Development of Soft Sensor Model Using Moving Window Approach

by

Lavaniya Rajan

Dissertation submitted in partial fulfillment

of the requirements for the

Bachelor of Engineering (Hons)

(Chemical Engineering)

MAY 2012

Universiti Teknologi PETRONAS
Bandar Seri Iskandar
31750 Tronoh
Perak Darul Ridzuan

CERTIFICATION OF APPROVAL

Development of Soft Sensor Using Moving Window Approach

by

Lavaniya Rajan

A project dissertation submitted to the

Chemical Engineering Programme

Universiti Teknologi PETRONAS

in partial fulfillment of the requirement for the

BACHELOR OF ENGINEERING (Hons)

(CHEMICAL ENGINEERING)

Approved by,

07/09/2012

(Dr. Ramasamy Marappagounder)

UNIVERSITI TEKNOLOGI PETRONAS

TRONOH, PERAK

May 2012

CERTIFICATION OF ORIGINALITY

This is to certify that I am responsible for the work submitted in this project, that the original work is my own except as specified in the references and acknowledgements, and that the original work contained herein have not been undertaken or done by unspecified sources or persons.

LAVANIYA RAJAN

Table of Contents

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Soft sensors are used broadly in the industries to predict the process variables which are not measurable by sensors. The objective of this project is to develop a data-driven soft sensor using Moving Window approach with the selective regression techniques and to evaluate and validate the advantages and performances of Moving Window approach over the traditional soft sensor models. Time invariant and stationary process conditions are those assumptions made in developing soft sensors, and these assumptions causes degradations and limitations to the soft sensors in estimating process variables. Degradations of soft sensors are caused by process shift, catalyst performance lost and et cetera. Besides that, the restrictions of sensors in estimating difficult-to-measure variables and the delays during the laboratory tests have become one of the factors in developing soft sensor. This paper presents a study regarding the multivariate statistical process control techniques that can be used in developing soft sensors such as Least Square Regression method, Partial Least Square Regression method and Principle Component Analysis. The scope of study for the project includes understanding the concept and what are the adaptive schemes available to construct the soft sensors. Besides that further research on Moving Window approach together with MSPC techniques will be carried out which can be adapted into the adaptive models to develop the soft sensors. Systematic approach will be presented through this project in using Moving Window approach to construct the soft sensors and this includes an analysis of an appropriate case study where the approach can be implemented.

Keywords: Multivariate Statistical Process Control techniques, Least Square Regression method, Partial Least Square Regression method and Principle Component Analysis

ACKNOWLEDGEMENT

Enrolment in Final Year Project has been truly valuable knowledge and experience to me. I am grateful in acknowledging the helpful and supportive parties who were with me throughout the completion of the project. This dissertation would not have been possible unless guidance from my supervisor, Dr. Ramasamy Marappagounder, who has given me an opportunity and trust in carrying out this project. The guidance and support from him have helped me tremendously in achieving the goals and objectives of the project. His ever willingness to teach and guide me despite all his other works has been a great assistance throughout the achievement of the project.

Besides that, I would like to thank my Final Year Project coordinators, Puan Anis, Puan Asna and Puan Norhayati for the supervision throughout the final year project. Their guidance has enabled me to prepare quality documents such as Interim Report, Dissertation and Extended Proposal. The lectures and writing workshops, which were initiated by them, have been a great experience on writing thesis and technical papers.

Furthermore, I owe sincere and earnest thankfulness to my colleague, who supported and helped me in investigating the engineering problems related to the project. The team works in dealing with software such as MATLAB, HYSYS, and SIMCA-P have aided me a lot in completing the project and widen my knowledge about the software which also built the confidence in using that software in the future.

Finally, I feel it have been a great opportunity in finalizing the project as I have a remarkable time working with the aforementioned parties who boosted me morally and given me a great exposure in the field of research and development.

Lavaniya Rajan

CHAPTER 1
INTRODUCTION

1.1 Background

Sensors can be classified as instruments, which aid the technical personnel to observe the trend or pattern of specific processes in the industries. Sensors are converters that measure physical quantities or parameters and convert it into signals; these signals can be read by personnel or controller. Apparently, sensors incapable in giving sufficient information regarding the difficult-to-measure variables such as concentration (Gonzalez, 1999). In order to curb that, soft sensors are developed based on correlations between difficult-to-measure and easy-to-measure variables (Okada, 2011). Figure 1.1 shows the concept of soft sensor.

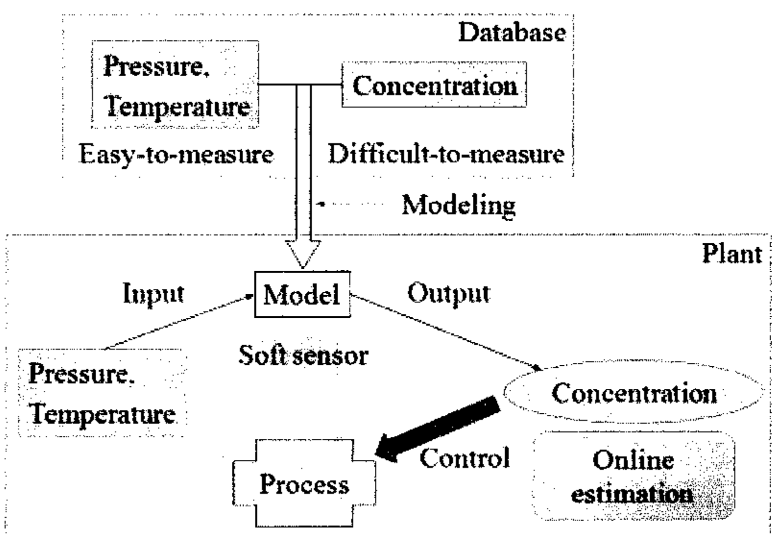

Figure 1.1: Basic Concept of Soft Sensor (Okada, 2011).

However, soft sensors tend to deteriorate due to modifications in the state of chemical plants, catalyst performance loss, sensor and process drift and et cetera (Kaneko. H &Funatsu.K, 2011). The rate of deterioration of soft sensors can be categorized into three main mechanisms which are gradual deterioration, instant deterioration and rapid deterioration. For example, poor catalytic performance and shift in the internal temperature cause gradual deterioration; in contrast to rapid degradation which happens due to sudden change in the raw material. On the other hand, regular maintenance of the plant will cause the degradation of the soft sensor to happen instantly (Kaneko. H &Funatsu.K, 2011). As a result, it could tarnish the performance of soft sensors in estimating the process variables. The suggested solution to handle such a situation is by updating the soft sensor model to handle the variations in the process characteristics. Updating the soft sensor models can be done using two main approaches which are first-principle models and data-driven models (Liu. J et.al, 2008). However first-principle models could be fuzzy because of the complexity which requires a lot of effort and time to develop or either it could be very simple to be accurate and precise in predicting the values. Data-driven models can provide reasonably accurate information which uses regression methods such as PCR (Principle Component Regression), PLS (Partial Least Square) and CCR (Canonical Coordinates Regression). There are soft sensors which being developed using nonlinear models such as ANN (Artificial Neural Network), SVM (Support Vector Machines) and KPLS (Kernel Partial Least Square). After analysis of these tools/techniques, appropriate technique will be selected in order to be adapted into the adaptive schemes or approaches to be used in updating the soft sensors.

1.2 Problem Statement

It has been assumed that the industrial processes are stationary and time-invariant; these assumptions have led to development of static soft sensor models or also known as traditional soft sensors. But the **time-variant and dynamic characteristics of industrial processes** violate the assumptions, and as a result the static soft sensor models are incapable in estimating the process variables (He & Yang, 2007). The time-variant and non-stationary characteristics of the industrial processes have resulted in the degradation of the soft sensors. Quality of the products has been maintained by using online sensors and by performing laboratory tests. The **malfunctions of the sensors** where it could not gain information on parameters in determining the quality of the products and also **significant delays during laboratory tests (data processing)** have unable the personnel to determine the quality of the product instantaneously (Liu. J et.al, 2008).

1.2.1 Problem Identification

- Existence of gap between data acquisition from plant, data processing and process control
- Inability of the soft sensors to estimate the process variables accurately in all conditions
- Deterioration of soft sensors by some factors

1.2.2 Significant of the Project

- The data acquisition from the plant can be done instantaneously which can lead to estimation of difficult-to-measure variables
- The adaptive model can adopt to the changes of the plant's characteristics and enable the soft sensor to provide accurate estimations

1.3 Objectives

There are two main objectives to be achieved at the end of this project, such as:

- To develop data-driven soft sensors using Moving Window approach with the selective regression technique.
- To evaluate advantages of Moving Window approach over the traditional soft sensor models

1.4 Scope of Study

The research for this project will be on understanding the concept of soft sensors and what are the multivariate statistical process control methods are available to be adapted into the adaptive schemes/models to avoid the degradation of the soft sensors. An appropriate case study will be selected from the literature. This includes understanding the concept of Moving Window approach associated with mathematical equations, which will be translated into a programming language (e.g. MATLAB coding). Based on the case study, it is possible to observe the performance of Moving Window approach in re-constructing the soft sensor model.

1.5 Relevancy and Feasibility of the Project

Relevancy

- Related to the area of Advanced Process Control and Optimization.

Feasibility

- Approximately the time frame to complete the FYP 1& 2 is 8 months
- Usage of available software (MATLAB and HYSYS) in UTP enables for the completion
- Besides that usage of SIMCA-P software by Umetrics enables the development of soft sensor model within a short period of time

CHAPTER 2

LITERATURE REVIEW

To study regarding this topic, several researches, journal and conference papers are reviewed. Soft sensor models are extensively used for prediction of quality measurements which normally determined through irregular sampling and offline analysis ((Lin, Knudsen, & Jorgensen, 2007). The research that have been carried out can be divided mainly into 2 components or criteria which are regarding the availability of types of Multivariate Statistical Process Control methods or techniques and also about the adaptive models to re-construct the soft sensor to be used for chemical processes.

Primarily for the multivariate statistical process control techniques that have been applied in the industries so far can be divided into 2 main groups which are first-principle method and data-driven method (Liu. J et.al, 2008).The widely used method will be data-driven method which comprises Principal Component Analysis (PCA), Partial Least Square (PLS) and Canonical Coordinates Regression (CCR) because of the accuracy of the information given by this method compare to first-principle method which require lots of effort and time to develop due to its complexity. All the algorithms are studied in order to understand the working principle of the techniques. The selected algorithms are PCA and PLS, where further analysis is done to compare the performance of both techniques in the adaptive models for the soft sensors. Basically PCA can be said the simpler version of the PLS method where it minimizes the usage of many parameters and also simpler technique where it analyzes the recorded variables directly (Wang et al., 2005).The main objective of PCA is to identify the outliers in the observations besides reveal the relationships between observations and variables. According to Jeng (2010), PCA is used to remove the collinearity and noise between the variables and reserve vital information of the original data. PCA technique is used widely in process monitoring and inability in adapting the time varying characteristic of the process has been the major drawback for PCA (Lu. B et al., 2009). On the other hand, PLS method is used widely in modeling a soft sensor. PLS method can be said a method that relates the X (input) and Y variables (output) by a linear multivariate modeling and computation will be carried out using the traditional regression calculations (Kaneko & Funatsu, 2011).

Secondly is all about the adaptive models for the soft sensor development by using the appropriate multivariate statistical process control (MSPC) methods. Furthermore, soft sensor which been developed using traditional method of MSPC have some shortcomings where the models could not accommodate the deviation of the process due to time-variant factor of the process where it interprets the slow changes of the process as faults (Bo & Xianhui, 2011) and increases the number of false alarms (He & Yang, 2007). Furthermore, the inferential model developed using PCA or PLS algorithms degrade due to existence of outliers and abnormal observations (Lin, Knudsen, & Jorgensen, 2007).Solution for the stated problem above are given by some of the researches who suggested a large number of adaptive PCA or PLS algorithms as shown below (Bo & Xianhui, 2011):

- (Svante, 1994) suggested Exponentially Weighted Moving Average (EWMA) filter using PCA method.
- (Cervantes.V et al., 2000) suggested Recursive-PCA algorithm
- (Choi et al., 2006) introduced a forgetting factor into Recursive-PCA algorithm.
- (Wang et al., 2005) proposed a Fast Moving Window-PCA algorithm.
- (Liu et al., 2009) suggested variable Moving Window-PCA algorithm

For this research paper, Moving Window model will be used and PLS algorithm is chosen after some analyses. Generally the principle of operation of Moving Window model is the window will be moving or sliding along the data, as the new data will be adapted into the window and new process model will be generated while discarding the oldest data which no longer representing the current process. Figure 2.1 shows the concept of Moving Window model. For process monitoring purposes, mainly Moving Window approach will be used along with PCA algorithms. According to Wang et al., 2005, most suitable approach for process monitoring will be combination of Moving Window and Recursive with PCA algorithms. This is because the Moving Window technique will update the window with new data points and discard the old data points, on the other hand, recursive technique will help the computation by updating the process model using the former model rather than developing it from the original data points.

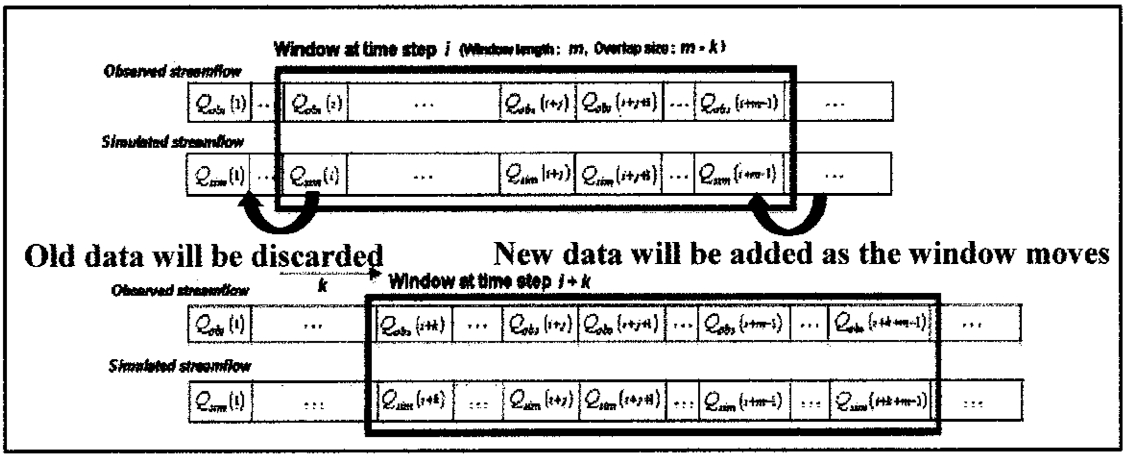

Figure 2.1: Basic Concept of Moving Window Model.

There is drawback using conventional method of Moving Window approach where the usage of constant number of samples (window size) influents the predictability of the process variables. Constant length of window will cause problem when the window has to cover larger number of data in order to calculation in predicting the variable Y (difficult-to-measure variables) (Wang et al., 2005). There are some papers which arguing and discussing about variable Moving Window length by combining the advantages of both Recursive model and Moving Window model for the soft sensor development (Wang et al., 2005). On the other hand, recursive technique also can be complex due to time varying processes in the industries and selection of forgetting factor is difficult without an aforementioned knowledge (Wang et al., 2005).

CHAPTER 3
METHODOLOGY

3.1 Introduction

First step is identifying the type of algorithms available for the soft sensor development and analyze the mathematical model/equations behind those techniques. Right after that, the appropriate algorithm will be selected to be adapted into the models/approaches to observe the performance of each model. This will be done by translating the mathematical equations into computer codes using MATLAB. Next, an appropriate case study will be selected after reviewing the journals and modeling/simulation will be done for the selected case study and hence data required will be fitted into Moving Window approach. Validation of the soft sensor model will be carried out to analyze and observe the performance of the developed soft sensor. Figure 3.1 shows the flow chart for the methodology as explained above. Steps 1 to 6 addresses the Objective #1 which been mentioned in earlier chapter whereas steps 7 to 9 addresses the Objective #2 of the project.

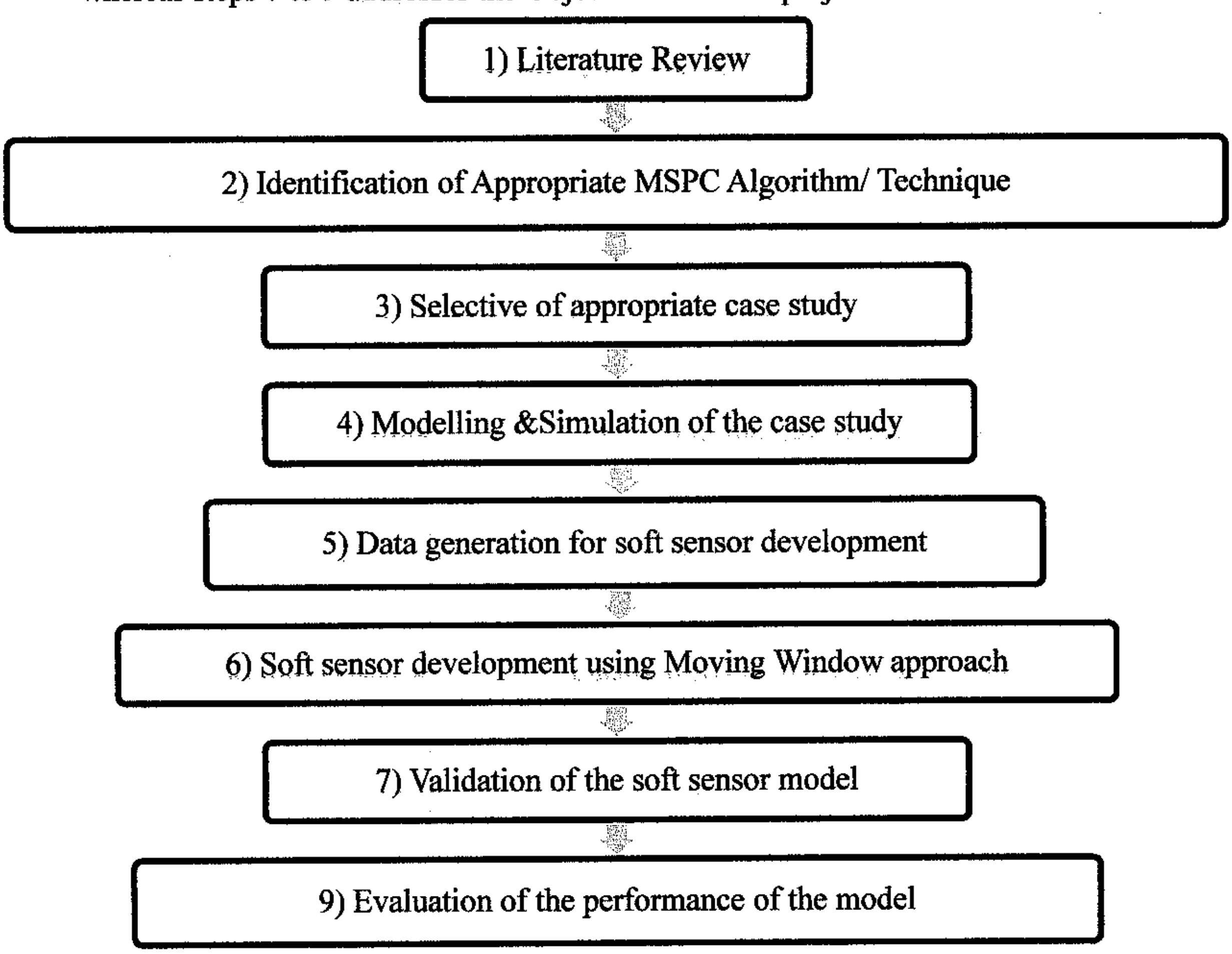

Figure 3.1: Flow chart of the project methodology.

