables} - \#\,\text{Equations} \\ &= \text{number of independent variables} \\ &\quad - \text{number of independent equations} \end{aligned} \qquad (3.1)$$

For a multi-component distillation column, the degree of freedom analysis is presented in Table 3.2 and Table 3.3 as in Luyben (1963).

a) Number of trays = 28, 1 total reboiler. Total trays = 29 (*N* trays)

b) Number of components = 9 (*M* components)

Table 3.2: Degree of freedom analysis – number of variables

Variables	**General Form**	**No. of Variables**
Tray composition (vapor and liquid)	2*MN*	2 x 9 x 29
Tray liquid flow, *L*	*N*	29
Tray vapor flow, *V*	*N*	29
Tray hold up, molar	*N*	29
Reflux flow rate	1	1
Distillate flow rate	1	1
Sidedraw flow rate	1	1
Pumparound flow rate	1	1
Liquid stream flow rate	1	1
Pumparound point hold-up	1	1
Reflux point hold-up	1	1
Side draw composition	2*M*	2 x 9
Bottom flow rate	1	1
Bottom hold-up	1	1
Vapor flow from reboiler	1	1
Bottom composition	2*M*	2 x 9
Trays temperature and pressure	2*N*	2 x 29
Pumparound point temperature and pressure	2	2
Reflux point temperature and pressure	2	2
Pumparound stream temperature	1	1
Reboiler duty	1	1
Cooler duty	1	1
Total		**720**

Table 3.3: Degree of freedom analysis – number of equations

Equations	**General Form**	**No. of Equations**
Total mass balance (tray, reflux point, pumparound point)	$N + 2$	31
Component Balance, liquid composition	$(N + 2)(M - 1)$	31 x 8
Summation Equation for vapor and liquid composition	$2(N + 2)$	2 x 31
Hydraulic (liquid flow rate)	$N + 2$	31
Equilibrium Equation (vapor composition)	$M(N + 2)$	9 x 31
Heat Balance	$N + 2$	31
Total		**682**
Degrees of Freedom =	**720 – 682 = 38**	

The degrees of freedom calculated can be further reduced by specifying the value of independent variables or introducing more independent equations. The distillation column is installed with several control loops to ensure stable operation of the column. Figure 3.1 shows the distillation column with all the control loops. The general properties of each control loops are explained in Table 3.4. The controller of each control loop is tuned using the Cohen-Coon method (Stephanopolous, 1984). The method in determining the controller sampling time and controller tuning procedure and the details of the controllers are given in Section 3.7.4.

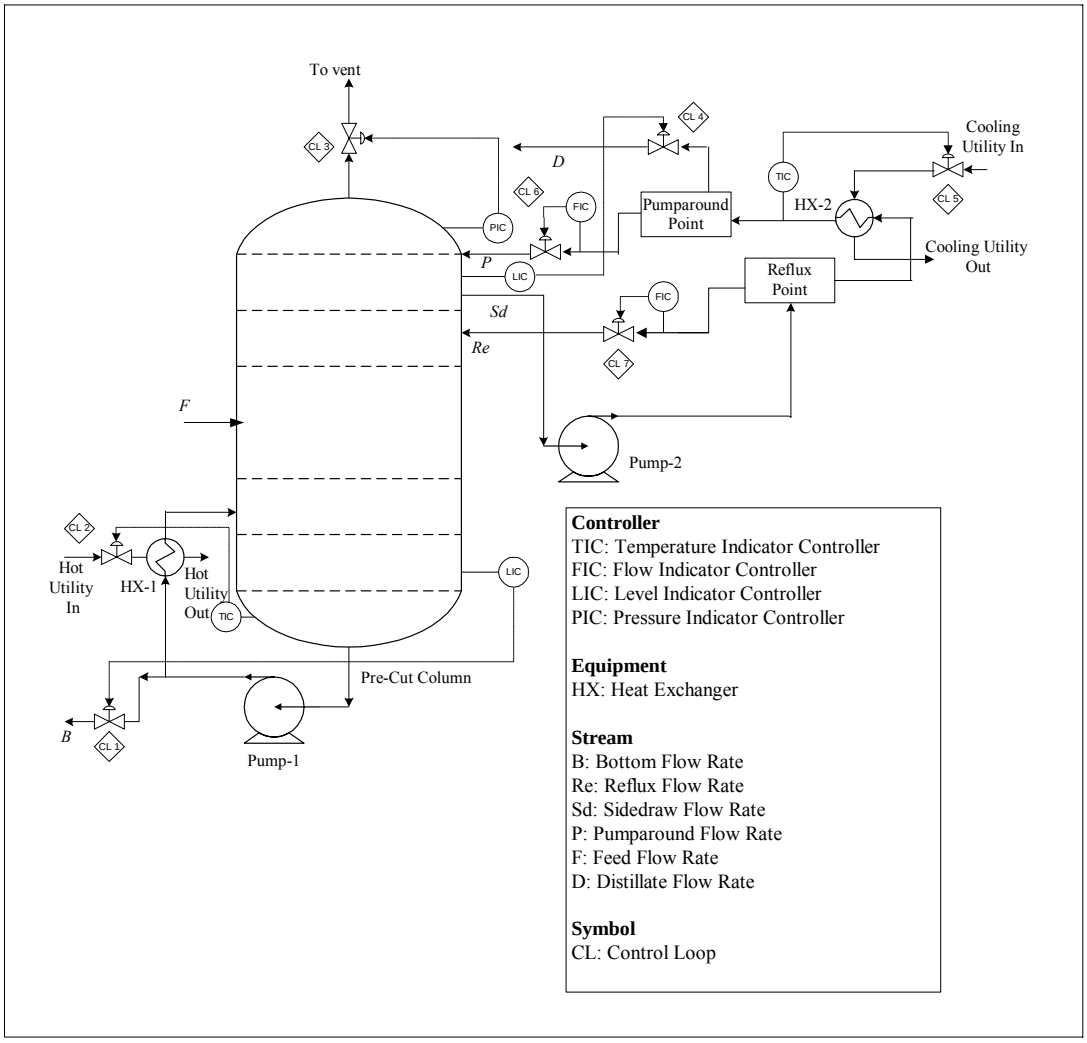


Figure 3.1: The control structure of the distillation column

Table 3.4: Properties of each control loop

Control Loop	Local Disturbance*	Control Variable	Manipulated Variable	Type of Control Loop
1	Liquid flow rate from tray 28, L_{28}	Liquid level at bottom column, $L_{H,\,B}$	Bottom flow rate, B	Feed back
2	Liquid flow rate from tray 28, L_{28}	Bottom temperature	Hot utility flow rate	Feed back
3	Vapor flow rate from tray 2, V_2	Top column pressure	Vapor flow rate to vent	Feed back
4	Liquid flow rate from tray 1, L_1	Side draw tray liquid level	Distillate flow rate, D	Feed back
5	Liquid stream flow rate, L_{Dp}	Pumparound temperature	Cooling utility flow rate	Feed back
6	Liquid stream flow rate, L_{Dp}	Pumparound flow rate, P	Flow rate prior to pumparound stream	Feed back
7	Sidedraw flow rate, *Sd*	Reflux flow rate, *Re*	Flow rate prior to reflux stream	Feed back

*The local disturbance of each control loop in Table 3.4 refers to main local disturbance among various local disturbances on each control loop. The location of the variables (control variables, manipulated variables and local disturbance) are shown in Figure 3.2.

The degrees of freedom for the system are 31 (38 – 7) since the 7 control loops are equivalent to 7 independent equations. In order to define the system completely, 31 independent variables are needed to be specified. The defined independent variables are shown as follow:

a) 28 trays pressure. Pressure varies linearly from top tray to bottom column,
b) Reflux flow rate,
c) Pumparound flow rate,
d) Reboiler duty.

3.5 Distillation Column Models Formulation

The mathematical models used in dynamic simulation should be as general as possible so that it could handle wide range of problems and processes. All models start from basic mass and energy balance as well as the equilibrium relationship between phases. Various forms of MESH equations have appeared in the literature. The arrangement of the sets of equations depends on the methods adopted in solving the problem. In addition, the equations can be written in transient form or in steady state form. Other terms like tray efficiency, chemical reaction and thermal efficiency can also be introduced in these MESH equations. Beside these MESH equations (Wang and Henke, 1966), correlations are also needed to determine equilibrium ratio, *K*'s; liquid enthalpy, *h*'s; vapor enthalpy, *H*'s, Francis weir equations and hydraulic equations. The following section presents the distillation column dynamic modeling systematically.

3.5.1 Mass Balance Equations

The mass balance equations involve total flow rate balance and components flow rate balance for each tray. The equations are written in derivative form in order to reveal the dynamic behavior of the column. In order to get a better picture of mass balance for reflux point, pumparound point, normal tray, sidedraw tray, reflux tray, pumparound tray, feed tray and column bottom, a schematic diagram for the distillation column was presented in Figure 3.2.

Distillation Column Representation

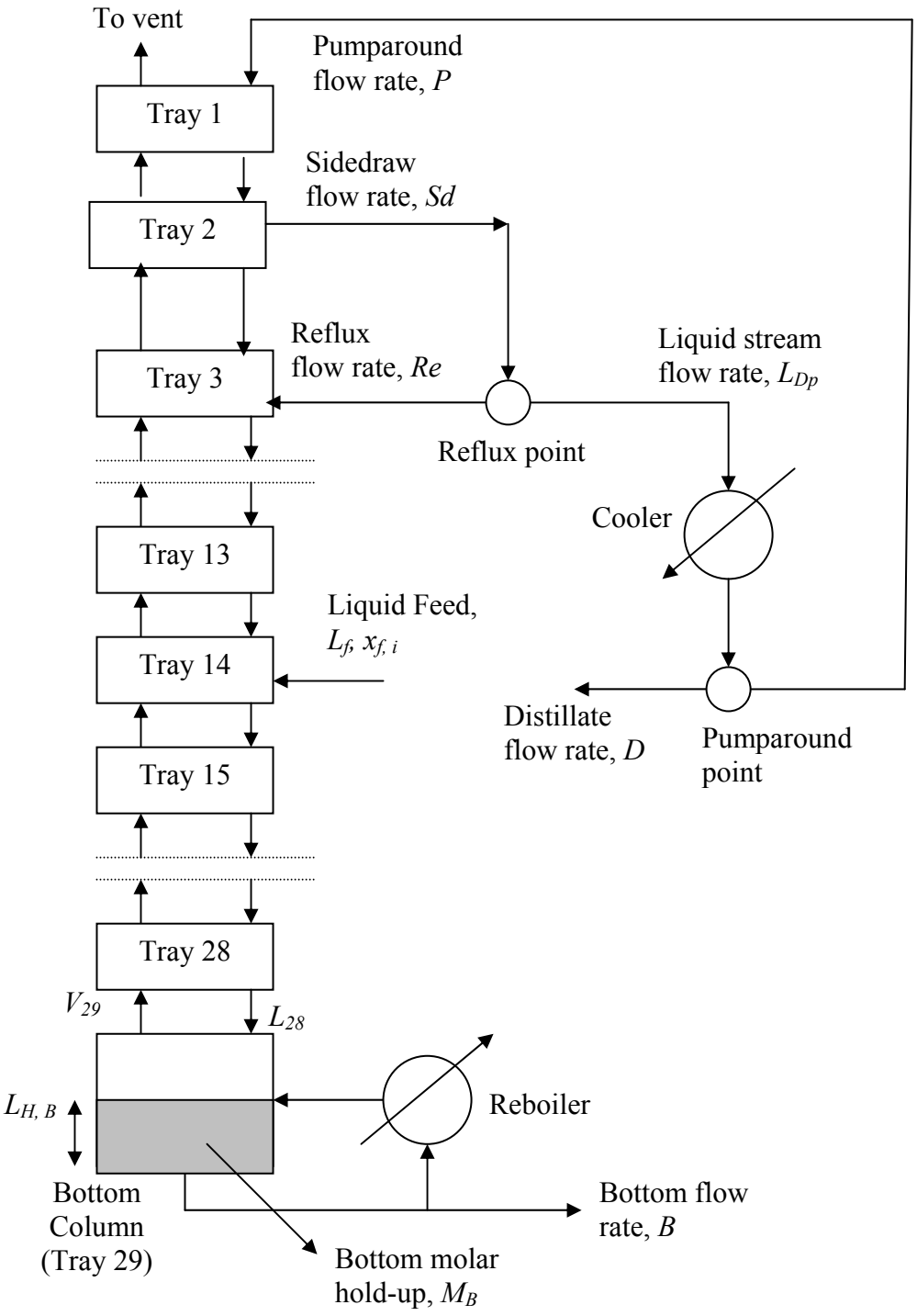


Figure 3.2: Distillation column schematic diagram

3.5.1.1 Bottom Column Model

The bottom column (tray 29) schematic diagram is shown in Figure 3.3.

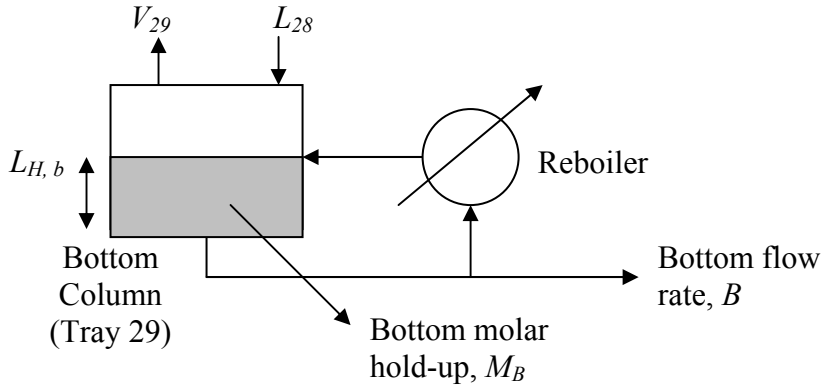


Figure 3.3: Bottom column schematic diagram

I) Bottom flow rate has a very close relationship with the bottom liquid level. The equation relating these two variables (Douglas *et al*., 1996) is shown in Equation3.2.

$$\text{Bottom flow rate, } B = \frac{C_v \rho L_{H,B}^{0.5}}{MW} \qquad (3.2)$$

The bottom molar hold-up is related to the bottom liquid level as shown in Equation 3.3.

$$\text{Bottom molar hold-up, } M_B = \frac{L_{H,B} A_B \rho}{MW} \qquad (3.3)$$

$$\text{Liquid density, } \rho = \sum_{i=1}^{9} x_i \rho_i \qquad (3.4)$$

II) Bottom height derivative expression.

$$\frac{dL_{H,B}}{dt} = \frac{L_{N-1} - V_N - B}{cons\tan t} \qquad (3.5)$$

III) Components mass balance in term of mole fraction derivative expression.

$$\frac{dx_{N,i}}{dt} = \frac{L_{N-1}x_{N-1,i} - V_N y_{N,i} - Bx_i}{L_{H,B}(cons\tan t)} \qquad (3.6)$$

$$\text{constant} = \frac{A_b \rho}{MW}$$

where:

ρ = liquid density
MW = molecular weight
$L_{H,B}$ = bottom liquid level
A_b = bottom column area
C_v = discharge coefficient
B = bottom flow rate
M_B = bottom molar hold-up
M_N = molar hold-up at tray-N
M_P = pumparound point molar hold-up
M_{Re} = reflux point molar hold-up
x_i = component liquid mole fraction
y_i = component vapor mole fraction
N = number of trays
L_N = liquid flow rate at tray-N
V_N = vapor flow rate at tray-N
x_f = liquid feed mole fraction
x_{Re} = liquid reflux mole fraction
x_P = liquid pumparound mole fraction
x_{Sd} = liquid sidedraw mole fraction

3.5.1.2 Tray Model

The normal trays (N = 4 to 13 and 15 to 28) schematic diagram are presented in Figure 3.4.

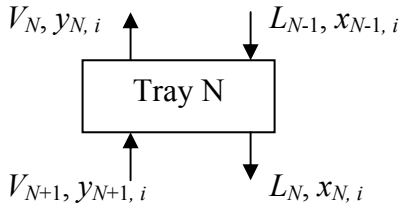


Figure 3.4: Tray schematic diagram

I) Total mass balance

$$\frac{dM_N}{dt} = L_{N-1} + V_{N+1} - L_N - V_N \qquad (3.7)$$

II) Component mass balance

$$\frac{dx_i}{dt} = \frac{L_{N-1}x_{N-1,i} + V_{N+1}y_{N+1,i} - V_N y_{N,i} - L_N x_{N,i}}{M_N} \qquad (3.8)$$

3.5.1.3 Feed Tray Model

The feed tray (N = 14) schematic diagram is presented in Figure 3.5.

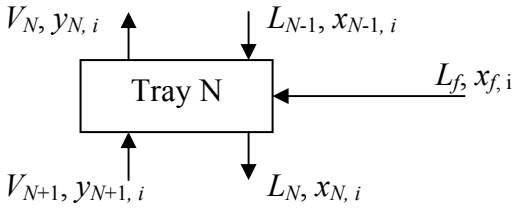


Figure 3.5: Feed tray schematic diagram

I) Total mass balance

$$\frac{dM_N}{dt} = L_{N-1} + V_{N+1} + L_f - L_N - V_N \qquad (3.9)$$

II) Component mass balance

$$\frac{dx_i}{dt} = \frac{L_{N-1}x_{N-1,i} + V_{N+1}y_{N+1,i} + L_f x_{f,i} - V_N y_{N,i} - L_N x_{N,i}}{M_N} \qquad (3.10)$$

3.5.1.4 Reflux Point Model

The reflux point model is similar to the bottom column model. The schematic diagram for the reflux point is shown in Figure 3.6.

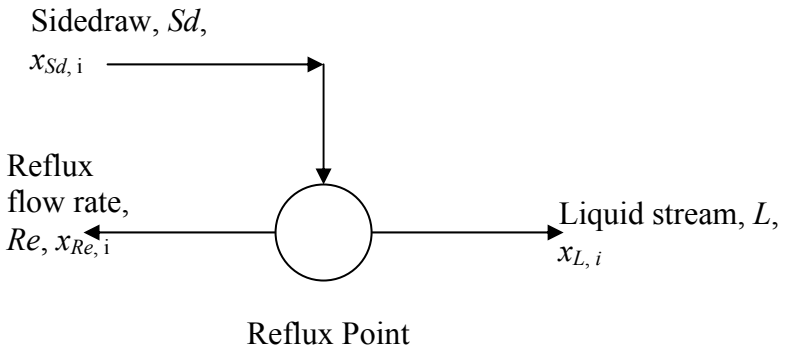


Figure 3.6: Reflux point schematic diagram

I) The liquid stream flow rate is calculated using Equation 3.11 based on (Douglas *et al*., 1996).

$$\text{Liquid stream flow rate, } L = \frac{C_v\,\rho\left(\frac{M_{\text{Re}}MW}{A_{\text{Re}}\rho}\right)^{0.5}}{MW} \qquad (3.11)$$

II) The liquid density is calculated as shown in Equation 3.12.

$$\text{Liquid density, } \rho = \sum_{i=1}^{9} x_i \rho_i \qquad (3.12)$$

III) Components mass balance in term of mole fraction derivative expression.

$$\frac{dx_{N,i}}{dt} = \frac{Sdx_{Sd,i} - \text{Re}\, x_{\text{Re},i} - Lx_{L,i}}{M_{\text{Re}}} \qquad (3.13)$$

where A_R = cross-sectional area of reflux drum

3.5.1.5 Pumparound Point Model

The schematic diagram for the pumparound point is shown in Figure 3.7.

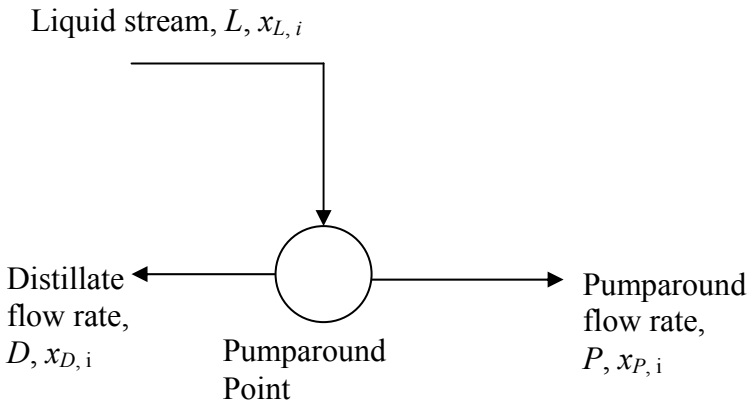


Figure 3.7: Pumparound point schematic diagram

I) The distillate flow rate has a very close relationship with the pumparound point molar hold-up.

$$\text{Distillate flow rate, } D = \frac{C_v \rho (\frac{M_P MW}{A_P \rho})^{0.5}}{MW} \qquad (3.14)$$

II) The liquid density is calculated as shown in Equation 3.15.

$$\text{Liquid density, } \rho = \sum_{i=1}^{9} x_i \rho_i \qquad (3.15)$$

III) Components mass balance in term of mole fraction derivative expression.

$$\frac{dx_{N,i}}{dt} = \frac{Lx_{L,i} - Px_{P,i} - Dx_{D,i}}{M_P} \qquad (3.16)$$

where A_P = cross-sectional area of pumparound drum

3.5.1.6 Pumparound Tray Model (N = 1)

The schematic diagram for pumparound tray (tray 1) is shown in Figure 3.8. The major assumption in this tray is that the operating temperature and pressure of this tray is such that the flow rate of V_1 is very small and very little product is lost in V_1 as it is vent.

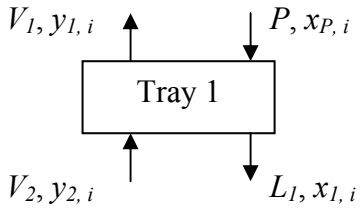


Figure 3.8: Pumparound tray schematic diagram

I) Total mass balance

$$\frac{dM_1}{dt} = P + V_2 - L_1 - V_1 \qquad (3.17)$$

II) Component mass balance

$$\frac{dx_i}{dt} = \frac{Px_{P,i} + V_2 y_{2,i} - V_1 y_{1,i} - L_1 x_{1,i}}{M_1} \qquad (3.18)$$

3.5.1.7 Sidedraw Tray Model (N=2)

The schematic diagram for the side draw tray (tray 2) is shown in Figure 3.9.

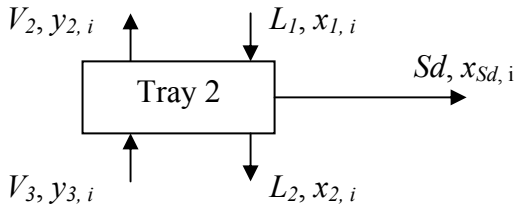


Figure 3.9: Sidedraw tray schematic diagram

I) Total mass balance

$$\frac{dM_2}{dt} = L_1 + V_3 - Sd - L_2 - V_2 \qquad (3.19)$$

II) Component mass balance

$$\frac{dx_i}{dt} = \frac{L_1 x_{1,i} + V_3 y_{3,i} - Sdx_{Sd,i} - V_2 y_{2,i} - L_2 x_{2,i}}{M_2} \qquad (3.20)$$

3.5.1.8 Reflux Tray Model (N = 3)

The schematic diagram for the reflux tray (tray 3) is shown in Figure 3.10.

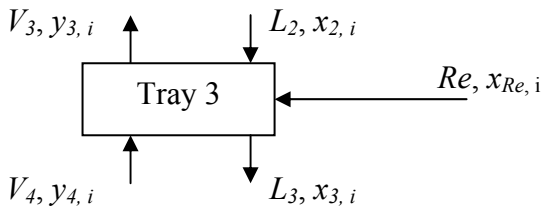


Figure 3.10: Reflux tray schematic diagram

I) Total mass balance

$$\frac{dM_3}{dt} = L_2 + V_4 + \mathrm{Re} - L_3 - V_3 \qquad (3.21)$$

II) Component mass balance

$$\frac{dx_i}{dt} = \frac{L_2 x_{2,i} + V_4 y_{4,i} + \mathrm{Re}\, x_{\mathrm{Re},i} - V_3 y_{3,i} - L_3 x_{3,i}}{M_3} \qquad (3.22)$$

3.5.2 Equilibrium Equations

Equilibrium equations involve using appropriate thermodynamic equations to determine the phase equilibrium between vapor and liquid phase. Therefore, E-equations are used to determine the liquid phase compositions, vapor phase compositions and temperature of the tray.

The study distillation column contains nine components: N-Hexanoic Acid, N-Octanoic Acid, N-Decanoic Acid, N-Dodecanoic Acid, N-Tetradecanoic Acid, N-Hexadecanoic Acid, Stearic Acid, Oleic Acid and Linoleic Acid. The

thermodynamic equations used for describing phase equilibrium and bubble point calculations are based on the Wilson correlation (Walas, 1985) for the liquid phase and the Virial Correlation (Van Ness *et al*., 1996) for the vapor phase.

The Virial Correlation is used to calculate the fugacity coefficient of the vapor phase because the system pressure is less than atmospheric pressure. This correlation gives a good approximation in this operating region (Van Ness *et al*., 1996). The Wilson correlation is used for estimating the activity coefficient of the liquid phase since the system exhibit non-ideality behavior (Reid *et al*., 1987). The involved equations for both the correlations are presented as follow:

Virial Correlation:

$$\ln \phi_k = \frac{P_{tot}}{RT}\left[B_{kk} + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j (2\delta_{ik} - \delta_{ij}) \right] \quad (3.23)$$

$$\delta_{ik} = 2B_{ik} - B_{ii} - B_{kk} \quad (3.24)$$

$$\delta_{ij} = 2B_{ij} - B_{ii} - B_{jj} \quad (3.25)$$

with $\delta_{ii} = 0$, $\delta_{kk} = 0$, etc., and $\delta_{ki} = \delta_{ik}$, etc.

$$B_{ij} = \frac{RT_{cij}}{P_{cij}}(B_{ij}^0 + \omega_{ij} B_{ij}^1) \quad (3.26)$$

$$B_{ij}^0 = 0.083 - \frac{0.422}{T_{rij}^{1.6}} \quad (3.27)$$

$$B_{ij}^1 = 0.139 - \frac{0.172}{T_{rij}^{4.2}} \quad (3.28)$$

$$T_{rij} = \frac{T}{T_{cij}} \qquad (3.29)$$

$$\omega_{ij} = \frac{\omega_j + \omega_i}{2} \qquad (3.30)$$

$$T_{cij} = (T_{ci}T_{cj})^{1/2} \qquad (3.31)$$

$$P_{cij} = \frac{Z_{cij}RT_{cij}}{V_{cij}} \qquad (3.32)$$

$$Z_{cij} = \frac{Z_{ci} + Z_{cj}}{2} \qquad (3.33)$$

$$V_{cij} = \left(\frac{V_{ci}^{1/3} + V_{cj}^{1/3}}{2} \right)^3 \qquad (3.34)$$

where:

P_{tot} = total pressure of the system
T = temperature of the system
R = universal gas constant, 83.14 bar.cm^3/mol.K
T_{ci} = critical temperature of component-*i*
P_{ci} = critical pressure of component-*i*
ω_i = critical accentric factor for component-*i*
Z_{ci} = critical compressibility factor for component-*i*
V_{ci} = critical volume for component-*i*
T_{ri} = reduced temperature of component-*i*
n = number of components
y_i = vapor mole fraction of component-*i*
ϕ_k = vapor fugacity coefficient for component-*i*

Wilson Correlation:

$$\ln \gamma_i = 1 - \ln \sum_{j=1}^{n} x_j \Lambda_{ij} - \sum_{k=1}^{n} \frac{x_k \Lambda_{ki}}{\sum_{j=1}^{n} x_j \Lambda_{kj}} \qquad (3.35)$$

$$\Lambda_{ij} = \frac{V_j}{V_i} \exp\left(\frac{-a_{ij}}{RT} \right) \qquad (3.36)$$

with $\Lambda_{ij} = 1$ for $i = j$, etc and $\Lambda_{ij} \neq \Lambda_{ji}$.

where:

γ_i = liquid phase activity coefficient for component-*i*

x_i = liquid mole fraction for component-*i*

Λ_{ij} = Wilson binary interaction parameters for component-*i* and *j*

V_i = liquid molar volume of component-*i*

a_{ij} = Wilson constant

The liquid mole fraction and vapor mole fraction for component-*i* are related by the following equation:

$$y_i \phi_i P_{tot} = x_i \gamma_i P_i^{sat} \qquad (3.37)$$

where:

P_i^{sat} = vapor pressure of component-*i*

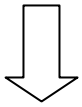
Equation 3.37 is the fundamental vapor-liquid equilibrium equation that will be used in this research. The bubble point calculation procedure that will be presented in the next section is based on this equation.

3.5.2.1 Bubble Point Calculations

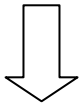
Liquid phase activity coefficient and vapor phase fugacity coefficient help in determining the vapor phase and liquid phase compositions. The tray temperature is depended on the tray pressure, liquid and vapor compositions. The bubble point calculations based on Equation 3.38 are known as the gamma-phi approach because the liquid phase coefficient and vapor phase coefficient are determined using two different thermodynamic models (Walas, 1985). The procedures of the bubble point calculations are presented in Figure 3.11:

Bubble Point Calculation Procedures by using Wilson Correlation and Virial Coefficient (gamma-phi approach)

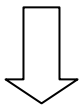
Objective- for the given pressure and liquid phase compositions, determine the vapor phase compositions and temperature.



Fix Pressure, P_{tot} and liquid phase compositions, x_i. Guess temperature, T



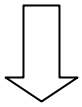
Set all $\phi_i = 1.0$ and all $\gamma_i = 1.0$. Identify component j.



Determine P_i^{sat} of each component using Antoine Equation:

$$\text{Log}_{10} P_i^{sat}(bar) = A_i - \frac{B_i}{T(K) + C_i} \qquad (3.38)$$

where A_i, B_i and C_i are Antoine constants.



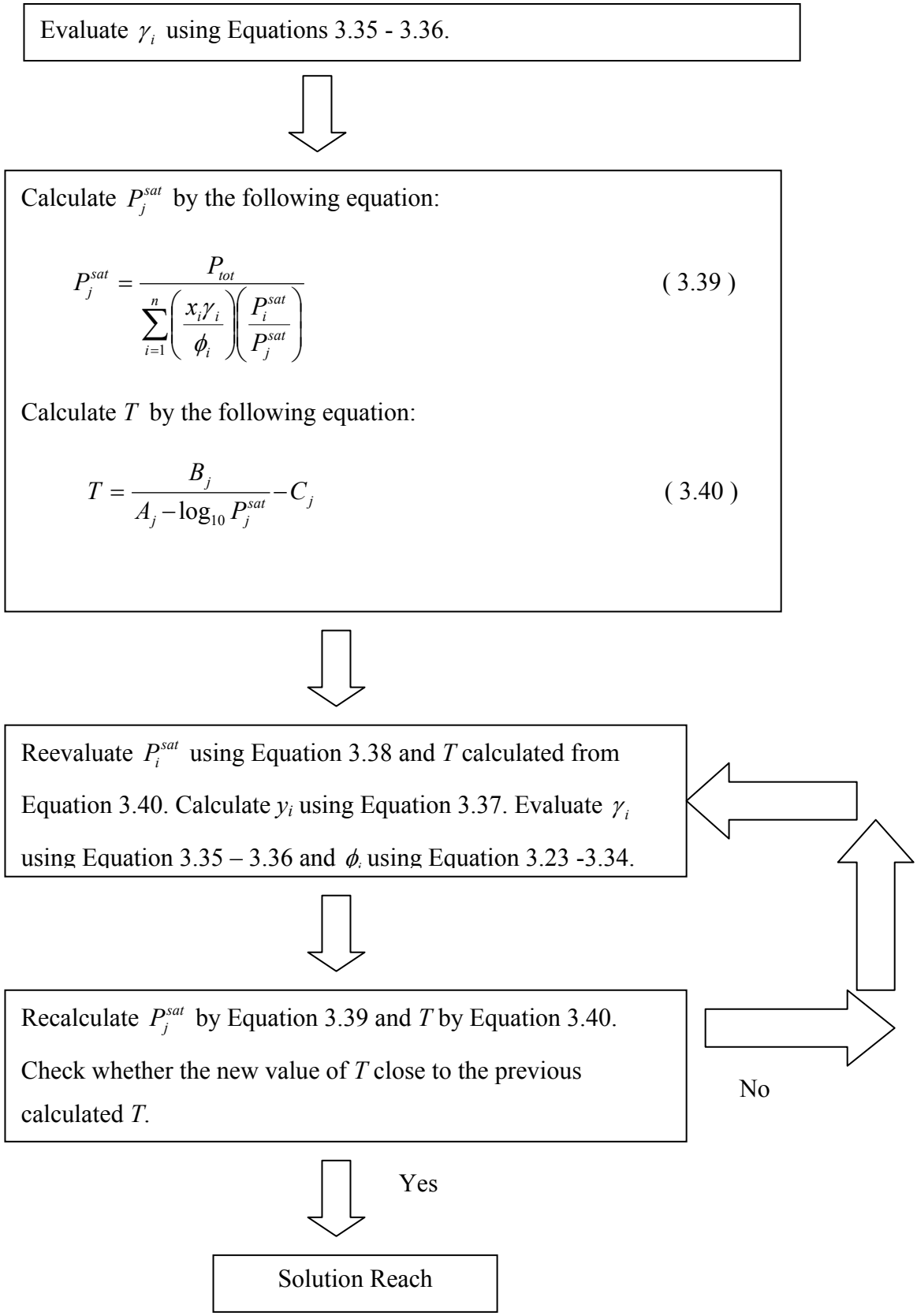


Figure 3.11: Bubble point calculation procedure

3.5.3 Summation Equations

Summation equations are used to determine the vapor phase and liquid phase compositions of the last component (Linoleic Acid).

$$x_9 = 1 - \sum_{i=1}^{8} x_i \qquad (3.41)$$

$$y_9 = 1 - \sum_{i=1}^{8} y_i \qquad (3.42)$$

3.5.4 Heat Balance Equations

Heat balance equations are used to determine the vapor flow rate leaving each tray. The reference temperature is set at 273 K. The liquid enthalpy and vapor enthalpy are calculated by using the following equations (Felder and Rousseau, 1986):

Liquid enthalpy, $h = \int_{273}^{T} A^* + B^*T + C^*T^2 + D^*T^3 \, dT$

$$= \left[A^*T + \frac{B^*T^2}{2} + \frac{C^*T^3}{3} + \frac{D^*T^4}{4} \right]_{273}^{T} \qquad (3.43)$$

Vapor enthalpy, $H = h + \Delta H_{vap}$ (3.44)

Heat of Vaporization, $\Delta H_{vap} = \Delta H_{vap,\,n} \left[\frac{T_c - T}{T_c - T_b} \right]^{0.38}$ (3.45)

Vapor leaving the reboiler, $V_{29} = \frac{Q_R + (L_{28} - B)h_B}{H_{29}}$ (3.46)

Vapor leaving each tray, $V_N = \dfrac{L_{N-1}h_{N-1} + V_{N+1}H_{N+1} - L_N h_N}{H_N}$ (3.47)

Where: A^*, B^*, C^*, D^* = constants for liquid heat capacity

ΔH_{vap} = heat of vaporization

$\Delta H_{vap,\,n}$ = heat of vaporization at normal boiling point

T_b = normal boiling point

L_{28} = liquid flow rate at tray-28

H_N = vapor enthalpy at tray-N

h_N = liquid enthalpy at tray-N

h_B = enthalpy for bottom stream

Q_R = reboiler duty

3.5.5 Hydraulic Equations

Besides MESH equations, Francis Weir hydraulic equations are used to determine the liquid flow rate leaving each tray and molar hold-up. Liquid flow rate leaving each tray is depended on the over weir height, h_{ow}. The involved hydraulic equations are shown as follow (Sinnot, 1984):

Liquid level, $L_H = \dfrac{M_N(MW)}{\rho(A_t)}$ (3.48)

Over weir height, $h_{ow} = L_H - W_L$ (3.49)

Tray liquid flow rate, $L_N = \left(\dfrac{h_{ow}}{750}\right)^{3/2}\left(\dfrac{\rho(3600)}{W_L(MW)}\right)$ (3.50)

Where: h_{ow} = over weir height

W_L = weir length

A_t = tray active area

After the column models and involved equations are formulated, the following task is to develop a proper algorithm that arranges the sets of equations in correct order so that dynamic simulation can be performed. The literature dynamic simulation algorithms are studied and surveyed in the following section.

3.6 Distillation Column Dynamic Simulation Literature Review

Multi-stage distillation column simulation basically is to solve the combinations of four sets of equations (MESH equations) and hydraulic equations. There are a lot of numerical algorithms discussed in the literature for solving MESH equations from dynamic state until steady state. The most widely used algorithms are proposed by Luyben (1963) and Gani *et al*. (1986). The algorithm developed in this research is based on the work of Lee and Kamarul (2001). The details for these algorithms are systematically presented in the following subsections.

3.6.1 Dynamic Simulation Algorithm Based on Luyben (1963)

The dynamic simulation algorithm proposed by Luyben (1963) is shown as follow:

1) Read data on the column size, components, physical properties, feeds and initial conditions (T, x, L).
2) Calculate the initial tray hold-up and the pressure profile.
3) Calculate temperature and vapor compositions from vapor-liquid equilibrium thermodynamic calculation method.
4) Calculate liquid and vapor enthalpies.
5) Calculate vapor flow rate leaving each tray using energy balance.
6) Evaluate all derivatives – total and component hold-ups.
7) Integrate.
8) Calculate new liquid rates from new hold-ups using Francis Weir equation.

9) Go back to step 3 and repeat.

3.6.2 Dynamic Simulation Algorithm Based on Gani *et al*. (1986)

The algorithm proposed by Gani *et al*. (1986) is also known as generalized model because the proposed algorithm allows the solution of a wide variety of problems, from open- and closed-loop responses of a single (and multiple) columns to operability studies (of feed changeover and start-up operations) and column instability studies (effect of plate hydraulics during transient operations). The simulation procedures are arranged as follow:

1) Specify the plate pressure. From initial components molar hold-up and total molar hold-up at time-t, determine the liquid composition on each plate.
2) From the plate pressure and liquid composition, determine plate temperature and vapor composition.
3) Determine the vapor and liquid enthalpy.
4) From the total molar hold-up and total pressure drop, determine the liquid and vapor flow across the plate, weeping, flooding and entrainment rates (if any).

