Online Composition Prediction of a Debutanizer Column Using Artificial Neural Network

Nasser Mohammed Ramli
Chemical Engineering Department, Universiti Teknologi Petronas, Bandar Seri Iskandar, 31750 Tronoh, Perak, MALAYSIA

Mohd Azlan Hussain*+; Badrul Mohamed Jan
Chemical Engineering Department, Faculty of Engineering, University of Malaya,50603 Kuala Lumpur, MALAYSIA

Bawadia Abdullah
Chemical Engineering Department, Universiti Teknologi PETRONAS, Bandar Seri Iskandar, 31750 Tronoh, Perak, MALAYSIA

ABSTRACT: *The current method for composition measurement of an industrial distillation column includes offline method, which is slow, tedious and could lead to inaccurate results. Among advantages of using online composition designed are to overcome the long time delay introduced by laboratory sampling and provide better estimation, which is suitable for online monitoring purposes. This paper presents the use of an online dynamic neural network to simultaneously predict n-butane composition of the top and bottom products of an industrial debutanizer columns. Principal component and partial least square analysis are used to determine the important variables surrounding the column prior to implementing the neural network. It is due to the different types of data available for the plant, which requires proper screening in determining the right input variables to the dynamic model. Statistical analysis is used as a model adequacy test for the composition prediction of n-butane in the column. Simulation results demonstrated that the Artificial Neural Network (ANN) can reliably predict the online composition of n-butane of the column. It is further confirmed by the statistical analysis with low Root Mean Square Error (RMSE) value indicating better prediction.*

KEYWORDS: *Principal component analysis; Partial least square analysis; Neural network; Debutanizer column.*

INTRODUCTION

Distillation column is one of the important unit operations in the downstream process industries. In the 20th century, distillations are widely used for separating liquid mixtures of chemical compounds. In petroleum downstream

* *To whom correspondence should be addressed.*
+ *E-mail: mohd_azlan@um.edu.my*
1021-9986/2017/2/153-174 22/$/7.20

process alone, an approximately two thirds of the energy for distillation is consumed where distillation is favorable for separating crude oil into petroleum fractions, light hydrocarbon and aromatic chemicals [1]. The separation of other chemical compounds, often in the presence of water, is a common practice in the chemical industry. The success of a distillation column as a method of separation is due to its operational flexibility. Inferential estimator is based on heuristic model of the process and can be based on availability of measurement and multivariate regression technique. This modeling approach is preferred since online prediction can provide a fast and accurate response.

In a distillation column, proper control strategies are selected through appropriate implementation. This is very important because controller has significant effect on product quality, production rate, and energy usage. In a distillation process, controlling a column is challenging since it involves nonlinearities, dual composition control and disturbances. In refining industries, the product quality of a debutanizer column is always the main focus of its operation [2, 3]. Several alternative column configurations have been developed, to control temperature of the column. For example, dual temperature control method is achieved by combining the middle and top temperature change to control the state switch of the total reflux and withdrawal during the operation [4]. Composition prediction based on tray temperatures is typically used in industries by controlling tray temperatures [5]. The steady state and dynamic process could be characterized for product variability prediction. These characteristics were used to generate a linear dynamic tray-to-tray model for a distillation column [6].

Dynamic Principal Component Analysis (PCA) consider the dynamic process by introducing time lagged variables into the inferential models. The dynamic methods are suitable for processes with long time delays and varying throughputs on process variables [7]. Secondary variables are sensitive to primary variables when implementing inferential estimator in achieving the column optimum performance. As there are many possible locations of the temperature sensor for a column, the selection of the secondary variables as inputs becomes difficult [8]. PCA properties are measured by extracting the input variable for the online implementation. Many approaches have been studied to build a secondary variable model from readily available measurements such as tray temperatures to replace the quality measurements. A secondary variables approach has been studied for use when quality measurements are not available and a multivariate statistical control proposed [9].

Composition of a distillation column requires an online estimation measurement. Temperature is the main variable that will affect the composition for the column. Therefore temperature gives tight control on product composition despite wide variation such as internal reflux ratio. A simple feedback loop will give good control when composition of only one stream is important. When monitoring or controlling both top and bottom compositions simultaneously, a simple feedback loop is unfavorable due to interactions of the two loops. With neural network, it is possible to monitor composition at both ends of the column. Neural network techniques have been increasingly used for a wide variety of applications where statistical method had been traditionally employed. Neural network is able to give better prediction of important parameters and be applied to wide range of problems [10].