3.2 Model Development

The typical methodology of soft sensor development is shown in Figure 3.2 below.

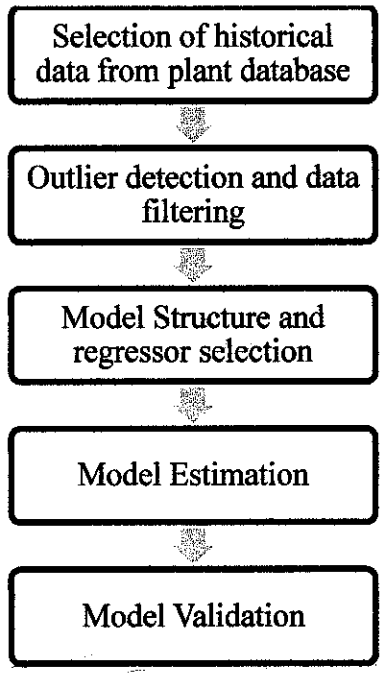

Figure 3.2: Block diagram of soft sensor designing.

Selection of historical data from the plant will be the first step in developing soft sensor. The historical data must represent the whole system dynamic and the high frequency disturbances should be removed. After the data is selected, second step will be carried out where the missing data or outliers will be detected and these phenomena can be caused by faults in measurements or in the instruments or unusual disturbances. Outliers are values which deviate from the ranges of measured values and processing data with outliers can affect the quality of soft sensors. Strategies that can be used in detecting the outliers are by using 3σ edit rule, Jolliffe parameters and residual analysis of linear regression (Fortuna et al., 2006). Besides existence of abnormalities in the historical data, the data measured in the process industries are also strongly co-linear. The two methods to deal the co-linearity is by transforming the input variables into a new reduced space with less co-linearity as it is done in the case of PCA or PLS (Kadlec et al., 2009). Figure 3.3 shows the outlier detected in a spectroscopy process. After the second step (pre-processing the data), third step will be carried out which is model structure and type of regression selection where appropriate MSPC technique will be selected. Then the model will be estimated and finally the performance of the model will be validated.

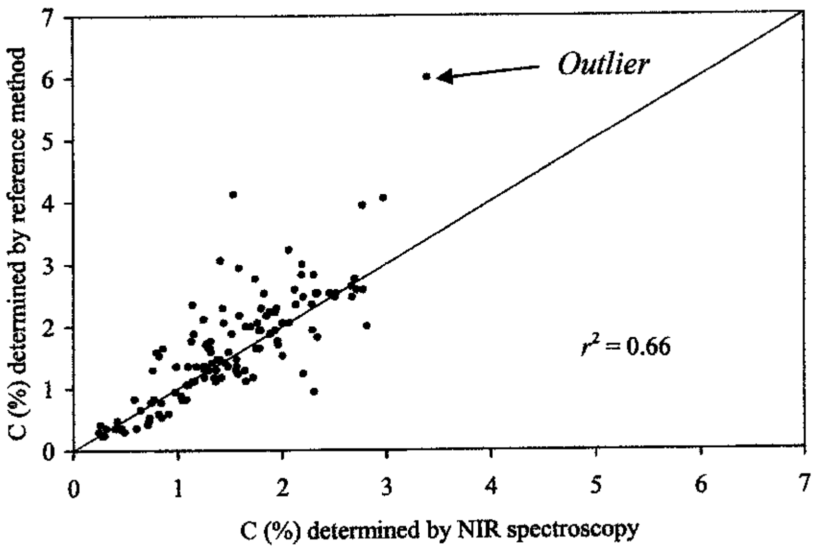

Figure 3.3: Example of outlier detected in spectroscopy experiment.

Based on the research that has been carried out, few of Multivariate Statistical Process Control (MSPC) techniques are identified to be used in the project. The techniques which are being studied are Least Square Regression Analysis, Partial Least Square Regression Analysis and Principle Component Analysis.

3.3 Regression Analysis

Generally regression analysis can be said as a method that comprises many modeling techniques in finding the relationship between one dependent variable with one or more independent variables. By using regression analysis, the characteristics of a dependent variable can be identified by analyzing the independent variables. Application of regression analysis focuses more on prediction and forecasting purposes especially for an industry which deals with processing and production activities where the quality of the product need to be computed and estimated using those variables. Figure 3.4 summarizes the types of regression method available and can be used for the project. Specifically for this semester studies on Linear Least Square Analysis, Partial Least Analysis and Principle Component Analysis have been performed.

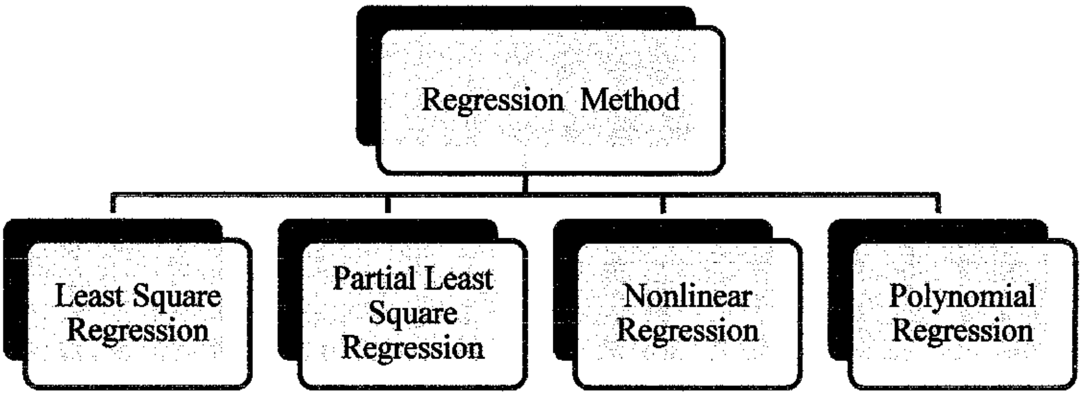

Figure 3.4: Types of Regression Analysis.

3.3.1 Least Square Regression Analysis

As mentioned above, Least Square is one of the methods in Regression analysis. Least square regression method is one of the ways to derive a curve which minimizes the discrepancy between the data points and the curve. On the other word, least square analysis can minimize the sum of squares of the deviations between the actual data and the estimated data. Least square analysis can be divided into 3 main categories which are:

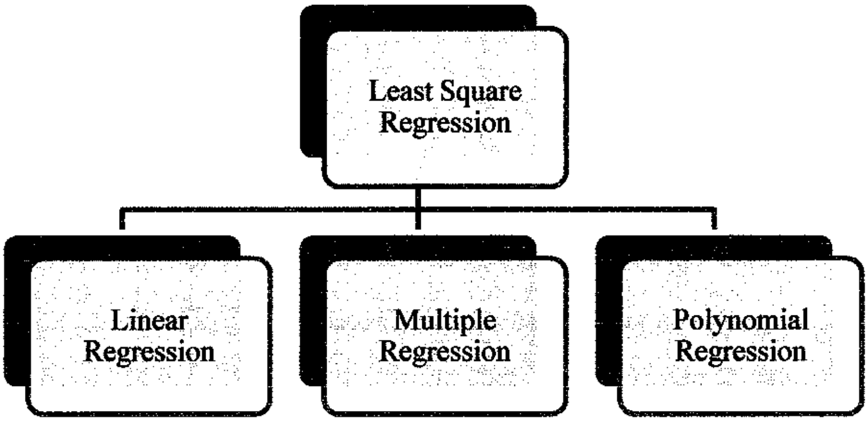

Figure 3.5: Categories of Least Square Regression Analysis.

3.3.1.1 Linear Least Square Regression Analysis

This analysis can be said as the simplest in Least Square Regression where a straight line is fitted to a set of paired observations. Equation below shows the general expression of linear least square regression:

$$\mathbf{y = a_0 + a_1 x + e} \quad \mathbf{(1)}$$

In the equation above, e represents the error between the actual value taken from the process and the value estimated by the mathematical function above. On the other hand, a_0 and a_1 represent the intercept and the slope respectively. Derivations below shows the procedures to minimize the sum of squares/discrepancy between the actual data and estimated values (curve).

From the Equation (1): $e = y - a_0 - a_1 x$

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} (a_0 - a_1 x_i) \qquad (2)$$

$$Sum\ of\ squares, S_r = \sum_{i=1}^{n} {e_i}^2 = \sum_{i=1}^{n} [y_i(measured) - y_i(predicted)]^2$$

$$S_r = \sum_{i=1}^{n} (y_i - a_o - a_1 x_i)^2 \qquad (3)$$

In order to minimize the sum of squares, the partial derivations of the Equation (3) is set to be zero:

$$\frac{\partial S_r}{\partial a_0} = -2 \sum (y_i - a_0 - a_1 x_i) \qquad (4)$$

$$\frac{\partial S_r}{\partial a_1} = -2 \sum [(y_i - a_0 - a_1 x_i)\, x_i] \qquad (5)$$

Equation (4) and (5) are further expanded and with an assumption of:

$$\sum a_0 = n a_0$$

$$\sum y_i = n a_0 + a_1 \sum x_i \qquad (6)$$

$$\sum y_i x_i = a_0 \sum x_i + a_1 \sum {x_i}^2 \qquad (7)$$

Equation (6) is multiplied with $\sum x_i$ and Equation (7) is multiplied with n:

$$\sum y_i \sum x_i = n a_0 \sum x_i + a_1 (\sum x_i)^2 \qquad (8)$$

$$n \sum y_i x_i = n a_0 \sum x_i + n a_1 \sum {x_i}^2 \qquad (9)$$

Equation (8) and (9) can be solved simultaneously to get the values of the parameter:

$$a_1 = \frac{n \sum y_i \sum x_i - \sum y_i \sum x_i}{n \sum {x_i}^2 - (\sum x_i)^2} \qquad (10)$$

$$a_0 = \bar{y} - a_1 \bar{x} \qquad (11)$$

$$Standard\ deviation\ for\ regression\ line, S_{y/x} = \sqrt{\frac{S_r}{n-2}}$$

Example below illustrates the analysis of linear least square regression method by using the equations which been derived above:

<u>Case Study 1</u>

Fit a straight line to the x and y values in the table below. Compute the total standard deviation, standard error of the estimate and the correlation coefficient for the data. Other than that, APPENDIX A shows the MATLAB coding for the Linear Least Square Regression analysis

Table 3.1: Computations for an error analysis of the linear fit

x_i	y_i
1	0.5
2	2.5
3	2
4	4
5	3.5
6	6
7	5.5

$n = 7$

$\sum y_i x_i = 119.5$

$\sum x_i^2 = 140$

$(\sum x_i)^2 = 784$

By substituting the values into Equation (10), the value of a_1 is known:

$$a_1 = \frac{7(119.5) - (28)(24)}{7(140) - 784} = 0.83929$$

By substituting the respective value into Equation (11), value of a_0 is known:

$$a_0 = 3.48571 - (0.83929)(4) = 0.07143$$

The least squares fit line is: y = **0.07143 + 0.83929 x**

The standard deviation of the data against the mean will be:

$$S_y = \sqrt{\frac{\sum(y_i - \bar{y})^2}{n-1}} = 1.9457$$

$$S_r = \sum_{i=1}^{n} (y_i - a_o - a_1 x_i)^2 = 2.9911$$

$$S_{\frac{y}{x}} = \sqrt{\frac{S_r}{n-2}} = 0.7735$$

$$Error\ reduction\ percentage, r^2 = \frac{S_y - S_r}{S_y} \times 100 = 86.6\ \%$$

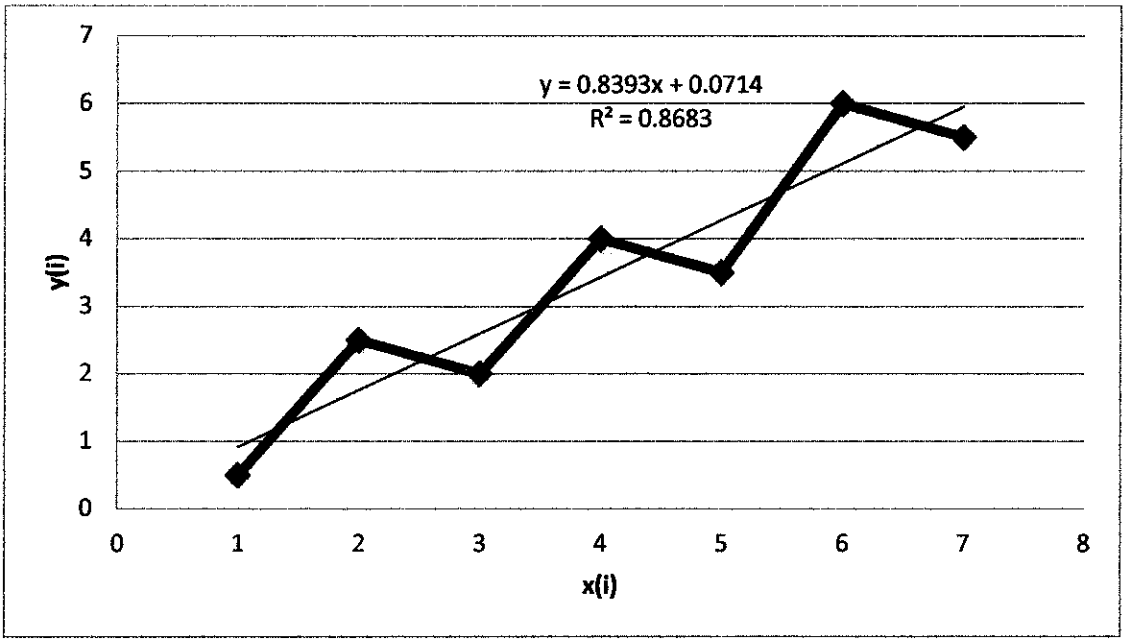

Figure 3.6: Graphical presentation of Linear Least Square Regression analysis.

3.3.1.2 Multiple Linear Least Square Regression Analysis

Similar like linear least square regression analysis, the objective of multiple linear least square regression analysis is to minimize the sum of squares of the discrepancy between the actual data and the values predicted by the model. In multiple linear regressions, the dependent variable (y) not only dependent to single variable (x) but it may depend on many independent variables. Here for the computation of multiple linear least square regressions, all the independent variables which contributes to the estimation of dependent variable will be take in account. Equation below shows the general equation for multiple linear regressions:

$$\mathbf{y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + \cdots + a_kx_k + e} \qquad \textbf{(12)}$$

The sum of squares of the Equation (12) is shown in the equation below:

$$\mathbf{\textstyle\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n}(a_0 - a_1x_{1i} - a_2x_{2i} - \cdots - a_kx_{ki})^2} \qquad \textbf{(13)}$$

In order to reduce the sum of squares, the partial derivation for the Equation (13) is set to be zero:

$$\frac{\partial S_r}{\partial a_0} = \frac{\partial S_r}{\partial a_1} = \cdots = \frac{\partial S_r}{\partial a_k} = 0$$

The partial derivation equations can be expanded with an assumption: $\sum a_0 = na_0$. The expanded equations can be expressed in the form of matrix for detail understanding about the equations:

$$\begin{bmatrix} n & \cdots & \sum x_{ki} \\ \vdots & \ddots & \vdots \\ \sum x_{ki} & \cdots & \sum x_{ki} \sum x_{ki} \end{bmatrix} \begin{bmatrix} a_0 \\ \vdots \\ a_k \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \vdots \\ \sum x_{ki}y_i \end{bmatrix}$$

To solve the equations involve in multiple linear regression method, Gauss Elimination method can be used unlike in linear regression method where the equations can be solve simultaneously (involve only 2 equations).

$$\mathit{Standard\ deviation\ for\ regression\ line}, S_{y/x} = \sqrt{\frac{S_r}{n-(m+1)}}$$

The case study below shows the application of multiple linear least square regression method in analyzing the data:

Case Study

Use multiple linear least square regression method to fit the data below. Example of MATLAB coding can be referred at APPENDIX B & C:

Table 3.2: Computations for an error analysis of the multiple linear fit

x_1	x_2	y
0	0	5
2	1	10
2.5	2	9
1	3	0
4	6	3
7	2	27

By setting the partial derivations of each term in the general equation, the equations obtained are converted into matrices as shown below:

$$\begin{bmatrix} n & \sum x_{1i} & \sum x_{2i} \\ \sum x_{1i} & \sum x_{1i} \sum x_{1i} & \sum x_{1i} \sum x_{2i} \\ \sum x_{2i} & \sum x_{1i} \sum x_{2i} & \sum x_{2i} \sum x_{2i} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_{1i} y_i \\ \sum x_{2i} y_i \end{bmatrix}$$

By computing all the terms as shown in the matrices above, following solution is obtained for the case study:

$$\begin{bmatrix} 6 & 16.5 & 14 \\ 16.5 & 76.25 & 48 \\ 14 & 48 & 54 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 54 \\ 243.5 \\ 100 \end{bmatrix}$$

$$\mathbf{6a_0 + 16.5a_1 + 14a_2 = 54} \quad \textbf{(14)}$$

$$\mathbf{16.5a_0 + 76.25a_1 + 48a_2 = 243.5} \quad \textbf{(15)}$$

$$\mathbf{14a_0 + 48a_1 + 54a_2 = 100} \quad \textbf{(16)}$$

By using gauss elimination, the equations above is solved and the solution for the case study is $\mathbf{y = 5 + 4\,x_1 - 3\,x_2}$

3.3.2 Partial Least Square Regression Analysis

PLS basically means the Projections to Latent Structures by means of Partial Least Squares. PLS generally finds the linear regression model by projecting the predicted variables and the observable variables into a new space. The data have to be pre-processed before using PLS modeling because PLS gives accurate prediction when the data are symmetrically distributed and contained fairly constant "error variance". The data will be centered and scaled to unit variance before using PLS for the analysis. In PLS the covariance between the score vector is maximize while the sum of squares is minimize by using the least square method. Furthermore, PLS gives higher predictive power than the multiple linear least square methods. Generally PLS contains the two equations as shown below (Kaneko & Funatsu, 2011):

$$\mathbf{X} = \mathbf{TP'} + \mathbf{E} \quad \mathbf{(17)}$$

$$\mathbf{y} = \mathbf{Tq'} + \mathbf{f} \quad \mathbf{(18)}$$

The X-loading matrix $= P \in R^{nxl}$

The Y-loading matrix $= q \in R^{1xl}$

X residual matrix $= E \in R^{mxn}$

Y residual matrix $= f \in R^{mx1}$

l is the number of components

Above equations shows the equations of PLS model whereas below equations shows the regression model for PLS:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{constant} \quad \mathbf{(19)}$$

$$\mathbf{b} = \mathbf{W(P'W)^{-1}q'} \quad \mathbf{(20)}$$

X-weight matrix $= W \in R^{nxl}$

Regression coefficient vector $= b \in R^{nx1}$

The basics of PLS is explained in the illustrations below where the amount of response used is 1 (M=1). Before the data is implemented into PLS, it is first treated as mentioned above. First the data will be scaled to be plotted because if the two variables are plotted with similar scale, the data might be spread in any of the axis (e.g. vertical axis). To avoid such a situation, the data will be scaled where one axis will be compressed whereas the other axis will be zoomed (expanded). There are many types of scaling but the most used will be unit variance (UV) scaling. After scaled the variables, the second part of pre-processing procedure will be carried out which is mean centering. Mean centering the average value of the each data will be subtracted from the data. Figure 3.7 shows the variables that been pre-treated with UV scaling and mean-centering, resulted in equal 'length" and mean value zero.