The above algorithm involves a lot of hydraulic parameters like total head loss, head loss due to vapor flow, dry-cab head loss coefficient and other hydraulic parameters, in determining the liquid and vapor flow rates. The hydraulic parameters, which depend on types of plate like bubble-cap or sieve plate, are hardly obtained from the literature. The proposed method relies on hydraulic equations to determine the liquid and vapor flow rate. Furthermore, the proposed algorithm has been demonstrated on five case studies. However, the results were not compared to any commercial dynamic simulator's results. Therefore, the demonstrated debutanizer column showed mass non-balance condition and showed infeasible column pressure operating condition.

3.6.3 Dynamic Simulation Algorithm Based on Lee and Kamarul (2001)

Lee and Kamarul (2001) proposed an algorithm based on the work of Grassi (1990) and Luyben (1963). The proposed algorithm was able to simulate the debutanizer column from Gani *et al*. (1986) dynamically. The proposed method is shown as follows:

1) From the tray hold-up composition and tray pressure, calculate the equilibrium vapor composition and temperature by bubble point calculation. Equilibrium calculations have to base on suitable thermodynamic method.
2) Calculate the vapor and liquid enthalpy from their composition and tray temperature.
3) Calculate the liquid rate leaving the stage from Francis Weir equation. Liquid flow rate of a tray depends on the height of over weir.
4) Calculate the vapor flow rate leaving the stage from energy balance.
5) Calculate the components and total mass derivatives for each stage starting with the bottom and working up the column.
6) Integrate all derivatives to get tray hold-up and components composition.
7) Increment time by the integration time step.

This method was developed to simulate a normal distillation column. Since the study column in this research involves a pumparound system at the top of the column, there will be some modifications on the algorithm by Lee and Kamarul (2001) before using it to the case study in this research.

3.7 Dynamic Simulation of the Study Column

Dynamic simulation activities can be divided into four sub-sections: pre-simulation, algorithm formulation, program development (Lee and Kamarul, 2001) and tuning of controllers for the study column.

3.7.1 Pre-Simulation

Pre-simulation activities consist of: understanding the assumptions that are used in simulating the column, acquiring the inlet stream conditions, simulate the study column using commercial simulator Design II and specify the desired separation flow rates and fix the degrees of freedom. The adopted assumptions in simulating the column are structured as follow:

1) Liquid on the tray is perfectly mixed and incompressible. Compositions of the liquid stream leaving the tray will be the same as the tray molar hold-up compositions.
2) Tray vapor hold-ups are negligible because the column operating pressures are less than 10 bars (Lee and Kamarul, 2001).
3) Vapor and liquid are in thermal equilibrium. The temperature for liquid stream leaving the tray and the vapor stream leaving the tray is the same.
4) Vapor and liquid are in the phase equilibrium.
5) Pressure is constant on each tray but varies linearly up the column from bottom pressure to the top column pressure. In another word, pressure drop for each tray is the same.
6) Coolant and steam dynamics are negligible.
7) Dynamic changes in internal energy on the trays are negligible compared with the latent heat effects. Therefore, energy balance on each tray is just algebraic.

After the assumptions that are used in simulating the column have been studied, the following task is to acquire enough information for the inlet feed stream. This information is obtained from the plant simulated data (**Appendix B**). The given data was sufficient enough for the inlet stream.

The study column is simulated using commercial simulator, Design II, to obtain a set of steady state results that comprises the following column parameters profile:

1) Column temperature profile.
2) Column liquid flow rate profile.

3) Column vapor flow rate profile.
4) Trays liquid compositions profile.
5) Trays vapor compositions profile.
6) Liquid enthalpy profile.
7) Vapor enthalpy profile.
8) Bottom stream flow rate and compositions.
9) Distillate stream flow rate and compositions.
10) Sidedraw stream flow rate and compositions.
11) Pumparound stream flow rate and compositions.
12) Reflux stream flow rate and compositions.

The Design II simulation results are used as the initial guess for the dynamic simulation of the column in Matlab. The parameter profiles that are going to be used as the initial guesses for the written dynamic simulation are the column temperature profile and the liquid compositions profile.

From the degree of freedom analysis, it is shown that there are 31 degrees of freedom. Therefore, 31 variables have to be fixed in order to achieve the desired compositions for the bottom and distillate streams.

These 31 variables are:

1) 28 tray pressure. The tray pressures vary linearly from top column to the bottom of the column. Each tray encounters 0.00095 bar pressure drop.
2) Reflux flow rate = 10 kmole/hr.
3) Pumparound flow rate = 49.6 kmole/hr.
4) Reboiler duty = 2.16×10^6 kJ/hr.

3.7.2 Algorithm Formulation

The column models that describe the column behavior are bottom column model, normal tray model, feed tray model, reflux point model, pumparound point model, sidedraw tray model, reflux tray model and pumparound tray model. These models are comprised of mass balance equations, equilibrium equations, heat balance equations, summation equations and hydraulic equations. Algorithm formulation involved solving these equations in the proper order and sequence so that the column behaviors are revealed. For this research, the dynamic simulation algorithm proposed by Lee and Kamarul (2001) are referred. However, since in this research the study column has a pumparound system incorporated at the top of the column; some modifications are made before applying the stated algorithm.

The systematic procedures in formulating this research's dynamic simulation algorithm are elaborated as follow:

The algorithm for the bottom column to the 4^{th} tray:

1) Initialize the following derivative variables:

i) Bottom liquid level and compositions – 9 derivative variables

ii) Each tray molar hold-up and compositions – 225 derivative variables

There are 234 derivative variables that have to be initialized since there are 234 derivative equations.

2) Calculate the bubble point from the bottom tray till the 4^{th} tray.

The tray pressure is fixed since the pressure drop for each tray is assumed to be the same. The bottom column pressure is fixed at 0.1333 bar and each tray will encounters a 0.00095 bar pressure drop. The tray temperature and liquid compositions are guessed. The guessing values are referred from the Design II steady state simulation results. The bubble point calculation procedures which were presented in Section 3.5.2.1 are referred. The purposes of bubble point calculations are to determine the tray temperature and vapor phase compositions.

3) Calculate the liquid and vapor enthalpy from bottom tray till the 4th tray. Based on the T_n, $x_{i,\ n}$ and $y_{i,\ n}$ as calculated in the previous step, liquid and vapor enthalpy for each tray are determined via Equation 3.43 to 3.44.

4) Calculate the liquid flow rate by Francis Weir equations.

From the initial bottom liquid level and tray's molar hold-up, the following variables are determined from the bottom stage till the 4th stage:

i) Bottom flow rate by Equation 3.2.

ii) Liquid leaving each tray by Equation 3.48 – 3.50.

5) Calculate the vapor flow rate from energy balance.

Based on the calculated liquid and vapor enthalpy, the vapor flow rate leaving each tray is determined by Equation 3.46 for the reboiler stage and Equation 3.47 for other trays.

6) Evaluate the molar hold-up and components composition derivative variables.

For every time increment h = 0.005 hour, the evaluation for each derivative variable is started from the bottom stage till the 4th stage using Runge-Kutta fourth order method.

The algorithm for the 3rd tray to 1st tray, pumparound point and reflux point:

7) Start at the reflux point. Initialize the derivative variables as in Step 1. Perform bubble point calculation as in Section 3.5.2.1, calculate the liquid and vapor enthalpy for all the streams and the flow rate of the liquid stream is determined using Equation 3.11. Repeat Step 4 to 6.

8) Pumparound point. Calculation procedure as Step 7 with the distillate flow rate calculated using Equation 3.14.

9) Tray-3 to Tray-1. The calculation procedure is as Step 1 to 6.

10) Increase time step.

After all the derivative variables at time = t are evaluated, the calculation steps are repeated until time = t_{final}.

3.7.3 Simulation Program Development

The developed algorithm served as the basis in formulating the simulation program to perform dynamic simulation. Matlab software is used to write the program. The written program has the following characteristics:

1) Process variables: feed stream flow rate, feed stream temperature, reflux flow rate, pumparound flow rate and reboiler duty are having randomly generated value. In the commercial simulators, these variables can only accept fixed value.
2) These process variables have autocorrelation effect, which is one of industrial data characteristics.
3) The developed program has a main program and sub-programs. The main program mainly contains the sub-programs executing sequence. The following calculations are written in sub-program:
    a) Bubble point calculations.
    b) Thermodynamic models for vapor-liquid equilibrium calculations.
    c) Liquid and vapor enthalpy calculations.
    d) Runge-Kutta fourth order calculations.

The developed dynamic simulation program is used to perform dynamic simulation. The performance of the developed dynamic simulation program is evaluated in the Section 3.8.

3.7.4 Controller Tuning

Prior to the tuning of controllers, the process time constant, τ, the dead time, t_d, and the static gain, K_p, of process sections (the process sections that are to be installed with controllers) have to be determined. The methods of determining these parameters are using the process reaction curve method (Stephanopolous, 1984) and the method found in Smith and Corripio (1985). Figure 3.12 shows the process reaction curve of a process in a step test.

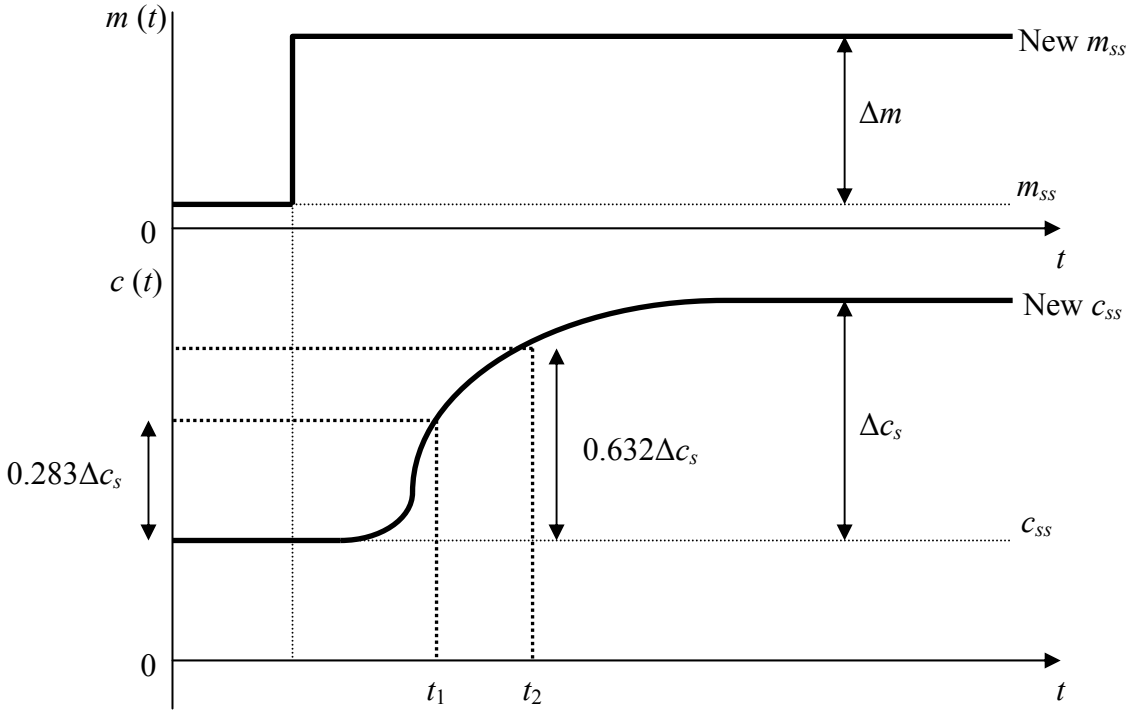


Figure 3.12: Process reaction curve of a process in a step test

In Figure 3.12,

t = time of the process

$m\ (t)$ = value of input variable at time t

$c\ (t)$ = value of output variable at time t

c_{ss} = value of output variable at steady-state

m_{ss} = value of input variable at steady-state

Δm = new $m_{ss} - m_{ss}$

Δc_s = new $c_{ss} - c_{ss}$

According to Smith and Corripio (1985),

$$K_p = \frac{\Delta c_s}{\Delta m} \qquad (3.51)$$

$$\tau = \frac{3}{2}(t_2 - t_1) \qquad (3.52)$$

$$t_d = t_2 - \tau_p \qquad (3.53)$$

In tuning the controllers of the column, the first parameter that was determined before tuning of any controllers are the controller sampling time, T_{APC}. The method in determining this parameter is shown in the next section.

3.7.4.1 Controller Sampling Time, T_{APC}

The controller sampling time, T_{APC}, of the study column is determined using the guideline proposed by Ogunnaike and Ray (1994). Ogunnaike and Ray (1994) proposed using one tenth to one twentieth of smallest time constant of a process for its controller sampling time. The bottom liquid level-bottom flow rate process section of the study column process has the smallest time constant, τ = 0.1 hour. Therefore, the controller sampling time of the study column is set at T_{APC} = 0.01 hour (one tenth of the smallest time constant of the study column process).

The controllers of the study column are tuned using the Cohen-Coon method (Stephanopolous, 1984). Equation 3.54 to Equation 3.59 show the formulas to calculate the controller gain, K_c, integral time constant, τ_I, and derivative time constant, τ_D, of different types of controllers.

For proportional controllers,

$$K_c = \frac{1}{K_p}\left(\frac{\tau}{t_d}\right)\left(1+\frac{t_d}{3\tau}\right) \quad (3.54)$$

For proportional-integral controllers,

$$K_c = \frac{1}{K_p}\left(\frac{\tau}{t_d}\right)\left(0.9+\frac{t_d}{12\tau}\right) \quad (3.55)$$

$$\tau_I = t_d\left(\frac{30+3t_d/\tau}{9+20t_d/\tau}\right) \quad (3.56)$$

For proportional-integral-derivative controllers,

$$K_c = \frac{1}{K_p}\left(\frac{\tau}{t_d}\right)\left(\frac{4}{3}+\frac{t_d}{4\tau}\right) \quad (3.57)$$

$$\tau_I = t_d\left(\frac{32+6t_d/\tau}{13+8t_d/\tau}\right) \quad (3.58)$$

$$\tau_D = t_d\left(\frac{4}{11+2t_d/\tau}\right) \quad (3.59)$$

There are seven control loops in the study column. The following sections show the tuning procedure and controller properties of each control loop.

3.7.4.2 Bottom Temperature Controller

The reboiler of the study column is assumed to be a total reboiler with hot oil as the heating fluid for the reboiler system. The control variable in the bottom temperature control loop is the bottom temperature and the proposed manipulated variable is the flow rate of the hot oil in the reboiler system. The hot oil is assumed to have a heat capacity, C_p = 7.5 kJ/kg.K (Geankoplis, 1995) and a change in

temperature, ΔT = 40K. These two parameters are assumed to be constant and the steady-state (SS) value of the hot oil flow rate is 7192 kg/hr. The flow rate of the hot oil is varied (various step changes) and the static gain, K_p, time constant, τ, and dead time, t_d, for each change of the value of the flow rate of the hot oil is recorded and shown in Table 3.5.

Table 3.5: The step change results for bottom temperature controller

Change in flow rate of hot oil	Static Gain, K_p, K/(kg/hr)	Time Constant, τ, hour	Dead Time, t_d, hour
+5% of SS value	0.03547	0.2	0.05
+10% of SS value	0.03673	0.2	0.05
-5% of SS value	0.03615	0.2	0.05
-10% of SS value	0.03662	0.2	0.05
Average Value	0.03624	0.2	0.05

The controller used for controlling the bottom temperature is Proportional-Integral-Derivative (PID) controller. Using the average value of K_p, τ, and t_d from Table 3.5, the controller parameters such as K_c, τ_I and τ_D are calculated using Equation 3.57 – Equation 3.59. The average values are used in order to check the linearity of the process response to step changes. The calculated parameters and information of the bottom temperature PID controller are shown in Table 3.6.

Table 3.6: Properties of bottom temperature PID controller

Controller Type	Control Variable	Manipulated Variable	Local Disturbance	K_c, (kg/hr)/K	τ_I, hour	τ_D, hour
PID	Bottom temperature, T_{bot}	Hot oil flow rate, F_{hot}	Liquid flow rate from tray 28, L_{28}	154.050	0.11167	0.0174

3.7.4.3 Bottom Liquid Level Controller

The control variable for the bottom liquid level control loop is the bottom liquid level while the proposed manipulated variable is the bottom flow rate. The value of the bottom flow rate is varied (various step changes) and K_p, τ, and t_d for each change of the value of the bottom flow rate is recorded and shown in Table 3.7. However, only small changes of the value of the bottom flow rate is carried out as the developed Matlab program will fail to converged for large changes in the value of the bottom flow rate.

Table 3.7: The step change results for bottom liquid level controller

Change in flow rate of bottom stream	Static Gain, K_p, m/(kmole/hr)	Time Constant, τ, hour	Dead Time, t_d, hour
+0.5% of SS value	-1.04750	0.1	0.01
+1.0% of SS value	-1.04754	0.1	0.01
-0.5% of SS value	-1.04753	0.1	0.01
-1.0% of SS value	-1.04754	0.1	0.01
Average Value	-1.04753	0.1	0.01

The controller used for controlling the bottom liquid level is Proportional-Integral (PI) controller. Using the average value of K_p, τ, and t_d from Table 3.7, K_c and τ_I are calculated using Equation 3.55 and Equation 3.56. The calculated parameter and information of the bottom liquid level P controller are shown in Table 3.8.

Table 3.8: Properties of bottom liquid level PI controller

Controller Type	Control Variable	Manipulated Variable	Local Disturbance	K_c, (kmole/hr)/m	τ_I, hour
PI	Bottom liquid level, $L_{H,\,b}$	Bottom flow rate, B	Liquid flow rate from tray 28, L_{28}	-8.6712	0.0275

3.7.4.4 Pumparound Temperature Controller

The control variable for the pumparound temperature controller is the temperature of the pumparound stream while the proposed manipulated variable is the flow rate of the cooling water in the pumparound cooler system. The water has a heat capacity C_p = 4.184 kJ/kg.K (Geankoplis, 1995) and a change in temperature, ΔT = 10K. These two parameters are assumed to be constant and the steady-state (SS) value of the cooling water flow rate is 411.57 kmole/hr. The value of the cooling water flow rate is varied (various step changes) and K_p, τ, and t_d for each change of the value of the cooling water flow rate is recorded and shown in Table 3.9.

Table 3.9: The step change results for pumparound temperature controller

Change in flow rate of cooling water	Static Gain, K_p, K/(kmole/hr)	Time Constant, τ, hour	Dead Time, t_d, hour
+5% of SS value	-0.30282	1.1	0.1
+10% of SS value	-0.30287	1.1	0.1
-5% of SS value	-0.30302	1.1	0.1
-10% of SS value	-0.30297	1.1	0.1
Average Value	-0.30292	1.1	0.1

The controller used for controlling the pumparound temperature is PID controller. Using the average value of K_p, τ, and t_d from Table 3.9, the controller parameters such as K_c, τ_I and τ_D are calculated using Equation 3.57 – Equation 3.59. The calculated parameters and information of the pumparound temperature PID controller are shown in Table 3.10.

Table 3.10: Properties of pumparound temperature PID controller

Controller Type	Control Variable	Manipulated Variable	Local Disturbance	K_c, (kmole /hr)/K	τ_I, hour	τ_D, hour
PID	Pumparound temperature, T_P	Cooling water flow rate, F_{cw}	Liquid flow rate, L_{Dp}	22.708	0.4579	0.0704

3.7.4.5 Sidedraw Tray Liquid Level Controller

The control variable for the sidedraw tray liquid level control loop is the sidedraw tray liquid level while the proposed manipulated variable is the distillate flow rate. The value of the distillate flow rate is varied (various step changes) and K_p, τ, and t_d for each change of the value of the distillate flow rate is recorded and shown in Table 3.11.

Table 3.11: The step change results for sidedraw tray liquid level controller

Change in flow rate of bottom stream	Static Gain, K_p, m/(kmole/hr)	Time Constant, τ, hour	Dead Time, t_d, hour
+5% of SS value	-0.02299	0.2	0.05
+10% of SS value	-0.02324	0.2	0.05
-5% of SS value	-0.02140	0.2	0.05
-10% of SS value	-0.02177	0.2	0.05
Average Value	-0.02235	0.2	0.05

The controller used for controlling the sidedraw liquid level is PI controller. Using the average value of K_p, τ, and t_d from Table 3.11, the K_c of this controller is calculated using Equation 3.55 and Equation 3.56. The calculated parameter and information of the sidedraw tray liquid level PI controller are shown in Table 3.12.

Table 3.12: Properties of sidedraw tray liquid level PI controller

Controller Type	Control Variable	Manipulated Variable	Local Disturbance	K_c, (kmole/hr)/m	τ_I, hour
PI	Sidedraw tray liquid level, $L_{H, Sd}$	Distillate flow rate, D	Liquid flow rate from tray 1, L_1	-164.802	0.1098

3.7.4.6 Top Column Pressure Controller

The control variable for the top column pressure controller loop is the top column pressure while the proposed manipulated variable is the top vapor flow rate. The value of the top vapor flow rate is varied (various step changes) and K_p, τ, and t_d for each change of the value of the top vapor flow rate is recorded and shown in Table 3.13.

Table 3.13: The step change results for top column pressure controller

Change in flow rate of top vapor stream	Static Gain, K_p, bar/(kmole/hr)	Time Constant, τ, hour	Dead Time, t_d, hour
+5% of SS value	-9.60095	1.0	0.1
+10% of SS value	-9.60120	1.0	0.1
-5% of SS value	-9.60194	1.0	0.1
-10% of SS value	-9.60169	1.0	0.1
Average Value	-9.60144	1.0	0.1

The controller used for controlling the top column pressure is Proportional-Integral (PI) controller. Using the average value of K_p, τ, and t_d from Table 3.13, the controller parameters such as K_c and τ_I are calculated using Equation 3.55 and Equation 3.56. The calculated parameters and information of the top pressure PI controller are shown in Table 3.14.

Table 3.14: Properties of top column pressure PI controller

Controller Type	Control Variable	Manipulated Variable	Local Disturbance	K_c, (kmole/hr)/Bar	τ_I, hour
PI	Top column pressure, P_{Top}	Top vapor flow rate, V_{top}	Vapor flow rate from Tray 2, V_2	-0.946	0.2755

3.7.4.7 Reflux Flow Rate Controller

The control variable for the reflux flow rate control loop is the reflux flow rate while the proposed manipulated variable is the liquid flow rate prior to the reflux stream. The valve in the line of flow of the reflux is modeled as a first order lag system with linear characteristics. The major assumptions in this valve model are the pressure drop across the valve is constant and a function of flow only. The value of the liquid flow rate prior to the reflux is varied (various step changes) and K_p, τ, and t_d for each change of the value of the flow rate of the liquid prior to the reflux is recorded and shown in Table 3.15.

Table 3.15: The step change results for reflux flow rate controller

Change in flow rate of liquid stream prior to reflux	Static Gain, K_p, bar/(kmole/hr)	Time Constant, τ, hour	Dead Time, t_d, hour
+5% of SS value	1.0003	0.1072	0.0378
+10% of SS value	1.0005	0.1096	0.0352
-5% of SS value	1.0002	0.1092	0.0372
-10% of SS value	1.0004	0.1069	0.0352
Average Value	1.0004	0.1082	0.0363

The parameter and information of reflux flow rate controller are shown in Table 3.16.

Table 3.16: Properties of reflux flow rate PI controller

Controller Type	Control Variable	Manipulated Variable	Local Disturbance	K_c	τ_I, hour
PI	Reflux flow rate, *Re*	Liquid flow rate prior to reflux stream	Side draw flow rate, *Sd*	2.7648	0.0716

3.7.4.8 Pumparound Flow Rate Controller

The control variable for the pumparound flow rate control loop is the pumparound flow rate while the proposed manipulated variable is the liquid flow rate prior to the pumparound stream. The valve in the line of flow of the pumparound is modeled as a first order lag system with linear characteristics. The major assumptions in this valve model are the pressure drop across the valve is constant and a function of flow only. The value of the liquid flow rate prior to the pumparound is varied (various step changes) and K_p, τ, and t_d for each change of the value of the flow rate of the liquid prior to the pumparound is recorded and shown in Table 3.17.

Table 3.17: The step change results for pumparound flow rate controller

Change in flow rate of liquid stream prior to pumparound	Static Gain, K_p, bar/(kmole/hr)	Time Constant, τ, hour	Dead Time, t_d, hour
+5% of SS value	3.8319	0.1023	0.0743
+10% of SS value	2.0069	0.1122	0.0836
-5% of SS value	3.5824	0.1245	0.0635
-10% of SS value	1.6431	0.1302	0.0770
Average Value	2.7661	0.1173	0.0746

The parameter and information of pumparound flow rate controller are shown in Table 3.18.

Table 3.18: Properties of pumparound flow rate PI controller

Controller Type	Control Variable	Manipulated Variable	Local Disturbance	K_c	τ_I, hour
PI	Pumparound flow rate, P	Liquid flow rate prior to pumparound stream	Liquid stream flow rate, L_{Dp}	0.5417	0.1096

3.8 Dynamic Simulation Program Performance Evaluation

The developed simulation program is dynamic in nature while the available commercial simulator (Design II) is steady state in nature. The performance of the developed dynamic simulation program is evaluated against the data from the plant simulated data (**Appendix B**). The performance evaluation is conducted using the following aspect:

1) The ability of the developed algorithm and program to achieve steady state results that close to the steady state results in the plant simulated data.

3.8.1 Steady State Results Comparisons

The simulation results generated from the developed program are presented in Table 3.19.

Table 3.19: Matlab simulation results for steady state

Parameters	Bottom Stream	Distillate Stream
Flow rate (kmole/hr)	37.5	4.0
Temperature (K)	508.05	344.17
Pressure (bar)	0.1333	0.11
Compositions (kmole/kmole)		
Hexanoic Acid	0	0.02165
Octanoic Acid	0	0.52351
Decanoic Acid	0.00036	0.44145
Dodecanoic Acid	0.57591	0.01395
Tetradecanoic Acid	0.17257	0
Hexadecanoic Acid	0.07469	0
Stearic Acid	0.01632	0
Oleic Acid	0.13419	0
Linoleic Acid	0.02282	0
Sidedraw flow rate (kmole/hr)	63.6	
Liquid stream flow rate (kmole/hr)	53.6	
Reflux flow rate (kmole/hr)	10.0	
Pumparound flow rate (kmole/hr)	49.6	
Reboiler duty (kJ/hr)	2.158×10^6	

The steady state results generated by Matlab simulator are compared to the plant simulated data. Results of these comparisons are shown in Table 3.20. The plant simulated data are used as the basis in the comparison. The percentage difference is calculated as follow:

Percentage Difference

$$= |(\text{Plant simulated data} - \text{Matlab result})/(\text{Plant simulated data})| \times 100\ \% \qquad (3.57)$$

Table 3.20: Comparison of simulator results between Matlab and Plant Simulated Data

Parameters	**Matlab Simulation Results**	**Plant Simulated Data**	**% Difference (%)**
Bottom Stream			
Flow rate (kmole/hr)	37.504	37.48	0.064
Temperature (K)	508.05	509.55	0.294
Compositions (kmole/kmole)			
Hexanoic Acid	0	0	0
Octanoic Acid	0	0	0
Decanoic Acid	0.00036	0.00034	5.882
Dodecanoic Acid	0.57591	0.57778	0.324
Tetradecanoic Acid	0.17257	0.17394	0.788
Hexadecanoic Acid	0.07469	0.07532	0.836
Stearic Acid	0.01632	0.01614	1.115
Oleic Acid	0.13419	0.13516	0.718
Linoleic Acid	0.02281	0.02273	0.352
Distillate Stream			
Flow rate (kmole/hr)	4.00	3.998	0
Temperature (K)	344.17	344.15	0
Compositions (kmole/kmole)			
Hexanoic Acid	0.02165	0.02132	1.547
Octanoic Acid	0.52351	0.52101	0.480
Decanoic Acid	0.44145	0.44283	0.312
Dodecanoic Acid	0.01395	0.01484	6.000
Tetradecanoic Acid	0	0	0
Hexadecanoic Acid	0	0	0
Stearic Acid	0	0	0
Oleic Acid	0	0	0
Linoleic Acid	0	0	0

From the comparison shown in Table 3.20, the maximum difference is about 6.0%. The written program is able to achieve the desired bottom and distillate compositions that are close to the simulator. The liquid and vapor flow rate profile of the Matlab program were also compared to the plant simulated data. The liquid profile of the Matlab program was very close to the liquid profile of the plant simulated data. The vapor profile exhibits some differences but the maximum difference is about 5% which is acceptable. This difference is caused by the different method in calculating the enthalpies between the two programs and also in the plant simulated data, the presence of neutrals and water was considered in the feed stream to the column while in the Matlab program, these components were ignored as their compositions in the feed stream is very small (around 0.004 for each component). However, the developed Matlab program was able to achieve the desired bottom and distillate flow rate and compositions. Therefore, the developed Matlab program is capable of generating the desired data for this research.

3.9 Chapter Summary

This chapter presents modeling of the study column, literature review on dynamic simulation algorithm, dynamic simulation algorithm formulation, dynamic simulation program development and program performance evaluation. The study column is a distillation column from literature. There are nine chemical components in this column.

There are 31 degrees of freedom for the study column. The variables that are fixed are the 28 trays pressure, the reboiler duty, reflux flow rate and pumparound flow rate. The modeling of the column requires MESH equations and hydraulic equations. Total mass balance and components balance are written for each model in the column. The equilibrium equations based on Wilson Correlation and Virial Correlation are used to determine the values of the activity coefficients and the fugacity coefficients of the liquid and vapor phase respectively. Heat balance equations are used to determine the vapor leaving each tray. Hydraulic equations are used to calculate the liquid flow rate leaving each tray and tray molar hold-up.

The dynamic simulation program built using Matlab software was developed based on the algorithm in Lee and Kamarul (2001). The algorithm proposed by Lee and Kamarul (2001) worked well in a normal distillation column. The study column of this research has a pumparound system on top of the column, some modifications on the algorithm by Lee and Kamarul (2001) was done to suit the needs of the study column. The initial values for the derivative variables in the dynamic simulation program were obtained using a commercial steady state simulator, Design II. The results obtained from Matlab were compared to the data from literature. The developed program was able to produce results close to the data from literature. As a conclusion, the developed program is ready to generate a set of normal operating condition (NOC) data and disturbance data for usage later on in this research. A summary of this chapter can be represented in Figure 3.13.

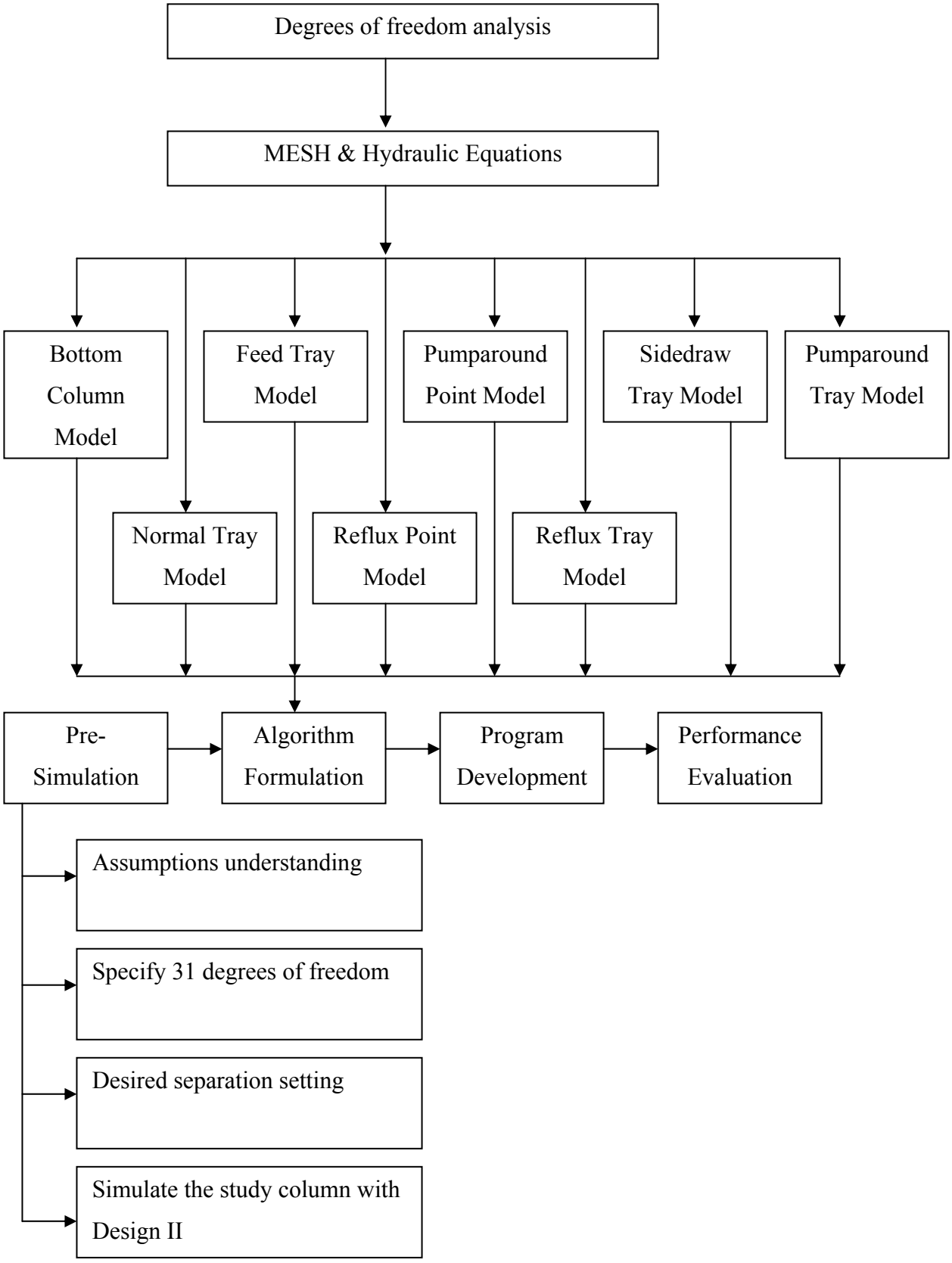


Figure 3.13: Summary of scope of work for distillation column dynamic modeling and simulation

CHAPTER IV

METHODOLOGY FOR IMPLEMENTATION OF FAULT DETECTION AND DIAGNOSIS

4.1 Introduction

This chapter contains four major sections: chapter introduction, development of the fault detection and diagnosis (FDD) algorithm based on Multivariate Statistical Process Control (MSPC), the performance evaluation method for the developed FDD algorithm and chapter summary.

The second section comprises of the methodologies in formulating the FDD algorithm based on Normal Correlation (NC), Principal Component Analysis (PCA) and Partial Correlation Analysis (PCorrA). The procedures include: generating a set of nominal operation condition (NOC) data, development of MSPC fault detection and diagnosis tools and applying the developed tools to detect fault situations and identify the fault causes.

The third section outlines the fault detection and diagnosis performance evaluation method.

4.2 Development of Fault Detection and Diagnosis Algorithm Based on Multivariate Statistical Process Control (MSPC)

In this research, a fault detection and diagnosis (FDD) algorithm is established in order to detect any non-statistical stable condition. The non-statistical-stable condition is also known as fault situation. The root causes for the fault situations are known as fault causes. The performance of the developed algorithm will be evaluated as in Section 4.3 of this chapter.

A distillation column is selected as the study unit operation. The developed FDD algorithm is applicable to any type of unit operation in the industry and not only on a distillation column as shown in this research. The first step in developing the FDD algorithm is the selection of quality variables of interest and the corresponding key process variables that are related with the selected quality variables of interest. The next section will explain how the selection of quality variables and key process variables are carried out. The employed procedures in formulating the MSPC FDD algorithm for this research are systematically presented in the following sections.

4.2.1 Selection of Variables

The study column is the first column from a series of column from a palm oil fractionation plant (**Appendix B**). This column is to separate the fatty acid components: N-Decanoic Acid and lighter components from N-Dodecanoic Acid and heavier components in the feed stream. The quality variables of interest in this research are the oleic acid and linoleic acid composition in the bottom stream. Oleic acid and linoleic acid are important products at the end of the fractionation plant. Table 4.1 shows the information on the selected quality variables of interest.

Table 4.1: Selected quality variables of interest

Quality Variable of Interest	Name of Quality Variable	Location of Variable
1	Linoleic acid mole fraction, x_8	Bottom stream of column
2	Oleic acid mole fraction, x_9	Bottom stream of column

After the selection of quality variables of interest, the key process variables that are related with the selected quality variables will be chosen from a list of available (measured) process variables. The correlation (normal correlation) between the process variables with the two quality variables will serve as a good indication of whether a process variables is key process variable that has major contribution to the variation of the quality variables or not. This is important as the proposed FDD algorithm in this research is based on the correlation between the selected key process variables and the quality variables of interest.

Small random noise (arithmetic average =0 and standard deviation =1) is added into process variables such as feed flow rate, feed temperature, cooler duty, reboiler duty, pumparound flow rate and reflux flow rate to generate a set of data with small fluctuation from their steady-state (SS) value (the quality variables have a fluctuation of ± 5% of their SS value). The correlation (normal correlation) between a list of process variables with the two quality variables of interest is determined. Process variables with high correlation (absolute correlation value ≥ 0.1) are chosen as the key process variables that have high contribution to the variation of the two quality variables of interest. The full list of the process variables and their correlation values are given in Section 5.2 of Chapter 5. Table 4.2 shows the information of the selected key process variables.