Artificial Neural Network (ANN) is a powerful tool to model non-linear process in refinery. ANN is a black box where no priori knowledge for a system is required. ANN has the ability to learn a relationship between the outputs and the inputs for a system. To develop a process using ANN, it requires suitable network architecture and appropriate data training. The neural network architecture makes use of many hidden layers for the column and the inputs only consider temperature surrounding the column [11]. To improve product quality in a debutanizer column, soft sensor design has been used. The dynamic neural model that has been implemented used three steps predictions to evaluate its top product concentration. The output from the first dynamic network are fed to input of the second dynamic network and the output from the second dynamic network are fed to the third dynamic network to obtain the desired product. The approach uses appropriate lagged inputs including composition to the neural network.

Real time estimation of plant variables and the composition are used for monitoring purposes and the number of neurons in the hidden layer for the neural

network is determined by trial and error method [12]. The Levenberg–Marquardt (LM) algorithm for neural network training has been used because it is suitable for binary as well as multi-component mixture. The LM algorithm is more suitable compared to Steepest Descent Back Propagation (SDBP) algorithm in both cases and give more accurate and sensitive results. The LM approach has proven worked efficiently in complex chemical plants, having hundreds of parameters [13]. Neural network gives better performance than the conventional control loop and inferential control by developing a model based on neural control for single composition. The strategy is used to compensate the upsets in the operating pressure, feed flow rate, and feed composition. To keep the content of the key component in the distillate stream, the performance of the neural network has to be precise [14]. ANN has also been applied in a crude fractionation section. Back propagation algorithm is used on real time data and the output of the neural network prediction is the naphtha temperature and not the composition prediction. However, the proposed method is only used for the product quality determination [15].

An optimization framework to obtain optimal operation of dynamic processes under process-model mismatches has been developed. In order to model these mismatches, neural network have been utilized. To demonstrate the technique, a batch distillation process is used. For simplicity, they consider binary batch distillation with only one specified product [16]. A framework is proposed to optimize the operation of batch columns. The proposed framework uses ANN based process model to be employed by the optimizer. The optimization of a pilot-plant middle-vessel batch column has been considered. The maximum-product problem is formulated and solved by optimizing the column operating parameters, such as the batch time, the reflux and reboil ratios. ANN based model is capable of reproducing the actual plant dynamics with good accuracy, and allows a large number of optimization studies to be carried out with little computational effort [17].

The main contribution of the work in this paper is the use of online closed loop and open loop data for training the neural network. The close loop data has been extracted to obtain the open loop data, which is then used for training, validation and testing. In this study, column information was obtained from the actual measurement over a 4 year period in the oil refinery industry. PCA is one of the criteria to describe a remarkably simple approach to multivariate analysis based on projection methods. The projection approach can be adapted to a variety of data analytical objectives such as summarizing and visualizing a data set, multivariate classification and discriminant analysis and finding quantitative relationship among the variables. Projection methods can be made robust to outliers, deal with non-linear relationship and adapt to drift in multivariate process data. Furthermore the application of principal component and partial least square analysis are used to determine the important variables surrounding the column prior to implementing the neural network which is vital. This is because there are different types of data available for the plant, which requires proper screening to determine the right input variables to the dynamic model. In addition, this work also involves a single dynamic neural network model with lagged inputs to predict the top and bottom composition simultaneously. Statistical analyses are used as a model adequacy test for composition prediction of n-butane in the column. Neural network modeling is a good strategy for large industrial application when online estimation is required for monitoring purposes. The online measurement provide good estimate to overcome delay introduced by laboratory sampling.

PLANT AND DEBUTANIZER COLUMN DESCRIPTION

The plant under study is for offshore crude oil to produce high value petroleum products for domestic and export markets. The plant consists of a refinery process and involves in condensate fractionation and reforming aromatics processes. The feed stocks of the oil refinery are crude oil while the products are petroleum products, liquefied petroleum gas, naphtha and low sulphur waxy residue. The refinery has two main process units, which are the Crude Distillation Unit (CDU) and Catalytic Reforming Unit (CRU). The crude oil is preheated using heat exchangers to 190^{o}C – 210^{o}C. Subsequently, the stream is then further heated in a furnace at 340^{o}C – 342^{o}C to achieve a desired temperature for better separation at the CDU. The crude is then routed to the CDU before being split into a number of fractions which are Heavy Straight Run Naphtha (HSRN) as overhead vapor, untreated kerosene, straight run kerosene and

straight run diesel. From the crude tower, there are 3 sides cut streams, which are drawn to a stripper column. The stripper consists of naphtha stripper, kerosene stripper and diesel stripper.