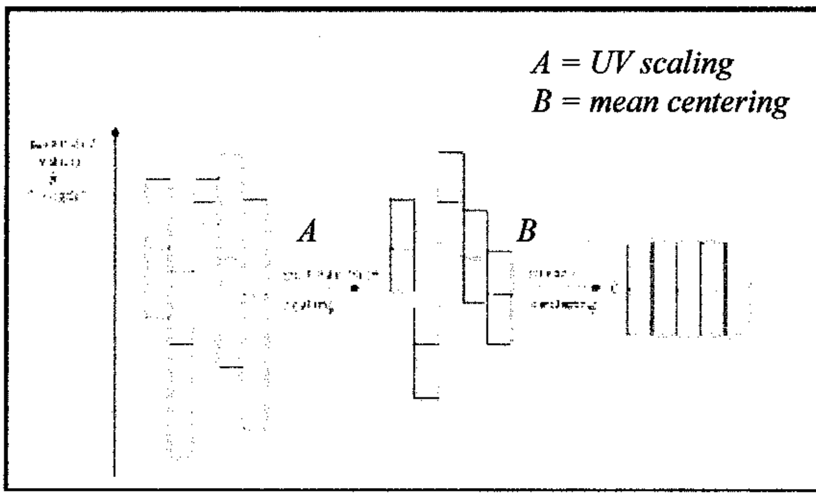

Figure 3.7: Plots of variables after mean-centering and unit variance scaling.

1. PLS basically finds the linear or polynomial relationship between the dependent (y) variables and independent (x) variables.

$$Y = f(x) + E$$

X = predictor variables

f (x) = linear or polynomial function

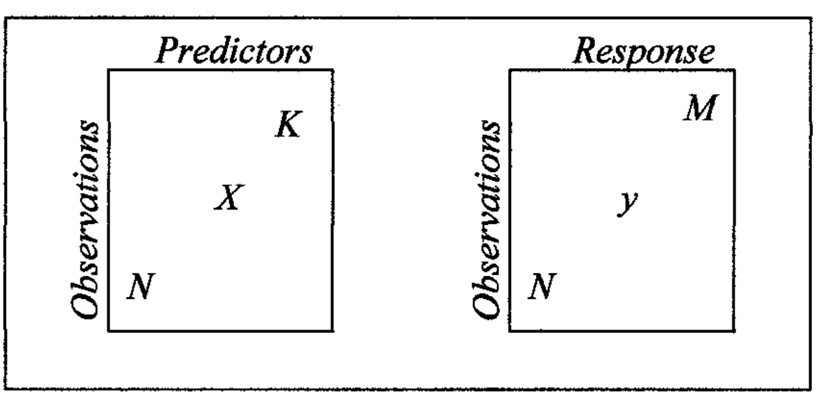

Figure 3.8: Matrix representation of X and Y variables.

2. The main objectives of PLS is to approximate the X and Y spaces precisely and maximize the correlation between X and Y. These objectives can be achieved by:

$X = (1 \times \overline{X}) + TP' + E$

$Y = (1 \times \overline{Y}) + UC' + F$

$U = T + H$ (inner relation where the coefficient is 1)

T = matrix of scores that summarizes the X variables

W = matrix of weights expressing the correlation between X and U(Y)

U = matrix of scores that summarizes the Y variables

C = matrix of weights expressing the correlation between Y and T(X)

E,F,H = matrices of residuals

Case Study

The case study of the Partial Least Square Regression analysis is performed using MATLAB by taking data on biochemical oxygen demand which is stored in moore.mat padded with noisy versions of the predictors to introduce correlations. Figure below shows the results obtained through the analysis using MATLAB. The size of the y matrix is (20 x 1) where the number of response is 1 whereas the numbers of observations are 20. On the other hand, the size of x matrix is (20 x 5) where the numbers of predictors are 5 whereas the numbers of observations are 20 as well.

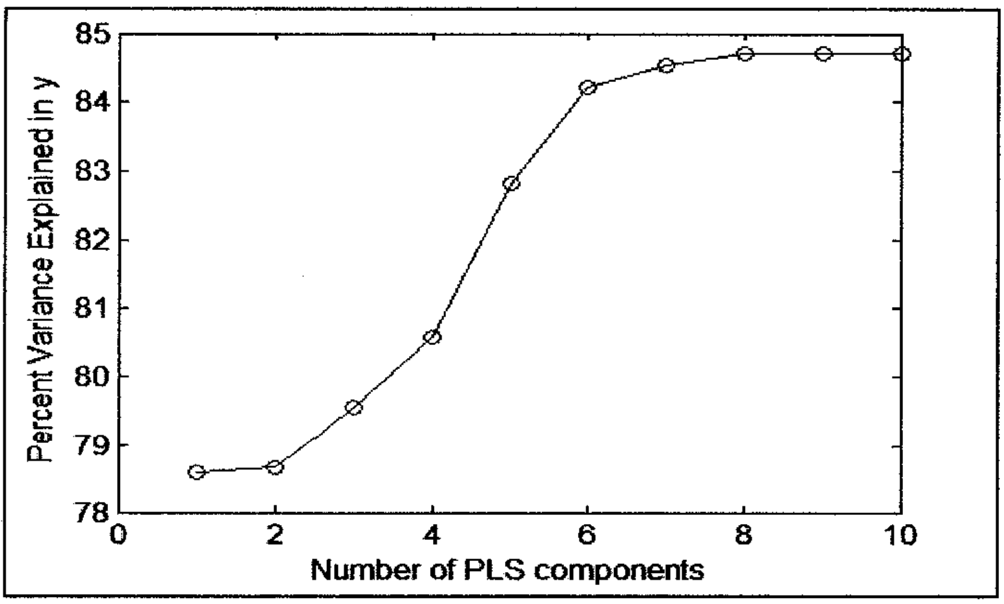

Figure 3.9: Graph of percentage variance explained in the response as a function of the number of components.

Selection of number of components in PLS is a decisive step. The figure below shows the correlation between the fitted and observed responses by computing the r^2 value which is approximately 0.8421.

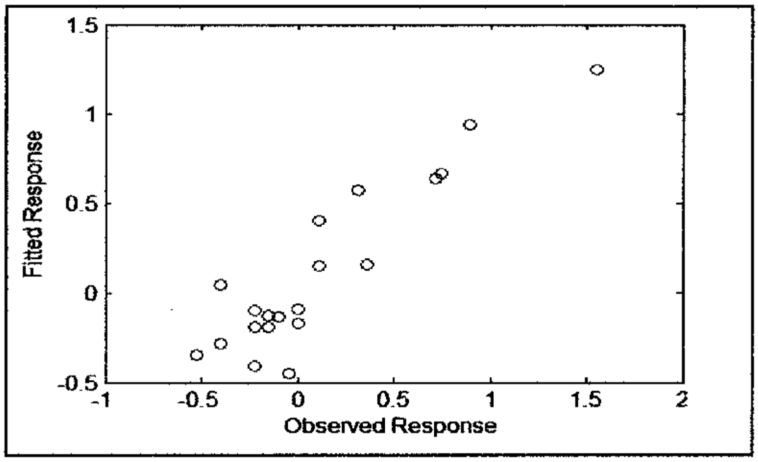

Figure 3.10: Scatter plot shows correlation between the fitted and observed responses.

Furthermore the number of components is increased to 6 components where the weights of the ten predictors in each of the six components show that the two of the components (the last two) describes the majority of the variance in X:

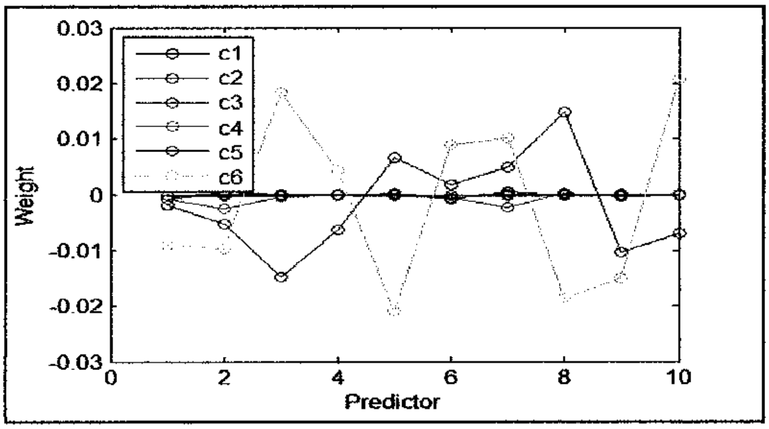

Figure 3.11: Plot of weights of the 10 predictors in each of the 6 components versus the predictors (K).

Finally a plot of the mean squared errors concludes that as few as 2 components may provide a satisfactory model and Figure 12 shows the mean squared errors plot:

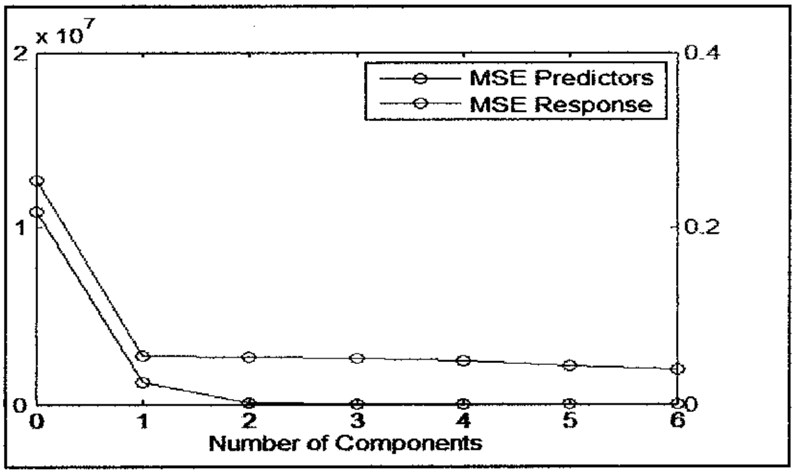

Figure 3.12: Plot of mean squared errors.

3.4 Gantt Chart & Key Milestone

Table 3.1: Gantt Chart and Key Milestone for Final Year Project 2012

No	Detail / Week	1	2	3	4	5	6	7	MID SEM BREAK	8	9	10	11	12	13	14	15
1	Selection of Project Topic																
2	Study on the topic- regarding the soft sensors																
3	Study on the MSPC techniques																
4	Study on the adaptive models for the soft sensors																
5	Submission of Extended Proposal																
6	Study & analyze the mathematical models/algorithms for the models																
7	Proposal Defense																
8	Translate the selected mathematical model into MATHLAB coding																
9	Introduction to the MATHLAB software																
10	Submission of Interim Report Draft																
11	Submission of Interim Report																
12	Analyze the simulation (Distillation Column)																
13	Extract the data from the simulation																
14	Progress Report																
15	Pre-SEDEX																
16	1^{ST} Draft Submission of Dissertation																
17	Submission of soft copy of Dissertation																
18	Submission of Technical Paper																
19	Oral Presentation																
20	Hard Bound Dissertation																

Processes Key Milestone

3.5 Tools Required

The tools required to accomplish this simulation project will be as shown below:

Table 3.4: Tools Required for the Research

TOOLS	DESCRIPTION
MATLAB	• It used to study and analyze the regression techniques such as linear regression, multilinear regression, partial least square regression and principal component analysis.
HYSYS	• Used to perform modification on the simulation of separation acetone from 2-propanol process. • Studied the effect of independent variables (input variables) on the dependent variable (output variable). • Generated data which is required for SIMCA-P software in developing soft sensor model based on PLS algorithm.
SIMCA-P	• Used to study the correlation between the input and output variables based on PLS algorithm. • Performed PCA analysis in order to find any outliers in the imported data. • Used to develop soft sensor model based on PLS algorithm by applying Moving Window approach on the generated data. • Performed validation of the soft sensor in predicting the y variable with higher precision.

CHAPTER 4

CASE STUDY: BINARY DISTILLATION COLUMN

4.1 Introduction

This chapter will discuss about the data analysis of the case study which been simulated in the HYSYS software. Separation of binary mixture of Acetone and 2-Propanol in distillation column (BDC) is selected as the case study of the project. Initially, the simulation model is generated using the calculated operating conditions as shown below and a dynamic mode was chosen to run the model in order to extract the data for further analysis.

Distillation column is a unit operation used to physically separate the mixtures based on the volatility of the components in the mixtures. There are three types of distillation processes which are Batch Distillation, Continuous Distillation and Azeotropic Distillation. The case study for this project is carried out in a continuous distillation which separates the Acetone and 2-Propanol. Below shows the schematic diagram of the selected distillation column:

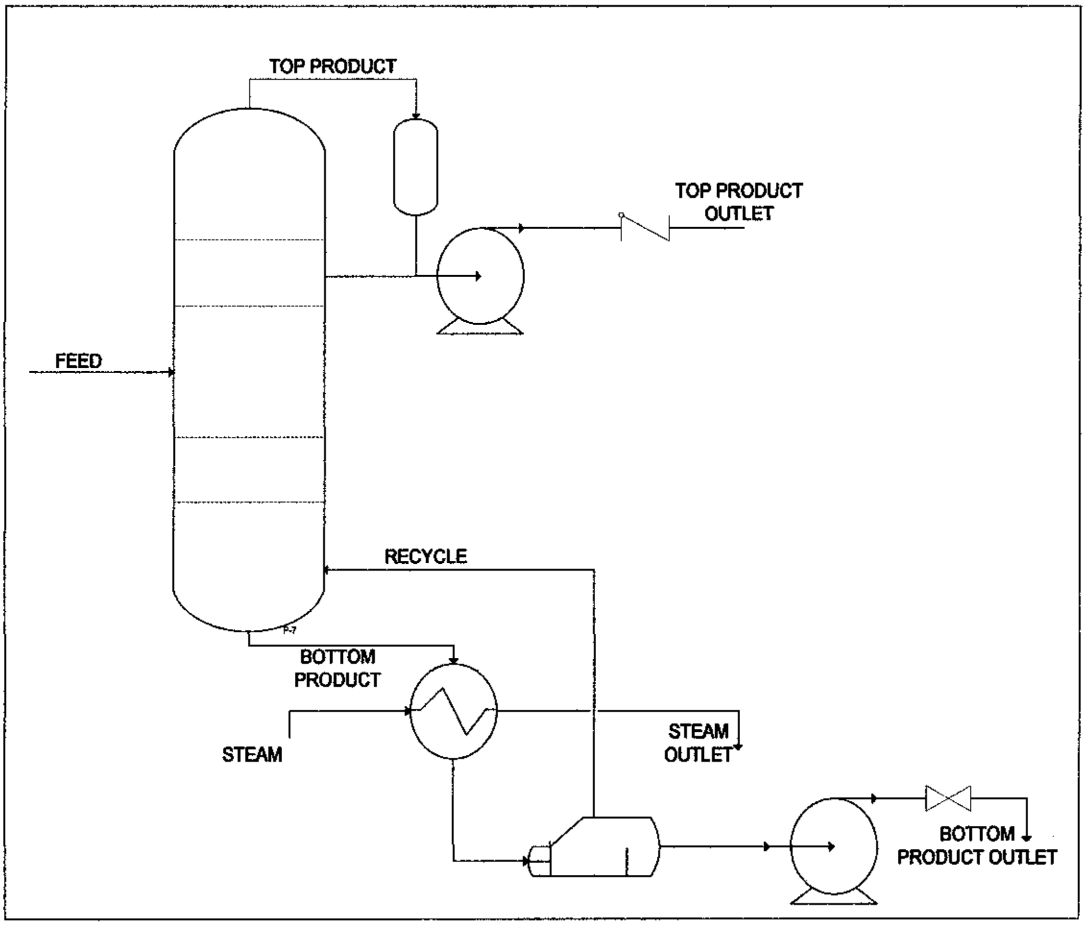

Figure 4.1: Schematic diagram of the distillation column.

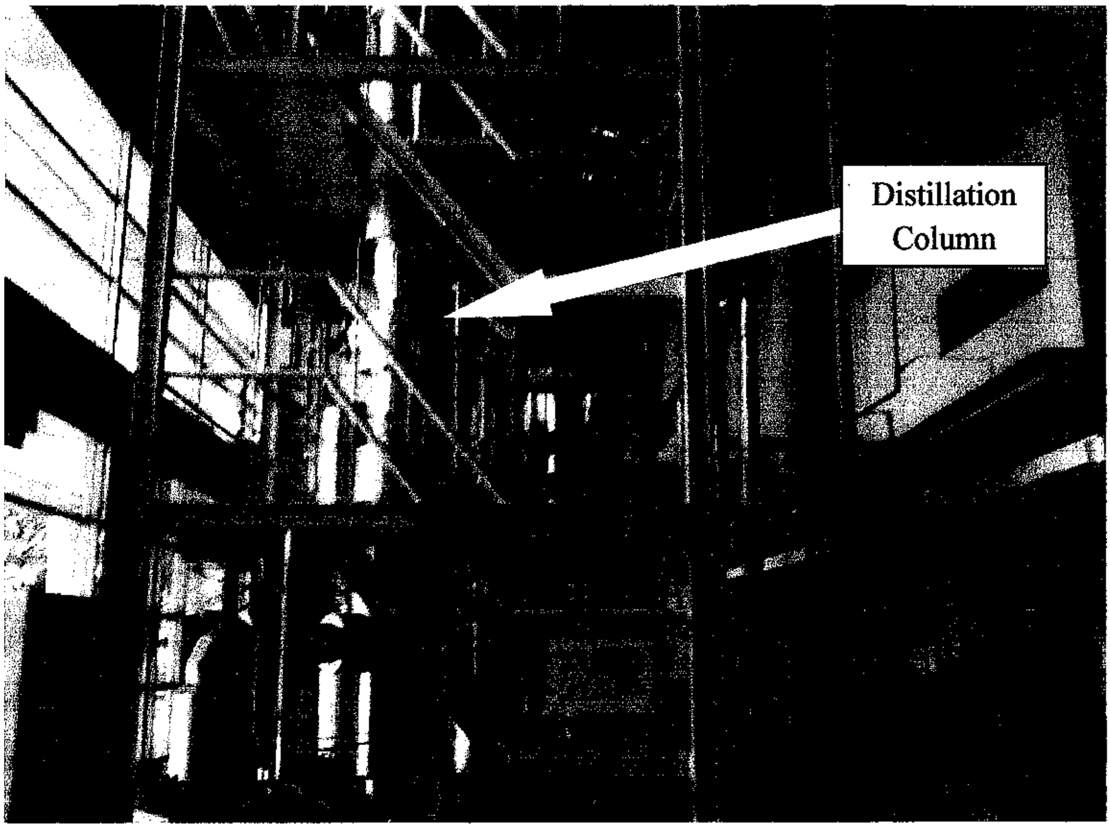

Figure 4.2: Snapshot of the distillation column in the lab.