Table 4.2: Selected key process variables

Key Process Variable	Name of Key Process Variable	Location of Key Process Variable
1	Feed flow rate, L_f	Feed stream of column
2	Feed temperature, T_f	Feed stream of column
3	Reflux flow rate, Re	Reflux stream of column
4	Pumparound flow rate, P	Pumparound stream of column
5	Reboiler duty, Q_r	Reboiler of column
6	Bottom temperature, T_{bot}	Bottom Column Temperature

Once the quality variables of interest and key process variables are selected, the column program is ready to generate the required data for the development of the proposed FDD algorithm. The process sampling time, T_{MSPC}, is an important parameter that needs to be determined before any data can be sampled from the column program. This process sampling time is for sampling data to be used in derivation of correlation coefficients between the selected key process variables and the quality variables of interest and is different from the controller sampling time, T_{APC}, which is used to sample data for control purposes. The next section gives the steps used to determine the process sampling time.

4.2.2 Process Sampling Time, T_{MSPC}

In a time series data, autocorrelation plays an important role in affecting the variation of the data. Autocorrelation is the correlation between successive data in a time series data. In order to find the true correlation between variables, the autocorrelation within the data series of a variable need to be omitted through the selection of a suitable process sampling time, T_{MSPC}. Wetherill and Brown (1991) proposed using an autocorrelation plot of a variable to determine the suitable T_{MSPC} of a process. Any value of autocorrelation outside the value of a selected threshold value is considered as significant autocorrelation. This threshold value is set as

$\pm 2/\sqrt{k}$ where k is the total number of observation in a time series of a variable. From Section 3.7.4 of this thesis, the pumparound cooler process section of the study column process has the highest time constant among the process sections of the study column process. Therefore, this process section will be used as the test for significant autocorrelation and the selection of the suitable T_{MSPC} of the study column process. The results of this autocorrelation test will be shown in Section 5.3 of Chapter 5. The T_{MSPC} is determined at a value of 4.6 hours. This value will be used to sample the desired data from the column program. This data are then used to derive the correlation coefficients between the selected key process variables and the quality variables of interest.

4.2.3 Generation of Nominal Operation Condition (NOC) Data

A set of nominal operation condition (NOC) data, which includes the quality variables of interest and the related key process variables are generated from the developed dynamic simulation program. Selection of key process variables into the NOC data matrix is important to ensure that the variation of the two quality variables is exhibited in the selected key process variables. In this research, NOC is referred to situation the two quality variables of interest (oleic acid and linoleic acid composition in the bottom stream) and the selected key process variables (the key process variables that are shown in Table 4.2) are within the control limit of their control charts (Statistical Control Charts: Shewhart Control Chart and Range Control Chart).

During the generation of the NOC data, the selected key process variables from Table 4.2 (except T_{bot}) are fed with small random noise to make the two selected quality variables of interest from Table 4.1 to fluctuate within ± 5% of their SS value. The term noise in this research refers to measurement noise. The definition of noise was given in Section 2.2.4. The quality variables of interest and the selected key process variables are collected to form the NOC data matrix for the process variables, **X** and for the quality variables, **Y**. These NOC data matrices are represented in Equation 4.1 and Equation 4.2.

$$\mathbf{X} = [\mathbf{F}\ \mathbf{T_f}\ \mathbf{Re}\ \mathbf{P}\ \mathbf{Q_R}\ \mathbf{T_{bot}}] \quad (4.1)$$

$$\mathbf{Y} = [\mathbf{X_8}\ \mathbf{X_9}] \quad (4.2)$$

Where: $\mathbf{F}$ = data matrix for feed flow rate

$\mathbf{T_f}$ = data matrix for feed temperature

$\mathbf{Re}$ = data matrix for reflux flow rate

$\mathbf{P}$ = data matrix for pumparound flow rate

$\mathbf{Q_R}$ = data matrix for reboiler duty

$\mathbf{T_{bot}}$ = data matrix for bottom temperature

$\mathbf{X_8}$ = data matrix for oleic acid mole fraction in the bottom flow

$\mathbf{X_9}$ = data matrix for linoleic acid mole fraction in the bottom flow

4.2.3.1 Standardization of NOC Data

The data matrices of **X** and **Y** are standardized. The standardization procedure is for each of the variables in both data matrices. The standardization of each variable is shown as in Equation 4.3.

$$z_i = \frac{x_i - \overline{x_i}}{s_i} \quad (4.3)$$

Where: z_i = standardized variable

x_i = variable from measurements in **X** or **Y**

$\overline{x_i}$ = arithmetic average of variable x_i

s_i = standard deviation of variable x_i

The reason for standardization is that the measurements are consists of pressure, temperature, flow rates, compositions and other variables that have different scales and units of measurements. The standardization process will yield standardized variables with equivalent variance and mean centered. After this

procedure, both the data matrices are ready for further analysis in order to obtain NOC data.

4.2.3.2 Normality Test of NOC Data

In Multivariate Statistical Process Control (MSPC), the NOC data has to follow the normal distribution before any further manipulation of the NOC data can be carried out. The NOC data are subjected to normality test to study the normality properties of the NOC data. The normality tests on the NOC data include the checking of value of skewness, kurtosis, standard deviation and arithmetic average of the NOC data. For data following normal distribution, the skewness, kurtosis, standard deviation and arithmetic average of the data must have the value of 0, 3, 1 and 0, respectively (Wetherill and Brown, 1991). From the results of the normality test, the NOC data follow the normal distribution. Therefore, further manipulation of the NOC data can be carried out. The results of the normality tests of the NOC data are given in Section 5.4 of Chapter 5. Sometimes, the collected data are skewed and have a skewness value of more or less than 0. In this case, more data should be collected in order to get the data to have a skewness value of very near to 0. This method is based on Central Limit Theorem (CLT), which states that when enough data are collected in a data set, the data set will follow the normal distribution.

4.2.3.3 Number of Measurements in NOC Data

The number of measurements that has to be collected is the critical parameter. In this research, the range of the value of the two quality variables of interest for NOC data is set at $\pm 3\sigma$ where σ is the standard deviation of the quality variable. The reason for choosing this range value is in such a way that during NOC, 99.7% of points on the control chart for the quality variables are within the control limit for normally distributed data (McNeese and Klein, 1991). The method in obtaining the NOC data set is shown in Figure 4.1. The steps involving the analysis of the

standardized NOC data for correlation using Normal Correlation (NC), Principal Component Analysis (PCA) and Partial Correlation Analysis (PCorrA) and the building of control limits for the statistical control charts will be discussed in detail in Section 4.2.4.

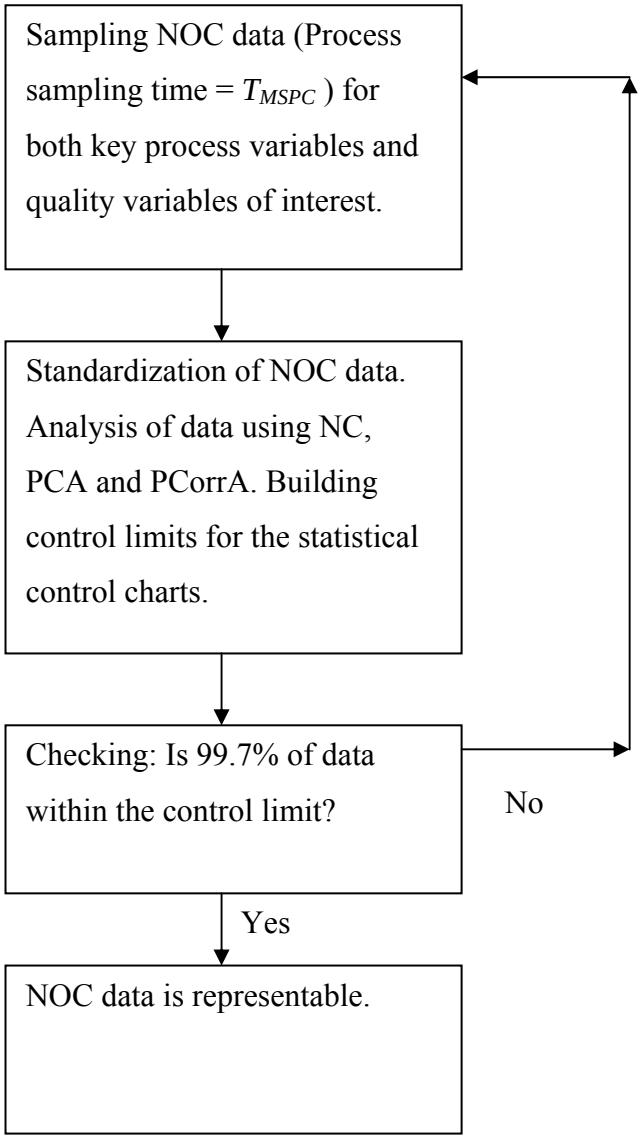


Figure 4.1: Procedure in obtaining the NOC data set

For 50 data points sampled with the process sampling time of 4.6 hours for each point sampled, all the data points of the quality variables and key process variables are within the control limits of their statistical control charts (more details of the statistical control charts are given in Section 4.2.5). Once the NOC data is obtained,

the correlation between the selected key process variables and the quality variables of interest will be determined.

4.2.4 Derivation of Correlation Coefficient

In this research, the fault detection and diagnosis algorithm will be built based on the development of a correlation coefficient, C_{ik}, between the quality variables of interest and the selected key process variables. This correlation coefficient will provide linear relationship between the selected key process variables and the quality variables. A selected key process variable is related to a quality variable of interest by the following equation:

$$x_i = \frac{y_i}{C_{ik}} \qquad (4.4)$$

where: y_i = quality variable

x_i = process variable

C_{ik} = correlation coefficient between y_i and x_i

These correlation coefficients are important as they relate the quality variables of interest with the selected key process variables. This aspect is important since the correlation coefficients will be used later on to relate the statistical control limits of the quality variables of interest and the selected key process variables.

There are three methods used in this research to derive the correlation coefficient, C_{ik}, from the two data matrices. Correlation coefficient via NC will be discussed in the next section followed by correlation coefficient using PCA and PCorrA in the two sections that follows.

4.2.4.1 Correlation Coefficient Derivation Using Normal Correlation

Normal Correlation (NC) between the quality variables and the selected key process variables are used to derive the correlation coefficient. The correlation between a quality variable and a process variable is normal one to one correlation. Equation 4.5 shows the calculation of correlation coefficient using NC.

$$C_{ik} = \frac{\sum_{j=1}^{n}(x_{ij} - \overline{x_i})(y_{ij} - \overline{y_i})}{\sqrt{\sum_{j=1}^{n}(x_{ij} - \overline{x_i})^2}\sqrt{\sum_{j=1}^{n}(y_{ij} - \overline{y_i})^2}} \qquad (4.5)$$

Where: x_{ij} = measurement j-th of process variable i-th

y_{ij} = measurement j-th of quality variable i-th

$\overline{x_i}$ = arithmetic average of process variable i-th

$\overline{y_i}$ = arithmetic average of quality variable i-th

n = number of measurements

4.2.4.2 Correlation Coefficient Derivation Using Principal Component Analysis

The derivation of correlation coefficient using Principal Component Analysis (PCA) will be based on the work of Lam and Kamarul (2002). Before applying PCA on the generated NOC data, a data matrix as shown in Equation 4.6 are formed.

$$\mathbf{Z} = [\mathbf{F}\ \mathbf{T_f}\ \mathbf{Re}\ \mathbf{P}\ \mathbf{Q_R}\ \mathbf{T_{bot}}\ \mathbf{X_8}] \qquad (4.6)$$

After PCA is applied on the **Z** data matrix, the principal components of this matrix are related to the original data matrix as shown in the following equations:

$$\mathbf{P} = \mathbf{ZV} \qquad (4.7)$$

$$\mathbf{P} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ v_{m1} & v_{m2} & \cdots & v_{mn} \end{bmatrix} \quad (4.8)$$

Where: $\mathbf{P}$ = principal component matrix

$\mathbf{V}$ = eigenvector matrix

x_{m1} = measurement m-th of variable x_1 in data matrix $\mathbf{Z}$

v_{m1} = value m-th of eigenvector v_1

m = number of measurements

n = number of variables

The Singular Value Decomposition (SVD) of the $\mathbf{Z}$ data matrix will yield the eigenvector matrices and singular value matrix of the covariance matrix of the original data matrix as shown in the following equations (Lam and Kamarul, 2002):

$$\mathbf{Z} = \mathbf{U}\mathbf{L}^{1/2}\mathbf{V}^{\mathbf{T}} \quad (4.9)$$

Where: $\mathbf{U}$ = eigenvector matrix of $\mathbf{Z}\mathbf{Z}^{\mathbf{T}}$

$\mathbf{V}$ = eigenvector matrix of $\mathbf{Z}^{\mathbf{T}}\mathbf{Z}$

$\mathbf{L}^{1/2}$ = diagonal matrix of positive square root of the eigenvalue of $\mathbf{Z}^{\mathbf{T}}\mathbf{Z}$

$\mathbf{V}^{\mathbf{T}}$ = transpose matrix of matrix $\mathbf{V}$

These matrices have the following properties (Geladi and Kowalski, 1986):

$$\mathbf{U}\mathbf{U}^{\mathbf{T}} = \mathbf{U}^{\mathbf{T}}\mathbf{U} = \mathbf{I} \quad (4.10)$$

$$\mathbf{V}\mathbf{V}^{\mathbf{T}} = \mathbf{V}^{\mathbf{T}}\mathbf{V} = \mathbf{I} \quad (4.11)$$

$$(\mathbf{L}^{1/2})(\mathbf{L}^{1/2})^{\mathbf{T}} = (\mathbf{L}^{1/2})^{\mathbf{T}}(\mathbf{L}^{1/2}) = \lambda \quad (4.12)$$

Where: $\mathbf{I}$ = identity matrix

λ = diagonal matrix of eigenvalue of $\mathbf{Z}^{\mathbf{T}}\mathbf{Z}$

By multiplying both side of Equation 4.7 with $\mathbf{V^T}$,

$$\mathbf{Z} = \mathbf{P}\mathbf{V^T} \qquad (4.13)$$

Representing quality variable 1, y_1 with x_k and process variable 1, x_1 with x_i, Equation 4.13 can be written for these two variables as follows:

$$x_i = \mathbf{P}\, v_i^T \qquad (4.14)$$

$$x_k = \mathbf{P}\, v_k^T \qquad (4.15)$$

Using substitution of $\mathbf{V^T} = [\, v_1 \;\; v_2 \;\; \ldots \;\; v_n \,]$, Equation 4.14 and 4.15 can be simplified to become:

$$x_i = \mathbf{P}\, v_i \qquad (4.16)$$

$$x_k = \mathbf{P}\, v_k \qquad (4.17)$$

By rearranging Equation 4.4 and combining Equation 4.16 and 4.17,

$$\begin{aligned} C_{ik} &= (\, x_i^T \,)(\, x_k \,) \\ &= (\mathbf{P}\, v_i \,)^T (\mathbf{P}\, v_k \,) \\ &= v_i{}^T \mathbf{P}^T \mathbf{P}\, v_k \end{aligned} \qquad (4.18)$$

Combining Equation 4.7 and 4.9,

$$\mathbf{P} = \mathbf{Z}\mathbf{V} = \mathbf{U}\mathbf{L}^{1/2} \qquad (4.19)$$

By combining results from Equation 4.10, 4.12 and 4.19,

$$\mathbf{P}^T\mathbf{P} = (\mathbf{U}\mathbf{L}^{1/2})^T(\mathbf{U}\mathbf{L}^{1/2})$$

$$= (\mathbf{L}^{1/2})^{\mathrm{T}} \mathbf{L}^{1/2} \mathbf{U}^{\mathrm{T}} \mathbf{U}$$
$$= \lambda \qquad (4.20)$$

Inserting this result into Equation 4.18,

$$C_{ik} = v_i^{\mathrm{T}} \lambda v_k \qquad (4.21)$$

Correlation coefficient between a quality variable and a process variable for *j* variables will be calculated using Equation 4.22 (Lam and Kamarul, 2002).

$$C_{ik} = \sum_{i=1}^{n} v_{ij} \lambda_i v_{ik} \qquad (4.22)$$

where: v_{ij} = eigenvector *i*-th for process variable *j*-th

v_{ik} = eigenvector *i*-th for quality variable *k*-th

λ_i = eigenvalue *i*-th

n = number of eigenvalues

In order to obtain the correlation coefficients for linoleic acid with the selected key process variables, $\mathbf{X_8}$ will be replaced with $\mathbf{X_9}$ in Equation 4.6. Then, SVD will be used to derive the eigenvectors and eigenvalues of the newly formed data matrix. Equation 4.22 will be used together with the newly derived eigenvectors and eigenvalues to calculate the correlation coefficients between linoleic acid and the selected key process variables.

PCA is a technique used to reduce the dimension of data while at the same time retaining the original variation of the original data to minimize the lost of information during dimension reduction. In this research, the correlation coefficient based on PCA will be comprised of two methods based on the retained variation of the original data. One method will retained 90% of the variation of the original data while the other method will retained 95% of the variation of the original data. PCA has advantage over Normal Correlation (NC) in term of data reduction while still retaining the information of the original data. The method based on NC does not

include data reduction and also the correlation developed is one variable to one variable.

4.2.4.3 Correlation Coefficient Derivation Using Partial Correlation Analysis

Partial Correlation Analysis (PCorrA) is used to develop correlation between two variables after taken into account the effect of the other variables. The application of PCorrA in this research on the development of the correlation coefficient, C_{ik} is based on the work of Kamarul (1997). In Kamarul (1997), the correlation coefficient developed was between multiple input variables and one quality variable of interest. In this research, there will be two quality variables, thus the application of correlation coefficient has been extended to multiple input multiple output correlation.

The method in developing the correlation coefficient between the two quality variables and the selected process variables will be as in Kamarul (1997). For quality variable 1, y_1 (oleic acid) and process variable 1, x_1, the correlation coefficient is calculated as shown in Equation 4.23.

$$C_{ik_{x_1 y_1}} = \frac{r_{x_1 y_1.(x_3,\ldots,x_n)} - r_{x_1 x_2.(x_3,\ldots,x_n)} r_{y_1 x_2.(x_3,\ldots,x_n)}}{(1 - r^2_{x_1 x_2.(x_3,\ldots,x_n)})^{1/2} (1 - r^2_{y_1 x_2.(x_3,\ldots,x_n)})^{1/2}} \qquad (4.23)$$

where: $C_{ik_{x_1 y_1}}$ = correlation coefficient between y_1 and x_1

$r_{x_1 y_1.(x_3,\ldots,x_n)}$ = partial correlation between y_1 and x_1 after the effect of $n-2$ variables

$r_{x_1 x_2.(x_3,\ldots,x_n)}$ = partial correlation between x_1 and x_2 after the effect of $n-2$ variables

$r_{y_1 x_2.(x_3,\ldots,x_n)}$ = partial correlation between y_1 and x_2 after the effect of $n-2$ variables

n = number of selected process variables

The correlation coefficients for quality variable 2, y_2 (linoleic acid) with the selected process variables are also determined using an equation similar to Equation 4.23.

4.2.5 Implementation of Correlation Coefficients

As stated in Chapter 2, the univariate statistical control charts used in this research are modified using the developed correlation coefficients in Section 4.2.4. The modification of these charts using correlation coefficients will be provided in the next section.

4.2.5.1 Constructing Statistical Control Charts

After the derivation of the correlation coefficients from the three analysis techniques, the control limits for the statistical control charts will be built. The statistical control charts used in this research are Shewhart Control Chart (SCC) and Range Control Chart (RCC). The control limits for these charts will be based on the calculated correlation coefficients. SCC has proven its ability to detect fault in univariate process. However, chemical processes being multivariate in nature make SCC perform well below the desired value (Manabu *et al.*, 2000). In this research, the Shewhart Control Chart used for normal univariate process will be modified to suit the multivariable nature of the case study.

The SCC for the quality variables has the same control limits since all these variables are in standardized form. The Upper Control Limit (UCL) and Lower Control Limit (LCL) for the quality variables are +3 and -3, respectively. The Center Line (CL) for these charts has a value of zero. The reason for the control limits having the stated value is the range of deviation of the quality variables for Nominal Operation Condition (NOC) has been set to $\pm 3\sigma$ where σ is the standard deviation of the quality variable. The control limits of the each quality variable of interest are translated into the control limits of the selected key process variables using the

information from Equation 4.4. Figure 4.2 shows how the translation is carried out on each quality variable and the corresponding key process variables.

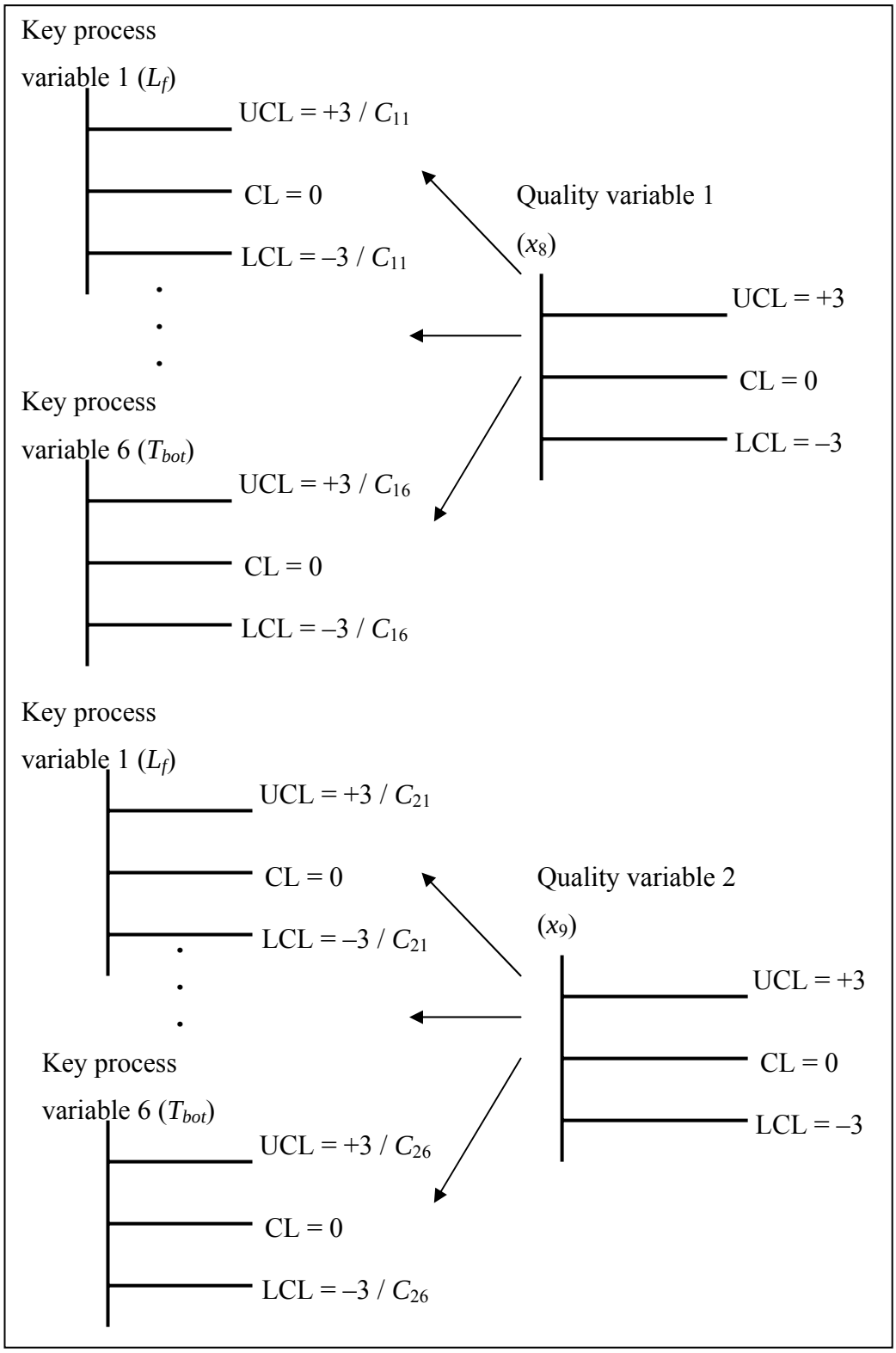


Figure 4.2: The control limits of each quality variable are translated to control limits of the selected key process variables

In Figure 4.2,

C_{1i} = correlation coefficient between quality variable 1 (linoleic acid composition in the bottom flow) and key process variable i (i = 1, 2, 3, 4, 5 and 6)

C_{2i} = correlation coefficient between quality variable 2 (oleic acid composition in the bottom flow) and key process variable i (i = 1, 2, 3, 4, 5 and 6)

The correlation coefficients for each quality variable with the corresponding selected key process variables are determined using Equation 4.5 (for NC), Equation 4.22 (for PCA) and Equation 4.23 (for PCorrA). Figure 4.3 shows the Shewhart Control Chart for quality variable 1 (x_8) and how the correlation coefficient is applied to translate the control limits for selected key process variable 1 (L_f) based on the control limits of the control chart of quality variable 1.

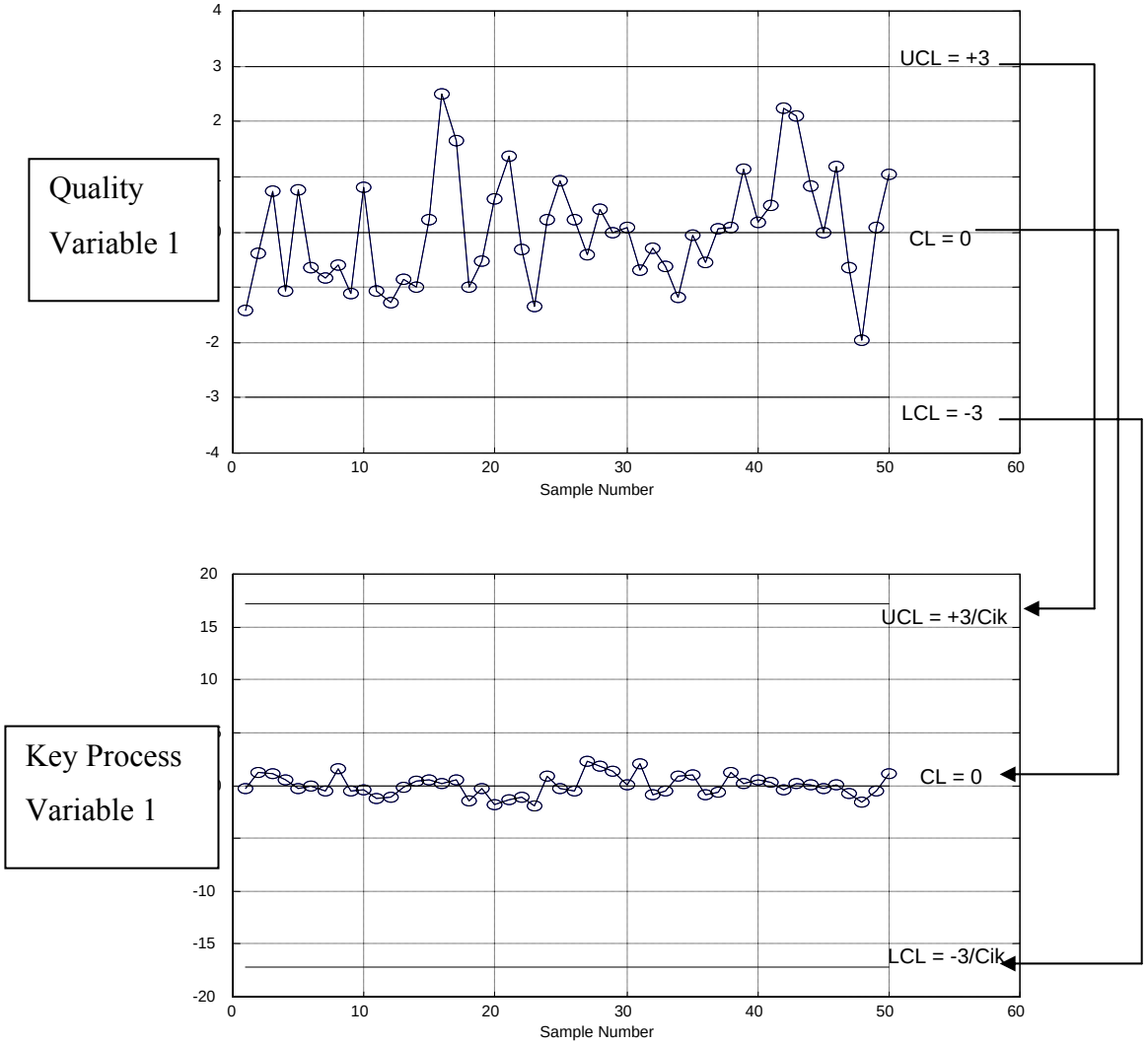


Figure 4.3: The control limits of quality variable 1 are translated to control limits of selected key process variable 1 (Shewhart Control Chart)

From Figure 4.3, the UCL, CL and LCL of a quality variable of interest are translated into UCL, CL and LCL of a selected key process variable. The UCL and LCL of a key process variable are $+3/C_{ik}$ and $-3/C_{ik}$, respectively. The UCL and LCL of all the selected key process variables are related to the UCL and LCL of the two quality variables of interest in this manner.

The Range Control Chart is also used in this research. Using a subgroup, m = 2, the range value for each of the variable and the control limits for the Range chart of each variable are calculated by the following equations:

First Range value, $R(1)$ $= 0$

Range value number i $= | z_i - z_{i-1} |$ $\quad i = 2, 3, 4,\ldots$ (4.24)

CL for Range chart, $\overline{R}$ = Arithmetic average of the range data set

$$= \frac{\sum_{i=1}^{n} R(i)}{n} \qquad (4.25)$$

LCL for Range chart $= 0$

UCL for Range chart $= d_2$*CL of Range chart

$$= d_2 * \frac{\sum_{i=1}^{n} R(i)}{n} \qquad (4.26)$$

where: d_2 = statistical parameter based on the subgroup m, for m =2, d_2 = 3.267 (McNeese and Klein, 1991)

n = number of Range values

z_i = standardized values for the selected process variables, measured values for the quality variables

Figure 4.4 shows how the control limits of each quality variable and the corresponding key process variables are related for Range Control Chart.

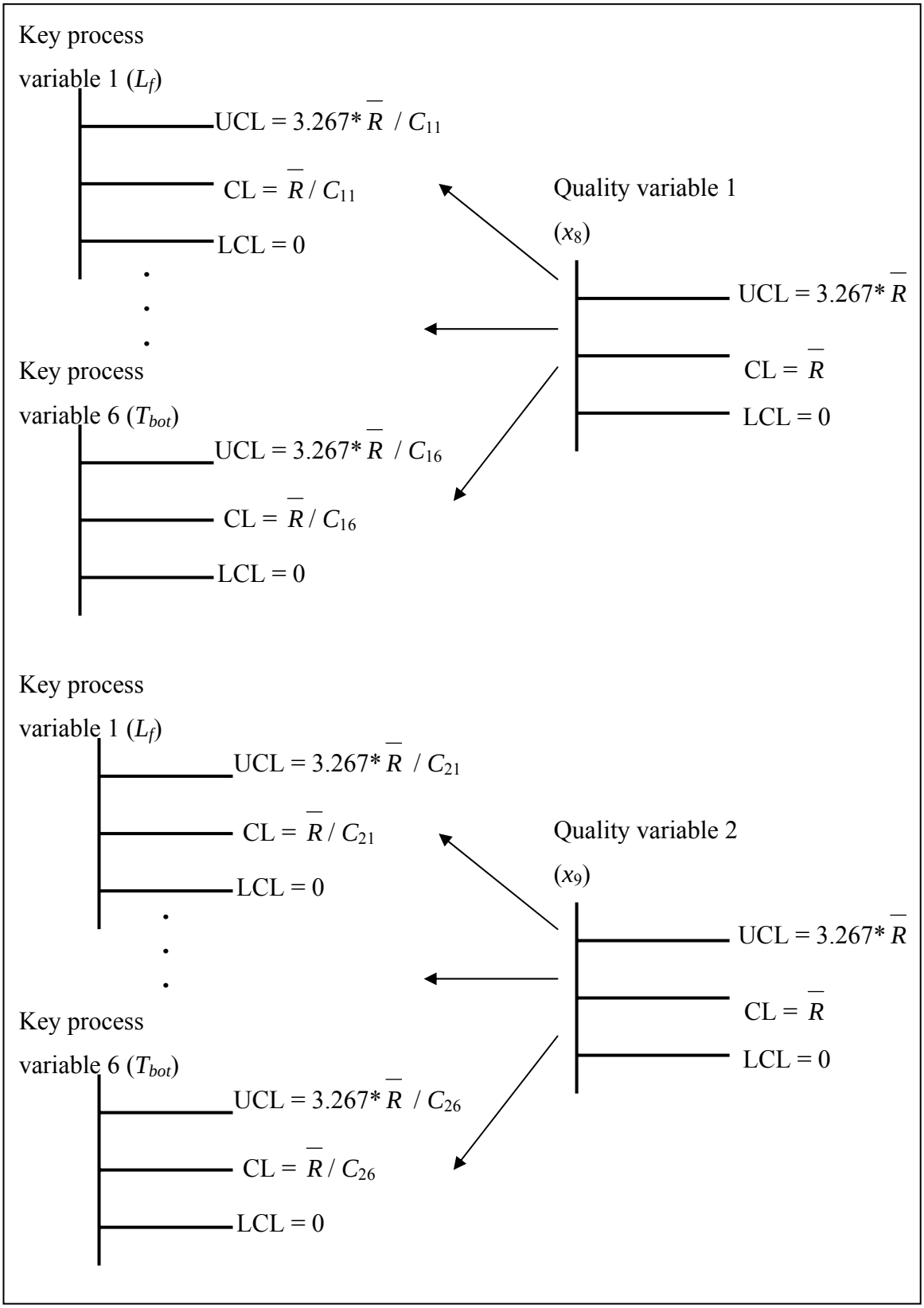


Figure 4.4: The control limits of each quality variable are translated to control limits of the selected key process variables

In Figure 4.4,

C_{1i} = correlation coefficient between quality variable 1 (linoleic acid composition in the bottom flow) and key process variable i (i = 1, 2, 3, 4, 5 and 6)

C_{2i} = correlation coefficient between quality variable 2 (oleic acid composition in the bottom flow) and key process variable i (i = 1, 2, 3, 4, 5 and 6)

Figure 4.5 shows the Range chart for quality variable 1 (x_8) and selected key process variable 1 (L_f) with their respective control limits.

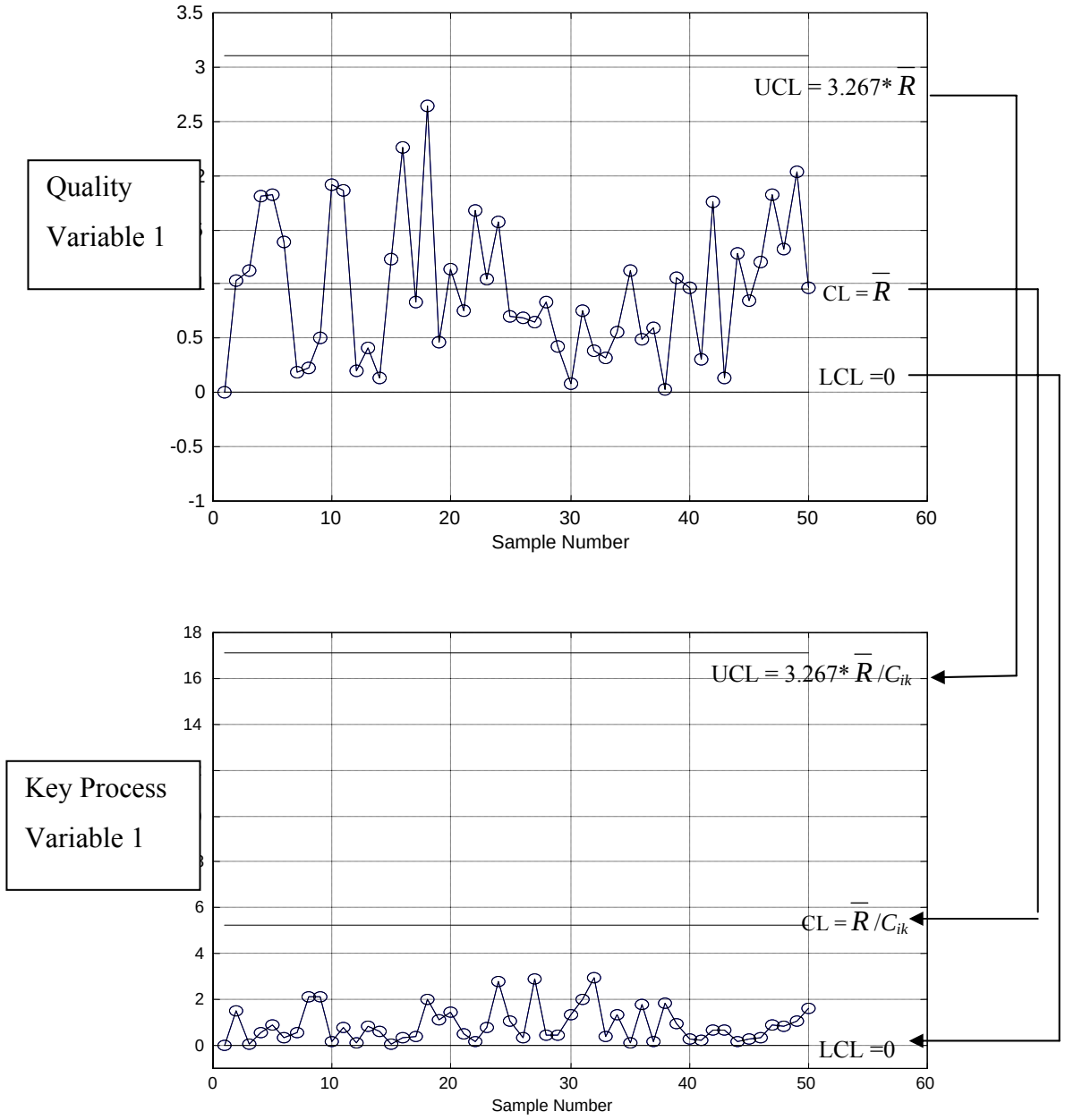


Figure 4.5: The control limits of quality variable 1 are translated to control limits of key process variable 1 (Range Control Chart)

From Figure 4.5, the UCL, CL and LCL of a quality variable of interest are translated into UCL, CL and LCL of a selected key process variable. After the statistical control charts (Shewhart Control Chart and Range Control Chart) are built and the limits are calculated, the algorithm for fault detection and diagnosis is ready to test its capability on a set of fault data. The definition of fault and generation of fault data will be presented in the next section.