4.2 Analysis of the Distillation Column

As the simulation generated in HYSYS software encountered few problems where it did not separate the components efficiently, analysis is performed on the simulation to identify the root of problem. Manual calculation is performed to calculate the theoretical values of the operating conditions of the distillation column for this specific case study.

4.2.1 Determination of Column Operating Conditions

To identify the theoretical values of the operating conditions, **Fenske-Underwood-Gilliland** method is chosen because ease of calculation and the method can be used for binary mixture. First of all the heavy and light key components are determined based on the objective of the separation where approximately 98% purity if Acetone is expected as the top product.

More volatile component : Acetone

Less volatile component : 2-Propanol

For further calculations the bubble point and dew point of the streams are determined using the following equation:

Bubble point: $\sum y_i = \sum K_i x_i = 1.0$ Dew point: $\sum x_i = \sum \frac{y_i}{K_i} = 1.0$

Here the value of P_i is the partial pressure of the individual component which will be determined by Antoine Equation:

$$Antoinne\ Equation: \log P = A - \frac{B}{T^{\circ}C + C}$$

Table below summarizes the respective Antoine parameters for the components which present in all the streams. The value of the Antoine parameter below is applicable for Pressure in mmHg and Temperature in degree Celsius.

Table 4.1: Antoine parameters for the components

No.	Component	Antoine Coefficients		
		A	B	C
1	Acetone	7.2316	1277.0300	237.2300
2	2-Propanol	8.1182	1580.9200	219.6200

Table below summarizes the temperature estimated using the above method for the feed, bottom and top streams of the distillation column (Calculation summary is shown in APPENDIX D).

Table 4.2: Summary of the estimated stream temperature

COMPONENT	T (°C)	BUBBLE POINT	DEW POINT
Feed (Liquid Feed)	76.740	1.000	-
Top (Vapor form)	65.048	-	0.9999
Bottom (Liquid form)	82.475	1.000	-

4.2.2 Calculation of Minimum Number of Stages

The equation used to calculate the minimum number of stages is as shown below:

$$N_m = \frac{\log \left[\frac{x_{LK}}{x_{HK}}\right]_d \left[\frac{x_{HK}}{x_{LK}}\right]_b}{\log \alpha_{LK}}$$

α_{LK} = average relative volatility of the light key with respect to heavy key

x_{LK} = light key mole fraction

x_{HK} = heavy key mole fraction

d = distillate (top product)

b = bottom product

The respective values of relative volatility and mole fractions are substituted into the equation for calculating the minimum number of stages and the answer obtained is:

$$\text{Nmin} = 4.8930 \sim \mathbf{5\ stages}$$

4.2.3 Minimum Reflux Ratio

In an operating column the effective reflux ratio will be increased by vapor condensed within the column due to heat leakage through the walls. Equation below is used to determine the minimum reflux ratio for the column:

$$\sum \frac{\alpha_i x_{i,f}}{\alpha_i - \theta} = 1 - q$$

q = thermal condition that depends on the condition of the feed

xi,f = the molar fraction of the component i in the feed

Basically the value of the theta should lie in between the values of relativity volatility of the light and heavy key. The feed is assumed to be saturated liquid (q=1) as it tends to decrease the minimum reflux ratio relative to a vaporized feed.

Table 4.3: Summary of the theta calculation

Component	α(feed)	x(f)	α(feed)*x(f)	α(feed)-θ	Underwood
Acetone	1.0000	0.3000	0.3000	0.3109	0.9650
2-Propanol	0.4164	0.7000	0.2915	-0.2727	-1.0687
				Sum	-0.1037

Value of theta, $\theta = \mathbf{0.68913}$ $(0.4164 < \theta < 1.0000)$

The second equation in order to find the minimum reflux ratio is shown below:

$$\sum \frac{\alpha_i x_{i,d}}{\alpha_i - \theta} = R_m + 1$$

αi = relative volatility of the component, with respect to heavy key component

Rm= minimum reflux ratio

Xi, d = molar fraction of component, I in the tops at minimum reflux

Table 4.4: Summary for the minimum reflux ratio calculation

Component	α(d)	y(d)	α(d)*y(d)	α(d)-θ	Underwood
Acetone	1.0000	0.9000	0.9000	0.3109	2.8951
2-Propanol	0.3612	0.1000	0.0361	0.3612	0.1000
				$R_m + 1$	2.9951

Minimum reflux ratio, Rm =**1.9951**

4.2.4 Estimate Optimum Reflux Ratio and Actual Number of Stages

Generally the reflux ratio and the number of theoretical stages should be greater than the minimum values that have been calculated in the early sections, this is to ensure the separation between the 2 key components is efficient. According to the rule of thumb, the value R is determined based on the following equation:

$$R = (1.2 - - - 1.5)Rm \text{ here factor of 1.5 is chosen to find the R value}$$

$$\mathbf{R = 2.9927 \sim 3.00}$$

An important shortcut method to determine the theoretical number of stages required for an operating reflux ratio, R is the empirical correlation of Erbar-Maddox correlation shown in below figure. This correlation is believed to give highly reliable predictions which give the ratio of number of stages required to the number of total reflux, as a function of the reflux ratio with the minimum reflux ratio as a parameter. According to the graph, the value obtained is:

$$\frac{Nm}{N} = 0.63 \text{ and } Nm = 9 \text{ stages}$$

$$N = \frac{Nm}{0.63} = 14.2118 \sim \mathbf{15\ stages} \text{ including reboiler}$$

Table below summaries the operating conditions of the distillation column which was generated in the HYSYS software to separate Acetone and 2-Propanol:

Table 4.5: Summary of the Design Operating Conditions

PARAMETERS	DESIGN VALUES
Height	5500 mm
Diameter	150 mm
Number of Trays (Actual)	15
Type of Trays	Bubble Cap
Tray Spacing	350 mm
Feed Tray Location	7
Column Pressure	101.32 kPa
Column Temperature	76.74 C
Reflux Ratio	3.0

The calculated design operating conditions were taken as input to HYSYS model which was run under dynamic mode. The following shows the summary of the parameters determined by the user to analyze the model:

Logger Size (# Sample)	: 4320
Integration Time	: 1440 minutes
Sample Interval	: 20 seconds

4.3 Model Setback

Before the model was used to perform testing, certain steady state problems were solved first. Those were:

PROBLEM	CONSEQUENCES	SOLUTIONS
Mole fraction of acetone in top product stream was not 98% purification	Efficient separation does not occur which incapable to reach the objective of the separation (complete separation of acetone from 2-propanol)	Performed step input change on process variables such as reflux flow rate, steam flow rate, and feed flow rate
Fluctuation in the feed flow controller (FIC-feed)	Affect other parameters such as composition of the products and feed flow rate	Modified the tuning parameter of the controller (K_c and T_i)
Tray efficiency was too low (6%)	In able to reach the desired composition of acetone in top product stream due to low column efficiency	Increased the column efficiency up to 85%
Reflux ratio was very small	Limited amount of flow back to the column (inefficient separation)	Performed changes to the reflux flow rate, feed flow rate and steam flow rate until desired reflux ratio is achieved (3.0)

Table 4.6: Steady State Operating Parameters

PARAMETER	STEADY STATE VALUES
Feed Flow Rate	0.6646 kmol/h
Reflux Flow Rate	1.0511 kmol/h
Distillate Flow Rate	0.1974 kmol/h
Bottom Flow Rate	0.4677 kmol/h
Steam Flow Rate	1.001 kmol/h
Reflux Ratio	5.325
Mole Fraction of Acetone (Top Product)	0.9843
Mole Fraction of 2-Propanol (Top Product)	0.0269
Reboiler Duty	40 720 kJ/hr

The modified simulation was used as the steady state model to perform step changes in the following input variables:

- Feed Flow Rate, kmol/h
- Feed Temperature, °C
- Steam Flow Rate, kmol/h
- Reflux Flow Rate, kmol/h

Figure 4.3 shows the snapshot of the PFD of the model generated in the HYSYS software for separation of Acetone and 2-Propanol whereas Figure 4.4 shows the method of step changes which been performed to the input variables:

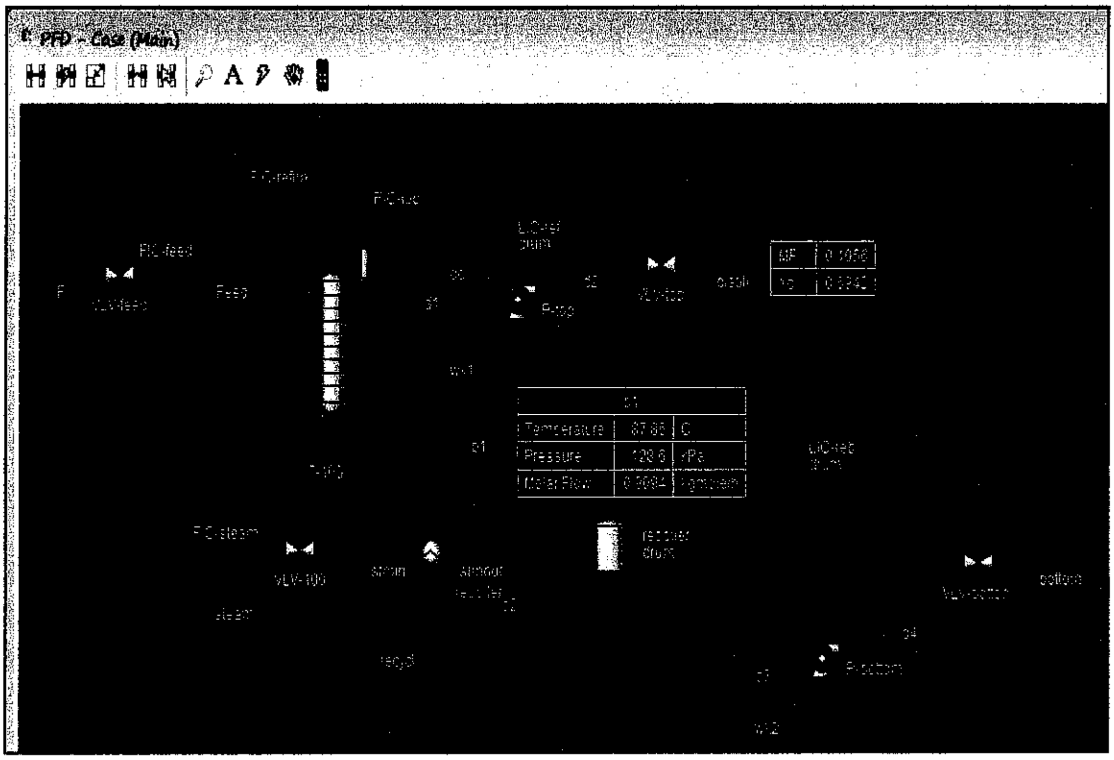

Figure 4.3: Generated model of the case study in HYSYS software.

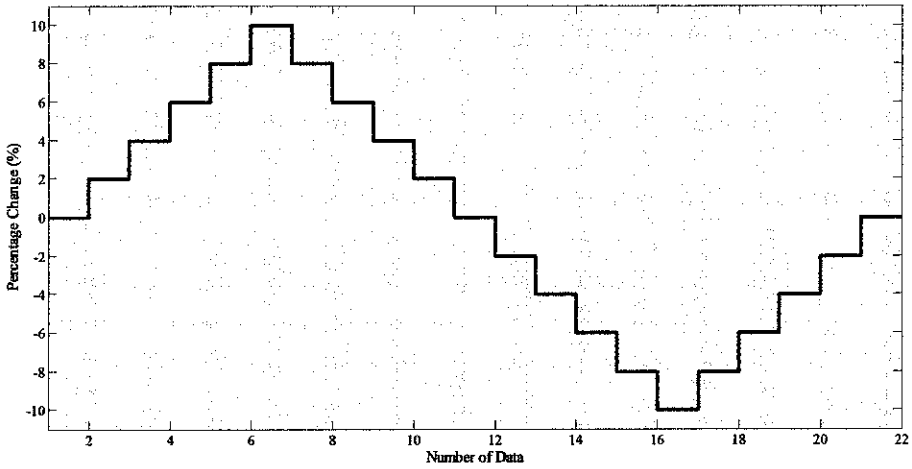

Figure 4.4: Method of step change on the input variables.

Conclusively, separation of acetone from the 2-propanol process is taken as the case study for the project and the operation conditions for the process were calculated before the simulation is run in HYSYS software to generate data. The input variables (feed flow rate, steam flow rate, reflux flow rate and feed temperature) were varied in the range of ±10% based on the method shown in Figure 4.4 and the corresponding data is recorded to study the impact of each input variable on the output variable and state variable. The generated data is used into SIMCA-P software to observe the correlation between the variables before due to some process drifts. This analysis is further explained in following chapter.

CHAPTER 5
RESULTS AND DISCUSSION

5.1 Testing Data

As explained in previous chapter, the calculated parameters of the operation conditions were used to perform simulation of the case study in HYSYS. Studies on the generated data are performed with an objective of knowing the relationship between the ranging input variables on the output variable (mole fraction of acetone). The impact of input variables on the output variable (mole fraction of acetone) and also on the state variable (temperature profile of the trays in BDC) were studied initially in this chapter. The state variable also been observed in this research because changes in the state variable will also indicate the changes in the output variable, this is explained in a mathematical function below:

$$\dot{x} = Ax + Bu \; and \; y = C\dot{x}$$

where $\dot{x}$ = variation in the state of the column

X = state variable and u = input variable

A, B, C = coefficients and y = output variable

As indicated in the equation above, variation in the state of the column not only includes the input variables (feed flow rate, steam flow rate, feed temperature and reflux flow rate) but also the state variable (temperature profile of the trays). Below shows the results of the analyzed data and this analysis covers the impact of varying input variables on the output variable and also on the state variable. Furthermore, PLS software provided by SIMCA-P is used to further data analysis.

5.1.1 Effect of Varying Feed Flow Rate

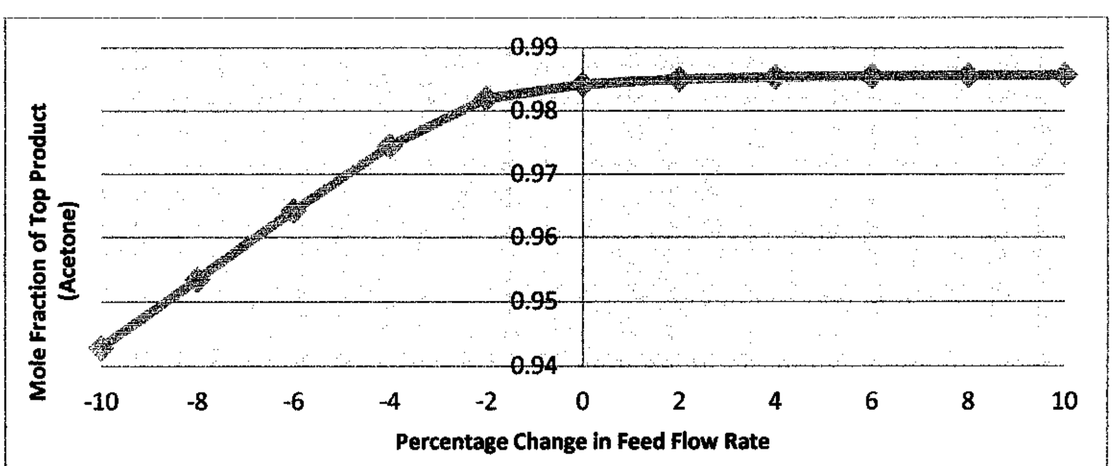

Figure 5.1: Relationship between Mole Fraction Acetone (Top) and Feed Flow Rate

Interpretation: As evident in Figure 5.1, the variation on the mole fraction is larger from the range of -10% up to the steady state condition. The feed reached its optimum separation (optimum purification) where further increment in the feed flow rate will not affect the mole fraction of acetone. The non-linearity characteristic of the variable can be observed where the output variable is not directly proportional to the input variable. Trays temperatures of BDC also taken into consideration for each step input change.

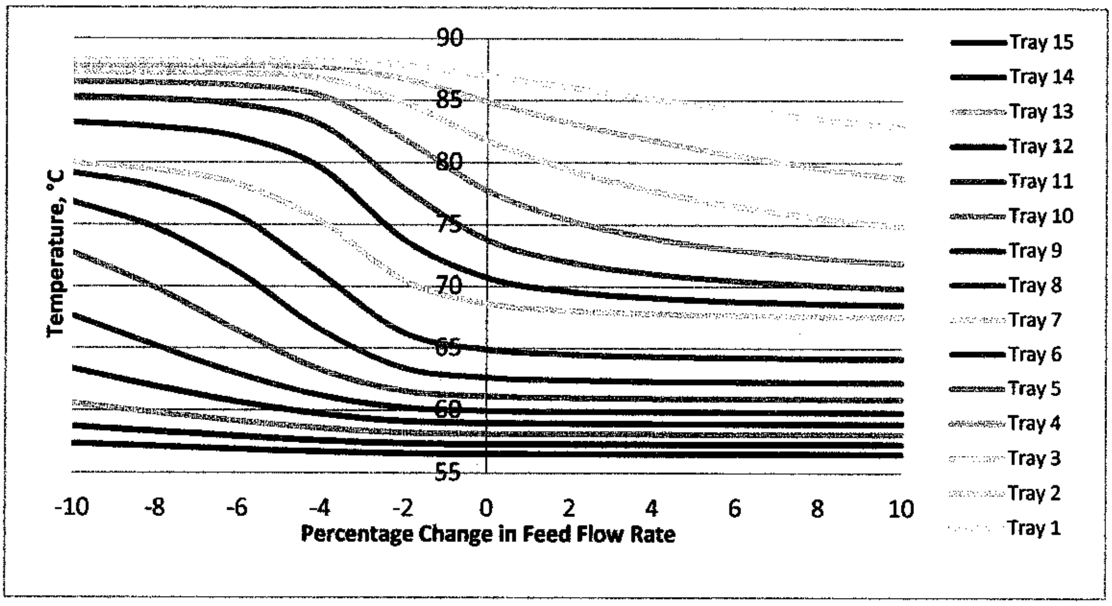

Figure 5.2: Relationship between Temperature of the Trays and Feed Flow Rate

Observation: Conclusively, the trays temperatures corresponded to the changes in the feed flow rate. The feed enters the distillation column at Tray 7 and temperature profile differs for the entire trays. The graph concludes that Tray 7 went through drastic change in the temperature due to the changes in feed flow rate. Trays in the stripping section of the distillation column exhibits lesser variations in the temperature profile. However, the case is different in the rectifying section where the temperature profile changes drastically.

5.1.2 Effect of Varying Steam Flow Rate

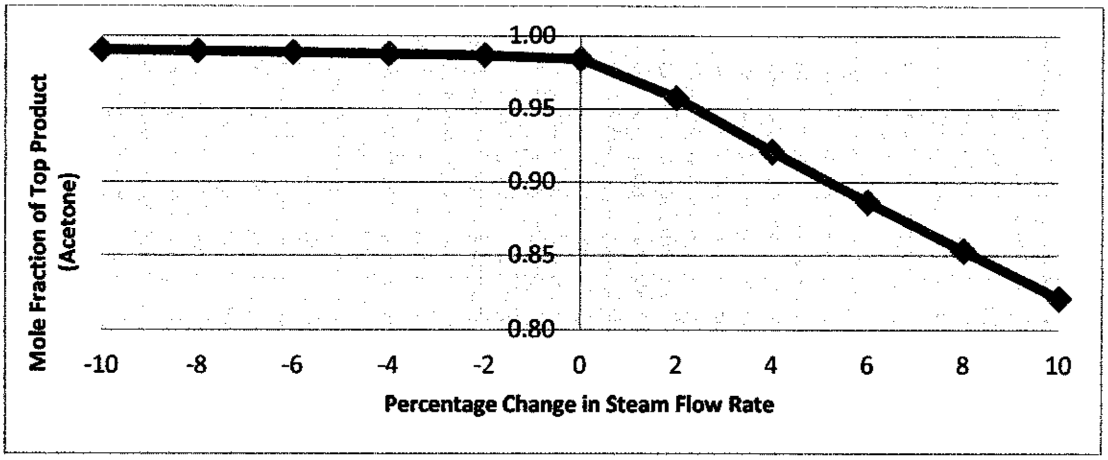

Figure 5.3: Relationship between Mole Fraction Acetone (Top) and Steam Flow Rate

Interpretation: In Figure 5.3, the variation of steam flow rate from 0% to 10% gave drastic affect to the mole fraction of acetone compared to the variation of steam flow rate from -10% to 0%. The less purity of acetone is obtained as the steam flow rate increases because of decrement in the reflux flow rate (decrement in flow of the fluid back to the column).

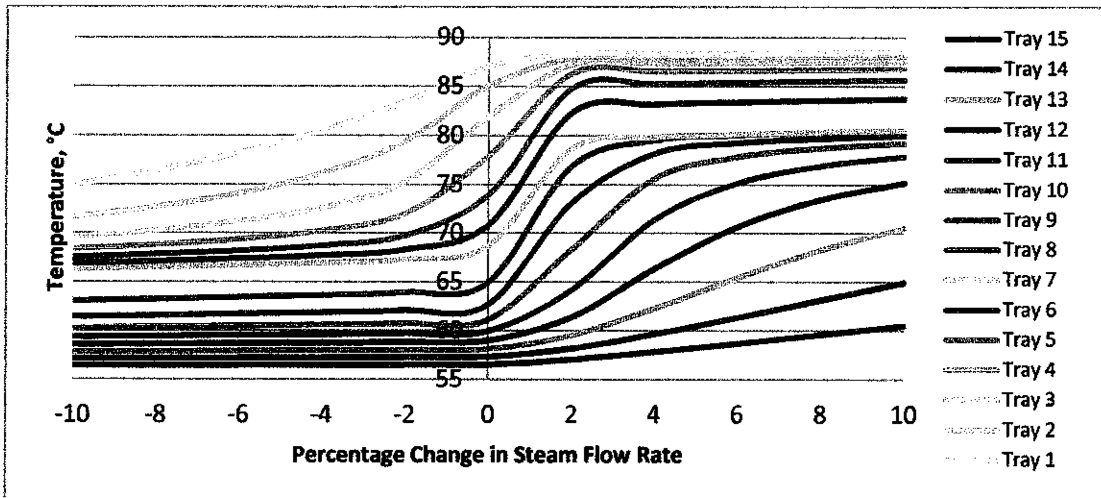

Figure 5.4: Relationship between Temperature of the Trays and Steam Flow Rate

Observation: The most affected tray due to the changes in the steam flow rate is Tray 10 where the temperature difference between the variations is larger than other trays (18.96 °C). Tray 15 was least affected by the changes in the steam flow rate with the temperature difference of 4.10 °C. Trays in the stripping section of the distillation column went through less variation compared to trays in the rectifying section. Rectifying section is where the concentration of acetone increases in liquid and vapor. On the other hand, stripping section is where the mole fraction of acetone decreases in liquid and vapor. The BDC operates under colder condition when the supply of steam decreases and vice versa.

5.1.3 Effect of Varying Reflux Flow Rate

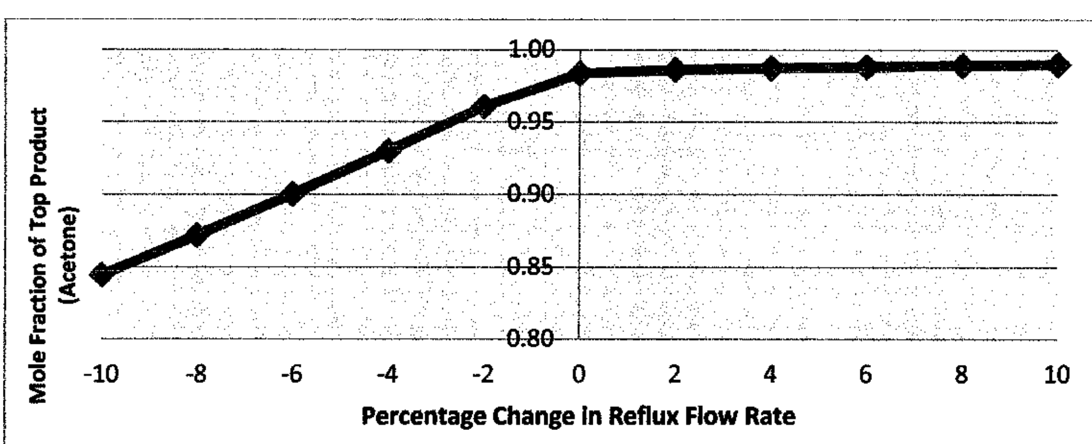

Figure5.5: Relationship between Mole Fraction Acetone (Top) and Reflux Flow Rate

Interpretation: Figure 5.5 reveals the drastic change of mole fraction of acetone from the range of -10% till 0% of reflux flow rate. Furthermore, acetone reached its optimum concentration when the reflux flow rate is varied from 0% to 10% and considerably the process has reached its optimum separation (optimum purification). Further increment in the percentage change in reflux flow rate, the mole fraction will not change in a larger amount because the changes almost reach stability (optimum). There is non-linearity characteristic existed in the relationship between input variable (reflux flow rate) and output variable (mole fraction of acetone). As the project emphasis on developing soft sensor for time-varying factor, reflux flow rate can be taken as the input variable or independent variable due to its effective impact on the mole fraction of acetone and also the non-linearity characteristic.

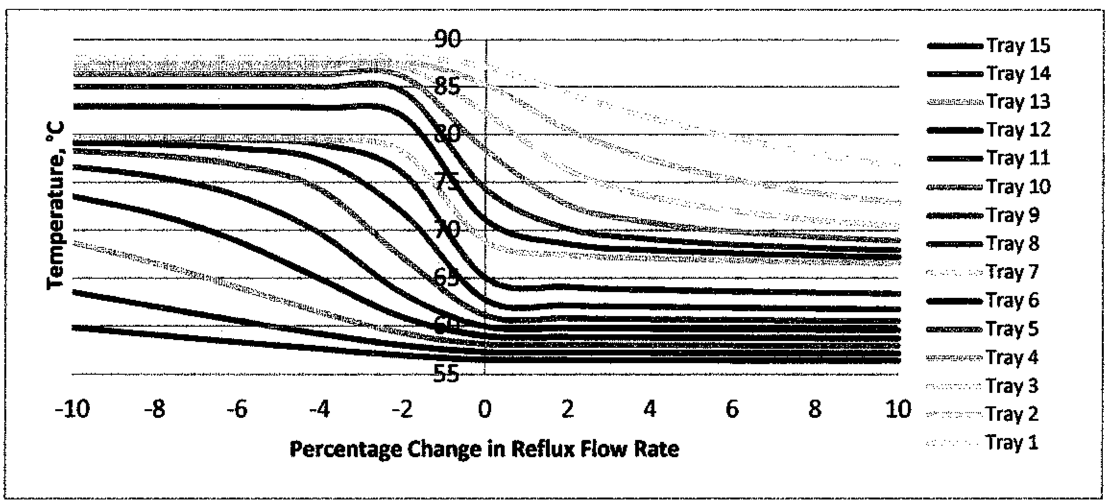

Figure 5.6: Relationship between Temperature of the Trays and Reflux Flow Rate

Observation: Convincingly, the temperature profile of the trays responded to the changes in the reflux flow rate. The most affected tray was the feed tray (Tray 7) where the slope of the graph is larger compared to others. The temperature profile differs for rectifying section (Tray 8 till Tray 15) and stripping section (Tray 1 till Tray 6). Trays in the stripping section of the distillation column displays smaller variations in the temperature profile meanwhile in the rectifying section, the temperature profile exhibits larger variations.

5.1.4 Effect of Varying Feed Temperature

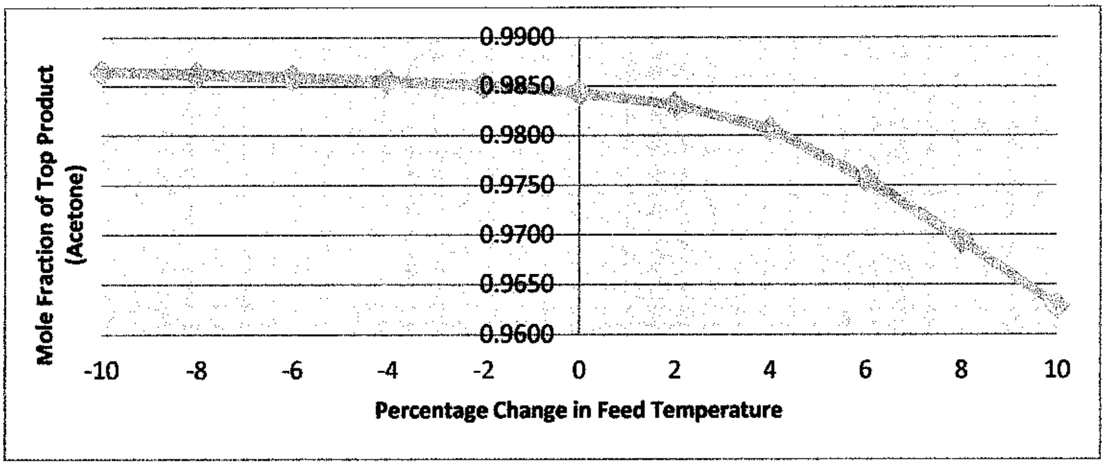

Figure 5.7: Relationship between Mole Fraction Acetone (Top) and Feed Temperature.

Interpretation: Effect of percentage change in feed temperature on mole fraction of acetone is inverse compared to feed flow rate and reflux flow rate. In Figure 5.7, the variation of steam flow rate from 0% to 10% gave drastic affect to the mole fraction of acetone compared to the variation of steam flow rate from -10% to 0%. Separation of acetone from 2-propanol reached its optimum when the feed temperature was decreased from the steady state condition. The feed temperature was not varied more than the range of ±10% because it can cause the column to operate under extreme conditions (hot or cold condition). Conclusively, the feed temperature gives a prior effect on mole fraction of acetone. For each step input change, the trays temperatures are taken for analysis and below shows the result obtained from the analysis.

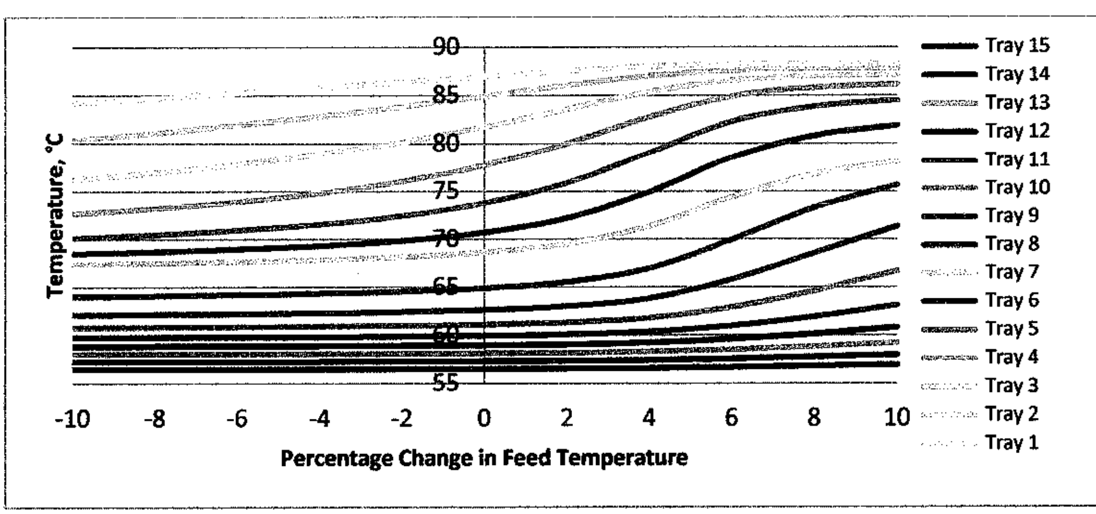

Figure 5.8: Relationship between Temperature of the Trays and Feed Temperature

Observation: As observed in Figure 5.8, the most affected tray was the feed tray (Tray 7) where the slope of the graph is larger compared to others. The temperature profile differs for all the 15 trays. Trays in stripping section (tray 1 till tray 6) did not go through much variation in the temperature profile compared to the trays in rectifying section (tray 8 till tray 15). From the analyses of trays temperature profiles, the impact of each tray on mole fraction of acetone can be determined. Convincingly, tray 8 till tray 15 gave larger impact on the concentration of acetone and as a result those data can be used as an input variable in developing correlation between independent variables and dependent variable.

Upon the analysis, the selected data is used to develop correlation between input variables and output variables using PLS algorithm. SIMCA-P software established by Umetrics AB is used to develop the model.

5.2 Data Analysis

5.2.1 Development of Model

In SIMCA-P, predictors (x) and dependent variable (y) are specified earlier in order to find relationship between those variables by PLS algorithm. Table below shows the selected Predictors (x) and Dependent variable (y) for the separation of acetone from 2-propanol process.

Table 5.1: Selected Dependent Variable and Predictors

Dependent Variable (Y)	Mole Fraction of Top Product (Acetone)	
Predictors (X)	Feed Temperature	Reflux Molar Flow Rate
	Feed Molar Flow Rate	Steam Molar Flow Rate
	Tray 1 Temperature	Tray 9 Temperature
	Tray 2 Temperature	Tray 10 Temperature
	Tray 3 Temperature	Tray 11 Temperature
	Tray 4 Temperature	Tray 12 Temperature
	Tray 5 Temperature	Tray 13 Temperature
	Tray 6 Temperature	Tray 14 Temperature
	Tray 7 Temperature	Tray 15Temperature
	Tray 8 Temperature	

As mentioned in Chapter 3, the generated data from HYSYS need to be pre-processed before import into SIMCA-P. Usage of raw data can cause an unequal spread of data in the scatter plot and the results would only reveal the deviation in one of the input variable (imbalance weightage of input variables on output variable). As a solution, the input variables were pre-treated using **Unit Variance Scaling (UV)** and mean-centering method. UV scaling provides an equal variance for all the variables via:

- Standard deviation for each variable is calculated (S_k)
- Obtained the scaling weight, ($1/S_k$)
- Each observation of the variables is multiplied with the scaling weight

Moreover, the scaled data will be **mean-centered** to improve the interpretability of the model. Here, the average value of each scaled variable will be computed and then subtracted from the data. After the pre-processing step, all the variables listed in Table 5.1 were given equal importance/weightage and can be used for further analysis and to develop the model. Below diagrams shows the analysis performed in developing soft sensor model using Moving Window approach by PLS algorithm.

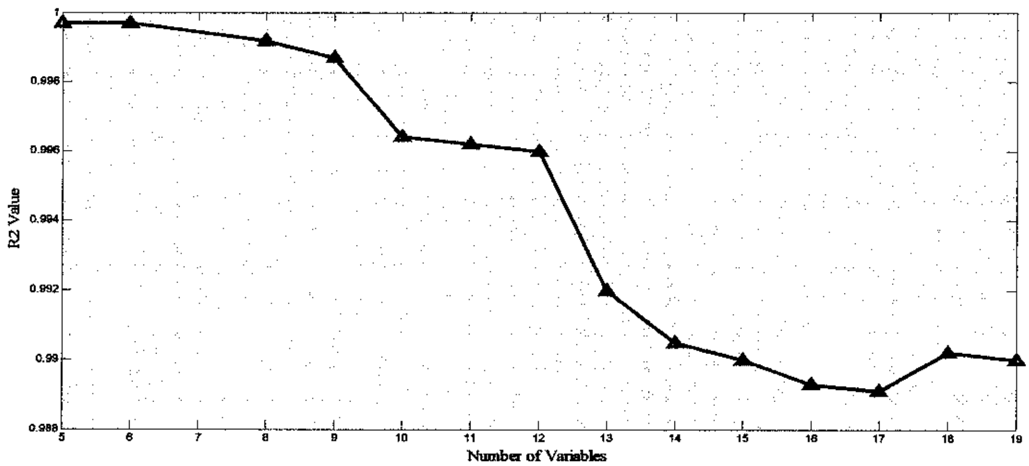

Figure 5.9: R^2 value vs. Number of Variables.