4.2.6 Fault Data Generation

In this section, the general definition of fault and the definition of fault in this research will be given. The types of faults considered and the generation of the fault data set will also be discussed in detail.

4.2.6.1 Fault Definition

In an equipment, fault is to designate the departure from an acceptable range of an observed variable or calculated parameter associated with the equipment (Himmelblau, 1978). The characteristics of the process chosen for measurement, their acceptable range of operation and the accuracy of the statistic used for classification of a potential fault are factors that influence the definition of fault.

A fault implies degradation of performance while failure means complete inoperability of equipment or the process (Himmelblau, 1978). Most chemical processes are sufficiently flexible and well organized that as soon as a fault shows up in any subsystem, the system compensates for the fault so as to continue operation. Thus, a fault will not be necessarily being a failure of equipment or process unable to continue operation. **In this research, the faults considered are those that do not lead to failure of the study equipment but rather those that cause the degradation of performance of the study column. When a fault is detected, the study column will fail to produce the desired linoleic acid (x_8) and oleic acid (x_9) compositions in the bottom flow rate**.

Thus, the definition of fault in this research is a fault occurs when any of the quality variables of interest (oleic acid and linoleic acid) exhibit an out-of-control limit signal in their statistical control charts (Shewhart Control Chart or Range Control Chart) and also an out-of-control limit signal in any of the statistical control chart (Shewhart Control Chart or Range Control Chart) of the selected key process variables. The out-of-control limit signal occurs when a value of a variable exceeds

the control limit of its statistical control chart. The type of faults considered and their properties will be presented in Section 4.2.6.2.

The term disturbance is often mentioned in works involving fault detection and diagnosis. Disturbance is referring to independent variable that causes the process in which this disturbance is interacting with to deviate from its desired operating condition. For example, the flow of liquid feed stream into a distillation column will cause the increase in the liquid level of the column bottom. The feed stream is the disturbance to the distillation column process. In this research, there are two types of disturbances considered. Overall disturbances are defined as independent variables that will cause the study column fail to produce the desired value for the two quality variables of interests (oleic acid and linoleic acid). Local disturbances are variables that will cause the control variable (not the quality variables of interest selected in Table 4.1) in which these variables are present in its control loop to deviate from its set point value. **The fault situations considered in this research do not include normal disturbance changes and set point changes**. **These two phenomena are considered to be normal operation of the study column**. The information on disturbances considered in this research is presented in Table 4.3.

The feed stream to the study column is assumed to be pumped from a storage tank. Therefore, the value of feed flow rate and feed temperature are assumed to be fixed (having normal random variation). Any changes in the value of these two variables will be considered due to faults (sensor faults or valve faults) and not caused by change in the properties of the raw material in the feed stream.

Table 4.3: Information on types of disturbances

Type of Disturbance	Name of Disturbance	Location of Disturbance	Location of Control Loop
Overall Disturbance	Feed flow rate, L_f	Feed stream to column	-
Overall Disturbance	Feed temperature, T_f	Feed stream to column	-
Local Disturbance*	Sidedraw flow rate, Sd	Sidedraw stream	Reflux flow rate control loop
Local Disturbance*	Liquid stream flow rate, L_{Dp}	Liquid stream after reflux point	Pumparound flow rate control loop
Local Disturbance*	Liquid stream flow rate, L_{Dp}	Liquid stream after reflux point	Pumparound temperature control loop
Local Disturbance*	Vapor flow rate, V_2	Vapor stream from tray 2	Top column pressure control loop
Local Disturbance*	Liquid flow rate, L_{28}	Liquid stream from tray 28	Bottom liquid level control loop
Local Disturbance*	Liquid flow rate, L_{28}	Liquid stream from tray 28	Bottom temperature control loop
Local Disturbance*	Liquid flow rate, L_1	Liquid stream from tray 1	Sidedraw tray liquid level control loop

*The local disturbance considered for each control loop is one of the many local disturbances present in each control loop.

4.2.6.2 Fault Generation

Faults are introduced into the process after the NOC data matrix is collected and the fault detection and diagnosis (FDD) algorithm is developed. These faults will be tested against the developed FDD algorithm. The fault generation procedure involves feeding the values of **the selected key process variables (feed flow rate, feed temperature, reflux flow rate, pumparound flow rate and reboiler duty)** with random numbers. These random numbers are generated using the random number generator in the Matlab software. The arithmetic average and standard deviation of these random numbers are zero and one respectively. Although random in nature, these numbers are following the normal distribution as justified in Section 4.2.3.2.

There are two types of fault generated in the fault data set:

- Significant Fault
  - Large change in value of the **selected key process variable(s)** designed for fault;
  - The **selected quality variables of interest (Listed in Table 4.1)** will exhibit values **exceeding** $\pm 4\sigma$ (σ is the standard deviation of each quality variable of interest)

- Insignificant Fault
  - Small change in the value of **the selected key process variable(s)** designed for fault;
  - The **selected quality variables of interest (Listed in Table 4.1)** will exhibit values **not exceeding** $\pm 4\sigma$ (σ is the standard deviation of each quality variable)

These two types of fault contain both single and multiple fault causes. There will be both: **one selected key process variable and multiple selected key process variables causing out-of-control signal in the statistical control charts of the two quality variables of interest. The out-of-control signals in the selected key process variables and quality variables of interest will yield fault signals. In this research, disturbances to the process and normal set point changes in the**

control loops of the column are considered as normal changes in the variation of the process while faults are due to the abnormal changes in the variation of the process due to assignable fault causes. When an out-of-control signal is detected in the control charts of the selected key process variables and the quality variables of interest, this signal is checked to determined whether there is an assignable fault cause to the observed signal or it is due to normal disturbance (disturbances listed in Table 4.3) changes or normal set point changes in the control loops of the column. **The developed FDD algorithm will only be activated if there are assignable fault causes to an out-of-control signal detected in the control charts of the quality variables and the selected key process variables. The main assumption is that the disturbances listed in Table 4.3 are monitored.** The flowchart in Figure 4.6 represents the mechanism of the developed FDD algorithm.

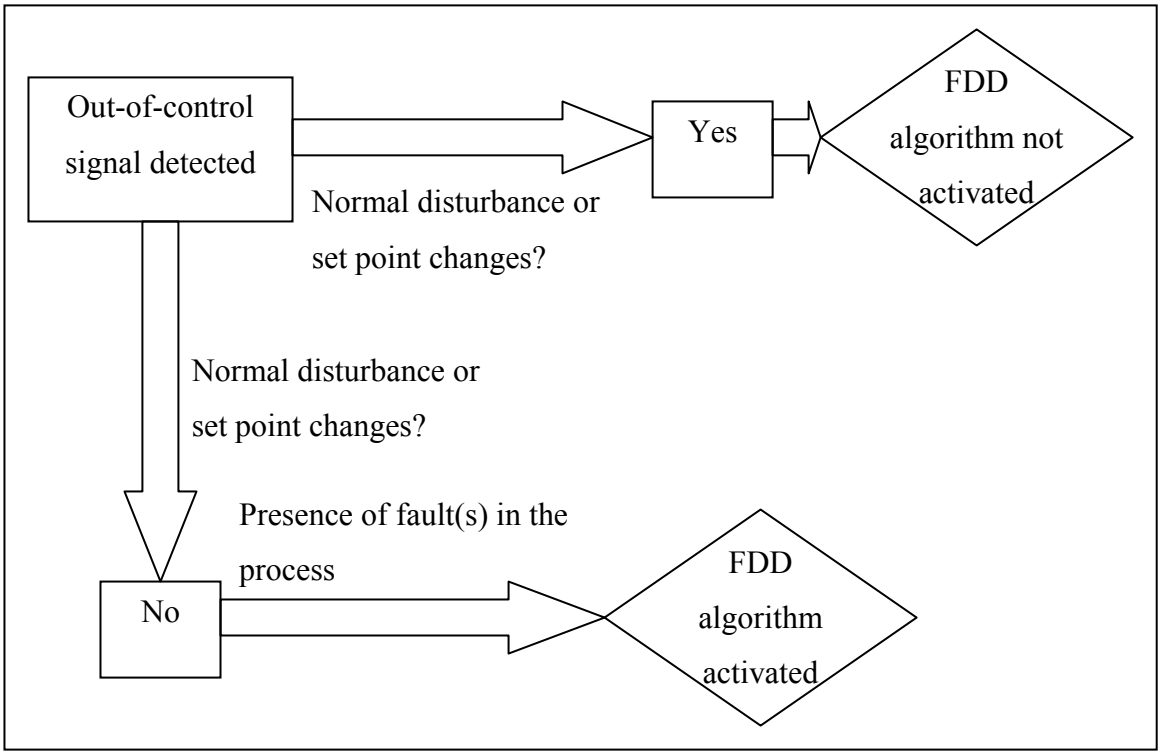
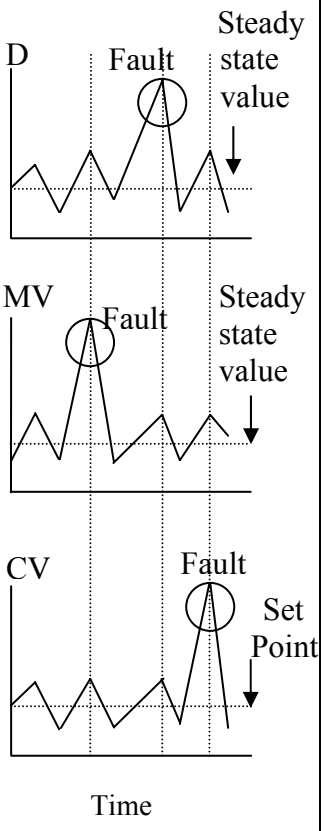
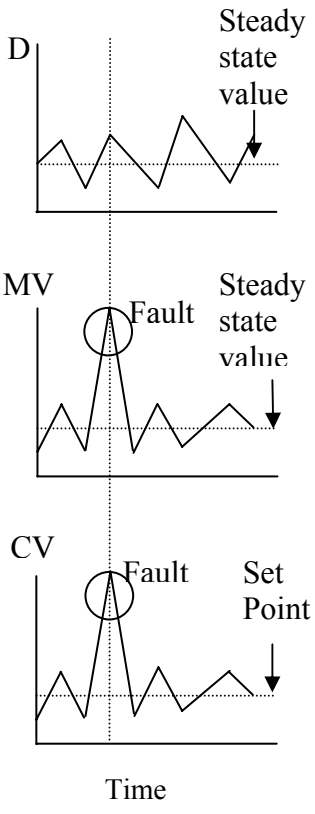
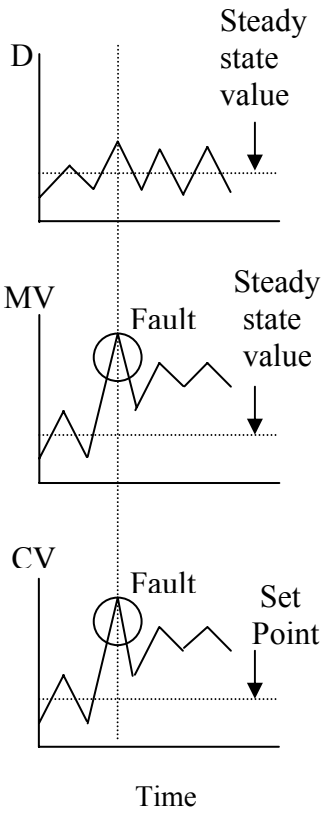


Figure 4.6: The mechanism of the developed FDD algorithm

The faults considered in this research involve fault in sensor, fault in valve and fault in controller in the study column. The description of each type of fault is presented in Table 4.4.

Table 4.4: Description of each type of fault

1) Sensor Fault	2) Valve Fault	3) Controller Fault
• For open loop variables, only the value of the variable changes abnormally. For closed loop variables, only the value of the disturbance (D) OR the manipulated variable (MV) OR the control variable (CV) changes abnormally.	• For open loop variables, only the value of the variable changes abnormally. For closed loop variables, both the value of manipulated variable (MV) AND control variable (CV) changes abnormally together.	• For closed loop variables, the value of manipulated variable (MV) AND control variable (CV) changes abnormally together.
		

From Table 4.4, each type of fault represents faults that occur in the chemical industries. Sensor faults representing situations where the sensors show bias (positive and negative) in their measurements. Valve faults represent situations where the valve is stuck (thus causing the value of the flow rate of the fluid in the line of this valve to suddenly increase or decrease). Controller faults are situations where a controller fails to function normally (only for controller having the integral mode failing to perform normally). The assumptions used in considering these faults are valve and sensor operations are assumed to be ideal (except when there are faults involving the sensors and valves) and controller are working fine throughout the process monitoring period (except when controller faults are present).

4.2.7 Working Procedure of Developed Fault Detection and Diagnosis Algorithm

In this section, the working procedure of the developed fault detection and diagnosis algorithm will be presented.

4.2.7.1 Fault Detection

The Shewhart Control Chart and Range Control Chart will be used together in fault detection. Any variable that observed a value exceeding the control limits in any of the control charts will yield an out-of-control limit signal. The process will be monitored using the control charts for the two quality variables (listed in Table 4.1). Any out-of-control limit signal detected by these control charts will lead the operator to check the control charts for the selected key process variables (listed in Table 4.2). The detected fault (both quality variable(s) and selected process variable(s) showing out-of-control limit signal) will be diagnosed by checking any faults observed in the selected process variables control chart. As stated earlier on, normal set point changes and normal disturbance changes are not considered as fault situations in this research. Figure 4.7 shows an example of fault detection involving a quality variable

of interest and its corresponding selected key process variables (only two key process variables are shown) for Shewhart Control Chart (for Range Control Chart, the procedure is similar).

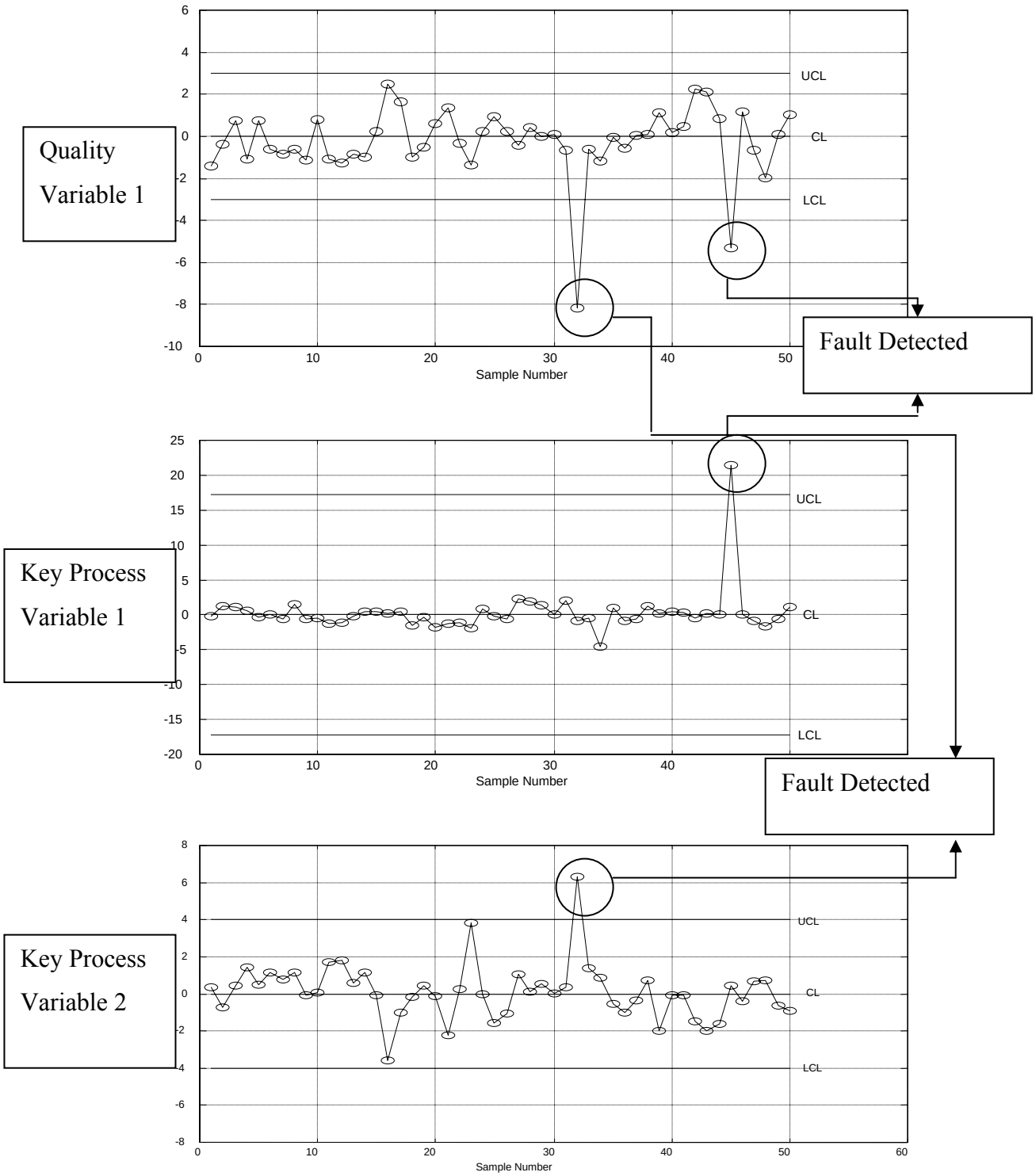


Figure 4.7: Fault detection procedure involving Shewhart Control Chart

Once a fault is detected, the statistical control charts (Shewhart Control Chart and Range Control Chart) will be checked to pinpoint the fault cause of the detected fault and further fault diagnosis will be carried out. The working procedure of fault diagnosis will be shown in the next section.

4.2.7.2 Fault Diagnosis

The fault diagnosis procedure involves checking the statistical control charts (Shewhart Control Chart and Range Control Chart) of the selected key process variables (listed in Table 4.2) that exhibit out-of-control limit signal for an observed fault. Once the fault cause (or fault causes) is determined, a run chart (value of variable versus time) of that variable will be plotted (charts shown in Table 4.4). The fault type will be determined using the information from this run chart and compared to the types of fault given in Table 4.4. In this research, selected key process variables that are not present in control loops are designed with sensor faults and valve faults while selected key process variable that are present in control loops (such as reboiler duty, Q_r) are designed with all three types of faults presented in Table 4.4. An example of fault diagnosis involving a quality variable of interest and its corresponding selected key process variables (only two key process variables are shown) for Shewhart Control Chart (for Range Control Chart, the procedure is similar) is shown in Figure 4.8.

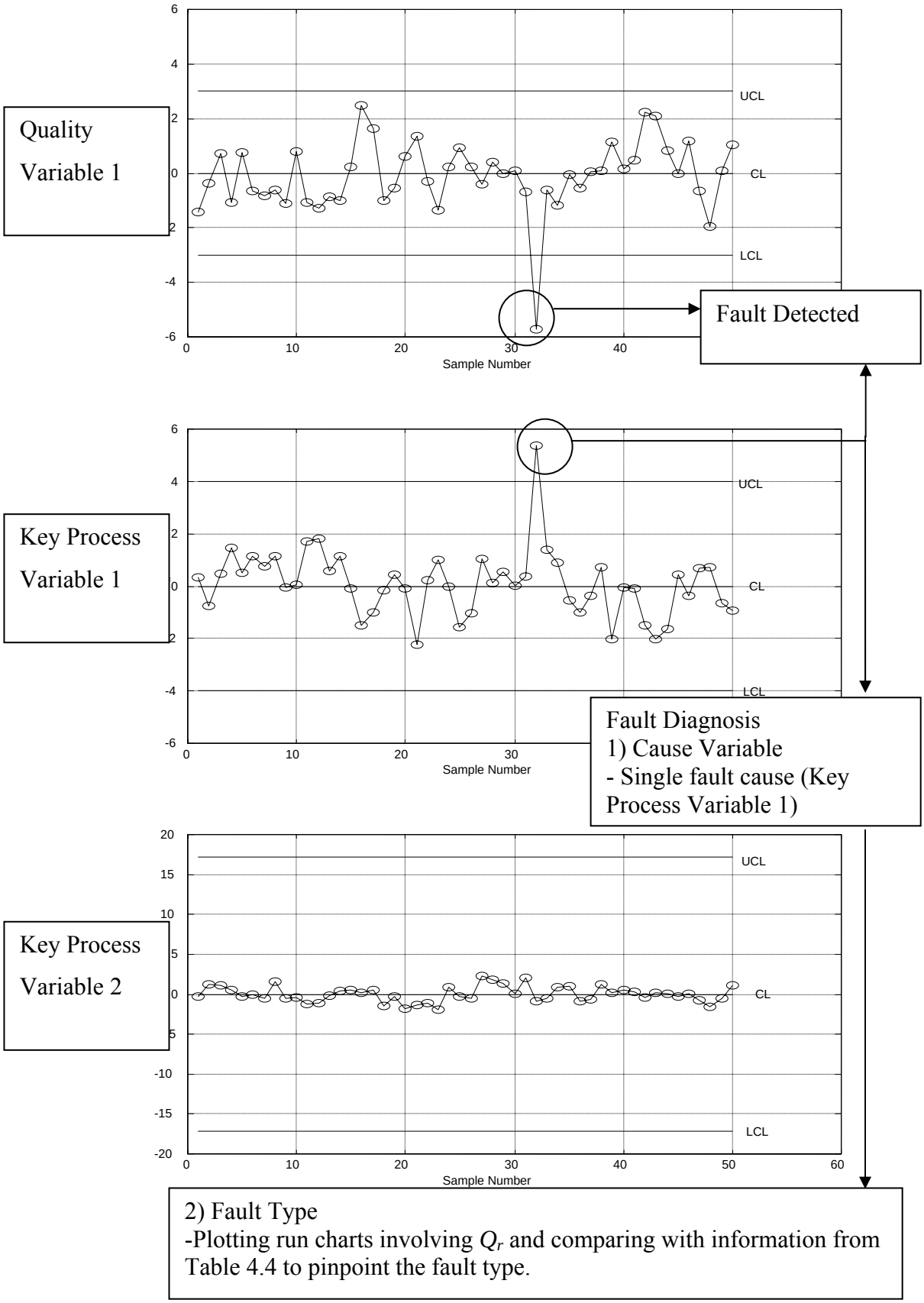


Figure 4.8: Fault diagnosis procedure involving Shewhart Control Chart

The results of the fault detection and diagnosis of the generated fault data will be given and discussed in detail in Section 5.6 of Chapter 5.

4.3 Performance Evaluation of the Developed Fault Detection and Diagnosis Algorithm

The evaluation of the developed fault detection and diagnosis (FDD) algorithm will be based on two aspects. The first aspect is how many of the generated faults in the process are successfully detected by the developed FDD algorithm. The second aspect is how many of the generated faults are successfully isolated or diagnosed the cause of the fault. Then, the overall performance of the developed FDD algorithm will be evaluated based on these two aspects. The method based on NC will served as the benchmark for comparison with the methods based on PCA and the method based on PCorrA.

4.3.1 Fault Detection Efficiency

The fault detection efficiency, η_{Det}, of the developed FDD algorithm will be evaluated using the following equation:

$$\eta_{Det} = \frac{number\ of\ \ faults\ \ detected}{number\ of\ \ faults\ generated\ in\ the\ process} \qquad (4.29)$$

4.3.2 Fault Diagnosis Efficiency

The fault diagnosis efficiency is not as straight forward as the fault detection efficiency. This is because not all detected fault will be successfully isolated or diagnosed. There will be successfully diagnosed faults and unsuccessful diagnosed

faults which are consist of partially diagnosed faults, over diagnosed faults and undiagnosed faults. These faults are explained as follow:

1) Successfully diagnosed fault

   The system is able to identify the exact causes of the fault without identifying other process variables as fault causes.

2) Unsuccessful diagnosed fault

   i) Partially diagnosed fault

      Several of the known fault causes are identified by the fault diagnosis system. There are still some fault causes not identified.

   ii) Over diagnosed fault

      The system over identifies other process variables that are not suppose to be the fault causes as fault causes.

   iii) Undiagnosed fault

      The system is unable to identify the correct fault causes.

Therefore, the fault diagnosis efficiency, η_{Dia} , will be calculated in the following manner:

Number of exact fault diagnosed
= total diagnosed fault situation – total successful diagnosed fault situation (4.30)

$$\eta_{Dia} = \frac{number\ of\ \ exact\ \ fault\ diagnosed}{number\ of\ \ faults\ generated\ in\ the\ process} \qquad (4.31)$$

4.3.3 Overall FDD Efficiency

The overall performance of the developed FDD algorithm is evaluated using the following equation:

$$\eta_{FDD} = \eta_{Det} * \eta_{Dia} * 100\% \quad (4.32)$$

The performance of the FDD algorithm based on PCA and PCorrA will be compared to the method based on NC. The basis of comparison is on the value of their overall FDD efficiency.

4.4 Chapter Summary

The FDD algorithm developed in this research are based on three analysis techniques: Normal Correlation (NC), Principal Component Analysis (PCA) and Partial Correlation Analysis (PCorrA). The procedures for developing the FDD algorithm are summarized by Figure 4.9.

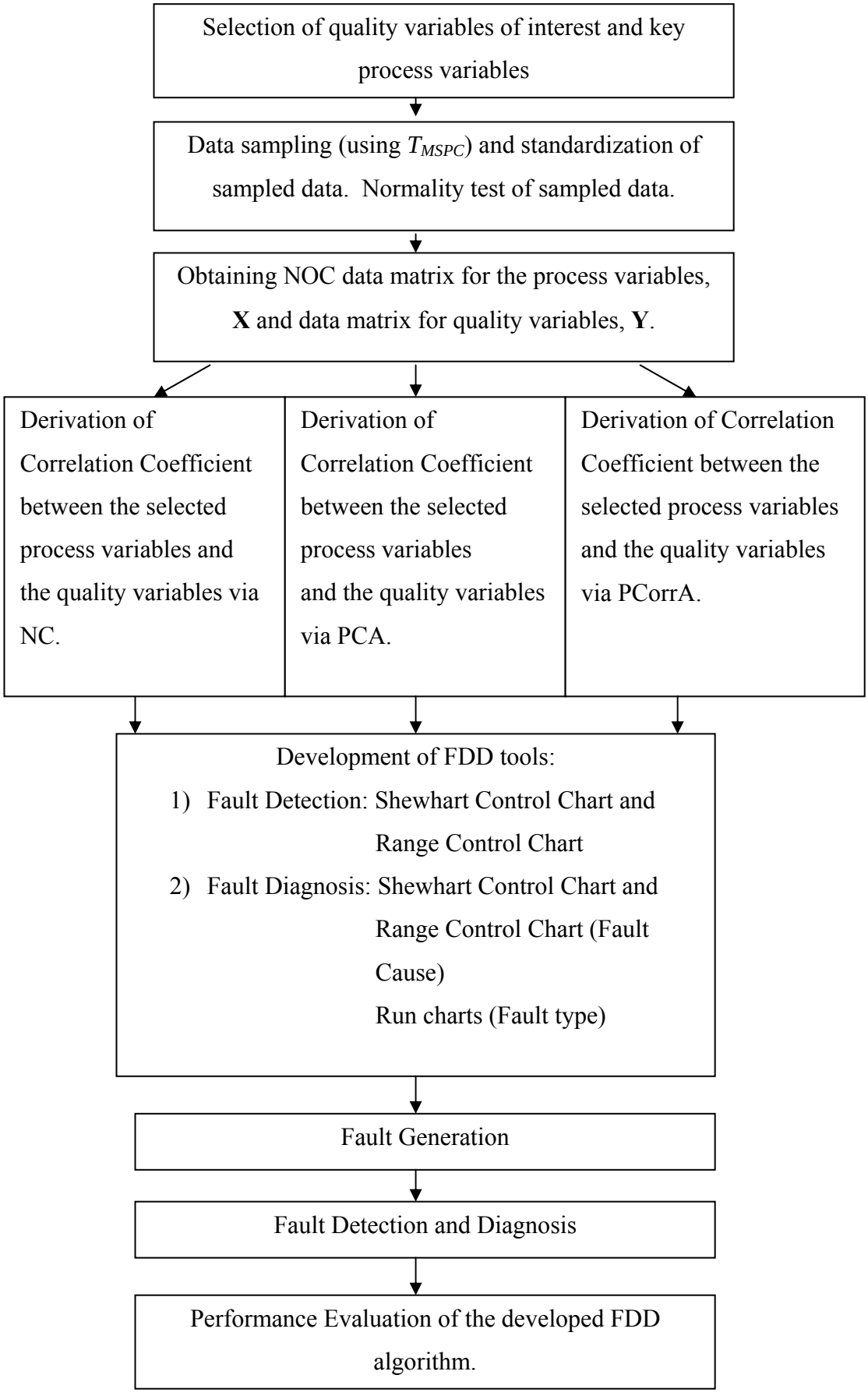


Figure 4.9: Summary of methodology of this research

CHAPTER V

RESULTS AND DISCUSSION

5.1 Introduction

This chapter will present the full results of this research. This chapter consists of seven main sections: introduction, selection of key process variables results, process sampling time results, normality test results, correlation coefficients derivation results, fault detection and diagnosis results and chapter summary.

The next section of this chapter gives the results of the selection of the key process variables. The third section presents the results obtained from the study of process sampling time. The fourth section presents the normality study results of the data obtained from the model.

The fifth section gives the results of correlation coefficient derivation from the generated Nominal Operation Condition (NOC) data. The sixth section consists of the results from the fault detection and diagnosis section of this research. The final section of this chapter is the chapter summary.

5.2 Results of Selection of Key Process Variables

From Section 4.2.1, the two quality variables of interest selected in this research are the linoleic acid composition (x_8) and oleic acid composition (x_9) in the bottom flow rate. The correlation (normal correlation) between a list of process variables with the two quality variables were determined to select key process variables that are having major contribution to the variation of the two quality variables (process variables that show absolute correlation value of equal or more than 0.1 are selected). Table 5.1 shows the results of the selection of key process variables.

Table 5.1: Results of selection of key process variables

Process Variable	Absolute Correlation value with Quality Variable 1 (x_8)	Absolute Correlation value with Quality Variable 2 (x_9)	Selection
Feed flow rate (L_f)	0.1712	0.1984	Selected
Feed Temperature (T_f)	0.7529	0.7219	Selected
Reflux flow rate (Re)	0.2122	0.2677	Selected
Pumparound temperature (T_p)	0.0114	0.0260	Not selected
Pumparound flow rate (P)	0.4594	0.4209	Selected
Reboiler duty (Q_r)	0.0718	0.2995	Selected*
Bottom temperature (T_{bot})	0.9987	0.9615	Selected

*Reboiler duty is selected since the absolute correlation value for this variable with Quality Variable 2 is more than 0.1.

5.3 Process Sampling Time (T_{MSPC}) Study Results

The sampling time of the Nominal Operating Condition (NOC) data is determined using an autocorrelation plot of the process data (using the process section model of a process with the highest time constant) based on the method proposed by Wetherill and Brown (1991). The total of observation used are 500 observations and the threshold value is 0.0894 (calculated using information from Section 4.2.2). Once the value of the autocorrelation is smaller than this threshold value, the corresponding first lag value when this value of autocorrelation is observed is chosen and the process sampling time will be calculated from this lag. Figure 5.1 shows the autocorrelation plot of the pumparound temperature (process section model with the highest time constant in the study column process) and how the MSPC sampling time is determined.

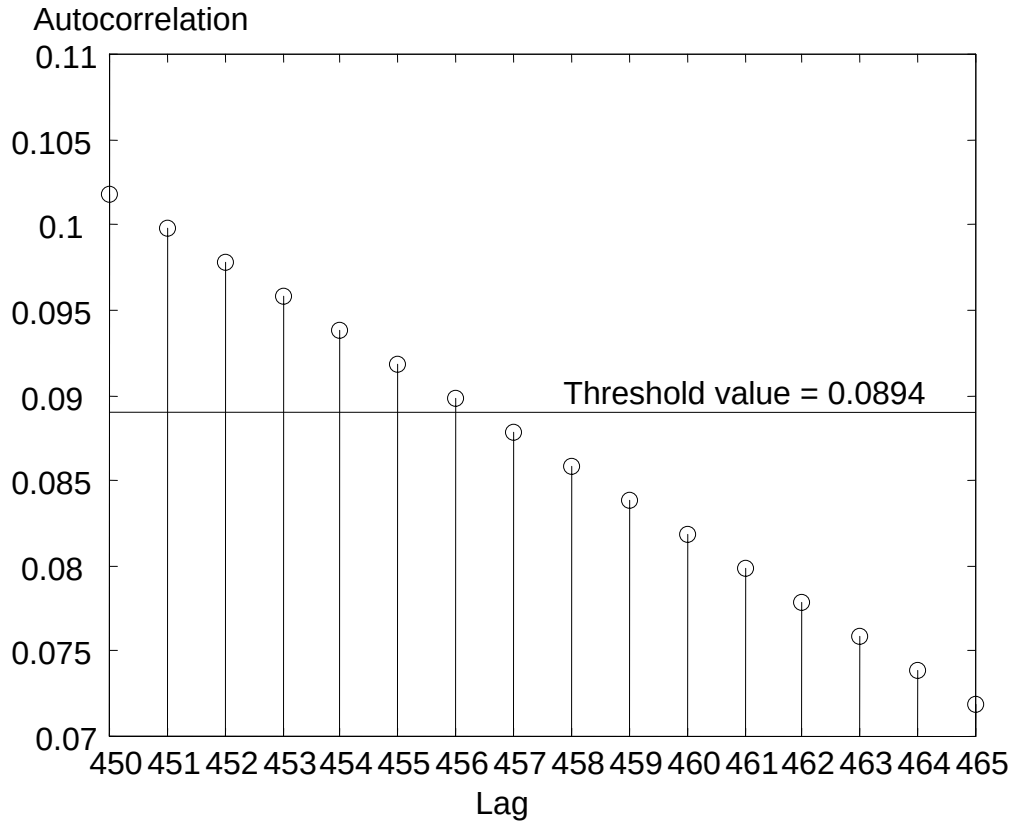


Figure 5.1: Autocorrelation plot of the pumparound temperature

From Figure 5.1, every one lag represents 0.01 hour and the first lag value that has an autocorrelation value of less than the chosen threshold value is around 457 lags. In

order to simplify the calculation of the process sampling time, a value of 460 lags are used. Therefore, the process sampling time (T_{MSPC}) is 4.6 hours (460 lags times 0.01hour/lag). **T_{MSPC} is used as the sampling time to sample data from the process for developing the fault detection and diagnosis algorithm and is different from the controller sampling time, T_{APC}, which is used to sample data from the process for control purposes**.

5.4 Normality Test Results

From Section 4.2.3.2, the sampled data (using T_{MSPC}) for establishing the NOC data are subjected to normality tests. There are 8 variables that are tested (two quality variables of interest and the six selected key process variables) and the results of their normality test are given in Table 5.2.

Table 5.2: Normality test results

Key Process Variable	Arithmetic average (x 10^{-9})	Standard deviation	Skewness	Kurtosis
Feed flow rate (L_f)	-0.0063	1.0000	0.0024	2.9603
Feed temperature (T_f)	0.0017	1.0000	-0.0042	2.9542
Reflux flow rate (Re)	0.0024	1.0000	0.0011	2.9464
Pumparound flow rate (P)	-0.0059	1.0000	0.0019	3.0350
Reboiler duty (Qr)	-0.0019	1.0000	0.0026	2.9660
Bottom temperature (T_{bot})	-0.0140	1.0000	0.0051	2.9821
Quality Variable				
Linoleic Acid mole fraction in bottom flow (x_8)	-0.0037	1.0000	0.0051	3.0412
Oleic Acid mole fraction in bottom flow (x_9)	0.0013	1.0000	0.0048	3.0127

For data following normal distribution, the mean, standard deviation, skewness and kurtosis of the data must have the value of 0, 3, 1 and 0, respectively (Wetherill and Brown, 1991). From the results of the normality test shown in Table 5.2, the small random numbers used to generate the NOC data follow the normal distribution. Therefore, the generated NOC data follow the normal distribution.

5.5 Correlation Coefficient Derivation Results

This section will show the results of the correlation coefficient derived from the analysis of NOC data using Normal Correlation (NC), Principal Component Analysis (PCA) and Partial Correlation Analysis (PCorrA). Table 5.3 – Table 5.5

shows the value of correlation coefficients between the selected key process variables (listed in Table 4.2) and the quality variables of interest (listed in Table 4.1).

Table 5.3: Correlation coefficient using Normal Correlation

Selected Key Process Variable	Correlation Coefficient with Quality Variable 1 (x_8)	Correlation Coefficient with Quality Variable 2 (x_9)
Feed flow rate (L_f)	0.1712	-0.1984
Feed temperature (T_f)	-0.7529	0.7219
Reflux flow rate (Re)	0.2122	-0.2677
Pumparound flow rate (P)	0.4594	-0.4209
Reboiler duty (Q_r)	-0.0719	0.2995
Bottom temperature (T_{bot})	0.9987	-0.9615

Table 5.4: Correlation coefficient using Principal Component Analysis

Selected Key Process Variable	Correlation Coefficient with Quality Variable 1 (x_8)		Correlation Coefficient with Quality Variable 2 (x_9)	
	90% of variation of original data retained	95% of variation of original data retained	90% of variation of original data retained	95% of variation of original data retained
Feed flow rate (L_f)	0.1744	0.1712	-0.2129	-0.1984
Feed temperature (T_f)	-0.7515	-0.7529	0.7198	0.7219
Reflux flow rate (Re)	0.2286	0.2122	-0.3268	-0.2677
Pumparound flow rate (P)	0.4565	0.4594	-0.4146	-0.4209
Reboiler duty (Q_r)	-0.0547	-0.0719	0.2423	0.2995
Bottom temperature (T_{bot})	0.9987	0.9988	-0.9672	-0.9616

Table 5.5: Correlation coefficient using Partial Correlation Analysis

Selected Key Process Variable	Correlation Coefficient with Quality Variable 1 (x_8)	Correlation Coefficient with Quality Variable 2 (x_9)
Feed flow rate (L_f)	0.9872	-0.9872
Feed temperature (T_f)	-0.9875	0.9874
Reflux flow rate (Re)	0.9882	-0.9881
Pumparound flow rate (P)	0.9876	-0.9875
Reboiler duty (Q_r)	-0.9956	0.9999
Bottom temperature (T_{bot})	0.4087	-0.3810

The results of the correlation coefficients for reboiler duty and bottom temperature in Table 5.3 – Table 5.5 shows that the feed temperature is between the boiling point of linoleic acid and oleic acid at the operating pressure of the bottom column. In this situation, when the feed temperature increases, the composition of linoleic acid in the bottom flow (x_8) will increase while the composition of oleic acid in the bottom flow (x_9) will decrease. If the feed temperature decreases, the opposite of the effect stated in the previous sentence will be observed in the bottom flow.