Initially, the model is developed by importing all the input variables generated from the HYSYS into SIMCA-P software. The PLS algorithm in SIMCA-P will identify the number of variables which gives larger impact on the output variable (mole fraction of acetone) and useful in predicting the y variable in future. The variables which are not affecting much on the y variable will be discarded one by one until higher accuracy model is obtained in predicting the mole fraction of acetone. As shown in Figure 5.9, fourteen models were developed with different number of input variables (some of variables were discarded due to its less contribution on y variable) in order to study the accuracy and precision of the model in estimating the mole fraction of acetone (y-predicted = y-observed). Table 5.2 shows the summary of the chosen cases to develop the PLS model:

Table 5.2: Selected Cases to develop the PLS Model

CASE	Number of Variables
1	19 variables
2	10 variables
3	5 variables

The accuracy of the model is studied by referring to the R^2 value of the regression line (Y-Predicted vs. Y-Observed). R^2 value basically provides info about the goodness of fit of the model and it measures how good the regression line can guesstimate the actual data points. Model with higher R^2 value is selected as the inferential model which will be used in studying the effect of Moving Window approach where few data points will be discarded and new set of data points will be added into the window (frame).

PLS Model of Case 1

Initially, the PLS model is developed by using all the 19 process variables as the predictors and 1 dependent variable and the results obtained are shown below:

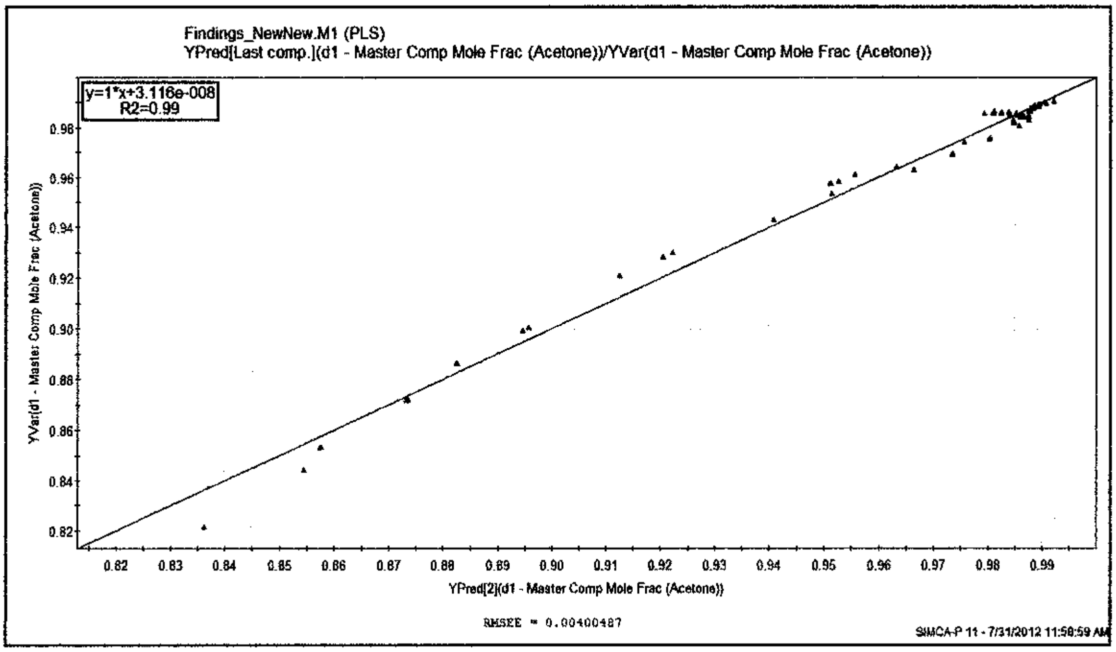

Figure 5.10: Regression line for the original data

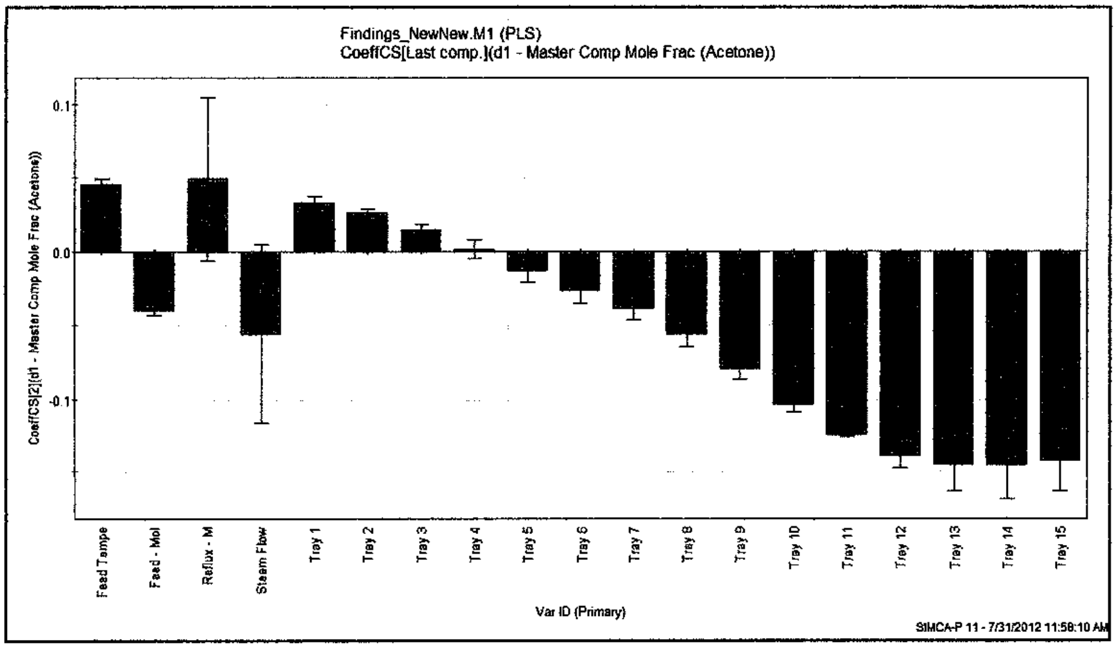

Figure 5.11: Coefficient Plot for the Original data

Interpretation: The R^2 shows the value of 0.99 but as observed in Figure 5.10, there are lots of data points which are away from the regression line. Furthermore, the Coefficient Plot in Figure 5.11 reveals the contribution of each variable on estimating the y-output (mole fraction of acetone). The model can be modified by excluding some of the variables which do not affect much the mole fraction of acetone such as feed flow rate, feed temperature, reflux flow rate, steam flow rate, and temperature profile of tray 1,3,4,5 and 6.

PLS Model of Case 2

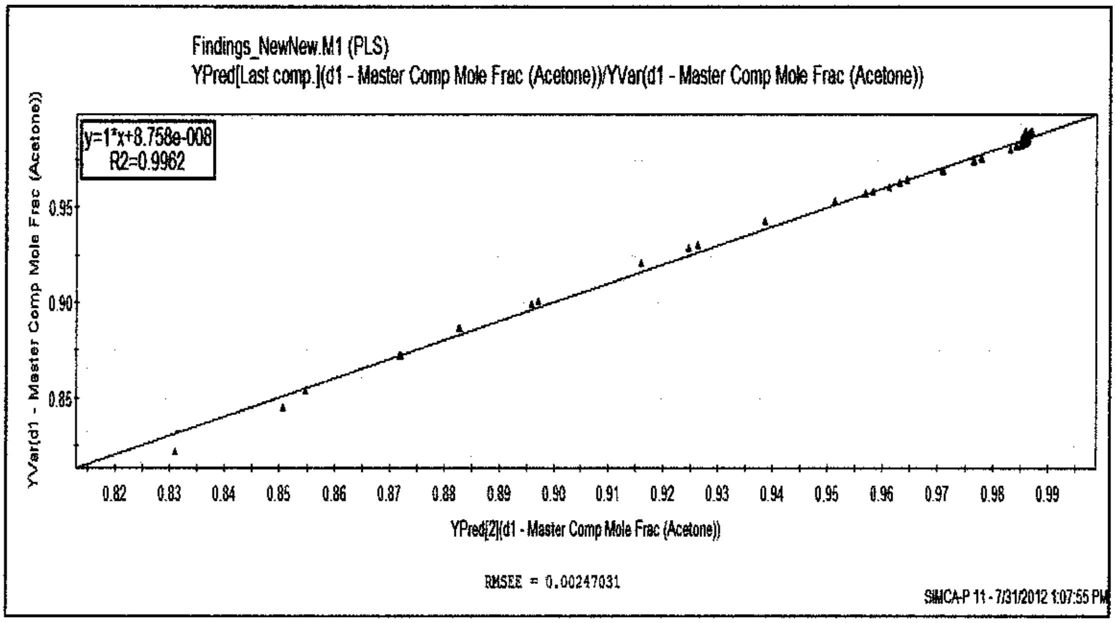

Figure 5.12: Y-Predicted vs. Y-observed plot after excluding 9 of the variable

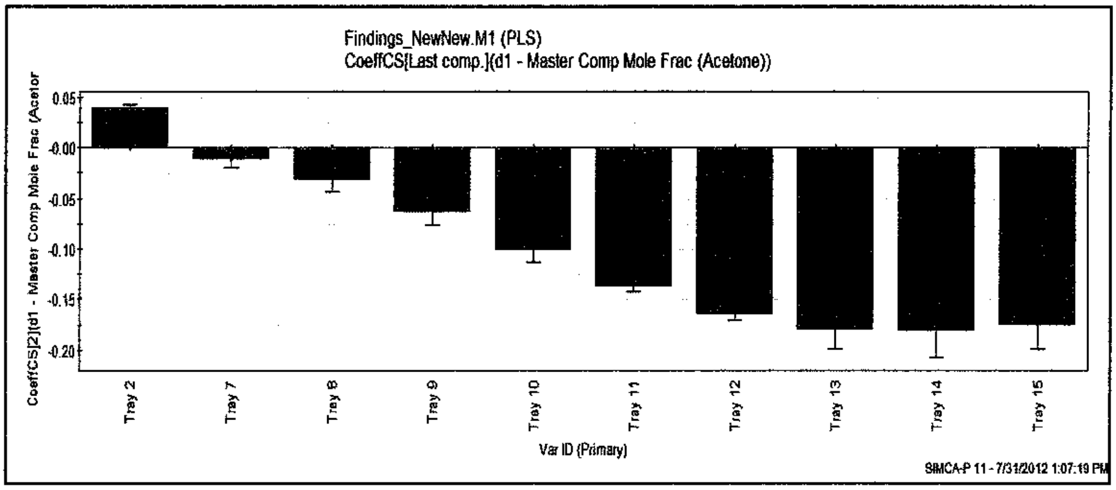

Figure 5.13: Coefficient Plot for the 10 variables.

Interpretation: Figure 5.12 shows the reconstructed regression line after removal of nine variables which the variation does not affect the mole fraction of acetone. The new R^2 shows the value of 0.9962, increased from the previous model. Compared to previous model, this model could predict better the y-output as can be seen the data points in Figure 5.12, are brought closer to the regression line. On the other hand, Figure 5.13, proves that all the selected variable do give an effective impact on the mole fraction of acetone. But the model can be transformed to a better version by removing certain variables such as temperature profile of tray 2, 7, 10, 11 and 12 due to smaller contribution on acetone's concentration. The main objective is to develop a model that can estimate the y-output (Y-Predicted) with higher accuracy so that the y-predicted by the model can reflect the y-actual of the process.

PLS Model of Case 3

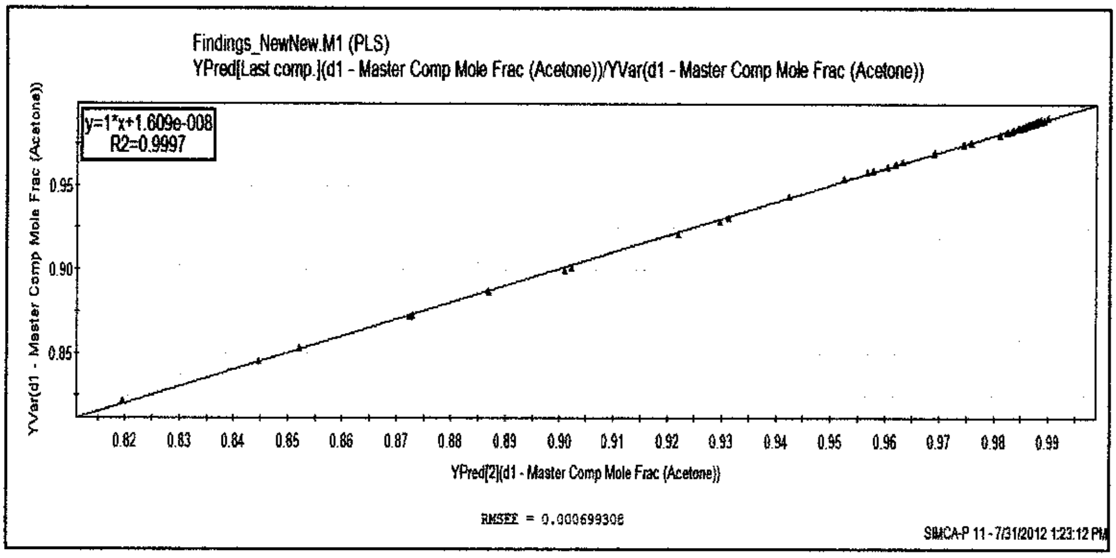

Figure 5.14: Y-Predicted vs. Y-observed plot after excluding 14 of the variable.

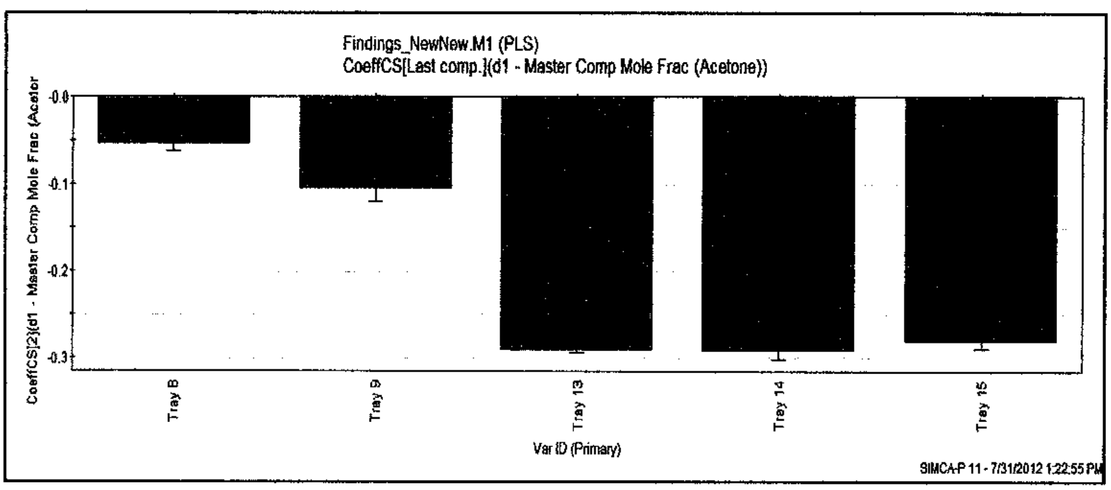

Figure 5.15: Coefficient Plot for the 5 variables

Interpretation: After removal fourteen variables, the model developed can be said satisfactory with the **R^2 value of 0.997** (almost reaching 1.0) and as observed in Figure 5.14, almost all the data points are on the regression line. The equation obtained from Figure 31 is $\boldsymbol{y = x + (1.609 \times 10^{-8})} \sim \boldsymbol{y = x}$and this indicates that y predicted is similar with y observed. Additionally, Figure 5.15 reveals the coefficient plot for the selected variables where the coefficients refer to the pre-processed input variables (UV scaling and mean centered) and output variable (UV scaling and not mean centered). Convincingly, all the variables gave almost equal contribution on mole fraction of acetone and from the coefficient plot the PLS model can be written as regression model as shown below:

$$Y = Y_{average} + XB + F$$

where F = model error and for this successful model the error is almost to the value of zero and it is omitted from the equation

X = input variables (independent variables)

B = coefficient of the respective input variables

Y, Y average = mole fraction of acetone

The regression model for the successfully developed model is as shown below:

$$\boxed{\boldsymbol{x_D = 24.3825 + [-0.0524 \quad -0.1038 \quad -0.291 \quad -0.2917 \quad -0.2817]\begin{bmatrix} T_8 \\ T_9 \\ T_{13} \\ T_{14} \\ T_{15} \end{bmatrix}}}$$

where T_i = Tray temperature of the selected trays in BDC

x_D = mole fraction of acetone in top product

As the developed model focuses on data generated when the efficiency of BDC is 0.85, it accommodates the time-invariant data only. In order to prove the model can be used for estimation of y-output for time-varying data, more data are generated from HYSYS by varying the efficiency of the column because efficiency reflects the characteristic of the column.

5.2.2 Validation of the Model

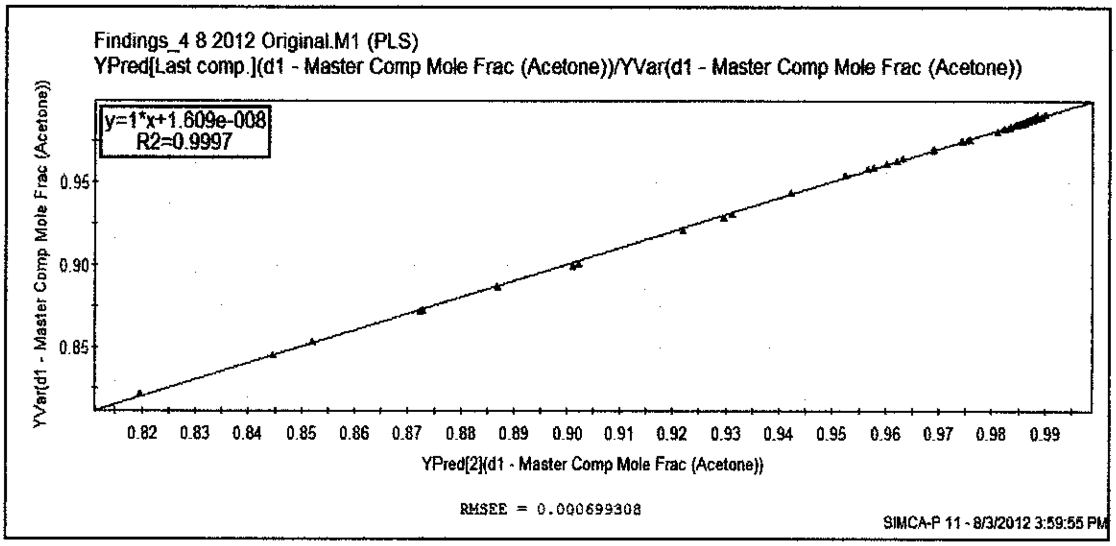

Figure 5.16: Regression Line Plot for the Model including New Set of Data.

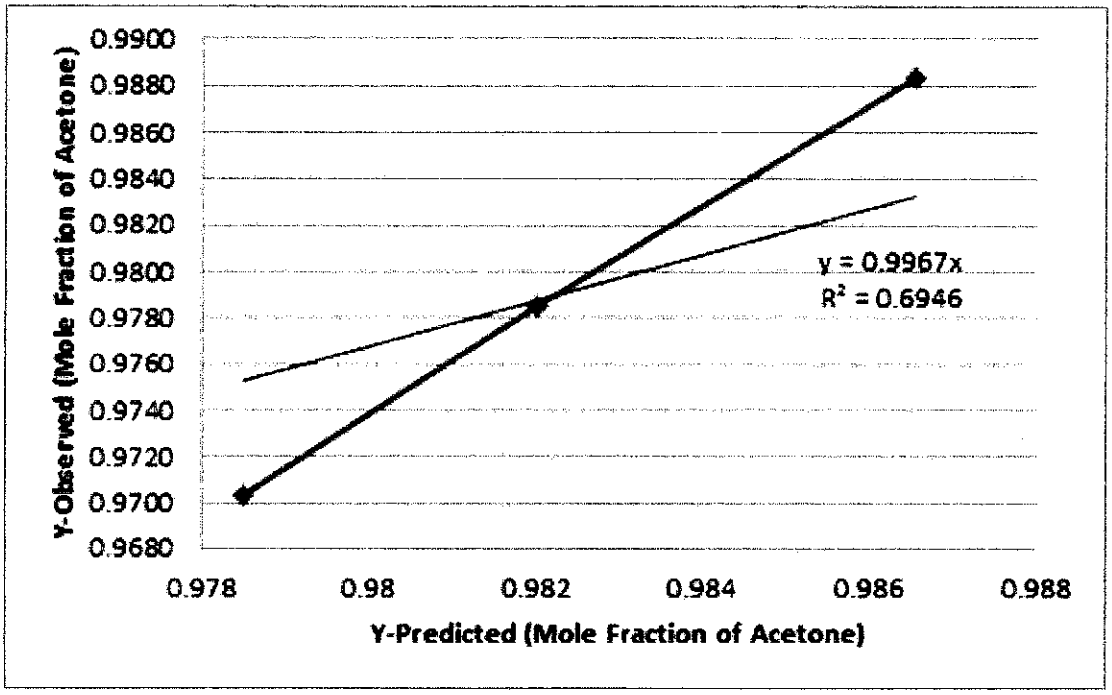

Figure 5.17: R^2 Value for the New Set of Data (3 Data Points).

Interpretation: Adaptation of the model to time-varying factor of the process is checked by validating it with different set of data. Changes in the efficiency of the trays also change the characteristic of the column which reflects the time varying factor. Essentially three test run is performed for efficiency of 0.95, 0.75 and 0.65 without varying any of the x variables. Prediction of new observations is performed by importing the new set of data into prediction set of the developed model in SIMCA-P. The raw data was pre-processed before estimating the y variable by fitting it into the model. The predicted values of y-variable (mole fraction of acetone) did not vary much from the observed values from the simulation. The R^2 values obtained after fitting the new three set of data are 0.995. This verifies the developed soft sensor model deteriorate with time varying factor of the process and can't accommodate the variation in the process by predicting or estimating the difficult-to-measure variable (concentration of acetone) with less precision. This validation is performed by adding new set of observation into the existing model and without discarding any of the observation. Figure 5.17, shows the R^2 value for the new set of data decreases to 0.6946 which proves the data did not fit perfectly on the line (Y-predicted not equal to Y- observed). This is because the developed model can't accommodate the time-varying phenomena of the process. Considerably it portraits the conventional soft sensor which still uses the historical data which no longer represent the current process condition and it tries to accommodate the time-varying data of the process. New sets of models are developed by applying the concept of Moving Window and the results are as shown below.

5.2.3 Development of Models based on Moving Window Approach

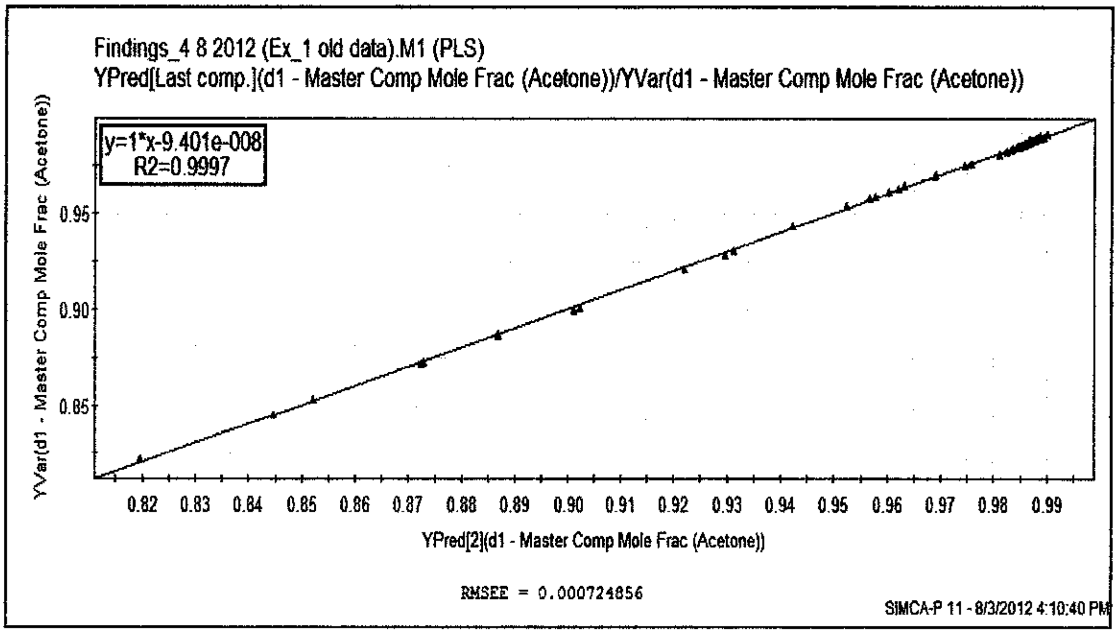

Figure 5.18: Model 1 by Discarding 1 Data Point.

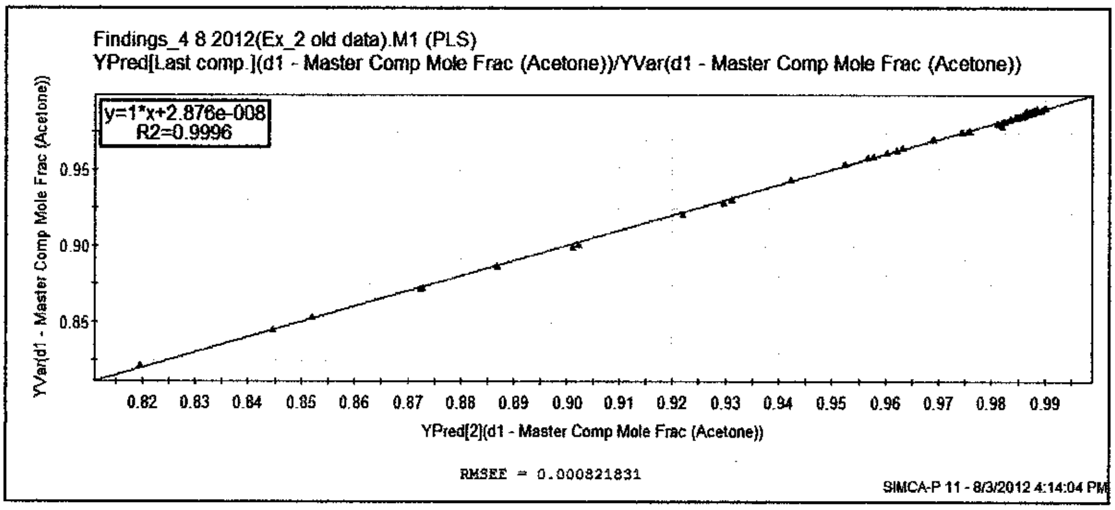

Figure 5.19: Model 2 by Discarding 2 Data Points.

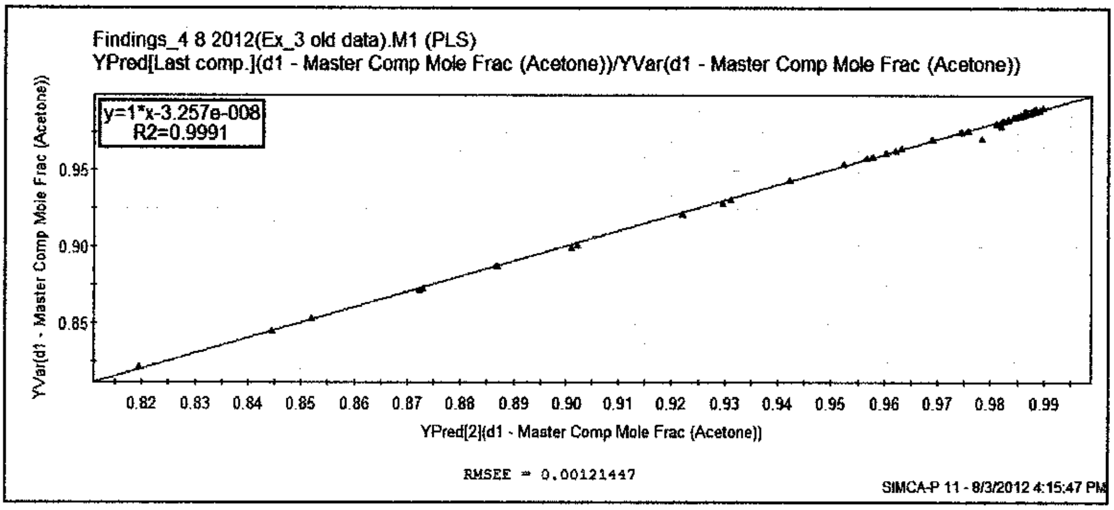

Figure 5.20: Model 3 by Discarding 3 Data Points.

Interpretation: Moving Window approach was applied on the developed model with the generated data (different efficiency of trays). The numbers of variables are fixed (temperature profile of tray 8, 9, 13, 14 and 15). For the first trial, the moving window will discard 1 observation from the frame and include 1 new observation (data generated for tray efficiency of 0.95) and perform the PLS computation. The new R^2 for the regression line is 0.9997 which proves the data fitted perfectly and y-predicted equal to y-observed. The similar procedures were carried out by discarding 2 and 3 observations alternately and the number of observation is fixed (84 observations). As observed in Figure 36, 37 and 38, the models can estimate the mole fraction of acetone with higher precision and accuracy. Conclusively, soft sensor model is developed which can accommodate the time varying factor of the process. Furthermore, Moving Window approach is used in the computation on estimating the output variable by discarding the historical data which no longer represent the process can be successfully applied on the developed soft sensor model.

Overwhelmingly, this chapter discussed about the data analysis of the project, from the point data is generated in HYSYS software until the soft sensor model is developed and validated using SIMCA-P software. From the research, it is proven that variables such as feed flow rate, steam flow rate, reflux flow rate, feed temperature, and reflux ratio and temperature profile of trays give an effect on the mole fraction/composition of the products. Due to this effect, a mathematical model can be built relating those variables with the composition of the products. This mathematical model can be used in estimating the product quality instantaneously without an effort of taking infrequent sampling to be tested in the laboratory which eventually changes the quality of the product caused by significant time delay. Input variables which affect the mole fraction as mentioned above are taken as predictors (X) while the composition of acetone is taken as the dependent variable. By using the PLS software provided by SIMCA-P; the mathematical model will be created using the generated data from HYSYS. After analysis, it is verified that soft sensor model with the highest accuracy (R^2 value of 0.9997) can be developed by using only five predictors, which gave larger impact on the dependent variable. Decrement of predictors from nineteen variables to five variables, helped to reduce the computation load to the model and can provide a faster response on estimating the y-variable. Moreover, the developed model is validated by importing time varying data and observed the accuracy of the model via R^2 value. Time varying factor can be represented by changing the characteristic of the column (changing the efficiency of the trays). The accuracy of the model is managed to maintain as the R^2 value is approximately 0.9997 even though time varying data is introduced into the model. This proves the developed model can perform well under non-stationary and time varying process. Finally, the Moving Window approach is introduced via discarding few data points and adapting new data points into the window (frame) by maintaining the size of the frame (N) to be 84 observations. The developed models could provide higher accuracy and precision in estimating the mole fraction of acetone with R^2 value approximately 0.9991 to 0.9997.

CHAPTER 6
CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

From the studies and research that have been carried out, the project's purposes have been identified and sufficient information regarding soft sensors has been collected. Besides that the availability of types of MSPC algorithms/techniques have been identified where widely used algorithm for developing a soft sensor will be PLS (Partial Least Square) and PCA (Principle Component Analysis) method. The concept and mathematical algorithm of those techniques are identified and studied. Other than that, studies have been carried out in identifying the types of adaptive models to be used in order to develop the soft sensor by adapting the algorithm and example of those models are Moving Window, Linear Recursive, Non-linear Recursive, Time Difference and Just-In-Time are identified. For this paper, Moving Window model will be used and the performance of this model will be monitored using MATLAB Software. The separation of acetone from 2-propanol process is studied and analyzed thoroughly where data were generated by using HYSYS software to be used in developing the soft sensor model. Firstly the simulation of the case study is corrected to give the preferred values of the composition and meet the objective of the project. After that, test run have been performed to the simulation by using step change method to the selected manipulated variables which are Feed Flow Rate, Steam Flow Rate, Feed Temperature and Reflux Flow Rate. The data was later analyzed to identify the effect of the independent variable on the dependent variable, so that the correlation between the data can be understood for development of inferential model which will be used for soft sensor modeling. The extracted data is used in SIMCA-P software to develop the correlation between the input and output variables by using PLS algorithm. The inferential model developed in SIMCA-P software represents the soft sensor and it was validated by importing new observations into the model which varies with time (by changing the characteristic of the column) and analyzed the y predicted value. Besides that Moving Window approach have been applied on the model by discarding one old observation and fixed the size of the frame to be 84 observations. Successfully model which can predict the difficult-to-measure variable instantaneously and represent the current process have been develop for the separation of acetone from 2-propanol process.

6.2 Recommendations

In this project, soft sensor model based on Moving Window approach is developed which, adapts the time-varying factors of the process. Further research can be done in studying the development of the soft sensor model based on different approach such as Time Difference, Just-In-Time and et cetera, which might provide a more accurate and precise model in estimating the difficult-to-measure variables. Besides that, the developed model can be implemented in the control system of the selected case study (separation of acetone from 2-propanol). Through this recommended action, the model can be further validated of its accuracy by changing not only its efficiency but some other parameters and variables as well. Furthermore, various types of case studies can be used in developing the model such as absorbers, which also requires similar input variables as discussed in this project. Finally, the developed model can be used in the pilot plants where the performance of the soft sensor can be validated with actual plant data, and further improvement can be made so that it can be released in the market.

REFERENCES

AB., U. (2006). Multi- and Megavariate Data Analysis: Basic principles and applications: Umetrics AB.

Bo, H., & Xianhui, Y. (2011, 6-8 June 2011). A model updating approach of multivariate statistical process monitoring. Paper presented at the Information and Automation (ICIA), 2011 IEEE International Conference on.

Chapra, S. C., & Canale, R. P. (2002). Numerical methods for engineers: with software and programming applications: McGraw-Hill.

Chen, X., Xu, O., & Zou, H. (2009). Recursive PLS soft sensor with moving window for online PX concentration estimation in an industrial isomerization unit. Paper presented at the Control and Decision Conference, 2009. CCDC '09. Chinese.

Choi, S. W., Martin, E. B., Morris, A. J., & Lee, I.-B. (2006). Adaptive Multivariate Statistical Process Control for Monitoring Time-Varying Processes. Industrial & Engineering Chemistry Research, 45(9), 3108-3118.

Freund, R. J., Wilson, W. J., & Sa, P. (2006). Regression Analysis: Statistical Modeling of a Response Variable: Elsevier Academic Press.

Fujiwara, K., Kano, M., Hasebe, S., & Takinami, A. (2009). Soft-sensor development using correlation-based just-in-time modeling. AIChE Journal, 55(7), 1754-1765.

Gonzalez, G. D. (1999, Jul 1999). Soft sensors for processing plants. Paper presented at the Intelligent Processing and Manufacturing of Materials, 1999. IPMM '99. Proceedings of the Second International Conference on.

He, X. B., & Yang, Y. P. (2007). Variable MWPCA for Adaptive Process Monitoring. Industrial & Engineering Chemistry Research, 47(2), 419-427.

Hiromasa Kaneko, K. F. (2011). Classification of the Degradation of Soft Sensor Models and Discussion on Adaptive Models. Paper presented at the TIChe International Conference 2011, Hatyai, Songkhla Thailand.

Jyh-Cheng, J. (2010). Adaptive process monitoring using efficient recursive PCA and moving window PCA algorithms. Festschrift Issue In memoriam: Cheng-Ching Yu, 41(4), 475-481.

Kaneko, H., & Funatsu, K. (2011). Development of Soft Sensor Models Based on Time Difference of Process Variables with Accounting for Nonlinear Relationship. Industrial & Engineering Chemistry Research, 50(18), 10643-10651.

Li-Jie, Z., Tian-You, C., & Qiu-Mei, C. (2004). Multivariate statistical modeling and monitoring of SBR wastewater treatment using double moving window MPCA. Paper presented at the Machine Learning and Cybernetics, 2004. Proceedings of 2004 International Conference on.

Lin, B., Knudsen, J. K., & Jorgensen, S. B. (2007). A Systematic Approach for Soft Sensor Development.

Li, W., Yue, H. H., Valle-Cervantes, S., & Qin, S. J. (2000). Recursive PCA for adaptive process monitoring. Journal of Process Control, 10(5), 471-486.

Li, X., Qian, Y., Wang, J., & Qin, S. J. Information criterion for determination time window length of dynamic PCA for process monitoring Computer Aided Chemical Engineering (Vol. Volume 14, pp. 461-466): Elsevier.

Liu, J., Chen, D.-S., & Shen, J.-F. (2010). Development of Self-Validating Soft Sensors Using Fast Moving Window Partial Least Squares. Industrial & Engineering Chemistry Research, 49(22), 11530-11546. doi: 10.1021/ie101356c

Liu, X., Kruger, U., Littler, T., Xie, L., & Wang, S. (2009). Moving window kernel PCA for adaptive monitoring of nonlinear processes. Chemometrics and Intelligent Laboratory Systems, 96(2), 132-143.

Luigi, F., Graziani, S., Alessandro, R., & Maria, G. X. Soft Sensors for Monitoring and Control of Industrial Processes (Advances in Industrial Control): Springer.