5.6 Fault Detection and Diagnosis Results

This section will show the results obtained from the fault detection and diagnosis algorithm on the generated fault data set. In the generated fault data set, there are 17 pre-designed faults including single cause faults, multiple cause faults, significant faults and insignificant faults. There are 14 significant faults and 3 insignificant faults. As stated earlier on, a fault is detected when one or both of the quality variables showing out-of-control limit signal in their statistical control chart (Shewhart Control Chart or Range Control Chart or both) and also one or more of the selected key process variables showing out-of-control limit signal in their statistical control chart (Shewhart Control Chart or Range Control Chart or both).

In this research, a variable (quality variable or selected key process variable) having value of more than the control limit of its statistical control chart will give an out-of-control limit signal and will be coded with a value of 1 while a variable (quality variable or selected key process variable) having value of less than the control limit of its statistical control chart will give an in-control limit signal and will be coded with a value of 0. Thus, for each sampling point (using T_{MSPC}), a series of codes of 0 and 1 will be given to the selected key process variables and the two quality variables of interest.

The completed detection and diagnosis profiles of the fault data using the developed fault detection and diagnosis (FDD) algorithm based on all three analysis methods (NC, PCA and PCorrA) will be given in **Appendix A**. Table 5.6 shows the total pre-designed faults generated in the fault data with their properties.

Table 5.6: Pre-designed faults generated in the fault data

Data Point	Cause Variable	Type of Fault	Size of Fault
2	L_f	Valve fault	Significant fault
5	*Re*	Sensor fault	Significant fault
8	T_f	Sensor fault	Significant fault
10	*P*	Valve fault	Significant fault
12	Q_r	Valve fault	Significant fault
14	*Re* and L_f	Sensor fault (*Re*), valve fault (L_f)	Significant fault
16	*Re* and T_f	Sensor fault (*Re* and T_f)	Insignificant fault
19	*P* and L_f	Sensor fault (L_f), valve fault (*P*)	Significant fault
23	T_f	Sensor fault	Insignificant fault
26	*P* and T_f	Sensor fault (*P* and T_f)	Significant fault
29	*Re* and L_f	Sensor fault (*Re*), valve fault (L_f)	Significant fault
32	Q_r and T_f	Sensor fault (T_f), valve fault (Q_r)	Significant fault
34	*P* and L_f	Sensor fault (*P*), valve fault (L_f)	Significant fault
37	*Re*, *P* and T_f	Sensor fault (*Re* and T_f), valve fault (*P*)	Significant fault
40	Q_r and L_f	Sensor fault (Q_r), valve fault (L_f)	Significant fault
42	*P*	Valve fault	Insignificant fault
45	Q_r, *Re*, *P* and L_f	Sensor fault (*Re*), valve fault (*P* and L_f), controller fault (Q_r)	Significant fault

(Feed flow rate (L_f), feed temperature (T_f), reflux flow rate (*Re*), pumparound flow rate (*P*), and reboiler duty (Q_r))

As stated earlier on, the types of faults considered in this research includes single cause faults, multiple cause faults, significant faults and insignificant faults. The FDD based on NC will serve as the benchmark for comparison. The total of faults incorporated into the Fault Data (OC) is 17 faults. For the NC method, 14 faults were detected out of the total 17 pre-designed faults. Out of these 14 detected faults, the NC method was able to successfully diagnose the cause of all the 14 detected faults. Figure 5.2 and Figure 5.3 show examples of fault detection and fault diagnosis of pre-designed faults using the FDD algorithm based on NC.

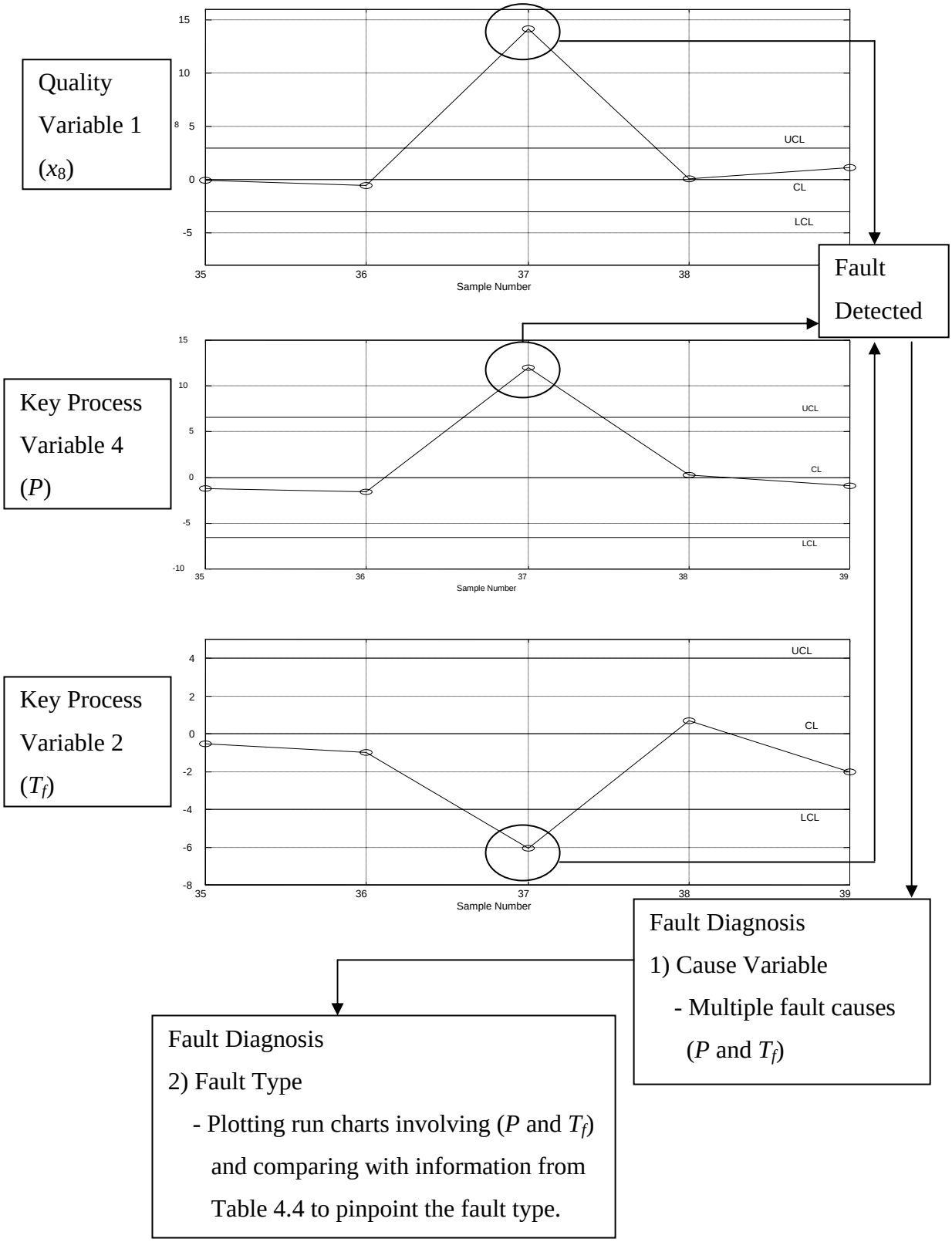


Figure 5.2: Example of fault detection and diagnosis of a multiple cause fault involving Quality Variable 1 (x_8) and Key Process Variable 4 (P) and Key Process Variable 2 (T_f) using FDD algorithm based on NC.

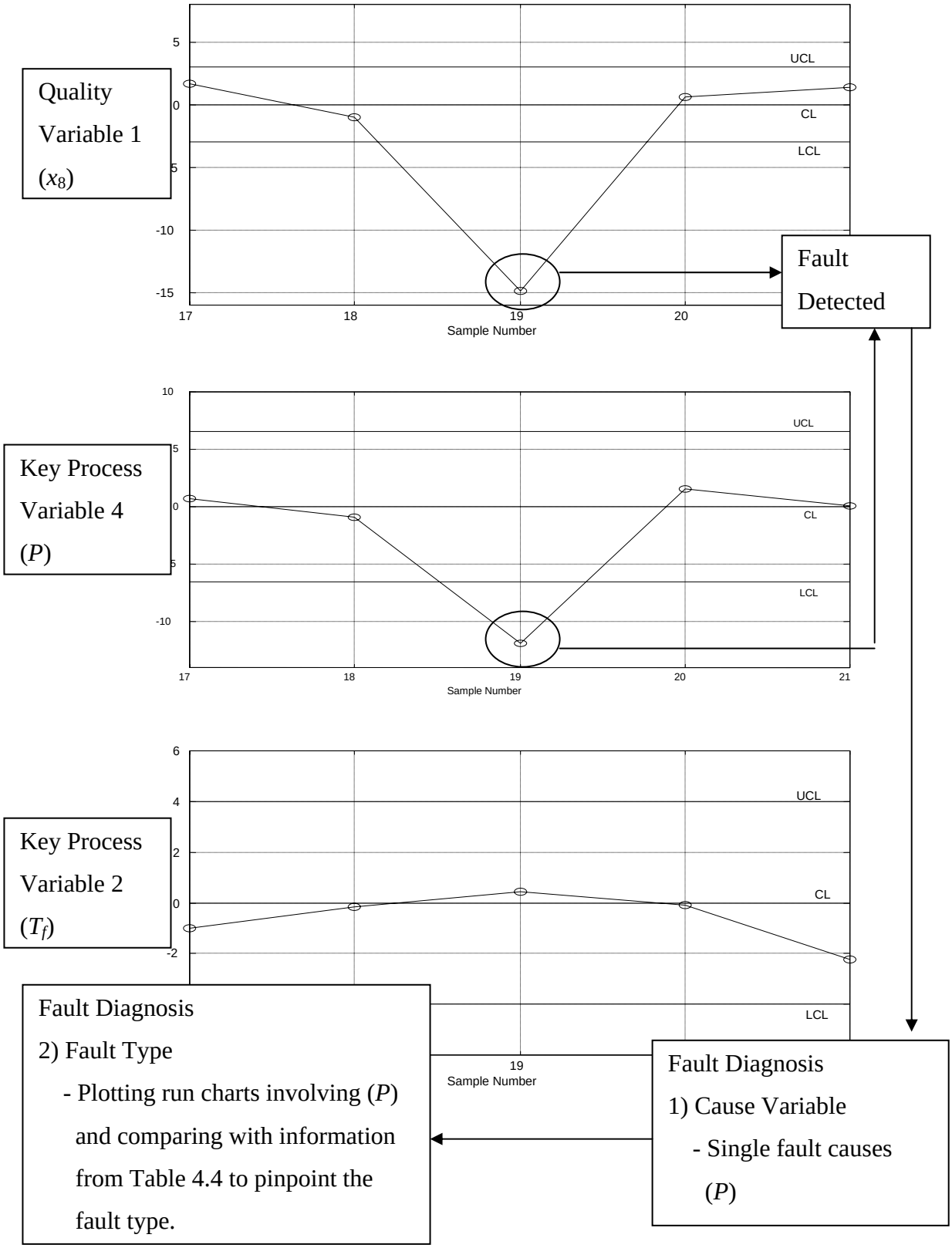


Figure 5.3: Example of fault detection and diagnosis of a single cause fault involving Quality Variable 1 (x_8) and Key Process Variable 4 (P) using FDD algorithm based on NC.

Figure 5.2 and Figure 5.3 show how fault detection and diagnosis was being conducted out using the developed FDD algorithm based on NC on Shewhart Control Chart. For Range Control Chart, the procedure is similar. From Figure 5.2 and Figure 5.3, only 2 key process variables were shown each time due to space constraint. The FDD algorithm based on NC did not manage to detect 3 faults (insignificant faults) present in the pre-designed faults incorporated into the process. This is because the statistical control limits for the control charts based on this NC method is insensitive to insignificant faults.

In this research, there are two PCA method used for developing the FDD algorithm based on the percentage of variation of the original Nominal Operation Condition (NOC) data retained. These PCA methods are PCA based on 90% of variation of original NOC data retained and PCA based on 95% of variation of original NOC data retained. The FDD algorithm based on both PCA methods managed to detect 14 faults and successfully diagnose the fault cause of each 14 detected faults. Figure 5.4 and Figure 5.5 show examples of fault detection and fault diagnosis of pre-designed faults using the FDD algorithm based on PCA (90% of variation of original NOC data retained).

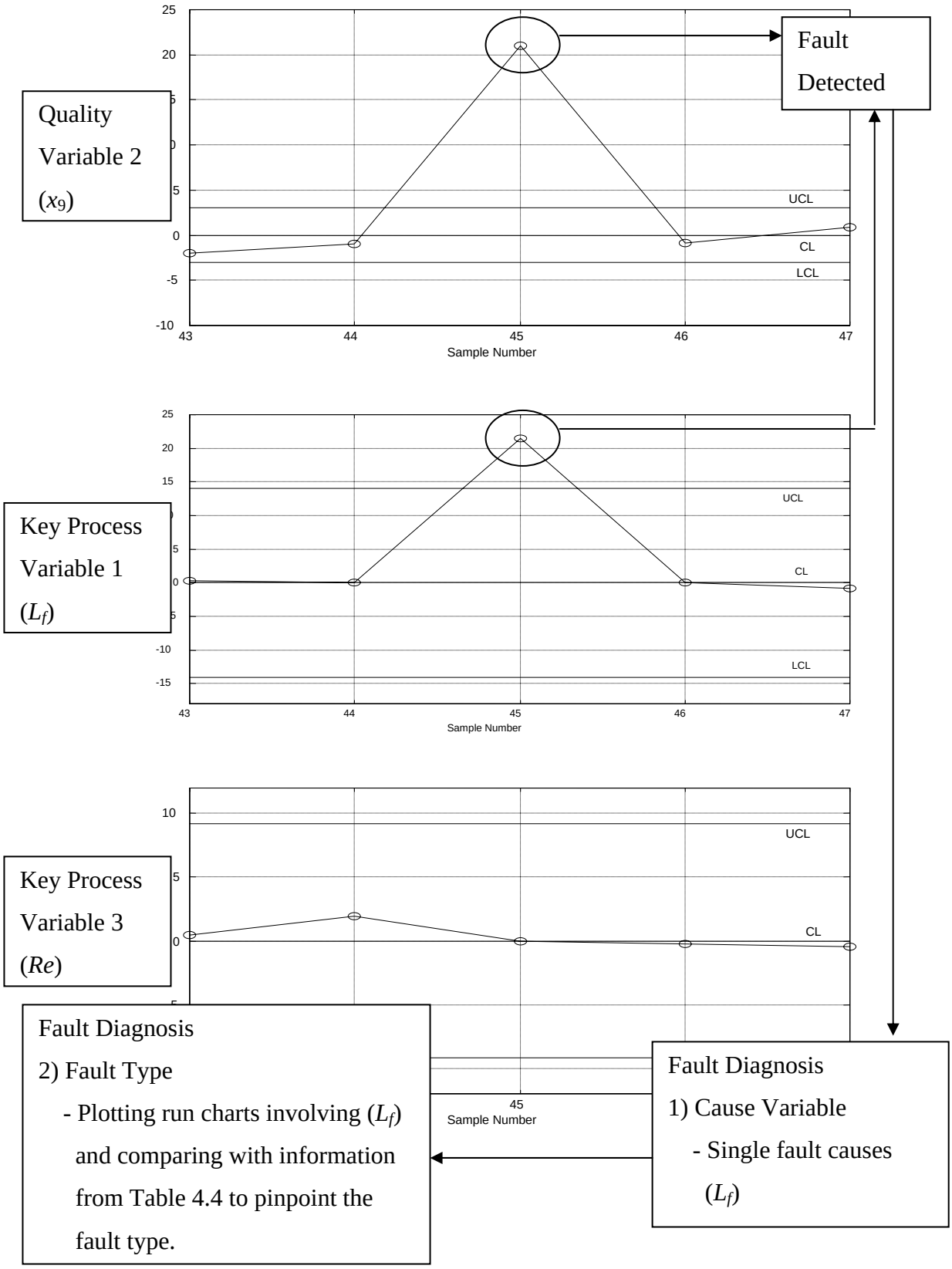


Figure 5.4: Example of fault detection and diagnosis of a single cause fault involving Quality Variable 2 (x_9) and Key Process Variable 1 (L_f) using FDD algorithm based on PCA (90% of variation of original NOC data retained).

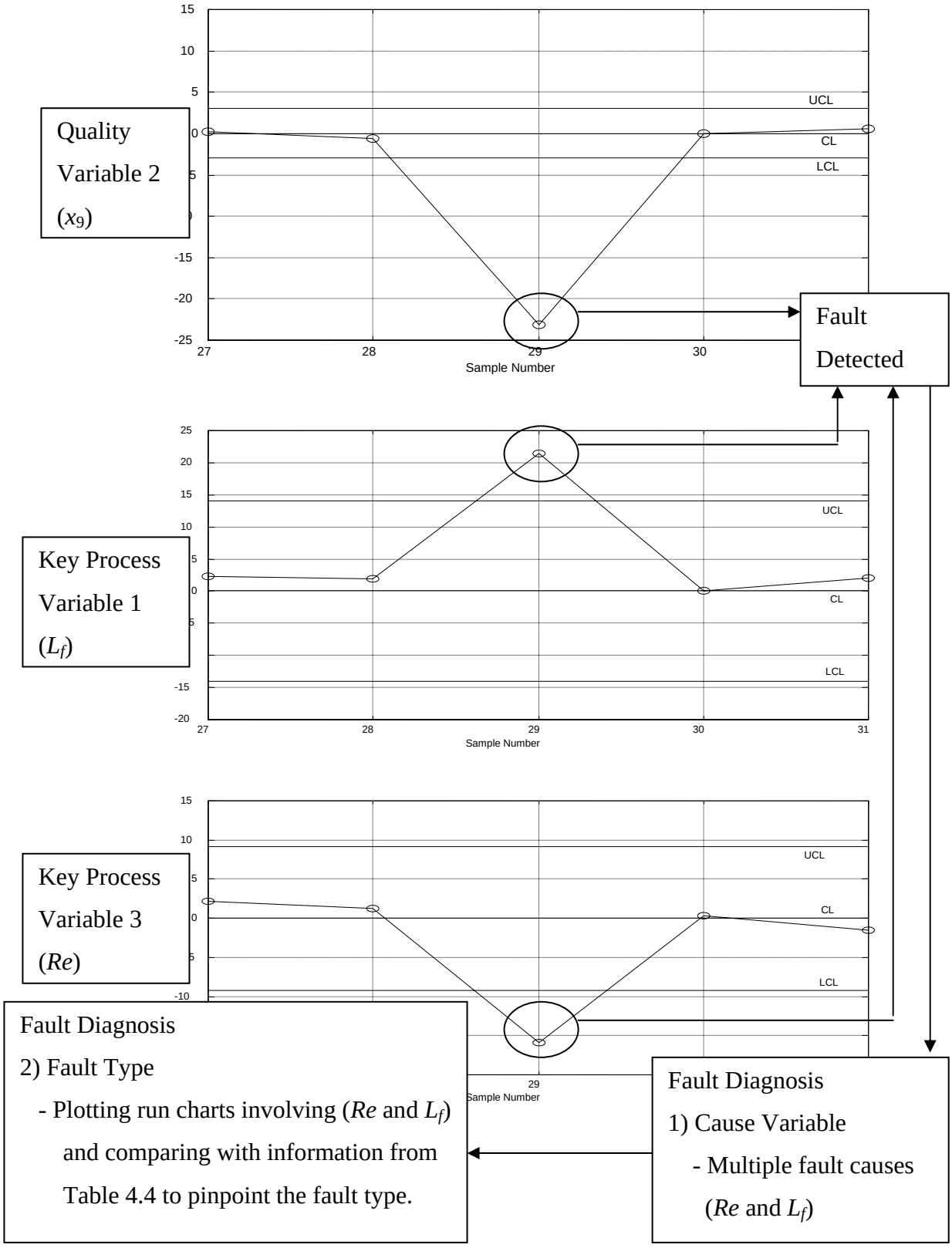


Figure 5.5: Example of fault detection and diagnosis of a multiple cause fault involving Quality Variable 2 (x_9) and Key Process Variable 1 (L_f) and Key Process Variable 3 (Re) using FDD algorithm based on PCA (90% of variation of original NOC data retained).

The FDD algorithm based on PCA (both method with 90% of variation of original NOC data retained and method with 95% of variation of original NOC data retained) also fail to detect the 3 insignificant faults in the pre-designed faults incorporated into the process. Although both PCA method show similar results, the method with 90% of variation of original NOC data retained is slightly better than the method with 95% of variation of original NOC data retained because the former method utilizes less variation of original NOC data compared to the latter method. PCA (both methods) also performed better than the NC although both PCA and NC show same results for FDD on the Fault Data (OC). This is because for PCA (both methods), there were dimension reduction of NOC data (not using 100% variation of the original NOC data) compared to NC which, utilizes all 100% variation of the original NOC data.

The algorithm based on NC and algorithm based on PCA (both method with 90% of variation of original NOC data retained and method with 95% of variation of original NOC data retained) show good results in detecting the multiple fault causes pre-designed in the fault data. These methods managed to detect 9 out of the total 10 multiple fault causes injected into the fault data. The data point in which the multiple fault causes were not detected is because this data point is an insignificant fault. Stated earlier on, these methods were insensitive to the 3 pre-designed insignificant faults.

The FDD algorithm based on PCorrA were able to detect all 17 pre-designed faults (significant faults and insignificant faults) and successfully diagnose the fault cause for all 17 detected faults. Figure 5.6 and Figure 5.7 show examples of fault detection and fault diagnosis of pre-designed faults using the FDD algorithm based on PCorrA.

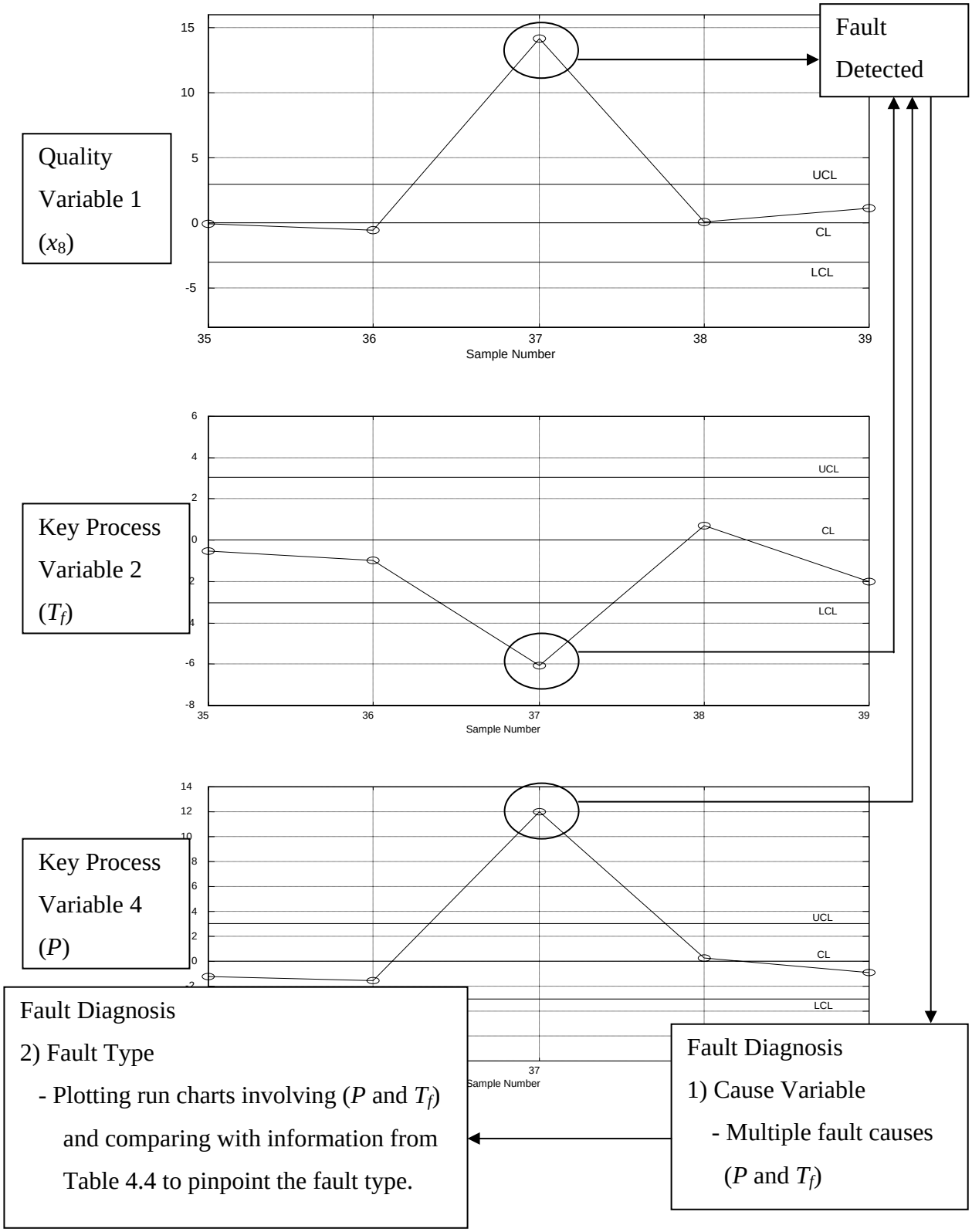


Figure 5.6: Example of fault detection and diagnosis of a multiple cause fault involving Quality Variable 1 (x_8) and Key Process Variable 2 (T_f) and Key Process Variable 4 (P) using FDD algorithm based on PCorrA.

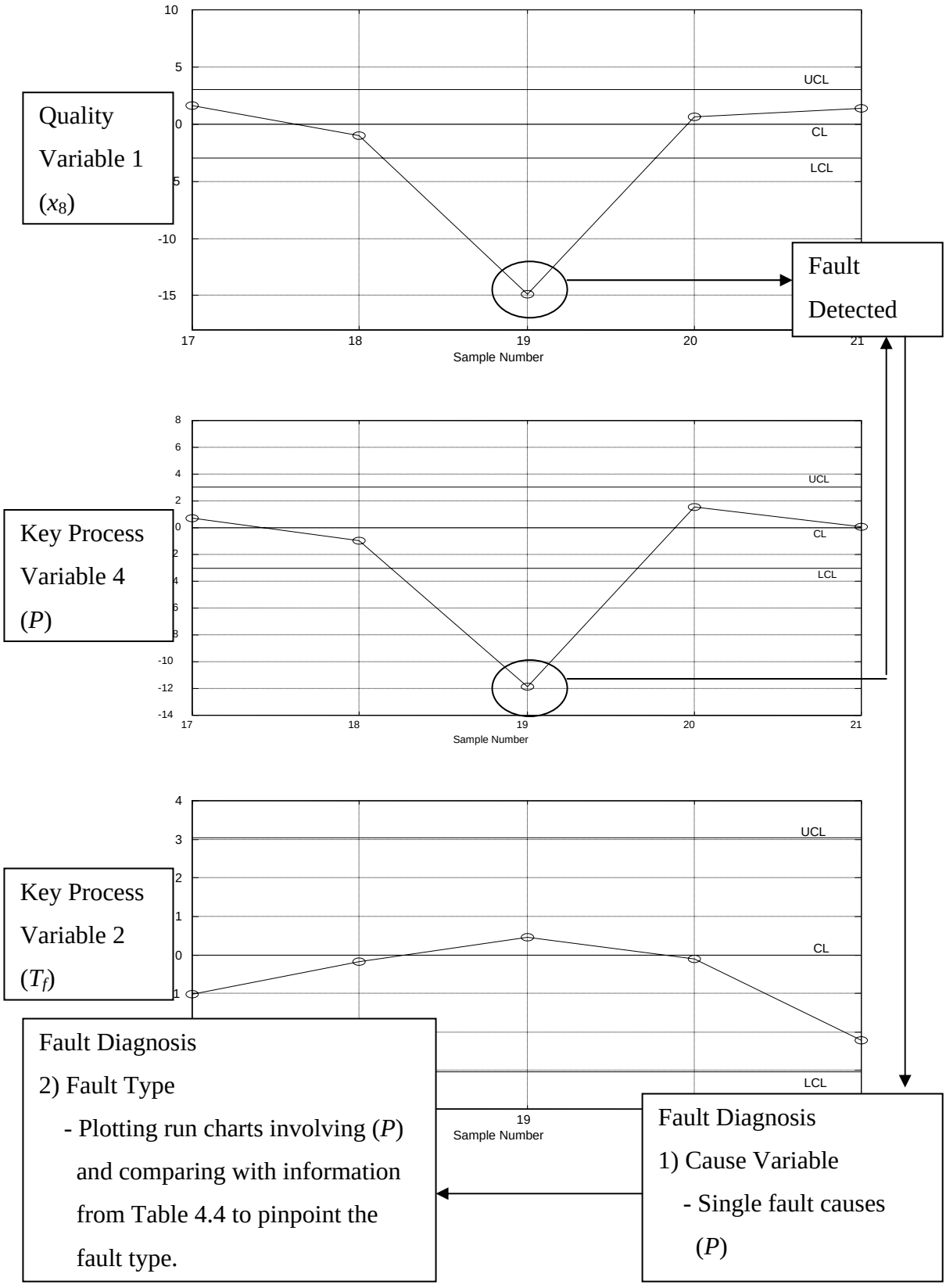


Figure 5.7: Example of fault detection and diagnosis of a single cause fault involving Quality Variable 1 (x_8) and Key Process Variable 4 (P) using FDD algorithm based on PCorrA.

The FDD algorithm based on PCorrA were able to detect the 3 insignificant faults in the pre-designed faults because the statistical control limits of control charts based on PCorrA are sensitive to insignificant faults. This method was also successfully in detecting and diagnosing all the 10 multiple fault causes injected in the fault data. Figure 5.8 and Figure 5.9 show examples when the FDD algorithm based on PCorrA managed to detect insignificant fault while the methods based on NC and PCA fail to do so.

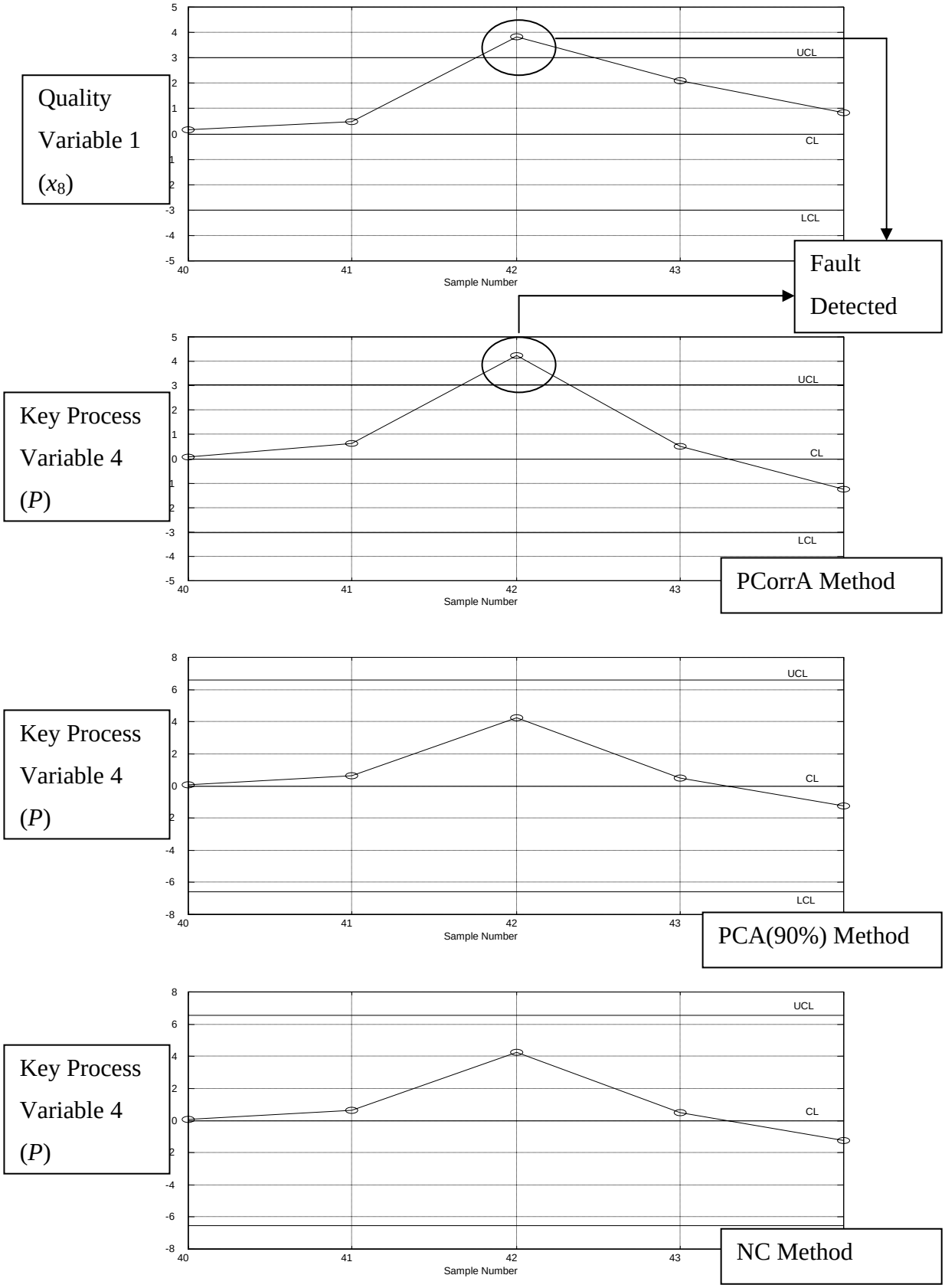


Figure 5.8: Fault detection of a single cause fault (insignificant fault) involving Quality Variable 1 (x_8) and Key Process Variable 4 (P) using FDD algorithm based on PCorrA.

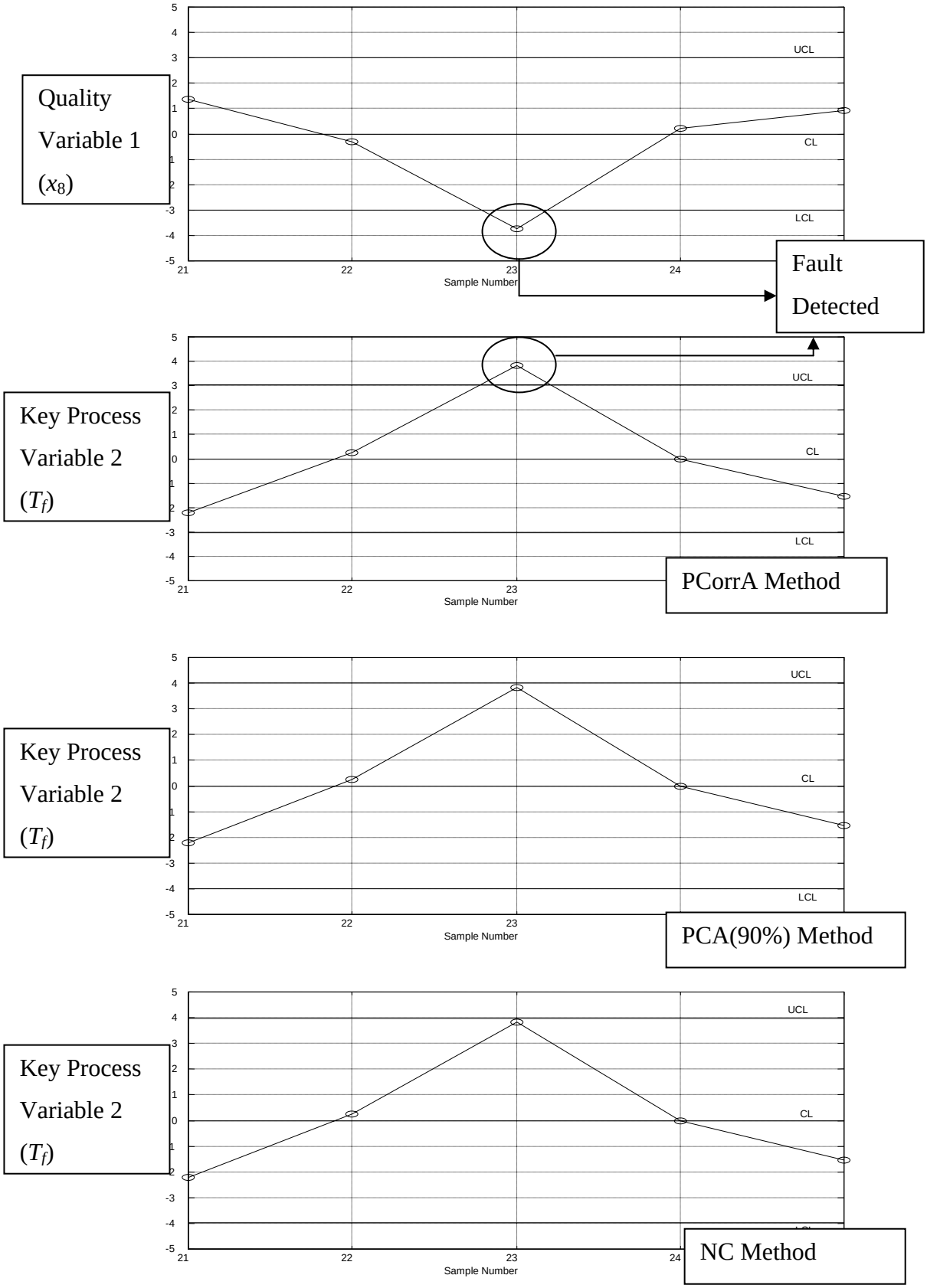


Figure 5.9: Fault detection of a single cause fault (insignificant fault) involving Quality Variable 1 (x_8) and Key Process Variable 2 (T_f) using FDD algorithm based on PCorrA.

Figure 5.10 – Figure 5.12 shows the performance of each method with their fault detection efficiency, fault diagnosis efficiency and overall FDD, respectively.

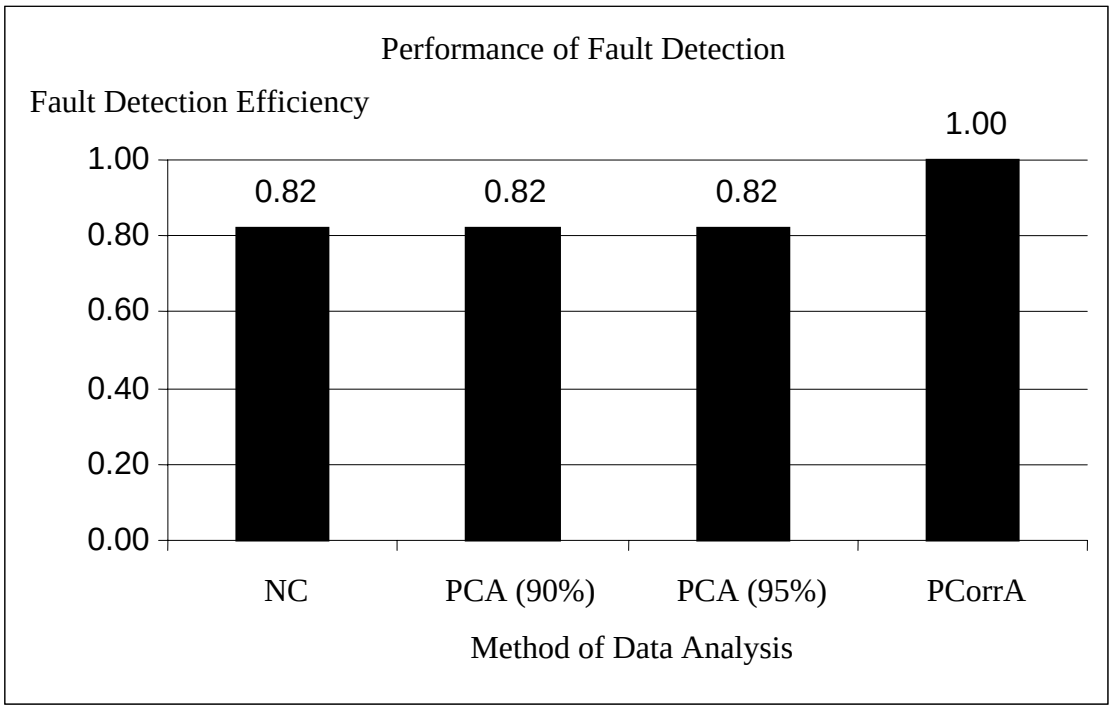


Figure 5.10: Fault detection efficiency of each method

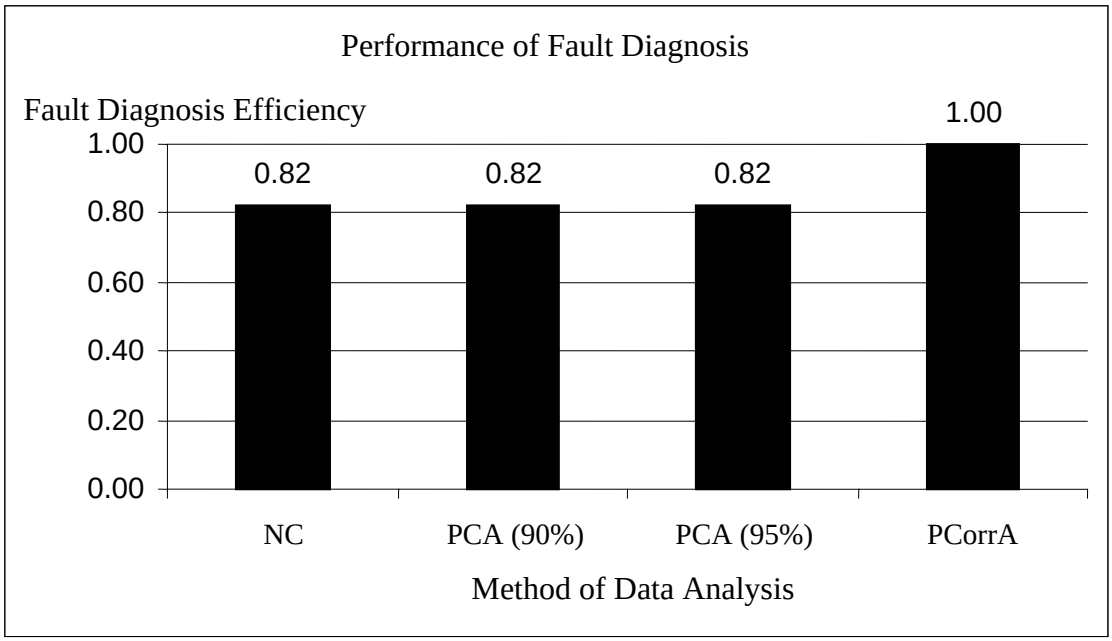


Figure 5.11: Fault diagnosis efficiency of each method

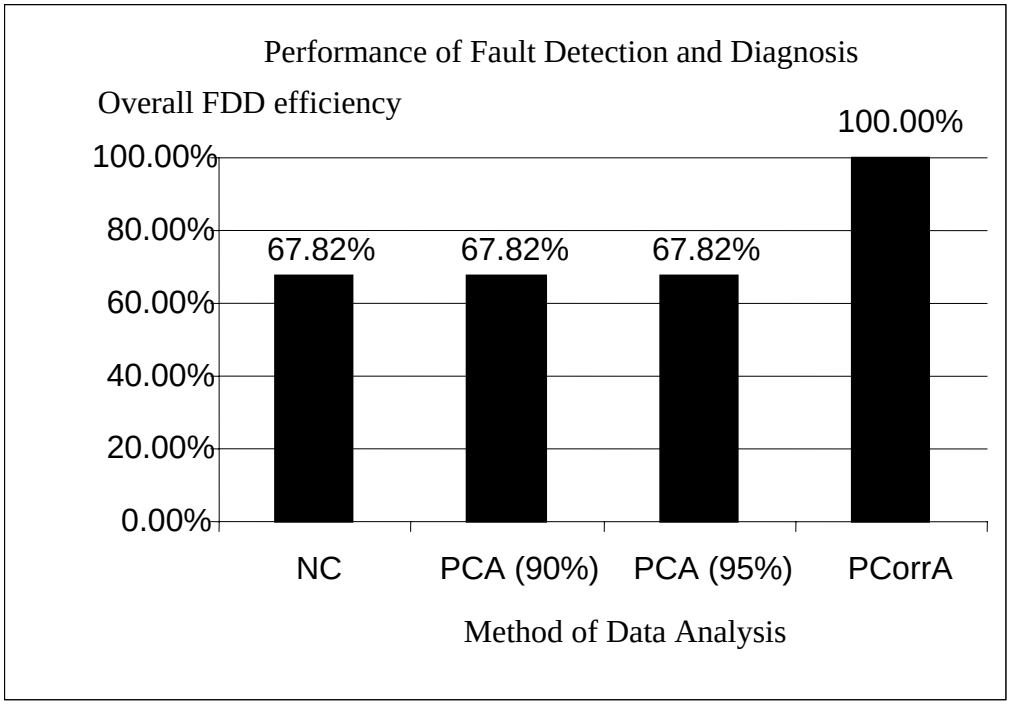


Figure 5.12: Overall FDD efficiency of each method

From Figure 5.10 – Figure 5.12, the FDD algorithm based on PCorrA perform the best among all the FDD algorithm based on different method of data analysis. The performance of FDD algorithm using correlation coefficients between the selected quality variables of interest and the selected key process variables very much depend on the value of the correlation coefficients. The closer the correlation coefficients represent the relationship between the selected key process variables with the selected quality variables of interest, the better the performance of the developed FDD algorithm.

The FDD algorithm based on NC although only manage to detect 13 faults out of total 17 pre-designed faults, this result certainly shows that the developed FDD algorithm are able to detect process faults that are present. The FDD algorithm based on PCA (both method with 90% of variation of original NOC data retained and method with 95% of variation of original NOC data retained) also show similar performance as of the FDD algorithm based on NC. However, the PCA method is slightly superior because this method takes into account the cross-correlation between the selected key process variables while determining their correlation with the two quality variables of interest while the NC method determine the correlation

coefficients by using only one to one regression. Another advantage of the PCA method is that PCA perform dimension reduction of the original NOC data during the determination of the correlation coefficients while the NC do not have dimension reduction in its operation. The advantage of dimension reduction is that less variation (fewer amounts of variables) of the original NOC data are needed to represent the original NOC data. This advantage will be more important if the study process involve high amount of variables.

The reason the FDD algorithm based on PCorrA perform the best among all the FDD algorithm was the value of correlation coefficients derived using PCorrA. PCorrA maintain the value of the other selected key process variable while determining the correlation coefficient between a selected key process variable with a quality variable of interest. This procedure not only takes account into the cross-correlation among the selected key process variables but also the effect of other selected key process on the correlation between the key process variable under study (the key process variable which its correlation with a quality variable of interest is to be determined) and the quality variable of interest. The derived correlation coefficients using PCorrA will better represent the relationship between the selected key process variables and the quality variables of interest. This will make the developed FDD algorithm using these correlation coefficients to be more sensitive to any changes in the variation of the quality variables of interest with the selected key process variables. Therefore, the FDD algorithm based on PCorrA were able to detect both significant faults and insignificant faults in the pre-designed faults incorporated into the Fault Data (OC).

The developed FDD algorithm using correlation coefficients (NC, PCA and PCorrA) were able to successfully diagnose the fault cause for each detected fault. These results show that the developed FDD algorithm using correlation coefficients is promising in diagnosing fault causes. The previous ambiguous nature (unavailability of control limit in contribution plots) of contribution plots use to diagnose fault causes (Yoon and MacGregor, 2000) is also overcome by the developed FDD algorithm using correlation coefficients as each statistical control chart (Shewhart Control Chart and Range Control Chart) have statistical control limits. Each situation can be easily assess whether the particular selected key

process variable is a fault cause or not due to the presence of statistical control limits in the statistical control charts.

The introduction of coding for fault situations or NOC conditions also make the developed FDD algorithm simple yet promising to be implemented in large scale chemical plant operations. As a conclusion, the developed FDD algorithm using correlation coefficients were able to detect the pre-designed faults and successfully diagnose the fault causes of the detected fault. The introduction of PCorrA as a method for data analysis in Multivariate Statistical Process Control (MSPC) was also successful as the FDD algorithm using PCorrA is the best method among all the developed FDD algorithms.

5.7 Chapter Summary

This chapter contains the results obtained from the work for the whole duration of this research. The results presented include the selection of key process variables results, process sampling time (T_{MSPC}) results, normality study of NOC data results, correlation coefficient derivation results and fault detection and diagnosis results. The FDD results obtained based on the methods of NC, PCA (90% and 95% of original variation of NOC data retained) and PCorrA were presented in detail and the reasons for the observed results were discussed.

CHAPTER VI

CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

The results of this research were presented in the previous chapter. There are few conclusions that can be drawn from the results obtained.

Firstly, an algorithm for fault detection, diagnosis and control system identification was successfully developed. The developed algorithm was able to detect and diagnose the pre-designed faults in the study process.

Secondly, the developed program package contains several analysis strategies (Normal Correlation, Principal Component Analysis (90%), Principal Component Analysis (95%) and Partial Correlation Analysis) and multiple types of monitoring charts (Shewhart Control Chart and Range Control Chart) for detecting and diagnosing faults in the study process.

6.2 Recommendations

In this research, the process is considered out-of-control when the value of a data point of a variable exceeds the control limit of its Shewhart Control Chart or Range Control Chart. The reason for using this rule is because the fault detection and diagnosis algorithm used in this research is only activated once there is an out-of-control limit signal present in one or both of the Shewhart Control Chart of the quality variables of interest. For normally distributed data, the probability of a data point exceeds the control limit of a quality variable of interest using a statistical control limit of $\pm 3\sigma$ is very low (around 0.27%). Therefore, by using only one rule (the value of a data point of a variable exceeds the control limit of its Shewhart Control Chart shows an out-of-control limit signal) will make the developed FDD algorithm able to detect any abnormal changes in the process without risking the increase in the number of false alarms (situations where there is no faults present in a process but the statistical control charts show otherwise).

There are numerous rules in which the data points can be considered to show an out-of-control limit signal in a Shewhart Control Chart (Quesenberry, 1997). By increasing the number of rules in the Shewhart Control Chart for an out-of-control limit signal will increase the robustness of the control chart, it will tend to increase the number of false alarms. On the whole, there must be a compromise between robustness of the statistical control charts and the number of false alarms that happens on the number of rules to be used when using Shewhart Control Chart.

Aside from Shewhart Control Chart and Range Control Chart, there are numerous other types of control charts that can be used with the proposed correlation coefficients in this research such as Exponentially-Weighted Moving Average Control Chart (EWMA), Cumulative-Sum Control Chart (CUSUM) and Moving Average Control Chart (MA) (Wachs and Lewin, 1999). By applying the developed correlation coefficients on these control charts, the results of the FDD of the fault data set will certainly be different.

In this research, three techniques of correlation analysis were used: NC, PCA and PCorrA. For future work, techniques such as Partial Least Squares (PLS) and

Independent Component Analysis (ICA) can be used to develop the correlation coefficients between the variables of the data matrix. Chemical processes tend to change continuously and the correlation coefficients developed from history data may not truly represent the relationship between process variables. Therefore, online updating of the data matrix used to develop the correlation coefficients can take account into process dynamics and give better FDD qualities of the developed FDD method based on correlation coefficients.

Finally, the author hopes that this research work can be a platform for future studies on the field of fault detection and diagnosis using Multivariate Statistical Process Control (MSPC) via correlation coefficients.

REFERENCE

Biran, A. dan Breiner, M. (1995). "MATLAB for Engineers." Addison-Wesley Publishing Company.

Chen, G. and McAvoy, T.J. (1998). "Predictive On-line Monitoring of Continuous Processes." *Journal of Process and Control*. **Vol.8**. 409-420.

Chen, J. and Kun, C.L. (2002). "On-line Batch Process Monitoring Using Dynamic PCA and Dynamic PLS Models." *Chemical Engineering Science*. **Vol.57**. 63-75.

Chen, J., Liao, C.M., Lin, F.R.J. and Lu, M.J. (2001). "PCA Based Control Charts with Memory Effect for Process Monitoring." *Industrial and Engineering Chemistry Research.* **Vol. 40**. 1516-1527.

Chen, J. and Liu, J. (2001). "Derivation of function space analysis based PCA control charts for batch process monitoring." *Chemical Engineering Science*. **Vol. 56**. 3289 – 3304.

Cho, J.H., Lee, J.M., Choi, S.W., Lee, D.K. and Lee, I.B. (2005). "Fault identification for process monitoring using kernel principal component analysis." *Chemical Engineering Science*. **Vol. 60**. 279 – 288.

Choi, S.W. and Lee, I.B. (2004). “Nonlinear dynamic process monitoring based on dynamic kernel PCA.” *Chemical Engineering Science*. **Vol. 59**. 5897 – 5908.

Choi, S.W. and Lee, I.B. (2005). “Multiblock PLS-based localized process diagnosis.” Journal of Process Control. **Vol. 15**. 295 – 306.

Cliff, A.D. and Ord, J.K. (1973). “Spatial Autocorrelations.” London: Pion.

Dash, S., Rengaswamy, R. and Venkata, V. (2003). “Fuzzy-logic based trend classification for fault diagnosis of chemical processes.” *Computers and Chemical Engineering*. **Vol. 27**. 347 – 362.

Dayal, B.S. and MacGregor, J.F. (1997). “Recursive Exponentially Weighted PLS and Its Applications to Adaptive Control and Prediction.” *Journal of Process Control*. **Vol.7 (3)**. 169-179.

Ding, A.C. and Nancy, P.L. (2000). “Partial Correlation of Fuzzy Sets.” *Fuzzy Sets and Systems*. **Vol.110**. 209-215.

Douglas, J.F., Gasiorek, J.M. and Swaffield, J.A. (1996). “Fluid Mechanics.” Longman Singapore Publishers (Pte) Ltd.

Dunia, R., Qin, S.J., Edgar, T.F. and McAvoy, T.J. (1996). “Use of Principal Component Analysis for sensor fault identification.” *Computers and Chemical Engineering*. **Vol. 20, supplement**. S713 – S718.

Ergon, R. (2004). “Informative PLS score-loading plots for process understanding and monitoring.” *Journal of Process Control*. **Vol. 14**. 889 – 897.

Felder, R.M. and Rousseau, R.W. (1986). “Elementary Principles of Chemical Processes.” John Wiley and Sons.

Gani, R., Ruiz, C.A. and Cameron, I.T. (1986). "A Generalized Model for Distillation Columns – I: Model Description and Applications." *Computers and Chemical Engineering*, **Vol.10 (3)**. 181-198.

Geankoplis, C.J. (1995). "Transport Processes and Unit Operations." Simon and Schuster (Asia) Pte Ltd.

Geladi, P. and Kowalski, B. (1986). "Partial Least Squares Regression: A Tutorial." *Analytical Chimica Acta*. **Vol.185**. 1-17.

Gertler, J. and Li, W.H. (1999). "Isolation Enhanced Principal Component Analysis." *AIChE Journal*. **Vol.45, (2)**. 323 – 334.

Grassi, V.G. (1990). "Rigorous Modeling and Conventional Simulation." In Luyben, W.L. "Practical Distillation Control." N.Y. Van Nostrand Reinhold. 29-47.

Haaland, P.M. and Thomas, E.V. (1988). "Partial Least Squares Methods for Spectral Analysis." *Analytical Chimica Acta*. **Vol.60**. 1193-1208.

Himmelblau, D.M. (1978). "Fault Detection and Diagnosis in Chemical and Petrochemical Processes." Elsevier Scientific Publishing Company.

Isermann, R. (1997). "Supervision, Fault Detection and Fault Diagnosis Methods – An Intorduction." *Control Engineering Practice*. **Vol. 5(5)**. 639 – 652.

Jade, A.M., Jayaraman, V.K., Kulkarni, B.D., Jog., J.F. and Priya, L. (2003). "Feature extraction and denoising using kernel PCA." *Chemical Engineering Science*. **Vol. 58**. 4441 – 4448.

Kamarul, A.I. (1997). "Application of Partial Correlation in Active Statistical Process Control." *Proceedings of RSCE 1997*. Johor, Malaysia. 434-439.

Kaspar, M.H. and Ray, W.H. (1993). “Partial Least Squares Modeling as SuccessiveSingular Value Decompositions.” *Computers and Chemical Engineering*. **Vol. 17 (10)**. 985-989.

Kourti, T. , Lee, J. dan MacGregor, J.F. (1996). “Experiences with Industrial Applications of Projection Methods for Multivariate Statistical Process Control.” *Computers in Chemical Engineering*, **Vol. 20**. **Suppl**. S745-S750.

Kourti, T. dan MacGregor, J.F. (1996). “Recent Developments in Multivariate SPC Methods for Monitoring and Diagnosing Process and Product Performance.” *Journal of Quality Technology*, **Vol.28 (4)**. 409-428.

Kresta, J., MacGregor, J.F. and Martin, T.E. (1991). “Multivariate Statistical MonitoringOf Process Operation Performance.” *Canadian Journal of Chemical Engineering.* **Vol.69.** 34-47

Kruger, U., Chen, Q., Sandoz, D.J. and McFarlane, R.C. (2001). “Extended PLS Approach For Enhanced Condition Monitoring of Industrial Processes.” *AIChE Journal*. **Vol.47 (9)**. 2076-2091.

Kruger, U., Zhou, Y.Q. and Irwin, G.W. (2004). “Improved principal component monitoring of large-scale processes.” *Journal of Process Control*. **Vol. 14**. 879 – 888.

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

Lam, H.L. and Kamarul, A.I. (2002) " Improved Multivariate Statistical Process Control for Chemical Process Fault Detection and Diagnosis via Cross Variable Correlation Coefficient Approach." *Regional Symposium on Chemical Engineering, RSCE/SOMChem 2002*, **Vol. 2.** 1637-1644.

Lee, J.M., Yoo, C.K., Choi, S.W. Vanrollegham, P.A. and Lee, J.B. (2004a). "Nonlinear process monitoring using kernel PCA." *Chemical Engineering Science*. **Vol. 59**. 223 – 234.

Lee, J.M., Yoo, C.K. and Lee, J.B. (2004b). "Statistical process monitoring with independent component analysis." *Journal of Process Control*. **Vol. 14**. 467 – 485.

Lee, L.E. and Kamarul, A.I. (2001). "Dynamic Modeling and Simulation of Distillation Column." *Proceedings of WEC (Biochemical and Environmental)*. 24.

Lennox, B., Goulding, P.R. and Sandoz, D.J. (2001). "Analysis of Multivariate Statistical Methods for Continuous Systems." Control Technology Center, School of Engineering, University of Manchester, United Kingdom.

Luyben, W.L. (1963). "Process Modeling, Simulation and Control for Chemical Engineer." McGraw-Hill. 152-159.

MacGregor, J.F., Kiparissides, C. and Skagerberg, B. (1992). "Multivariate Data Analysis Applied to Low Density Polyethylene Reactors." *Chemometrics and Intelligent Laboratory Systems*. **Vol.14**. 341-356.

Manabu, K., Koji, N., Shinji, H., Iori, H. Hiromu, O., Ramon, S. and Bakshi, B. (2000). "Comparison of Statistical Process Monitoring Methods: Application to the Eastmann Challenge Problem." *Computers and Chemical Engineering*. **Vol.24**. 175-181.

Manabu, K., Shouhei, T., Hiromu, O., Shinji, H. and Iori, H. (2002). " The Use of Independent Component Analysis for Multivariate Statistical Process Control." *Proceedings of AdCONIP'02*. 423-428.

Manish, M., Yue, H.H., Qin, S.J. and Ling, C. (2002). "Multivariate Process Monitoring and Fault Diagnosis by Multi-scale PCA." *Computers and Chemical Engineering*. **Vol.26**. 1281-1293.

McFarlane, R.C., Reineman, R.C., Bartee, J.F. and Georgakis, C. (1993). "Dynamic Simulator for a Model IV Fluid Catalytic Cracking Unit." *Computers and Chemical Engineering*. **Vol.17**. 275.

McNeese, W.H. and Klein, R.A. (1991). "Statistical Methods For the Process Industries." USA: ASQC Quality Press. 251.

Nomikos, P. and MacGregor, J.F. (1994). "Monitoring of Batch Processes Using Multi- Way PCA." *AIChE Journal*. **Vol.40 (8)**. 1361-1375.

Nomikos, P. and MacGregor, J.F. (1995). "Multi-Way PLS in Monitoring Batch Processes." *Chemometrics and Intelligent Laboratory Systems*. **Vol.30**. 97-108.

Ogunnaike, B.A. and Ray, W.H. (1994). "Process Dynamics, Modeling and Control." Oxford University Press, Inc.

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

Quemerais, B., Cossa, D., Rondeau, B., Pham, T.T. and Fortin, B. (1998). "Mercury Distribution in Relation to Iron and Manganese in the Waters of St. Lawrence River." *The Science of the Total Environment.* **Vol.213**. 193-201.

Quesenberry, C.P. (1997). "SPC methods for quality improvement." John Wiley and Sons, Inc. 142-144, 274.

Raich, A. and Cinar, A. (1996). "Statistical Process Monitoring and Disturbance Diagnosis in Multivariate Continuous Processes." *AIChE Journal*. **Vol.42**. 995-1009.

Randall, D.T. (1997). “An Introduction to Partial Least Squares Regression.”, SAS Institute Inc., United States of America.

Reid, R.C., Prausnitz, J.M. and Poling, B.E. (1987). “The Properties of Gases and Liquids.” 4th Edition. McGraw-Hill. 273-283.

Shinskey, F.G. (1984). “Distillation Control For Productivity and Energy Conservation.” 2nd Edition. McGraw-Hill. 94-100, 121-130.

Simoglou, A., Martin, E.B. and Morris, A.J. (2002). “Statistical Performance Monitoring of Dynamic Multivariate Processes Using State-Space Modeling.” *Computers and Chemical Engineering*. **Vol.26**. 909-920.

Sinnot, R.K. (1984). “Chemical Engineering Volume 6.” Pergamon.

Smith, C.A. and Corripio, A.B. (1985). “Principles and Practice of Automatic Process Control.” 2nd Edition, John Wiley and Sons, Inc. Canada.

Stephanopoulos, G. (1984). “Chemical Process Control: An Introduction to Theory and Practice.” 1st Edition, Englewood Cliffs, New Jersey, Massachussets Institute of Technology

Sourabh, D. and Venkata, V. (1999). “Challenges In the Industrial Applications of Fault Diagnostic Systems.” Laboratory for Intelligent Process Systems, School Of Chemical Engineering, Purdue University, United States of America.

Treasure, R.J., Kruger, U. and Cooper, J.E. (2004). “Dynamic multivariate statistical process control using subspace identification.” *Journal of Process Control*. **Vol. 14**. 279 – 292.

Van Ness, H.C., Smith, J.M. and Abbott, M.M. (1996). “Introduction to Chemical Engineering Thermodynamics.” 5th Edition. McGraw-Hill. 88-89, 434-442.

Vasilopoulos, A.V. and Stamboulis, A.P. (1978). “Modification of Control Chart Limits in the Presence of Data Correlation.” *Journal of Quality Technology*, **Vol.10 (1)**. 20-29.

Vedam, H. and Venkata, V. (1999). “PCA-SDG based process monitoring and fault diagnosis.” *Control Engineering Practice*. **Vol.** 7. 903 – 917.

Venkata, V., Rengaswamy, R., Yin, K. and Kavuri, S.N. (2003a). “A review of process fault detection and diagnosis Part I: Quantitative model-based methods.” *Computers and Chemical Engineering*. **Vol. 27**. 293 – 311.

Venkata, V., Rengaswamy, R. and Kavuri, S.N. (2003b). “A review of process fault detection and diagnosis Part II: Qualitative models and search strategies.” *Computers and Chemical Engineering*. **Vol. 27**. 313 – 326.

Venkata, V., Rengaswamy, R., Kavuri, S.N. and Yin, K. (2003c). “A review of process fault detection and diagnosis Part III: Process history-based methods.” *Computers and Chemical Engineering*. **Vol. 27**. 327 – 346.

Wachs, A. and Lewin, D.R. (1999). “Improved PCA Methods for Process Disturbance and Failure Identification.” *AIChE Journal*. **Vol.45**. 1688-1700.

Walas, S.M. (1985). “Phase Equilibria in Chemical Engineering.” Butterworth Publishers. 178-188.

Wang, J.C. and Henke, G.E. (1966). “Tridiagonal Matrix for Distillation.” *Hydrocarbon Processing*. **Vol.45 (8)**. 155-165.

Wang, F. and Chen, J. (2000). “Relation of Sediment Characteristics To Trace Metal Concentrations: A Statistical Study.” *Wat.Res*. **Vol. 34 (2)**. 694-698.

Wetherill, G.B. and Brown, D.W. (1991). “Statistical Process Control: Theory and practice.” 1st Edition. Chapman and Hall.

Wong, T.S. (2003). “Advanced Process Control of A Fatty Acid Distillation Column Using Artificial Neural Networks.” Unpublished Masters Thesis (Universiti Teknologi Malaysia), Malaysia.

Yoon, S.K. and MacGregor, J.F. (2000). “Statistical and Causal Model-Based Approach to Fault Detection and Isolation.” *AIChE Journal*. **Vol.46 (9)**. 1813-1824.

Yoon, S.K. and MacGregor, J.F. (2001). “Fault Diagnosis with Multivariate Statistical Models Part I: Using Steady-state Fault Signatures.” *Journal of Process Control*. **Vol.11**. 387-400.

Zhang, Y., Xu, B. and Kang, B. (1999). “Steady-state and Dynamic of Simulation of Distillation Column for Model-Based Operation Optimization.” Unpublished paper.

APPENDIX A

APPENDIX B

Improved Statistical Process Control, ISPC Chart, For Fault Detection and Diagnosis, FDD

Noorlisa Harun*, Kamarul Asri Ibrahim

Process Control and Safety Laboratory,
Department of Chemical Engineering,
Universiti Teknologi Malaysia,
81310 Skudai, Johor.
Email: noorlisa@kuktem.edu.my

Abstract

A new approach for detecting and diagnosing fault via correlation technique is introduced in this study. The correlation coefficient is determined using multivariate analysis technique, Principal Component Analysis (PCA) to improve the traditional SPC chart. Individual charting technique such as Shewhart, Exponential Weight Moving Average (EWMA), and Moving Average and Moving Range (MAMR) charts are used to facilitate the Fault Detection and Diagnosis (FDD). A precut multi component distillation is used as the case study in this work. Based on the result from this study Shewhart control chart gives the best performance with the highest FDD efficiency.

Keywords: Fault Detection and Diagnosis (FDD), Shewhart chart, Exponential Weight Moving Average (EWMA) chart, Moving Average and Moving Range (MAMR) chart, Principal Component Analysis (PCA)

1 Introduction

Malfunction of plant equipment, instrumentation and degradation in process operation increase the operating costs of any process industries. Venkat, et al., (2003) mentioned that the petrochemical industry annually losses approximately $20 billion due to poor management in abnormal detection events. Chen, et al., (2004) also highlighted that the US-based petrochemical industry could save up to $10 billion annually if abnormal process behavior could be detected, diagnosed and appropriately dealt with. Therefore, effective monitoring strategy for early fault detection and diagnosis is very important not only from a safety and cost viewpoints, but also for the maintenance of yield and the product quality in a process. Fault detection is to determine the occurrence of an abnormal event in a process, and that of fault diagnosis is to identify its reason or sources.

Statistical Process Control, SPC is an alternative approach in chemical process to detect and diagnose fault. The major benefits of this approach are that there is no need for a fundamental or causal model of the system. In chemical processes, data based approaches rather than model-based approaches have been widely used for process monitoring, because it is often difficult to develop detailed physical models (Kano, et al., 2000). SPC only requires a good database of normal historical data, and the models are quickly and easily built from this.

Individual Shewhart, Exponential moving average (EWMA), and Moving Average (MA) and Moving Range (MR) are Statistical Process Control, SPC charts. These traditional SPC chart are used to monitor and verify that the process remained in statistical control based on small number of variables. Normally the fault in the process is seek through the usage of SPC chart, i.e., the final product quality variables. Measuring quality variables alone are not enough to describe the process performance (Kourti, et al., 1996). In this study, the quality variables and some of the process variables that having a correlation between the former and the later variables are monitored. In traditional SPC, once the quality variables showed out of statistical control signal, it is then left up to process operators and engineers to try to diagnose the cause of out of control using their process knowledge and a one at a time inspection of process variables (MacGregor and Kourti, 1995).

To overcome this limitation, a new approach using multivariate analysis is applied in the SPC realm procedure to detect and diagnose the faulty condition. This new approach is called Improved SPC, ISPC. The correlation coefficient calculated from multivariate analysis technique is applied to improve the traditional

control chart rather than using Hotelling's T^2 and Q statistics which has been widely used in multivariate statistical process control (Wise and Gallegher, 1996; MacGregor and Kourti, 1995; Jackson, 1991; Kourti and MacGregor, 1996; Kresta, et al., 1991).

2 Data generation

Figure 1 shows the schematic diagram of dynamic simulated distillation column developed by Yee and Ibrahim (2003) that is used in this case study. The monitoring purpose of this column is to maintain the composition of oleic acid and linoleic acid at the range of 0.134 to 0.135 mole fraction and 0.024 to 0.025 mole fraction respectively. There are seven control loops installed in the study column. These control loops controlled the process by counteracts the effects of disturbances change and set point change to maintain the control variables at the set point values. This concept known as Automatic Process Control (APC), which uses continuous adjustment on manipulated variables to keep the process on target.

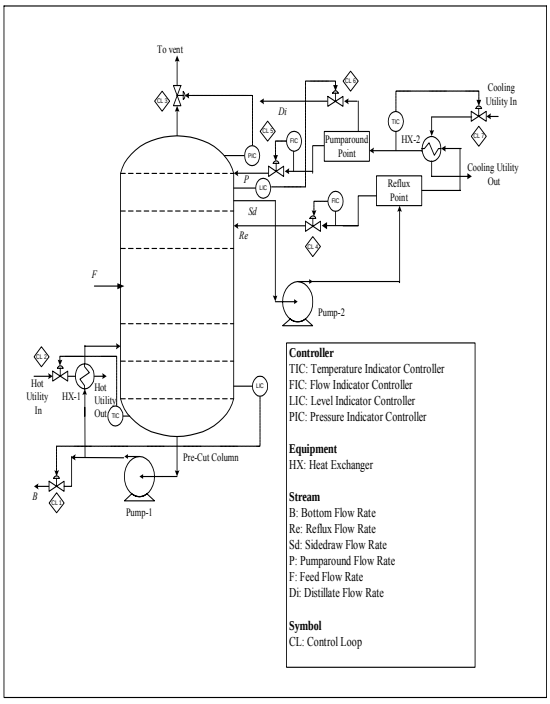


Figure1 *Schematic diagram of distillation column*

In contrast, SPC does not control the process but rather performs a monitoring function and gives signals when control is needed in the form of identification and removal the root causes. SPC accomplish the deviation in the process by detecting the changes in monitoring variables when the assignable causes occur in the process. It is important to note that the implementation of the SPC method is to detect and diagnose fault in the process such as line blockage, line leakage, sensor fault, valve fault and controller fault. This means that the controller in APC system are ready to counteracts the effect of disturbance change and set point change as long as the controller is function well.

SPC variable can be categorized into two i.e quality variables and process variables. Quality variables consist of oleic acid composition, x_{c8} and linoleic acid composition, x_{c9} at bottom stream. The process variables were selected from five locations i.e feed stream, pumparound stream, reflux stream and bottom stream. Feed flowrate and feed temperature of the study column are overall disturbances.Eventhough the study precut column has been installed with controllers, the probability of faults to occur is possible due to line blockage, line leakage and sensor fault. Pumparound flowrate and reflux flowrate are the controlled variables in APC system. These variables are controlled to maintain at the target values. The controllers will take action if any changes due to measured disturbances and set points. Both variables are set on the target values as long as the controllers are function well. The probability of faults to occur is possible due to line blockage, line leakage, sensor faults, valve sticking and controller fault. Therefore, these variables are included in monitoring process to identify the causes of faults. Reboiler heat duty of the study column does not measured directly and the value is calculated using Equation 1,

$$Qr = F_{hot}C_p\Delta T \qquad 1$$

where,

F_{hot} = Hot oil flowrate
= 8267 kg/hr
C_p = Specific heat capacity
= 7.5 kJ/kg.K
ΔT = Temperature difference
= 40 K

Hot oil flowrate is the manipulated variable in APC to maintain the bottom temperature at the desired value. This manipulated variable will change if the controller takes action to compensate the effect of disturbance (liquid flowrate at tray 28) change and set point (bottom temperature) change. Otherwise it is due to faults such as line blockage, line leakage, sensor faults, valve sticking and controller fault. Table 1 lists the selected process variables to be monitored in this study.

Table 1 *SPC variables*

Quality variable	Process variable
Oleic acid composition (x_{C8})	Feed flowrate, F
Linoleic acid composition (x_{C9})	Feed temperature, T_f
	Pumparound flowrate, P
	Reflux flowrate, Re
	Reboiler duty, Q_r

Two sets of data i.e Normal Operating Condition, NOC data and Out of Control, OC data is generated using simulated distillation column. The NOC condition exists when the process or quality variables remain close to their desired values or in a state of statistical control. In contrast, OC occurs when fault appears in the process. In general, fault is deviations from the normal operating behavior in the plant that are not due to disturbance and set point changes in the process.

NOC data that consist of quality variables and process variables were generated and arranged in the matrix form, **X** when the process is in statistical control or quality variables remain close to their desired values. The matrix data, **X** with m observations on p variables can be written as,

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mp} \end{bmatrix} \qquad 2$$

NOC data is very important in SPC methodology since it is used to predict the future behavior of the process. Some noises were imbedded into the process variables using Matlab simulator to create random process data with normally distributed. After NOC step done, faulty condition was introduced in the process by inserting deviations in the process variables and OC data was collected during this condition.

Both NOC and OC data are standardized before further analysis since the variables have different units and wide range of data measurements. Each variable is adjusted to zero mean by subtracting off the original mean of each column and adjusted to unit variance by dividing each column by its standard deviation. After the standardization, each variable have equal weights with zero mean and standard deviation equal to one (N (0, 1)). The linear relationship between quality variables and process variables is developed using multivariate analysis techniques, PCA during normal process operation. This relationship is interpreted in terms of correlation coefficient, C_{ik} which is used to diagnose the cause of the fault during the OC situation.