Macias-Hernandez, J. J., Angelov, P., & Xiaowei, Z. (2007). Soft sensor for predicting crude oil distillation side streams using evolving takagi-sugeno fuzzy models. Paper presented at the Systems, Man and Cybernetics, 2007. ISIC. IEEE International Conference on.

Piovoso, M. J., & Kosanovich, K. A. (1994). Applications of Multivariate Statistical-Methods to Process Monitoring and Controller-Design. International Journal of Control, 59(3), 743-765.

Svante, W. (1994). Exponentially weighted moving principal components analysis and projections to latent structures. Chemometrics and Intelligent Laboratory Systems, 23(1), 149-161.

T.Okada, H. K., K.Funatsu. (2011). Development of a Model Selection Method Based on Reliability of a Soft Sensor Model. Paper presented at the TIChe International Conference 2011, Hatyai, Songkhla Thailand.

Wang, X., Kruger, U., & Irwin, G. W. (2005). Process Monitoring Approach Using Fast Moving Window PCA. Industrial & Engineering Chemistry Research, 44(15), 5691-5702.

Yonggang, W., Jinliang, D., & Tianyou, C. (2008). Soft-sensor for alkaline solution concentration of evaporation process. Paper presented at the Intelligent Control and Automation, 2008. WCICA 2008. 7th World Congress on.

APPENDICES

APPENDIX A: MATLAB Coding for Linear Least Square Regression Analysis

```
% PROGRAM LREGES
% A LINEAR REGRESSION PROGRAM
% READ NUMBER OF DATA SETS, DATA OF X AND Y
n=input('\nEnter number data:');
forirow=1:n
x(irow)=input('\nEnter value of x:');
y(irow)=input('Enter value of y:');
end
% COMPUTE SUMMATION TERMS
sumx=0.0;
sumy=0.0;
sumx2=0.0;
sumxy=0.0;
for i=1:n
sumx=sumx+x(i);
sumy=sumy+y(i);
sumx2=sumx2+x(i)*x(i);
sumxy=sumxy+x(i)*y(i);
end
% SOLVE FOR COEFFICIENTS
det=n*sumx2-sumx*sumx;
A0=(sumy*sumx2-sumxy*sumx)/det;
A1=(n*sumxy-sumx*sumy)/det;
fprintf('\nCOEFFICIENT A0=%14.6e',A0)
fprintf('\nCOEFFICIENT A1=%14.6e',A1)
```

```
Enter number data:6

Enter value of x:10
Enter value of y:2.2

Enter value of x:15
Enter value of y:4.6

Enter value of x:20
Enter value of y:4.2

Enter value of x:25
Enter value of y:7.0

Enter value of x:30
Enter value of y:6.6

Enter value of x:35
Enter value of y:9.2

COEFFICIENT A0= 1.904762e-003
COEFFICIENT A1= 2.502857e-001>>
```

APPENDIX B: MATLAB Coding for Gauss Elimination

```
function x = gauss (n,a,b)
% FORWARD ELIMINATION: PERFORM ACCORDING TO THE ORDER OF
  'PRIME' FROM 1 TO
% N-1
forip = 1:n-1

% LOOP OVER EACH EQUATION STARTING FROM THE ONE THAT
  CORRESPONDS WITH THE
% ORDER OF 'PRIME' PLUS ONE
forie = ip+1:n
ratio  = a(ie,ip)/a(ip,ip)

% COMPUTE NEW COEFFICIENT OF THE EQUATION CONSIDERED
foric = ip+1:n
    a(ie,ic) = a(ie,ic) - ratio*a(ip,ic);
end
b(ie) = b(ie) - ratio*b(ip);
end

% SET COEFFICIENT ON LOWER LEFT PORTION TO ZERO:
forie = ip+1:n
a(ie,ip) = 0.;
end
end


% BACK SUBSTITUTION
% COMPUTE SOLUTION OF THE LAST EQUATION
x(n) = b(n)/a(n,n);

% COMPUTE SOLUTIONS FROM EQUATION N-1 TO 1
forie = n-1:-1:1
sum = 0.;
foric = ic+1:n
sum = sum + a(ie,ic) *x(ic);
end
x(ie) = (b(ie) - sum)/a(ie,ie);
end
```

APPENDIX C: MATLAB Coding for Multiple Linear Least Square Regression Analysis

```
% PROGRAM MREGRES
% A MULTIPLE LINEAR REGRESSION PROGRAM
% READ NUMBER OF DATA SETS N
% NUMBER OF INDEPENDENT VARIABLES K
% AND DATA OF X(I,K) AND Y(I)
fid = fopen('data.dat','r');
n = fscanf(fid,'%f',1);
k = fscanf (fid,'%f',1);
x = fscanf(fid,'%f',[3 6]);
x = x';
y = fscanf(fid,'%f',[n]);
b =zeros (k+1,k);
a =zeros (k+1,k+1);

% COMPUTE SQUARE MATRIX ON LHS AND VECTOR ON RHS OF SYSTEM
EQUATIONS
% CALL SUBROUTINE FOR SOLVING SYSTEM EQUATIONS
for i=1:n
forir=1:k+1
ifir==1
fr=1.;
end
ifir>1
fr=x(i,ir-1);
end
foric=1:k+1
ific==1
fc=1.;
end
ific>1
fc=x(i,ic-1);
end
a(ir,ic) = a(ir,ic) + fr*fc;
end
b(ir)=b(ir)+fr*y(i);
end
end
kp1 = k+1;
xx=gauss(kp1, a, b);
fprintf('\ncoefficient of fitted function are:')
for i = 1:k+1
     im1=i-1;
fprintf('\n A(%1d) = %13.7e',im1,xx(i));
end
```

```
ratio =

     1

ratio =

     1

ratio =

    0.5000

coefficient of fitted function are:

A(0) = 6.0000000e+000

A(1) = 3.5000000e+000

A(2) = 3.0000000e+000>>
```

APPENDIX C: MATLAB Coding for Partial Least Square Regression Analysis

```
loadmoore
y = moore(:,6)                  % Response
X0 = moore(:,1:5);               % Original predictors
X1 = X0+10*randn(size(X0))      % Correlated predictors
X = [X0,X1]

[XL,yl,XS,YS,beta,PCTVAR] = plsregress(X,y,10)
plot(1:10,cumsum(100*PCTVAR(2,:)),'-bo')
xlabel('Number of PLS components')
ylabel('Percent Variance Explained in y')
[XL,yl,XS,YS,beta,PCTVAR,MSE,stats] = plsregress(X,y,6)
yfit = [ones(size(X,1),1) X]*beta;
plot(y,yfit,'o')

TSS = sum((y-mean(y)).^2)
RSS = sum((y-yfit).^2)
Rsquared = 1 - RSS/TSS

plot(1:10,stats.W,'o-')
legend({'c1','c2','c3','c4','c5','c6'},'Location','NW')
xlabel('Predictor')
ylabel('Weight')
[axes,h1,h2] = plotyy(0:6,MSE(1,:),0:6,MSE(2,:))
set(h1,'Marker','o')
set(h2,'Marker','o')
legend('MSE Predictors','MSE Response')
xlabel('Number of Components')
```

```
stats =

             W: [10x6 double]

            T2: [20x1 double]

Xresiduals: [20x10 double]

Yresiduals: [20x1 double]

TSS =

    5.0679

RSS =

    0.7728

Rsquared =

    0.8475

axes =

170.0017  172.0013

h1 =

  171.0039

h2 =

  173.0018
```

APPENDIX D: Calculations for Binary Distillation Column

Table D.1: Dew Point Calculations for Top Product Stream

DEW POINT CALCULATIONS								
TOP PRODUCT								
Temperature	65.048	C						
Pressure	863.300	mmHg						
COMPONENT	A	B	C	Psat (Antoinne Equa)	Vapor Pressure (mmHg)	Ki = Pi/P	yd	yd/Ki
Acetone	7.2316	1277.0300	237.2300	3.0069	1016.0491	1.1769	0.9000	0.7647
2-Propanol	8.1182	1580.9200	219.6200	2.5646	366.9826	0.4251	0.1000	0.2352
							1.0000	**0.9999**

Table D.2: Bubble Point Calculations for Bottom Product Stream

BUBBLE POINT CALCULATIONS								
BOTTOM PRODUCT								
Temperature	82.475	C						
Pressure	863.300	mmHg						
COMPONENT	A	B	C	Psat (Antoinne Equa)	Vapor Pressure (mmHg)	Ki = Pi/P	xb	Ki*xb
Acetone	7.2316	1277.03	237.23	3.2372	1726.6496	2.0001	0.1000	0.2000
2-Propanol	8.1182	1580.92	219.62	2.8850	767.3953	0.8889	0.9000	0.8000
							1.0000	**1.0000**

Table D.3: Bubble Point Calculations for Feed Stream

BUBBLE POINT CALCULATIONS (FEED)								
FEED (Liquid Feed)								
Temperature	76.740	C						
Pressure	863.300	mmHg						
COMPONENT	A	B	C	Psat (Antoinne Equa)	Vapor Pressure (mmHg)	Ki = Pi/P	xf	Ki*xf
Acetone	7.2316	1277.03	237.23	3.1642	1459.6124	1.6907	0.3000	0.5072
2-Propanol	8.1182	1580.92	219.62	2.7837	607.7750	0.7040	0.7000	0.4928
							1.0000	**1.0000**

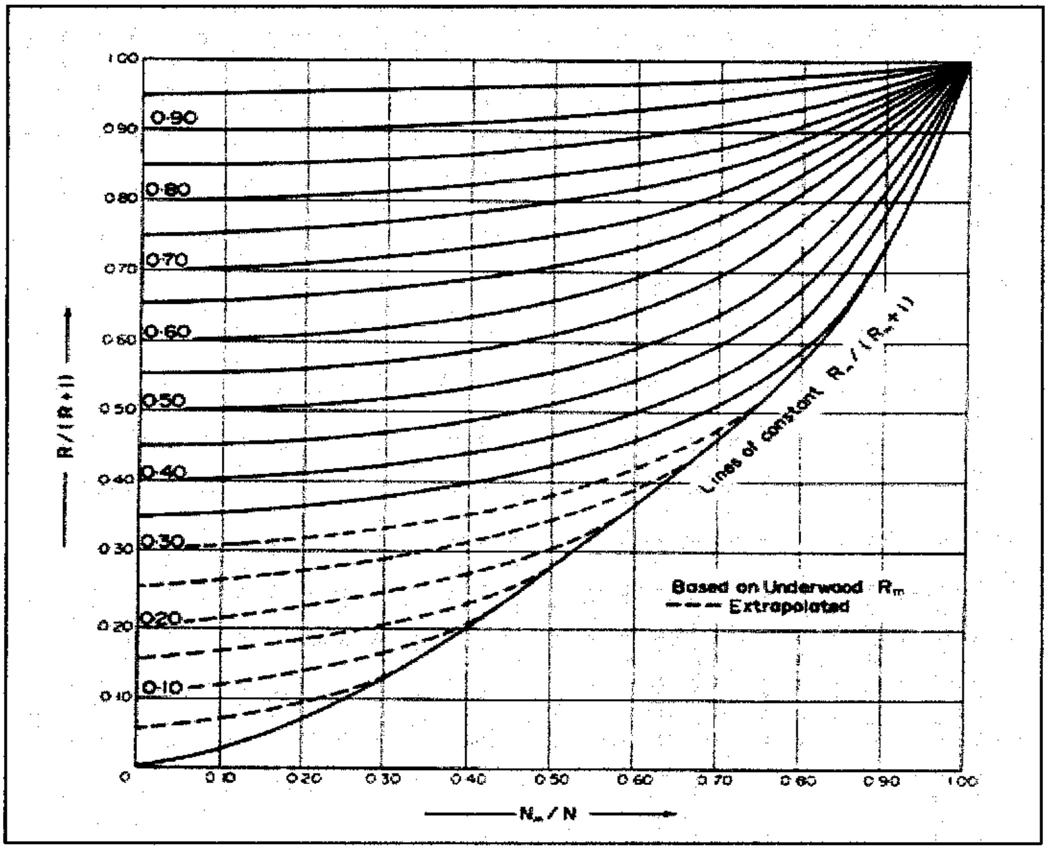

Figure D.1: Erbar-Maddox Correlation (Erbar and Maddox, 1961)

Table E.1: Data Generated from HYSYS for Percentage Change in Feed Flow Rate

0.00	0.6646	0.3000	0.7000	0.1976	0.9843	0.0157	0.4671	0.0106	0.9894	1.0511	0.9843	0.0157	5.3205
0.02	0.6779	0.3000	0.7000	0.1977	0.9850	0.0150	0.4803	0.0178	0.9822	1.0511	0.9850	0.0150	5.3163
0.04	0.6912	0.3000	0.7000	0.1977	0.9854	0.0146	0.4935	0.0252	0.9748	1.0512	0.9854	0.0146	5.3167
0.06	0.7044	0.3000	0.7000	0.1976	0.9856	0.0144	0.5069	0.0326	0.9674	1.0512	0.9856	0.0144	5.3208
0.08	0.7177	0.3000	0.7000	0.1973	0.9856	0.0144	0.5204	0.0398	0.9602	1.0512	0.9856	0.0144	5.3271
0.10	0.7310	0.3000	0.7000	0.1970	0.9857	0.0143	0.5341	0.0470	0.9530	1.0512	0.9857	0.0143	5.3366
0.08	0.7177	0.3000	0.7000	0.1974	0.9857	0.0143	0.5204	0.0400	0.9600	1.0512	0.9857	0.0143	5.3262
0.06	0.7044	0.3000	0.7000	0.1976	0.9856	0.0144	0.5069	0.0328	0.9672	1.0512	0.9856	0.0144	5.3208
0.04	0.6912	0.3000	0.7000	0.1977	0.9854	0.0146	0.4934	0.0254	0.9746	1.0512	0.9854	0.0146	5.3168
0.02	0.6779	0.3000	0.7000	0.1977	0.9851	0.0149	0.4802	0.0180	0.9820	1.0511	0.9851	0.0149	5.3164
0.00	0.6646	0.3000	0.7000	0.1976	0.9843	0.0157	0.4671	0.0108	0.9892	1.0511	0.9843	0.0157	5.3199
-0.02	0.6513	0.3000	0.7000	0.1970	0.9820	0.0180	0.4542	0.0043	0.9957	1.0509	0.9820	0.0180	5.3342
-0.04	0.6380	0.3000	0.7000	0.1958	0.9745	0.0255	0.4422	0.0012	0.9988	1.0503	0.9745	0.0255	5.3652
-0.06	0.6247	0.3000	0.7000	0.1941	0.9642	0.0358	0.4306	0.0006	0.9994	1.0496	0.9642	0.0358	5.4068
-0.08	0.6114	0.3000	0.7000	0.1922	0.9535	0.0465	0.4192	0.0004	0.9996	1.0488	0.9535	0.0465	5.4569
-0.10	0.5981	0.3000	0.7000	0.1903	0.9429	0.0571	0.4079	0.0004	0.9996	1.0480	0.9429	0.0571	5.5080
-0.08	0.6114	0.3000	0.7000	0.1922	0.9535	0.0465	0.4192	0.0004	0.9996	1.0488	0.9535	0.0465	5.4575
-0.06	0.6247	0.3000	0.7000	0.1941	0.9642	0.0358	0.4306	0.0006	0.9994	1.0496	0.9642	0.0358	5.4074
-0.04	0.6380	0.3000	0.7000	0.1956	0.9744	0.0256	0.4423	0.0012	0.9988	1.0503	0.9744	0.0256	5.3690
-0.02	0.6513	0.3000	0.7000	0.1972	0.9818	0.0182	0.4540	0.0041	0.9959	1.0509	0.9818	0.0182	5.3294
0.00	0.6646	0.3000	0.7000	0.1975	0.9842	0.0158	0.4670	0.0104	0.9896	1.0511	0.9842	0.0158	5.3220

Table E.1: Data Generated from HYSYS for Percentage Change in Feed Flow Rate

0.3000	0.7000	0.1976	0.9843	0.0157	0.4670	0.0106	0.9894	1.0511	0.9843	0.0157	5.3261
0.3000	0.7000	0.2081	0.9575	0.0425	0.4571	0.0006	0.9994	1.0491	0.9575	0.0425	5.0403
0.3000	0.7000	0.2166	0.9210	0.0790	0.4490	0.0004	0.9996	1.0463	0.9210	0.0790	4.8310
0.3000	0.7000	0.2251	0.8864	0.1136	0.4405	0.0004	0.9996	1.0437	0.8864	0.1136	4.6367
0.3000	0.7000	0.2338	0.8532	0.1468	0.4316	0.0003	0.9997	1.0412	0.8532	0.1468	4.4532
0.3000	0.7000	0.2429	0.8214	0.1787	0.4224	0.0003	0.9997	1.0388	0.8214	0.1787	4.2771
0.3000	0.7000	0.2339	0.8532	0.1468	0.4316	0.0003	0.9997	1.0412	0.8592	0.1468	4.4507
0.3000	0.7000	0.2251	0.8865	0.1135	0.4405	0.0004	0.9996	1.0437	0.8865	0.1135	4.6364
0.3000	0.7000	0.2166	0.9209	0.0791	0.4490	0.0004	0.9996	1.0463	0.9209	0.0791	4.8306
0.3000	0.7000	0.2081	0.9575	0.0425	0.4572	0.0006	0.9994	1.0491	0.9575	0.0425	5.0405
0.3000	0.7000	0.1974	0.9842	0.0158	0.4671	0.0102	0.9898	1.0511	0.9842	0.0158	5.3239
0.3000	0.7000	0.1828	0.9868	0.0132	0.4818	0.0391	0.9609	1.0513	0.9868	0.0132	5.7521
0.3000	0.7000	0.1663	0.9828	0.0122	0.4987	0.0703	0.9297	1.0513	0.9878	0.0122	6.3217
0.3000	0.7000	0.1484	0.9887	0.0113	0.5167	0.1019	0.8981	1.0514	0.9887	0.0113	7.0869
0.3000	0.7000	0.1292	0.9896	0.0104	0.5361	0.1335	0.8665	1.0515	0.9896	0.0104	8.1398
0.3000	0.7000	0.1089	0.9904	0.0096	0.5566	0.1646	0.8354	1.0515	0.9904	0.0096	9.6550
0.3000	0.7000	0.1292	0.9895	0.0105	0.5361	0.1341	0.8659	1.0515	0.9895	0.0105	8.1386
0.3000	0.7000	0.1483	0.9887	0.0113	0.5167	0.1026	0.8974	1.0514	0.9887	0.0113	7.0915
0.3000	0.7000	0.1662	0.9876	0.0124	0.4987	0.0710	0.9290	1.0513	0.9876	0.0124	6.3252
0.3000	0.7000	0.1827	0.9867	0.0133	0.4820	0.0400	0.9600	1.0513	0.9867	0.0133	5.7530
0.3000	0.7000	0.1976	0.9843	0.0157	0.4670	0.0110	0.9890	1.0511	0.9843	0.0157	5.3199