3 Correlation coefficient via Principal Component Analysis, PCA

Algebraically, PCA relies on *eigenvector* decomposition of the covariance or correlation matrix of the process variables. A set of original variables x_1, x_2, …, x_p is transform to a set of new variables PC_1, PC_2, …, PCp. The mathematical representations that describe the transformation of a data matrix, **X** (m,p) consist of m observations on p variables is shown as follow,

$$\mathbf{PC}_{p,m} = \mathbf{V}_{p,p}\mathbf{X}^{T}_{p,m} \qquad 3$$

V is the eigenvector matrix, which consists of *eigenvector* $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_p}$. Singular Value Decomposition (SVD) technique is used to decompose data matrix, X (m,p) into a product of the *eigenvectors* of $\mathbf{XX}^{\mathrm{T}}$, the *eigenvectors* of $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ and a function of their *eigenvalue*. The fundamental identity of SVD is shown by the following equation,

$$\mathbf{X}_{(m,p)} = \mathbf{U}_{(m,p)}\mathbf{L}^{1/2}{}_{(p,p)}\mathbf{V}^{\mathrm{T}}{}_{(p,p)} \qquad 4$$

The diagonal elements of **L** (p,p), $\lambda_1, \lambda_2, \ldots, \lambda_p$ are called *eigenvalues* of **X** while the columns vector of **U** (m,p) , $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p$ and the columns vector of **V** (p,p), $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ are called *eigenvectors* of **X** and both *eigenvector* are orthonormal. Matrices of **U**, **V** and $\mathbf{L}^{1/2}$ have the following properties (Nash and Lefkovitch, 1976):

$$\mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{U}\mathbf{U}^{\mathrm{T}} = \mathbf{I} \qquad 5$$

$$\mathbf{V}^{\mathrm{T}}\mathbf{V} = \mathbf{V}\mathbf{V}^{\mathrm{T}} = \mathbf{I} \qquad 6$$

$$(\mathbf{L}^{1/2})(\mathbf{L}^{1/2})^{\mathrm{T}} = (\mathbf{L}^{1/2})^{\mathrm{T}}(\mathbf{L}^{1/2}) = \mathbf{L} \qquad 7$$

The correlation between the variable k, x_k and variable i, x_i if j variables are involved can be written as following,

$$C_{ik} = \sum_{j=1}^{n} v_{ij} v_{kj} \lambda_j \qquad 8$$

n = Number of retained *eigenvectors*

v_k = *Eigenvectors* of process variable

v_i = *Eigenvectors* of quality variable

The number of retained eigenvector can be determined using Scree plot (Ralston, et al., 2001). Detailed derivation of this equation can be found in (Loong and Ibrahim, 2002).

4 Improved SPC, ISPC charts

The function of ISPC chart is to compare the current state of the process against Normal Operating Condition, NOC. ISPC chart is used to determine whether the process in a state of statistical control or in a state of out of statistical control using a set of historical data. There are two types of causes, which contribute to the existing of faults in the process. Chance, or common causes are small random changes in the process that cannot be avoided. Variation of this type is only removable by making a change in the existing process. Assignable causes, on the other hand, are large variation in the process that can be identified as having specified cause. Assignable causes are causes that are not part of the process on a regular basis. This type of variation arises because of specific circumstances. Sources of variation can be found in the process itself, the material used, the operator's actions, or the environment.

Control charts approach is based on the assumption that a process subject to common cause variation will remain in a state of statistical control under which process remain close to target which is known as NOC data for this study. By monitoring the process over time, OC events known as assignable cause can be detected as soon as they occur. If the causes for such events can be diagnosed and the problem can be corrected, the process is driven back to its normal operation.

Correlation coefficient from the multivariate analysis technique that is PCA, is used to relate the quality variables with the process variables. Let x_k is the quality variable and x_i is the process variable. The relationship between standardized quality variable, x_k^s and standardized process variable, x_i^s can be written as,

$$x_k^s = C_{ik} x_i^s \qquad 9$$

where $x_k^s = (x_k - \bar{x}_k)/s_k$, $x_i^s = (x_i - \bar{x}_i)/s_i$. s is the standard deviation, while $\bar{x}_k$ and $\bar{x}_i$ is quality variable mean and process variable mean respectively. The control limit for quality variable in general is

$$LCL < x_k^s < UCL \qquad 10$$

where UCL and LCL is upper control limit and lower control limit respectively. Substitute equation 9 into equation 10 and rearrange the equation 10. The control limits for corresponding process variable is

$$LCL/C_{ik} < x_i^s < UCL/\ C_{ik} \qquad 11$$

Equation 11 is used to calculate the control limits for process variables. These limits are calculated based on the NOC data. Figure 1 shows the used of correlation coefficient corresponding to process variables.

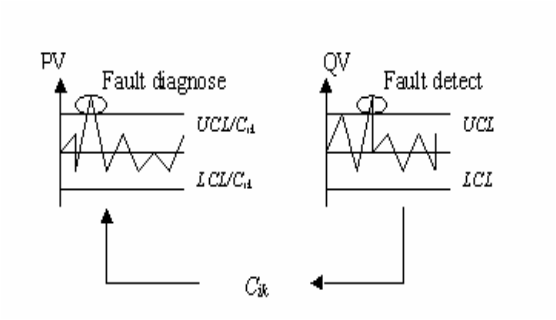


Figure 2 *The implementation of correlation coefficient, C_{jk} in ISPC chart*

The quality variable control limits and process variables control limits for each ISPC chart are accordingly presented in Table 2 and Table 3.

Table 2 *Quality variable control limits*

Control chart	Control limit
Shewhart Individual	$UCL = 3s, LCL = -3s$
Shewhart Range	$UCL = D'_{.001}\ \bar{R}$, $LCL = D'_{.999}\ \bar{R}$
EWMA	$UCL = +Ls\sqrt{\frac{\lambda}{2-\lambda}}\ [1-(1-\lambda)^{2i}]$ $UCL = -Ls\sqrt{\frac{\lambda}{2-\lambda}}\ [1-(1-\lambda)^{2i}]$
MA	$UCL = +A_2 \bar{R}$, $LCL = -A_2 \bar{R}$
MR	$UCL = D'_{.001}\ \bar{R}$, $LCL = D'_{.999}\ \bar{R}$

Table 3 *Process variable control limits*

Control chart	Control limit
Shewhart Individual	$UCL = 3s/C_{ik}$, $LCL = -3s/C_{ik}$
Shewhart Range	$UCL=D'_{.001}\ \bar{R}\ /C_{ik,}$ $LCL=D'_{.999}\ \bar{R}\ /C_{ik}$
EWMA	$UCL=+(Ls\sqrt{\frac{\lambda}{2-\lambda}}\ [1-(1-\lambda)^{2i}])/C_{ik}$ $UCL=-(Ls\sqrt{\frac{\lambda}{2-\lambda}}\ [1-(1-\lambda)^{2i}])/C_{ik}$
MA	$UCL=+A_2\ \bar{R}\ /C_{ik,}\ LCL=-A_2\ \bar{R}\ /C_{ik}$
MR	$UCL=D'_{.001}\ \bar{R}\ /C_{ik,}$ $LCL=D'_{.999}\ \bar{R}\ /C_{ik}$

Table 4 shows the equation to determine the statistical data for each control charts.

Table 4 *Statistical data for control charts*

Control chart	Chart Statistics
Shewhart Individual	$x = x_i$
Shewhart Range	$R_i = max\ [x_{i-1+1}] - min\ [x_{i-1+1}]$
EWMA	$z_i=\lambda x_i + (1-\lambda)z_{i-1}$
MA	$MA_i =(x_i + x_{i-1} + \ldots + x_{i-w+1})/w$
MR	$MR_i = max\ [x_{i-w+1}] - min\ [x_{i-w+1}]$

where

$\bar{R}$ = Average of range

λ =Weighting factor

A_2 , $D'_{.001}$, $D'_{.999}$ = Constant

L = Width of the control limit

z = EWMA statistic

w = window size for moving chart

The correlation coefficient is used to translate the control limits of SPC charts from quality variables into process variables, which is used to perform fault diagnosis for the process operation. The quality variables data is incorporate in the control chart during the faulty condition for fault detection purpose, while the process variables which has been correlate with the quality variables is used for fault diagnosis. Therefore, the Improve SPC charts applied not only for quality variables but also for process variables.

5 FDD efficiency using Improved SPC, ISPC chart

The efficiency of the FDD method using improved Shewhart, EWMA and MAMR chart is evaluated based on two aspects i.e the successful of ISPC chart to detect the fault and the successful of ISPC chart to identify the correct process variable as fault cause for each fault situation. The efficiency of fault detection, $\eta_{Fdetect}$ and the efficiency of the fault diagnosis, $\eta_{FDiagnose}$ is determined using the following equation,

$\eta_{Fdetect}$= [Number of faults detected / Total of faults generated in the process] x 100 10

$\eta_{FDiagnose}$ = [Number of faults diagnosed / Total of diagnosed fault] x 100 11

80 fault locations consist of 50 single fault and 30 multiple faults were introduced into the process. Figure 3 and figure 4 show the efficiency of FDD on quality variables, oleic acid, C8 and linoleic acid, C9 respectively using different ISPC charts. Shewhart chart give 100% performance in fault detection, which is better than EWMA and MAMR for both quality variables. Shewhart chart used 100% current data but MAMR statistic is calculated using window size of four, which consist of 25% current data and 75% previous data. This caused detection delay using MAMR. EWMA statistic exponential weighted average of all prior data, including the most recent data. The weighted average depends on weighting factor, λ. Small λ will give less weight to current data and more weight to previous data and vice versa. In this study, λ = 0.4 is used. This give the EWMA statistic consist of 40% current data 60% previous data. This caused EWMA is more efficient in detecting shift in the process compared to MAMR since EWMA statistic gives more weight to current data.

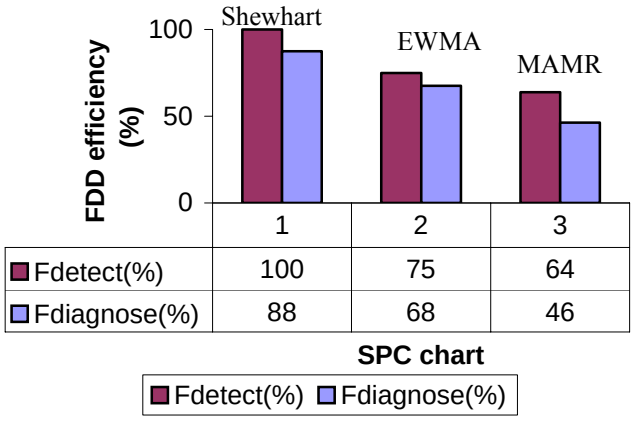


Figure 3 Oleic acid *FDD efficiency*

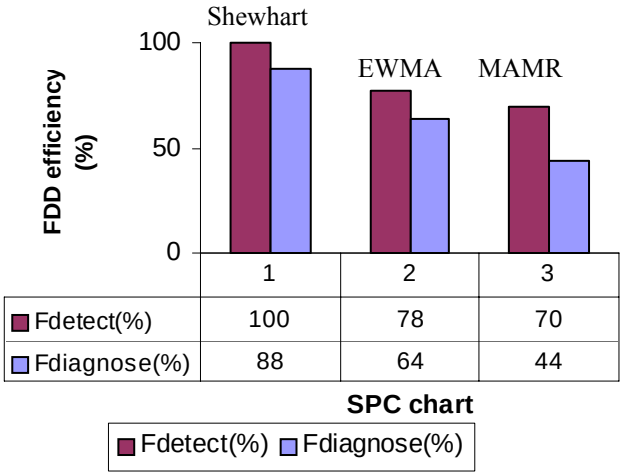


Figure 4 *Linoleic acid FDD efficiency*

Figure 5 and figure 6 show the fault detection efficiency using Shewhart, EWMA and MAMR in three different regions based on single fault data. The OC data that is greater than $\pm 3s$ is divided into three regions. Region 1, region 2, and region 3 refer to mean $\pm 4s$, mean $\pm 5s$ and over mean $\pm 5s$ respectively. Refer to figure 5, both Shewhart chart and EWMA can detect OC data in region 1 but Shewhart (100% fault detection efficiency) give better performance than EWMA (9.1% fault detection efficiency). MAMR chart only can detect deviation that is greater than mean $\pm 4s$ for oleic acid while one fault cases is detected by MAMR for linoleic acid.

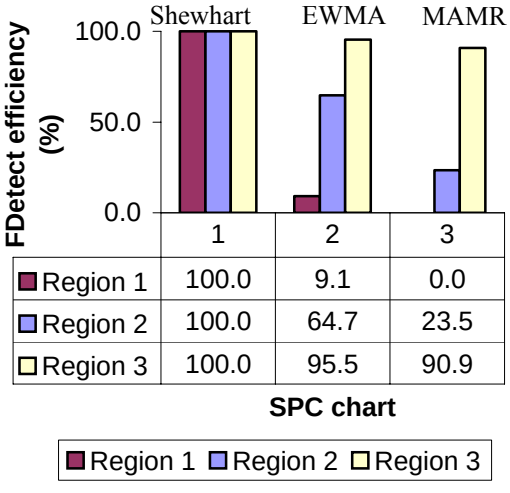


Figure 5 *Oleic acid FDD efficiency based on different region*

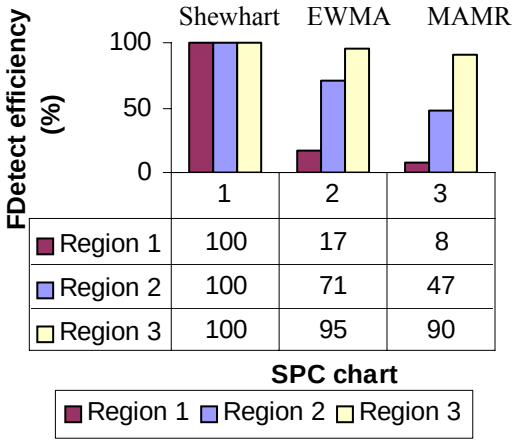


Figure 6 *Linoleic acid FDD efficiency based on different region*

Based on fault diagnosis result (not shown in the figure), the smallest deviation (0.24%) from the target in process variables, which contribute to the out of control situation in the quality variables, can be identified by shewhart. EWMA can diagnose for 0.58% deviation in process variable while MAMR can diagnose for 0.86% deviation in process variables. The result obtained shows that Shewhart chart was able to detect small shift in the process. EWMA chart can be used to detect moderate shift in the process while MAMR for large shift in the process.

Conclusion

Early fault detection and diagnosis is important in chemical industries for safety, maintaining product quality and reduce the cost. The potential of ISPC to detect and diagnose faults in simulated distillation column is shown. PCA technique is used to develop the correlation between quality variables and process variables in order to improved the traditional SPC chart for FDD. One major advantage of the correlation coefficients method is process monitoring for fault diagnosis can be done using process variables. Process fault diagnosis can be done in straightforward manner. The simplicity of the presentation and interpretation of the ISPC charts based on multivariate analysis technique makes these charts attractive to the plant engineers and operators to identify the out of statistical control in the process. Performance of Shewhart chart which used 100% current data is the best compared to EWMA and MAMR in detecting and diagnosing faults. Both EWMA and MAMR chart incorporate previous data in calculating the control limits. FDD result using EWMA

is better than MAMR to detect moderate shift in the process.

References

[1] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S. N. Kavuri, A review of process fault detection and diagnosis Part I: Quantitative model-based methods, *Computers and Chemical Engineering,* vol. 27, pp. 293-311, 2003.

[2] Q. Chen, U. Kruger, M. Meronk and A.Y.T. Leung, Synthesis of T^2 and Q statistic for process monitoring, *Control Engineering Practice*, vol. 12, pp. 745-755, 2004.

[3] M. Kano, K. Nagao, S. Hasebe, I. Hashimoto, H. Ohno, R. Strauss and B. R. Bakshi, Comparison of statistical process monitoring methods: application to the Eastman challenge problem, *Computers and Chemical Engineering*, vol. 24, pp.175-181, 2000.

[4] T. Kourti, J. Lee and J. F. Macgregor, Experiences with industrial applications of projection methods for multivariate statistical process control, *Computers chemical engineering*, vol. 20, pp. 745-750, 1996.

[5] J. F. MacGregor and T. Kourti, Statistical process control of multivariate processes, *Control engineering practice*, vol. 3, no. 3, 403-414, 1995.

[6] B. M. Wise, and N. B. Gallagher, The process chemometrics approach to process monitoring and fault detection, *Journal of Process Control*, vol. 6, 329-348, 1996.

[7] J. E. Jackson, *A User's Guide to Principal Components*, John Wiley & Sons, New York, 1991.

[8] T. Kourti and J.F. McGregor, Multivariate SPC methods for process and product monitoring, *Journal of Quality Technology*, vol. 28, no. 4, 409–428, 1996.

[9] J. Kresta, J.F MacGregor and T.E. Marlin, Multivariate statistical monitoring of process operating performance, *Canada journal chemical engineering*, vol. 69, 35-47, 1991.

[10] M. W. Yee and K. A, Ibrahim, "Fault Detection for Distillation Column Using Multivariate Statistical Process Control (MSPC)", *Proceedings of The 17th Symposium of Malaysian Chemical Engineers (SOMChE 2003)*, pp. 92-9, 2003.

[11] J. C. Nash and L. P. Lefkovitch, Principal component and regression by SVD on a small computer, *Apply Statistics*, vol. 25, no. 3, 210-216, 1976.

[12] Ralston, P., DePuy, G., and Graham, H., Computer-based monitoring and fault diagnosis: a chemical process case study, *ISA transaction*, vol. 40, 85-98, 2001.

[13] Loong, L. H. and K. A, Ibrahim, "Improved Multivariable Statistical Process Control (MSPC) for Chemical Process Fault Detection and Diagnosis (PFDD)–Cross-Variable Correlation Approach", *Regional Symposium on Chemical Engineering*, vol. 2, pp. 1637-1644, 2002.

Fault detection and diagnosis using Multivariate Statistical Process Control (MSPC)

Mak Weng Yee[1], Kamarul Asri Ibrahim[1]

[1]Faculty of Chemical Engineering and Natural Resources Engineering (FKKKSA), Universiti Teknologi Malaysia, 81310 Skudai, Johor Bahru, Johor.

Introduction

Currently, chemical plants face numerous challenges like stringent requirements are needed on the desired final product quality, utilization of a lot of energy, must be environmentally friendly and fulfill safety requirements. High operation cost is needed in order for chemical plants to overcome the stated challenges. Any faults that are present in a chemical process will yield higher operation cost on the plant due to increase in production of waste, re-work, re-processing and consumption of utilities. Therefore, accurate process fault detection and diagnosis (FDD) on a chemical process at an early stage is important to reduce the cost of operation due to present of faults.

The important task of detecting and diagnosing abnormal process behavior (faults) has led to the evolution of a range of statistically based condition monitoring approaches (Treasure *et al*., 2004). These approaches are collectively referred to as Multivariate Statistical Process Control (MSPC) and have gained attention over the past decades noticeable by the large number of publications in this area (MacGregor and Kourti, 1995). Application of MSPC as a fault detection tool in previous works was based on two conventional control chart: Hotelling's T^2 Statistic control chart and Square Prediction Error Statistic control chart (SPE) (Wachs and Lewin, 1999). These two control charts have shown good fault detection performance for simulated model unit operations (Wachs and Lewin, 1999). MSPC using the two stated conventional control charts is a very powerful tool for fault detection but its main limitation lies in the ability to isolate or diagnose the actual causes of the detected faults. The main fault diagnosis tool used together with the two control charts is the Contribution Plots (CP) (Wachs and Lewin, 1999). Although CP is used to diagnose the cause of the detected faults, they tend to be noisy and ambiguous. These plots also do not have confidence limit/control limit, thus making it difficult to determine whether a situation is normal or abnormal (Yoon and MacGregor, 2000).

The present fault diagnosis tool using CP has limited usage in diagnosing causes of detected faults. Faults that have effect propagated into other variables are hard to be isolated using CP. In enhancing the fault isolation ability of MSPC and overcoming the ambiguity of CP, fault signatures have been proposed. Faults from process data are collected and fault signatures are developed using Principal Component Analysis (PCA). Any new detected faults will exhibit certain fault signature and this signature will be compared to the database of fault signatures developed earlier on. Good results were obtained for the application of the proposed method (Yoon and MacGregor, 2001). Although the fault signature method shown better fault diagnosis ability compared to the previous Contribution Plots, there are several weaknesses of the former method. The fault signature database needs to be as comprehensive as possible to cover all possible faults in a process and great amount of computer calculation is needed in diagnosing a fault for highly multivariable processes. The present work focuses on overcoming the ambiguity nature of fault isolation using MSPC through contribution plots and also the need for big database of faults signatures by introducing fault diagnosis using correlation coefficients of process variables and quality variables. The proposed FDD method in this paper is an extension of fault detection using correlation coefficients (Mak and Kamarul, 2003).

Correlation coefficients between key process variables and quality variables of interest are used as fault detection and diagnosis tools. These coefficients are developed from nominal operating condition (NOC) data using multivariate projection techniques such as PCA and Partial Correlation Analysis (PCorrA). PCorrA has been applied in many applications (Ding and Nancy, 2000) and hardly been used in MSPC as a method for determining correlation

between variables. The developed correlation coefficients will be used together with conventional Shewhart Control Chart and Range Control Chart as FDD tools. The proposed method is applied to a simulated industrial column model (Wong, 2003).

Methodology

Process modeling and data generation

The most important part in obtaining an accurate correlation between the process variables and quality variables is the data mining section. In this research, data is obtained from a simulation model. A distillation column from a Palm Oil Fractionation Plant is selected as the case study. The model of this column is developed based on the model from literature with slight modifications to suit the present work (Wong, 2003). Figure 1 shows the distillation column with the key variables of the process. From the column model, two sets of process operating data were generated. For NOC data, some noises with zero mean were imbedded into the simulation program. The noises considered are small random change in selected key variables such as feed flow rate, feed temperature, reboiler duty, cooler duty, reflux flow rate and pumparound flow rate. While for Out-of-Control (OC) data, some large changes (significant faults) and moderate changes (insignificant faults) were purposely added into the process model as faults. These faults represent valve faults, sensor faults and controller faults. The description of each type of fault is described in Table1. The feed flow rate and feed temperature to the study column are assumed to be fixed. Any abnormal changes of the value of these two variables are due to faults as shown in Table 1 and not due to common cause variation (NOC). The generated NOC and OC data are mean-centered and variance scaled. The NOC data will be subjected to analysis using PCA and PCorrA for deriving the correlation coefficients between the selected process variables with the selected quality variables. The two quality variables of interest in this research are the oleic acid mole fraction, x_8, and linoleic acid mole fraction, x_9, in the bottom flow rate. The objective of the proposed FDD tools is to maintain the value of these two variables at their steady-state value through detection and diagnosis of faults present in the process.

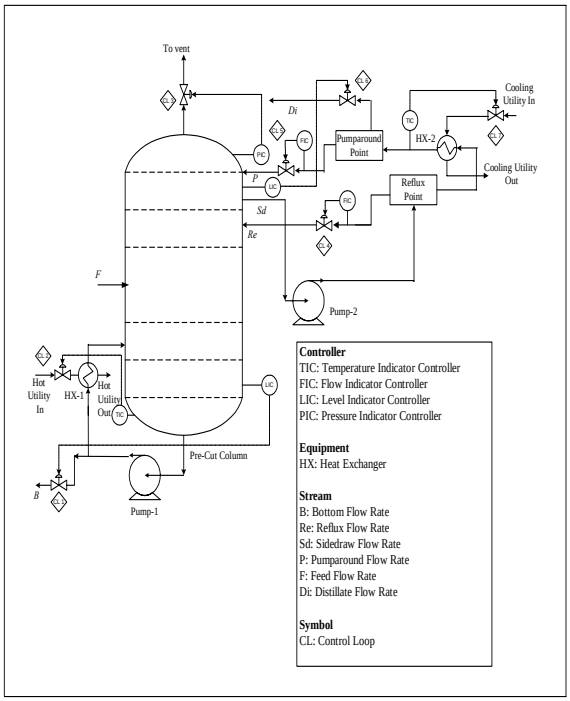


Figure 1 Distillation column model

Derivation of correlation coefficients

After the NOC data are obtained, the correlation coefficients between the selected key process variables and the quality variables of interest are determined using PCA and PCorrA. Method for obtaining correlation coefficients between the variables, C_{ik}, using PCA was based on previous PCA work (Lam and Kamarul, 2002). Correlation coefficients using PCA are calculated as in Equation 1.

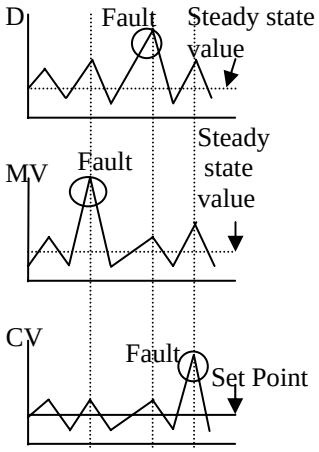
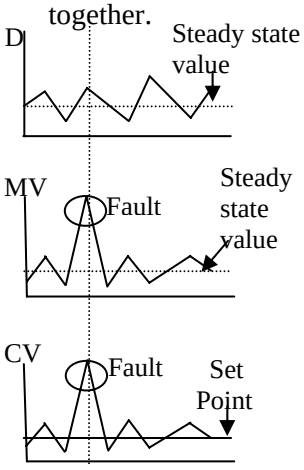
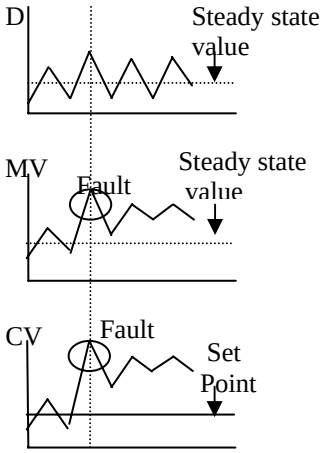
$$C_{ik} = \sum_{j=1}^{n} v_{ij} v_{kj} \lambda_j \qquad \text{(Eq.1)}$$

Where:

v_{ij}, v_{kj} = eigenvectors obtained from process data using PCA

λ_j = eigenvalue obtained from process data using PCA

Table 1 Fault Descriptions

Sensor Fault	Valve Fault	Controller Fault
-For open loop variables, only the value of the variable changes abnormally. For closed loop variables, only the value of the disturbance (D) OR the manipulated variable (MV) OR the control variable (CV) changes abnormally. 	-For open loop variables, only the value of the variable changes abnormally. For closed loop variables, both manipulated variable (MV) AND control variable (CV) changes abnormally together. 	-For closed loop variables, the value of manipulated variable (MV) AND control variable (CV) changes abnormally together. 

PCorrA determines the correlation between two variables while allowing the effect of other correlated variables on these two variables. For calculating correlation coefficient, C_{ik}, for variable 1 and 2 using PCorrA after allowing the effect of j-2 variables is as shown in Equation 2 (Cliff and Ord, 1973).

$$C_{ik_{12}} = \frac{r_{12.(4,\ldots,j-2)} - r_{13.(4,\ldots,j-2)} r_{23.(4,\ldots,j-2)}}{(1-r^2_{13.(4,\ldots,j-2)})^{1/2}(1-r^2_{23.(4,\ldots,j-2)})^{1/2}} \quad \text{(Eq.2)}$$

Where:

r_{12} = correlation between variable 1 and 2

$r_{12.3}$ = partial correlation between variable 1 and 2 after the effect of variable 3

$r_{12.(3,4,\ldots,j\text{-}1)}$ = partial correlation between variable 1 and 2 after the effect of j-2 variables

Development of FDD Tools

C_{ik} relates a process variable, x_i, with a quality variable, y_i, in the following way:

$$x_i = \frac{y_i}{C_{ik}} \quad \text{(Eq.3)}$$

For conventional Shewhart Control Chart, the Upper Control Limit (UCL), Center Line (CL) and Lower Control Limit (LCL) for mean-centered and variance-scaled variables are +3, 0 and -3 respectively (McNeese and Klein, 1991). Using the information from Equation 3, the UCL, CL and LCL for quality variables and process variables will be +3, 0 and -3 and +3/C_{ik}, 0 and -3/C_{ik} respectively. After the NOC control charts are established, they are used for fault detection of the OC data.

The UCL, CL and LCL for conventional Range Control Chart for mean-centered and variance-scaled variables are mean of the range values, R_{mean} multiplied by a constant, d_2, R_{mean} and 0 respectively (McNeese and

Klein, 1991). The constant, d_2, is determined by the number of subgroup used in calculating the range values. In the present work, d_2 is 3.267 for a subgroup, n = 2 (McNeese and Klein, 1991). R_{mean} is determined as shown in Equation 4.

$$R_{mean} = \frac{\sum_{i=1}^{n} R_i}{n} \qquad \text{(Eq.4)}$$

Where:
R_i = i-th Range value
R_{mean} = mean of the range values
n = number of range values

For the present work, the UCL, CL and LCL for the Range Control Chart of quality variables will be of the conventional Range Control Chart. For the selected process variables, the UCL, CL and LCL will be $(R_{mean}* d_2)/C_{ik}$, $(R_{mean})/C_{ik}$ and 0 respectively.

The major assumption in the proposed method is that all key process variables are measured. The process variables that are major contributors to the variation of the process are included into the correlation analysis. In this way, the behavior of the process will be well represented by the correlation determined from the selected key process variables and the developed fault detection and diagnosis method will suit the dynamic behavior of the process. From Figure 1, the study column is installed with several control loops to ensure the stable operation of the column. Any common cause changes in the column either through load problem (disturbance changes) or servo problem (set point changes) will be taken care of through these controllers. The causal cause changes of interest in this work are those involving abnormal changes in the values of the variables of the process not through the two mentioned problems rather through faults in sensors, valves or even controllers. For NOC data, only common cause variation is present in the process. While for OC data, the observed causal cause variation is caused by faulty operation of the process sensors, valves and controllers.

When a process variable changed from its normal steady-state value, the variable of that control chart will be checked whether it is a closed loop variable or open loop variable. A fault signal is observed only when either the Range Control Chart or Shewhart Control Chart of one or more quality variable show value that exceeds its control limit AND one or more process variable observed a value out of its control limit either in its Shewhart Control Chart or Range Control Chart. For open loop variable, the fault will be of sensor fault or valve fault as pre-designed while fault for closed loop variable can be of valve fault, sensor fault or controller fault. The cause variable(s) of each detected fault is diagnosed by checking the control charts of the process variables. Process variables that show value exceeding its control limit (either in Shewhart Control Chart or Range Control Chart) are diagnosed as the cause of the observed fault. To determine which type of fault is detected, the method used is as the previous paragraph.

Results and Discussions

Figure 2 shows an example of the fault detection and diagnosis using the proposed method based on PCA. For the PCorrA method, a similar plot of graphs will be observed as well. Due to space limitation, only the Shewhart Control Chart for the 6 selected key process variables (feed flow rate (*Lf*), feed temperature (*Tf*), reflux flow rate (*Re*), pumparound flow rate (*P*), reboiler duty (*Qr*) and bottom column temperature (*Tbot*)) and quality variable 1 (oleic acid mole fraction in the bottom flow rate (x_8) were shown in Figure 2. Similar results will also be observed through the Range Control Chart of these variables. The performance of PCA and PCorrA in detecting the faults and diagnosis the cause of each detected fault is shown in Figure 3 and Figure 4.

Both methods based on PCA and PCorrA were able to diagnose the cause of each fault detected. Out of the 17 faults in the fault data, 13 faults (both single fault and multiple faults) were successfully detected by the PCA method (Using data reduction with 95% of the variation of the original data retained). The 4 faults that were not detected by the PCA method were insignificant faults (moderate changes in the values of the process variables). The method based on PCorrA performed better than the PCA method by successfully detecting all the 17 pre-designed faults (both significant faults and insignificant faults). The PCorrA method performed better because the correlation coefficients developed by this method are closer to the actual value of the correlation coefficients representing the

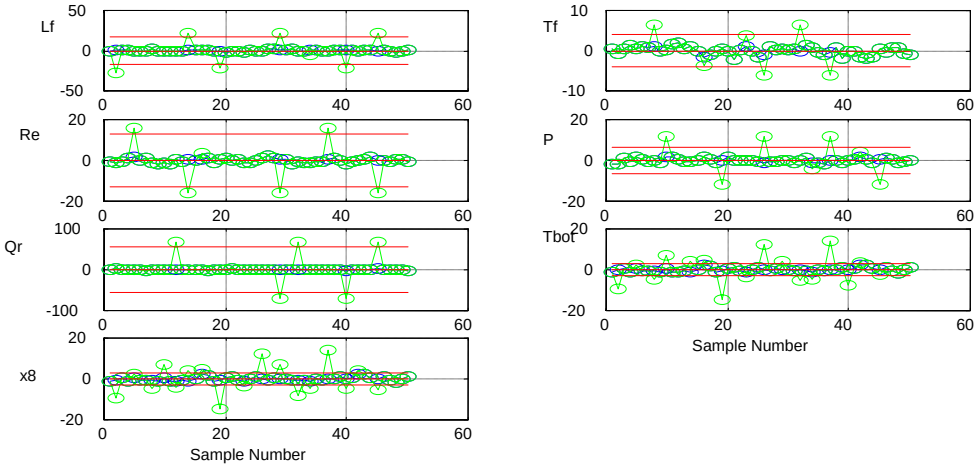


Figure 2 Example of Fault Detection and Diagnosis based on PCA

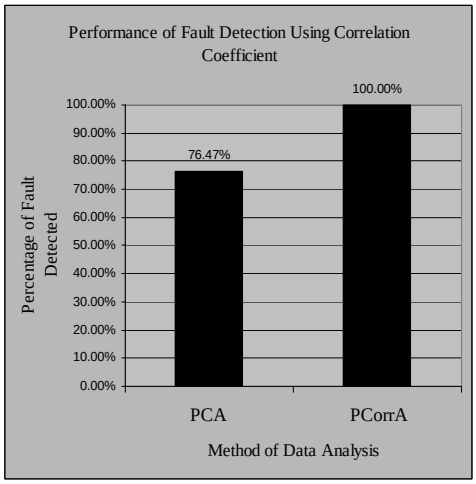


Figure 3 Performance of fault detection using correlation coefficient

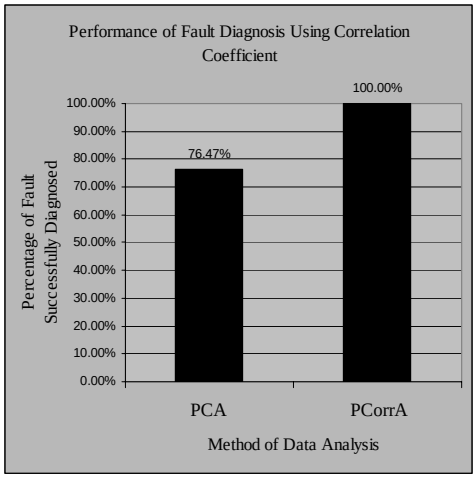


Figure 4 Performance of fault diagnosis using correlation coefficient

correlation between the selected process variables with the quality variables of interest. This is because the PCorrA method sets other selected process variables at constant values when calculating the correlation between a selected process variable with a quality variable. The PCA method calculates the cross-correlation between variables (interaction between variables) when determining the correlation coefficients between the process variables and quality variables. However, the PCorrA method was superior in determining the correlation between variables judging from the observed fault detection and diagnosis results of the study column.

One major advantage of the correlation coefficients method is the simplicity in determining the fault cause(s) of a detected fault. The control charts of the selected process variables will trigger alarm if any of them exhibit value out of their control limits and the charts that triggers an alarm will be determined as the root causes of the detected fault. Furthermore, the availability of control limits in these control charts will shed away any ambiguities of whether a change in value of the selected process variables are due to common cause (NOC) or causal cause (OC). For online process monitoring, the data that are used for calculating the correlation coefficients can be updated with dynamic data to take account into the changes of the process due to change in raw material, fouling in heat exchangers and other changes in the process parameters. This area can be further

researched and are a research problem for future work. The application of the developed FDD tools on a multiple unit operation case study is also a research work for the future.

Conclusion

An approach for fault detection and diagnosis using correlation coefficients based on PCorrA and PCA was presented. The performance of the approach was studied on an industrial distillation column. The results show that the fault detection and diagnosis method using cross correlation coefficient was able to detect the faults and diagnose the fault cause of each detected fault (both single fault cause and multiple fault causes). Although both methods based on PCA and PCorrA were successful in diagnosing the cause of each fault detected, PCorrA managed to detect all the pre-designed faults (both significant faults and insignificant faults) while PCA only managed to detect the significant fault. This is due to the fact that PCorrA determines the correlation between two variables after taken account into the effect of other variables that are correlated with the two variables of interest. Therefore, the correlation coefficients developed using the PCorrA method was better in representing the correlation between the selected process variables and the quality variables of interest.

Acknowledgements

The present work was funded by the National Science Fellowship (NSF) scholarship. Special thanks to Mr. Wong Teck Siang for providing the basic model for the case study of this work. Mr. Lam Hon Loong's help in providing the information on the derivation of the correlation coefficient via eigenvector-eigenvalue approach for the PCA method is gratefully acknowledged.

References

Cliff, A.D. and Ord, J.K. (1973). Spatial Autocorrelations. London: Pion.

Ding, A.C. and Nancy, P.L. (2000). Partial Correlation of Fuzzy Sets. *Fuzzy Sets and Systems* Vol. 110: 209-215.

Lam, H.L. and Kamarul, A.I. (2002). Improved Multivariate Statistical Process Control for Chemical Process Fault Detection and Diagnosis via Cross Variable Correlation Coefficient Approach. *Regional Symposium on Chemical Engineering, RSCE/SOMChem 2002, Hotel Hilton, Petaling Jaya, Selangor, Malaysia, 28-30 October 2002*. Vol. 2: 1637-1644.

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

Mak, W.Y. and Kamarul, A.I. (2003). Fault Detection for Distillation Column Using Multivariate Statistical Process Control (MSPC). *Proceedings of The 17th Symposium of Malaysian Chemical Engineers (SOMChE 2003), 29-30 December 2003, Hotel Copthorne, Penang, Malaysia*: 92-96.

McNeese, W.H. and Klein, R.A. (1991). Statistical Methods for the Process Industries. USA: ASQC Quality Press.

Treasure, R.J., Kruger, U. and Cooper, J.E. (2004).Dynamic Multivariate Statistical Process Control Using Subspace Identification. *Journal of Process Control* Vol. 14: 279-292.

Wachs, A. and Lewin, D.R. (1999). Improved PCA Methods for Process Disturbance and Failure Identification. *AIChE Journal* Vol. 45: 1688-1700.

Yoon, S.K. and MacGregor, J.F. (2000). Statistical and Causal Model-Based Approach to Fault Detection and Isolation. *AIChE Journal* Vol. 46 (9): 1813-1824.

Yoon, S.K. and MacGregor, J.F. (2001). Fault Diagnosis with Multivariate Statistical Models Part I: Using Steady-state Fault Signatures. *Journal of Process Control* Vol. 11 (4): 387-400.

Wong, T.S. (2003). Advanced Process Control of A Fatty Acid Distillation Column Using Artificial Neural Networks. Unpublished Masters Thesis (Universiti Teknologi Malaysia), Malaysia.

FAULT DETECTION AND DIAGNOSIS, FDD VIA IMPROVED UNIVARIATE STATISTICAL PROCESS CONTROL CHARTS, USPC

Noorlisa Harun*, Kamarul Asri Ibrahim

Process Control and Safety Laboratory,
Department of Chemical Engineering,
Universiti Teknologi Malaysia,
81310 Skudai, Johor.
Email: noorlisa@kuktem.edu.my

ABSTRACT

A new approach for detecting and diagnosing fault via correlation technique is introduced in this study. The correlation coefficient is determined using multivariate analysis technique, Partial Correlation Analysis (PCorrA). Individual charting technique such as Shewhart, Exponential Weight Moving Average (EWMA), and Moving Average and Moving Range (MAMR) charts are used to facilitate the Fault Detection and Diagnosis (FDD). A precut multi component distillation is used as the case study in this work. Based on the result from this study Shewhart control chart gives the best performance with the highest FDD efficiency.

Keywords: Fault Detection and Diagnosis (FDD), Shewhart chart, Exponential Weight Moving Average (EWMA) chart, Moving Average and Moving Range (MAMR) chart, Partial Correlation Analysis (PcorrA)

1 INTRODUCTION

Malfunction of plant equipment, instrumentation and degradation in process operation increase the operating costs of any process industries. More serious are a gross accident such as explosion. Even major catastrophes and disasters from chemical plant failures may be infrequent, minor accidents are very common, occurring on a day to day basis, resulting in many occupational injuries, illnesses, and costing the society billions of dollars. Venkat, et al*.,* (2003) mentioned that the petrochemical industry annually losses approximately $20 billion due to poor management in abnormal detection events. Chen, et al*.,* (2004) also highlighted that the US-based petrochemical industry could save up to $10 billion annually if abnormal process behavior could be detected, diagnosed and appropriately dealt with. Therefore, effective monitoring strategy for early fault detection and diagnosis is very important not only from a safety and cost viewpoints, but also for the maintenance of yield and the product quality in a process.

Statistical Process Control, SPC is an alternative approach in chemical process to detect and diagnose fault. The major benefits of this approach are that there is no need for a fundamental or causal model of the system. In chemical processes, data based approaches rather than model-based approaches have been widely used for process monitoring, because it is often difficult to develop detailed physical models (Manabu et al**.,** 2000). SPC only requires a good database of normal historical data, and the models are quickly and easily built from this.

SPC chart is the most technically sophisticated tool to monitor the performance of any given process. The function of this control chart is to compare the current state of the process against Normal Operating Condition, NOC. The NOC condition exists when the process or product variables remain close to their desired values or in statistical control. In contrast, the Out of Control, OC occurs when fault appears in the process. In general, fault is deviations from the normal operating behavior in the plant that are not due disturbance and set point changes in the process. Fault detection is to determine the occurrence of an abnormal event in a process, and that of fault diagnosis is to identify its reason or sources.

Traditional SPC methods assume that process data is statistically independent and stationary (Nong et al*.,* 2000) and ignoring the cross correlation between the variables. This can lead to faulty interpretation

during process monitoring. To overcome this limitation, a multivariate analysis approach is applied in the USPC realm procedure to detect and diagnose the faulty condition. Multivariate analysis method that is Partial Correlation Analysis, PCorrA is used to develop the control limits of USPC charts. Ibrahim (1997) has introduced PCorrA method to be applied in chemical process data to develop Multivariate Statistical Process Control (MSPC) scheme known as Active SPC.

2 DATA GENERATION

Figure 1 shows the schematic diagram of dynamic simulated distillation column developed by Mak and Kamarul (2003) that is used in this case study. The monitoring purpose of this column is to maintain the composition of oleic acid and linoleic acid at the range of 0.134 to 0.135 mole fraction and 0.024 to 0.025 mole fraction respectively. This column is used to generate two sets of data i.e NOC data and OC data.

SPC variable is categorized into two i.e quality variables and process variables. The quality variables acted as an indicator variables to show that the process in the state of statistical control or in the state of out of control. If any of the points of these variables are fall out of control limits, it shows that the fault situation is taken place. On the other hand, the process variables used to find the causes of the fault situation. Table 1 show the list of quality variables and process variables used in this study.

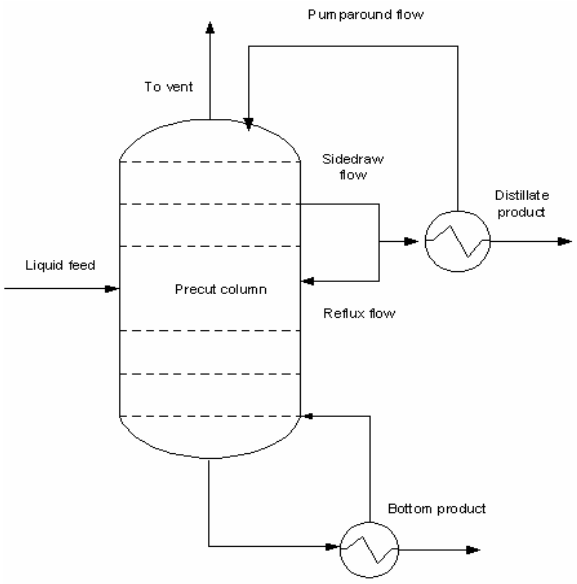


Figure1. Schematic diagram of distillation column

Table 1. SPC variables

Quality variable	Process variable
$xC8$	F
$xC9$	T_f
	P
	Re
	Q_r

where

F = Feed flowrate
T_f =Feed temperature
P =Pumparound flowrate
Re =Reflux flowrate
Q_r = Reboiler duty
$xC8$ = Oleic acid composition
$xC8$ = Linoleic acid composition

NOC data that consist of quality variables and process variables were generated and arranged in the matrix form, **X** when the process is in statistical control or quality variables remain close to their desired values. The matrix data, **X** with m observations on p variables can be written as,

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mp} \end{bmatrix} \quad (1)$$

NOC data is very important in SPC methodology since it is used to predict the future behavior of the process. Some noises were imbedded into the process variables using Matlab simulator to create random process data with normally distributed. After NOC step done, faulty condition was introduced in the process by inserting deviations in the process variables and OC data was collected during this condition.

Both NOC and OC data are standardized before further analysis since the variables have different units and wide range of data measurements. Each variable is adjusted to zero mean by subtracting off the original mean of each column and adjusted to unit variance by dividing each column by its standard deviation. After the standardization, each variable have equal weights with zero mean and one standard deviation (N (0, 1)). The linear relationship between quality variables and process variables is developed using multivariate analysis techniques, PCorrA during normal process operation. This relationship is interpreted in terms of correlation coefficient, C_{ik} which is used to diagnose the cause of the fault during the OC situation.

3 PARTIAL CORRELATION ANALYSIS, PCORRA

Partial correlation coefficient is defined as a correlation of quality variable, x_k and process variable, x_i when the effects of other process variable(s) have been removed from x_k and x_i. If the two variables of interest are $x_{k,1}$ and $x_{i,1}$ and the controlled variables are x_{i2}, x_{i3} ... x_{in}, then the corresponding partial correlation coefficient is

$$c_{xk1,xi1|xi2,xi3,\ldots xin} = \frac{c_{xk1,xi1|xi3,xi4,\ldots,xin} - c_{xk1,xi2|xi3,xi4,\ldots,xin} c_{xi1,xi2|xi3,xi4,\ldots,xin}}{\sqrt{1-c^2_{xk1,xi2|xi3,xi4,\ldots,xin}}\sqrt{1-c^2_{xi1,xi2|xi3,xi4,\ldots,xin}}} \quad (2)$$

As shown mathematically above, PCorrA is done by separating the group of process variables into subgroup in which one or more variables are held constant before determining the correlation among the other variables.

4 INDIVIDUAL SPC CHARTS

SPC chart is used to monitor the performance of any given process. There are two types of causes, which contribute to the existing of faults in the process. Chance, or common causes are small random changes in the process that cannot be avoided. Variation of this type is only removable by making a change in the existing process. Assignable causes, on the other hand, are large variation in the process that can be identified as having specified cause. Assignable causes are causes that are not part of the process on a regular basis. This type of variation arises because of specific circumstances. Sources of variation can be found in the process itself, the material used, the operator's actions, or the environment.

Control charts approach is based on the assumption that a process subject to common cause variation will remain in a state of statistical control under which process remain close to target which is known as NOC data for this study. By monitoring the performance of a process over time, OC events known as assignable cause can be detected as soon as they occur. If the causes for such events can be diagnosed and the problem can be corrected, the process is driven back to its normal operation. Individual Shewhart, Exponential moving average (EWMA), and Moving Average (MA) and Moving Range (MR) are USPC charts. They are used for individual data. Correlation coefficient from the multivariate analysis technique is used to relate the quality variables with the process variables. This correlation coefficient is used to translate the control limits of USPC charts from quality variables into process variables, which is used to perform fault diagnosis for the process operation. The OC situations will be considered whenever a point fall outside the control limits with 99.73% confidence limits for all these charts.

5 IMPLEMENTATION OF PCORRA IN USPC CHART

PCorrA method is used to determine the correlation coefficient between quality variable and process variables. Let x_k is the quality variable and x_i is the process variable. The relationship between standardized quality variable, x_k^s and standardized process variable, x_i^s can be written as,

$$x_k^s = C_{ik} x_i^s \quad (3)$$

where $x_k^s = (x_k - \bar{x}_k)/s_k$, $x_i^s = (x_i - \bar{x}_i)/s_i$. s is the standard deviation, while $\bar{x}_k$ and $\bar{x}_i$ is quality variable mean and process variable mean respectively. The control limit for quality variable in general is

$$LCL < x_k^s < UCL \quad (4)$$

where *UCL* and *LCL* is upper control limit and lower control limit respectively. Substitute equation 3 into equation 4 and rearrange the equation 4. The control limits for corresponding process variable is,

$$LCL/C_{ik} < x_i^s < UCL/\ C_{ik} \quad (5)$$

Equation 5 is used to calculate the control limits for process variables. These limits are calculated based on the NOC data. Table 2 shows both of the limits for all control charts as FDD tools in this study.

Table 2.Control limits for quality variable and process variable

Control chart	Quality variable control limit	Process variable control limit
Shewhart Individual	$UCL = 3s, LCL = -3s$	$UCL = 3s/C_{ik}$, $LCL = -3s/C_{ik}$
Shewhart Range	$UCL=D'_{.001}\ \bar{R}$, $LCL=D'_{.999}\ \bar{R}$	$UCL=D'_{.001}\ \bar{R}\ /C_{ik,}\ LCL=D'_{.999}\ \bar{R}\ /C_{ik}$
EWMA	$UCL=+Ls\sqrt{\frac{\lambda}{2-\lambda}}\ [1-(1-\lambda)^{2i}]$ $UCL=-Ls\sqrt{\frac{\lambda}{2-\lambda}}\ [1-(1-\lambda)^{2i}]$	$UCL=+(Ls\sqrt{\frac{\lambda}{2-\lambda}}\ [1-(1-\lambda)^{2i}])/C_{ik}$ $UCL=-(Ls\sqrt{\frac{\lambda}{2-\lambda}}\ [1-(1-\lambda)^{2i}])/C_{ik}$
MA	$UCL=+A_2\bar{R}$, $LCL=-A_2\bar{R}$	$UCL=+A_2\bar{R}\ /C_{ik,}\ LCL=-A_2\bar{R}\ /C_{ik}$
MR	$UCL=D'_{.001}\ \bar{R}$, $LCL=D'_{.999}\ \bar{R}$	$UCL=D'_{.001}\ \bar{R}\ /C_{ik}$, $LCL=D'_{.999}\ \bar{R}\ /C_{ik}$

where

$\bar{R}$	= Average of range
λ	=Weighting factor
A_2 , $D'_{.001}$, $D'_{.999}$	=Constant
L	= Width of the control limit

Table 3 shows the equation to determine statistical data for each control charts.

Table 3.Statistical data for control charts

Control chart	Quality variable	Process variable
Shewhart Individual	$x = x_i$	$y = y_i$
Shewhart Range	$R_i = max\ [x_{i-1+1}] - min\ [x_{i-1+1}]$	$R_i = max\ [y_{i-1+1}] - min\ [y_{i-1+1}]$
EWMA	$z_i=\lambda x_i + (1-\lambda)z_{i-1}$	$z_i=\lambda y_i + (1-\lambda)z_{i-1}$
MA	$MA_i = (x_i + x_{i-1} + \ldots + x_{i-w+1})/w$	$MA_i = (y_i + y_{i-1} + \ldots + y_{i-w+1})/w$
MR	$MR_i = max\ [x_{i-w+1}] - min\ [x_{i-w+1}]$	$MR_i = max\ [y_{i-w+1}] - min\ [y_{i-w+1}]$

6 FDD EFFICIENCY USING DIFFERENT CONTROL CHARTS

Both quality variable and process variable were monitored continuously using three types of control charts. The efficiency of the FDD method using Shewhart, EWMA and MAMR chart is evaluated based on two aspects i.e the successful of SPC chart to detect the fault and the successful of SPC chart to identify the correct process variable as fault cause for each fault situation. The efficiency of fault detection, $\eta_{Fdetect}$ and the efficiency of the fault diagnosis, $\eta_{FDiagnose}$ is determined using the following equation,

$\eta_{Fdetect}$= [Number of faults detected / Total of faults generated in the process] x 100
$\eta_{FDiagnose}$ = [Number of faults diagnosed / Total of diagnosed fault] x 100

The overall performance is,

$$\eta_{FDD} = \eta_{Fdetect} \times \eta_{Fdiagnose} \times 100\%$$

80 fault locations consist of 50 single fault and 30 multiple faults were introduced into the process. Figure 2 and figure 3 show the efficiency of FDD on quality variables, oleic acid, C8 and linoleic acid,C9 respectively using different SPC charts. Shewhart chart give 100% performance in FDD on oleic acid and linoleic acid, which is better than EWMA and MAMR, which result 90% fault detected on oleic acid and 93% fault detected on linoleic acid. Shewhart chart is plotted using 100% current data but MAMR statistic is calculated using window size of four, which consist of 25% current data and 75% previous data. This caused detection delay using MAMR. EWMA statistic exponential weighted average of all prior data, including the most recent data. The weighted average depends on weighting factor, λ. Small λ will give less weight to current data and more weight to previous data and vice versa. In this study, λ = 0.4 is used. This give the EWMA statistic consist of 40% current data 60% previous data. This shows that FDD efficiency decreases as the percentage of previous data involved in calculates the statistics value for each charts increases. False alarm (action is taken due to signal but in fact the process does not change at all) rate using MAMR is about 10% for C8 and 7% for C9. This is higher than EWMA, which has 3% false alarm rate for C8.

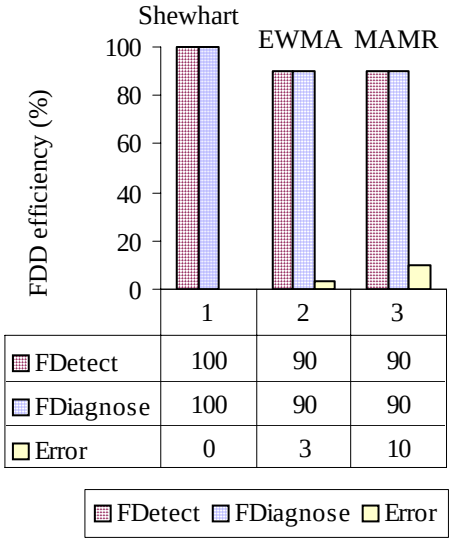


Figure 2. Oleic acid FDD effieciency

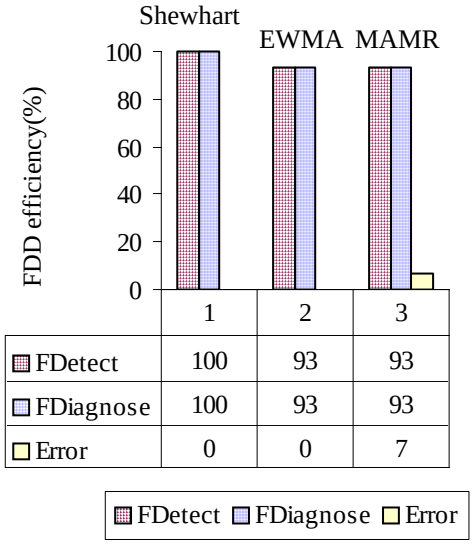


Figure 3: Linoleic acid FDD efficiency

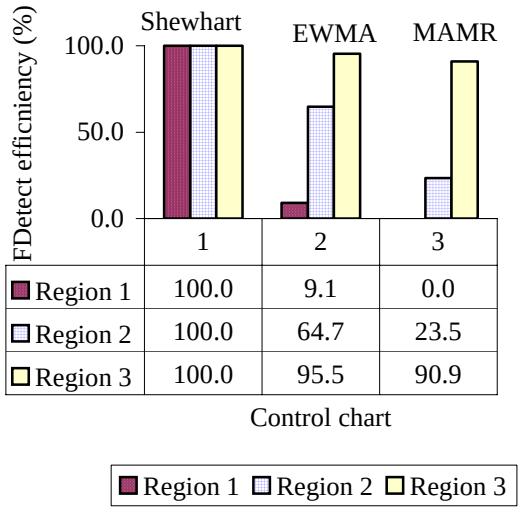


Figure 4 shows the fault detection efficiency using Shewhart, EWMA and MAMR in three different regions using single fault data. The OC data that is greater than $\pm$ 3s is divided into three regions. Region 1, region 2,and region 3 refer to mean $\pm$ 4s, mean $\pm$ 5s and over mean $\pm$ 5s respectively. Both Shewhart chart and EWMA can detect OC data in region 1 but Shewhart (100% fault detection efficiency) give better performance than EWMA (9.1% fault detection efficiency). MAMR chart only can detect deviation that is greater than mean $\pm$ 4s. The result obtained shows that Shewhart chart was able to detect small shift in the process.

Figure 4. Fault detection efficiency in different region

7 CONCLUSIONS

Improved SPC chart for FDD using PCorrA technique, which is used to develop correlation between quality variables and process variables have been presented. Process monitoring for fault diagnosis can be done using process variables with the implementation of correlation coefficient. Performance of Shewhart chart is the best compared to EWMA and MAMR in detecting and diagnosing faults. FDD result using EWMA is better than MAMR because high false alarm rate using MAMR shows the risk of taking wrong action on the process.

REFERENCES

Chen, Q., Kruger, U., Meronk, M. and Leung, A.Y.T. (2004). "Synthesis of T^2 and Q statistic for process monitoring." *Control Engineering Practice*, 12, 745-755.

Ibrahim, K.I. (1997). "Application of Partial Correlation Analysis in Active Statistical Process Control." Proceedings of Regional Symposium of Chemical Engineering, 434-438, Johor.

Mak, W.Y. and Kamarul, A.I. (2003). "*Fault Detection for Distillation Column Using Multivariate Statistical Process Control (MSPC)*". Proceedings of The 17th Symposium of Malaysian Chemical Engineers (SOMChE 2003), 29-30 December 2003, Hotel Copthorne, Penang, Malaysia. 92-96.

Manabu K., Koji N., Shinji H., Iori H., Hiromu, O., Ramon S., and Bhavik B. (2000). "Comparison of statistical process monitoring methods: application to the Eastman challenge problem." *Computers and Chemical Engineering*, 24, 175-181.

Nong Y., Qiang C., Syed Masum E., and Kyutae N. (2000). "Chi-square Statistical Profiling for Anomaly Detection." Proceedings of the 2000 IEEE Workshop on Information Assurance and Security United States Military Academy, West Point, NY, 187-193.

Venkat Venkatasubramanian, Raghunathan Rengaswamy, Kewen Yin, Surya N. Kavuri (2003). "A review of process fault detection and diagnosis Part I: Quantitative model-based methods, *Computers and Chemical Engineering,* 27, 293-311.

FAULT DETECTION FOR DISTILLATION COLUMN USING MULTIVARIATE STATISTICAL PROCESS CONTROL (MSPC)

Mak Weng Yee and Kamarul Asri Ibrahim

Faculty of Chemical Engineering and Natural Resources Engineering,
Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia.
Email: mak980176@lycos.com

ABSTRACT

Chemical process is inclined to be a large-scale, complex and having stringent requirements on the desired quality. It also utilizes a lot of energy, must be environmentally friendly and fulfill safety requirements. Accurate process fault detection at an early stage of the process is important to modern chemical plant in achieving the above requirements. This paper focuses on the application of Multivariate Statistical Process Control (MSPC) as a fault detection tool. An industrial distillation column is modelled and chosen as the case study for this research. Principal Component Analysis (PCA) and Partial Correlation Analysis (PCorrA) are used to develop the correlation coefficients between the variables of the process. Faults considered in the research are sensor failures, valve failures and controller malfunctions. Shewhart Control Chart with the developed correlation coefficients are used for detecting the faults. Results show that both methods based on PCorrA and PCA are able to detect the pre-designed faults.

Keywords: correlation coefficient; partial correlation analysis; principal component analysis

1.0 INTRODUCTION

Currently, many chemical processes are becoming increasingly measurement rich. Large volume of highly correlated data is always recorded. This large volume of data can be very useful for process monitoring if an appropriate analysis method is applied (Lam and Kamarul, 2002a). Multivariate Statistical Process Control (MSPC) is a method that is able to extract the desired information from the data by carrying out data reduction without losing the original information. Many industrial processes involve a set of input variables and quality variables, which are highly correlated. If one of the variable changes, it will affect the other correlated variables (Lam and Kamarul, 2002b). Thus, ignoring the cross-correlation between the variables can lead to misinterpretation of the process behaviour.

One advantage of MSPC is that this method could reduce the complexity of online process monitoring with its ability to detect process abnormalities that are difficult to notice. Principal Component Analysis (PCA) is used to extract the required information for process monitoring from the data of the process. Partial Correlation Analysis (PCorrA) will also be used in this work for information extraction of the original data. PcorrA is a method that is able to determine the correlation between two variables while maintaining other correlated variables at a constant value (Kamarul, 1995). In MSPC, the correlation between variables is the major information needed for good process monitoring performance. In Lam and Kamarul (2002b), the cross-correlation coefficients between process variables were introduced as tool for process monitoring for fault. In this research, PCA and PCorrA will derive the cross-correlation coefficients from data collected from the simulation model.

Shewhart Control Chart is plotted using the developed correlation coefficients from PCA and PCorrA for process monitoring of the process. The function of these control charts is to compare the current state of the process with “Normal Operating Condition (NOC)”. NOC exists when the process variables and quality variables remain close to their desired values. In contrast, “Out of Control (OC)” occurs when fault appears in the process. OC exists when one or more value of the quality variable and the input variable are outside the control limit of their respective control chart.

2.0 PROCESS MODELLING AND DATA GENERATION

Data mining is the most important part in obtaining an accurate correlation between the process variables and quality variables. In this research, data is obtained from a simulation model. A distillation column from a Palm Oil Fractionation Plant is selected as the case study. The model of this column is developed based on the model by Wong (2003). Figure 1 shows the process and the key variables.

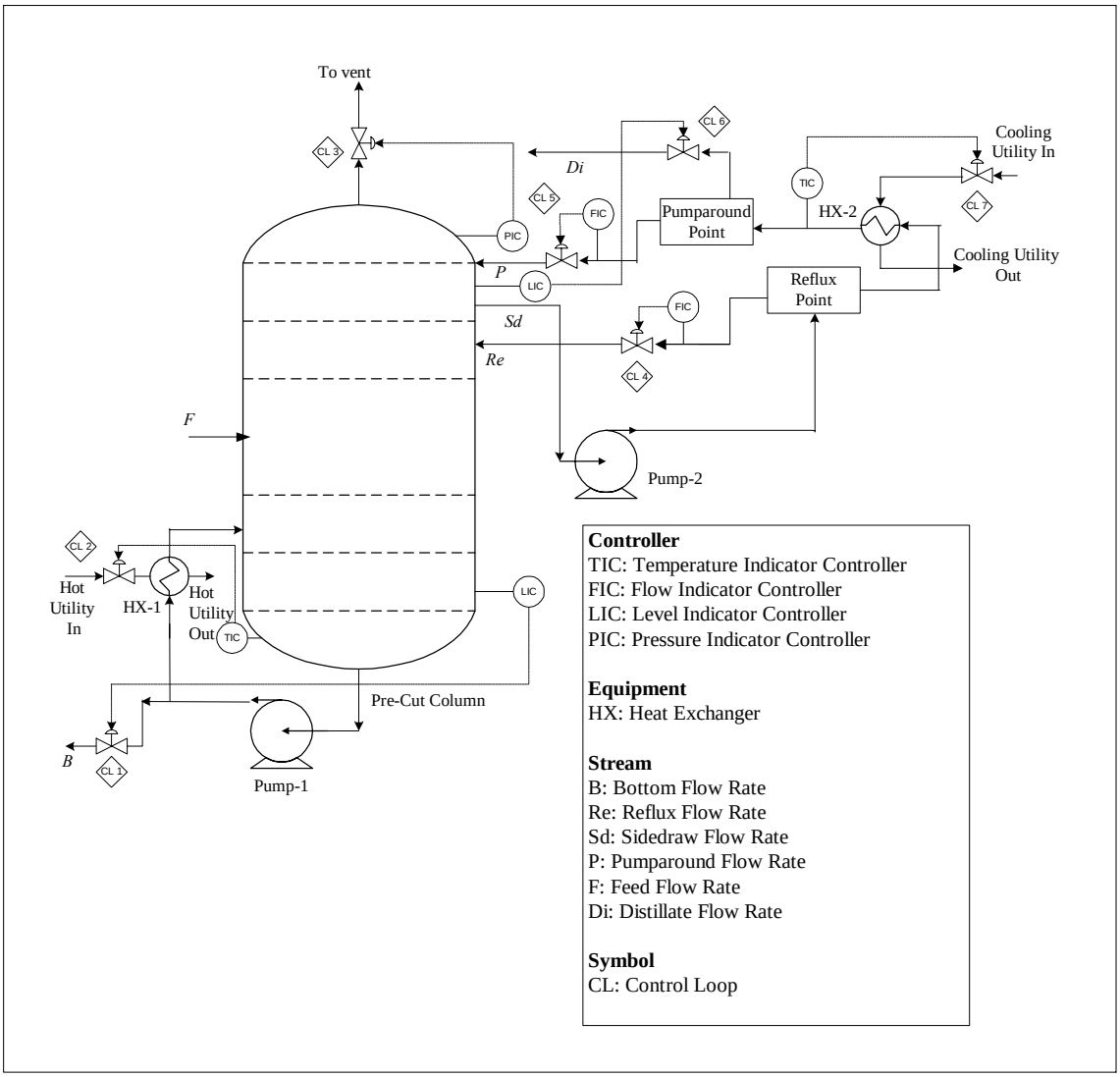
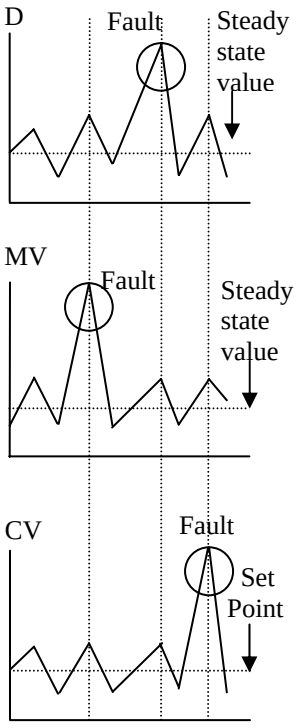
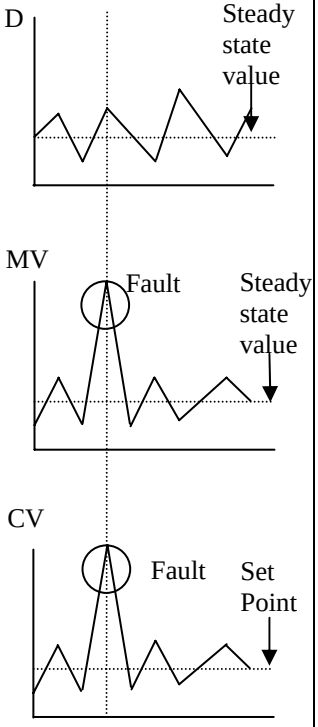
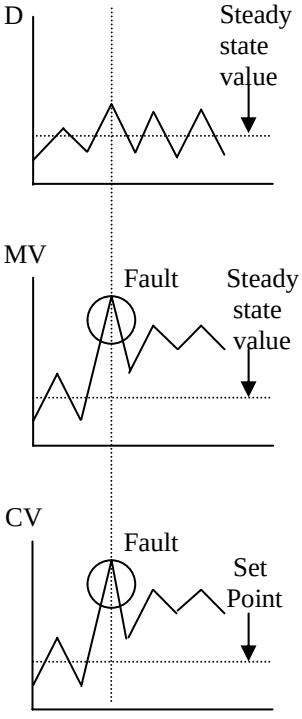


Figure 1: Distillation Column Model

The state equations for the distillation column were derived based on first principal equations. Ordinary Differential Equations (ODE) for state equations were formed and solved using 4th Order Runge-Kutta method. The MATLAB® software was used for the whole simulation program.

Based on the column model, two sets of process operating data were generated. For NOC data, some noises with zero mean were imbedded into the simulation program. The noises considered are small random change in selected key variables such as feed flow rate, feed temperature, reboiler duty, cooler duty and side draw flow rate. On the other hand, for OC data, some significant changes were purposely added into the process model as faults. These faults represent valve failures, sensor failures and controller malfunctions. The description of each type of fault is listed in Table1. The generated NOC and OC data are mean-centered and variance scaled. The NOC data will be subjected to multivariate analysis using PCA and PCorrA for deriving the correlation coefficients between the process variables and the selected quality variables.

Table 1: Fault Descriptions

Sensor Failure	Valve Failure	Controller Malfunction
• For open loop variables, only the value of the variable changes abnormally. For closed loop variables, only the value of the disturbance (D) OR the manipulated variable (MV) OR the control variable (CV) changes abnormally.	• For open loop variables, only the value of the variable changes abnormally. For closed loop variables, both the value of manipulated variable (MV) AND control variable (CV) changes abnormally together.	• For closed loop variables, the value of manipulated variable (MV) AND control variable (CV) changes abnormally together.
		

3.0 CORRELATION COEFFICIENTS OF NOC DATA

3.1 Correlation Coefficients Using PCA

Method for obtaining correlation coefficients, C_{ik}, using PCA was based on the work by Lam and Kamarul (2002b). Correlation coefficients using PCA are calculated as in Equation 1.

$$C_{ik} = \sum_{j=1}^{n} v_{ij} v_{kj} \lambda_j \qquad \text{(Eq. 1)}$$

Where: v_{ij}, v_{kj} = eigenvectors obtained from process data using PCA
λ_j = eigenvalue obtained from process data using PCA

3.2 Correlation Coefficients using PCorrA

Partial Correlation Analysis calculates the correlation between two variables while allowing the effect of other correlated variables on the two variables. For calculating correlation coefficient, C_{ik}, for variable 1 and 2 using PCorrA after allowing the effect of j-2 variables is as Equation 2.

$$C_{ik_{12}} = \frac{r_{12.(4,\ldots,j-2)} - r_{13.(4,\ldots,j-2)} r_{23.(4,\ldots,j-2)}}{(1 - r^2_{13.(4,\ldots,j-2)})^{1/2} (1 - r^2_{23.(4,\ldots,j-2)})^{1/2}} \quad \text{(Eq.2)}$$

Where: r_{12} = correlation between variable 1 and 2
$r_{12.3}$ = partial correlation between variable 1 and 2 after the effect of variable 3
$r_{12.(3,4,\ldots,j-1)}$ = partial correlation between variable 1 and 2 after the effect of j-2 variables

4.0 PROCESS FAULT DETECTION USING C_{IK} BASED ON PCA AND PCorrA

C_{ik} relates a process variable, x_i with a quality variable, y_i in the following way:

$$y_i = C_{ik} x_i \quad \text{(Eq.3)}$$

For conventional Shewhart Control Chart, the Upper Control Limit (UCL) and Lower Control Limit (LCL) for mean-centered and variance-scaled variables are +3 and -3 respectively (McNeese and Klein, 1991). Using the information from Equation 3, the UCL and LCL for quality variables and process variables will be +3 and -3 and +3/C_{ik} and -3/C_{ik} respectively. After the NOC control charts are established, they are used for fault detection of the OC data.

When a fault is detected, the variable of that control chart will be checked whether it is a closed loop variable or open loop variable. For open loop variable, the fault will be of sensor failure or valve failure as pre-designed while fault for closed loop variable can be of valve failure, sensor failure or controller malfunction. The performance of the fault detection method using correlation coefficients based on PCA and PCorrA is shown in Figure 2.

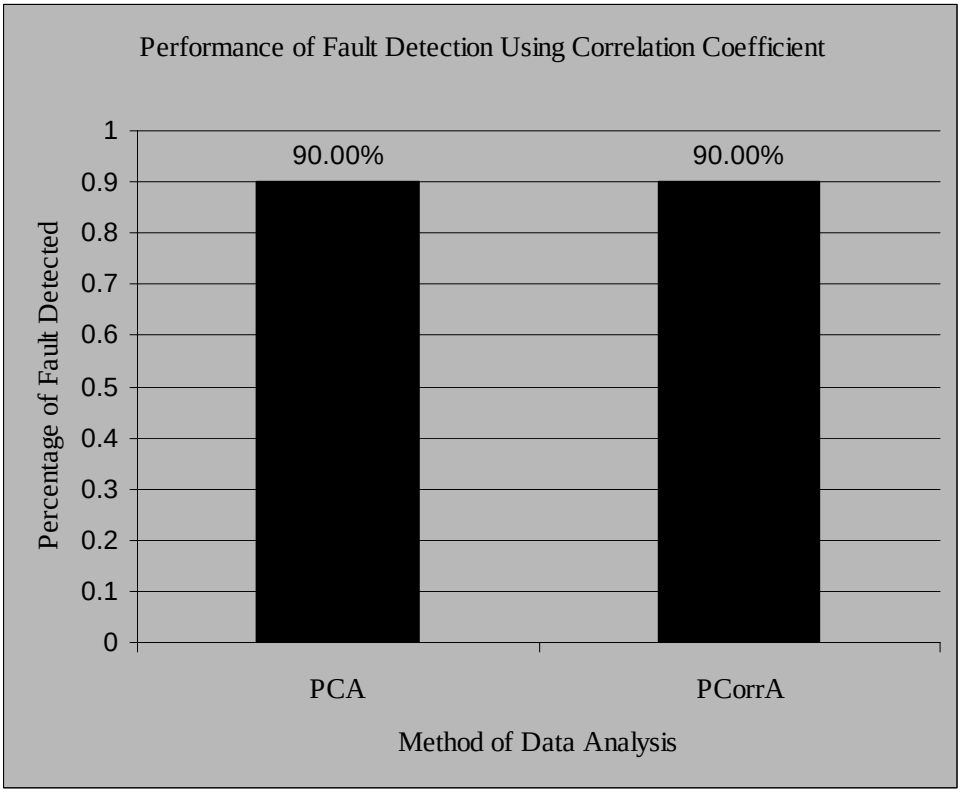


Figure 2: Performance of Fault Detection Using Correlation Coefficients

Both fault detection method using correlation coefficients based on PCA and PCorrA were able to detect the pre-designed faults. Out of the 10 faults in the fault data, 9 faults (both single fault and multiple faults) were successfully detected. These results show that the developed correlation coefficients were able to relate the key process variables to the quality variables of interest in the case study.

5.0 CONCLUSION

An approach for fault detection using correlation coefficients based on PCorrA and PCA was presented. The performance of the approach was studied on an industrial distillation column. The results show that the fault detection method using cross correlation coefficient was able to detect the faults present in the process. The cause of each fault can be diagnosed by checking the control charts of the key process variables in which a fault is detected. Fault diagnosis using method based on correlation coefficients is a research problem for future work.

6.0 NOTATION

C_{ik}: Correlation Coefficient
CV: Control variable
D: Disturbance
LCL: Lower Control Limit
MSPC: Multivariate Statistical Process Control
MV: Manipulated variable
NOC: Normal Operating Condition
OC: Out of control
ODE: Ordinary Differential Equation
PCorrA: Partial Correlation Analysis
PCA: Principal Component Analysis
v_{ij}, v_{kj}: Eigenvectors obtained from process data using PCA
UCL: Upper Control Limit
x: Process variable
y: Quality variable
λ_j: Eigenvalue obtained from process data using PCA

7.0 ACKNOWLEDGEMENTS

National Science Fellowship (NSF) scholarship is gratefully acknowledged.

8.0 REFERENCES

.

Kamarul, A.I., *Active Statistical Process Control*. Unpublished Ph.D. Thesis (University of Newcastle Upon Tyne), UK, (1996).

Lam, H.L. and Kamarul, A.I., *Improved MSPC for Chemical Process Fault Detection and Diagnosis via Outline Analysis Approach*. World Engineering Congress, Sarawak. 39-44, (2002a).

Lam, H.L. and Kamarul, A.I., *Improved Multivariate Statistical Process Control for Chemical Process Fault Detection and Diagnosis via Cross Variable Correlation Coefficient Approach*. *Regional* Symposium on Chemical Engineering, RSCE/SOMChem 2002, **Vol.(2.)**: 1637-1644, (2002b).

McNeese, W.H. and Klein, R.A., *Statistical Methods For the Process Industries*. USA: ASQC Quality Press. 251, (1991).

Wong, T.S., *Advanced Process Control of A Fatty Acid Distillation Column Using Artificial Neural Networks*.Unpublished Masters Thesis (Universiti Teknologi Malaysia), Malaysia, (2003).