The feed of Heavy Straight Run Naphtha (HSRN) from the CDU is mixed with hydrogen from the reformer and heated up to the reaction temperature using a heater and fed into a pretreater catalytic reactor. The products from the reactor is sent to the pretreater stripper. The bottom product of the stripper is sent to the reforming unit as feed. The treated naphtha is heated to reaction temperature and fed to the reforming reactors. Effluent from the reactor is cooled and collected in a reformer separator. One part of the gas, which is separated, is sent to an absorber. In the absorber, hydrogen gas is purged and recycled to the pretreater heater as the hydrogen make-up for the raw naphtha feed. The liquid fraction is pumped into a stabiliser. The reformate is withdrawn from the stabiliser bottom and cooled before being transferred to storage. Overhead vapor from the stabiliser are cooled, condensed and recovered from the stabiliser reflux drum. In the current work, we are focusing on the debutanizer column.

The debutanizer column is located at the CDU section. The unit is used to recover light gases and LPG from the overhead distillate before producing light naphtha. The light gases mainly C_2 are used to refine fuel gas and mixed LPG to LPG storage. The unit operation at the LPG section includes the Deethanizer and Debutanizer columns. The recovery unit is used to process unstable product from the top unit section together with unstable LPG from the reformer. Deethanizer bottom product enters as feed to the debutanizer column. The debutanizer column has 35 valve trays (one liquid pass). Low boiling point components rise up the tower in contact with the internal reflux. The high boiling point of heavy component flow down in contact with vapor produced in the debutanizer reboiler. Overhead vapor is then condensed by the debutanizer condenser. The overhead system is set and controlled by the debutanizer overhead pressure control valves which have two split ranges control.

Part of the collected condensed hydrocarbon is routed to the top of the debutanizer as reflux. The flow is measured by the reflux flow meter. The debutanizer bottom section is provided by the debutanizer reboiler to strip the light component. The reboiler temperature is controlled by the debutanizer reboiler control valve while the bottom product level is controlled by the debutanizer bottom level controller. Table 1 outlines the column specification and Table 2 describes the tag name surrounding the column. Fig. 1 shows the flow chart for the refinery process and Fig. 2 shows the column configuration for the debutanizer column. The objective of this case study is to estimate the top and bottom compositions of n-butane for the debutanizer column using a single neural network model based on a mix of industrial and simulation data. The methodology applied will be described in the next section.

Table 1: Column specification.

Number of tray of the column	35
Feed tray - stage number	23
Type of tray used	Valve
Column diameter	1.3 meter
Column height	23.95 meter
Condenser type	Partial
Feed mass flowrate	44106 $kghr^{-1}$
Feed temperature	113 ^0C
Feed pressure	823.8 kPa
Overhead vapor mass flowrate	11286 $kghr^{-1}$
Overhead liquid mass flowrate	5040 $kghr^{-1}$
Condenser pressure	823.8 kPa
Reboiler pressure	853.2 kPa

METHODOLOGY

Data generation

Although most online open loop response data from the plant surrounding the column is available, some of the variables in open loop surrounding the column are not available. Such data include Temp 5, Pressure 1 and compositions at both ends of the column. Therefore plant process simulation of the debutanizer column is performed

Table 2: Tag name description of the column.

Tag	Description	Units
Temp 1	Debutanizer top temperature	°C
Temp 2	Debutanizer bottom temperature	°C
Temp 3	Debutanizer receiver bottom temperature	°C
Temp 4	Light Naphtha temperature after condenser E 1	°C
Temp 5	Reboiler outlet temperature to column	°C
Temp 6	Debutanizer feed temperature	°C
Level 1	Debutanizer level	%
Level 2	Debutanizer condenser level	%
Level 3	Debutanizer level indicator	%
Level 4	Condenser level indicator	%
Flow 1	Light Naphtha flow to storage	m^3/hr
Flow 2	LPG flow to storage	m^3/hr
Pressure 1	Debutanizer receiver overhead pressure	kPa

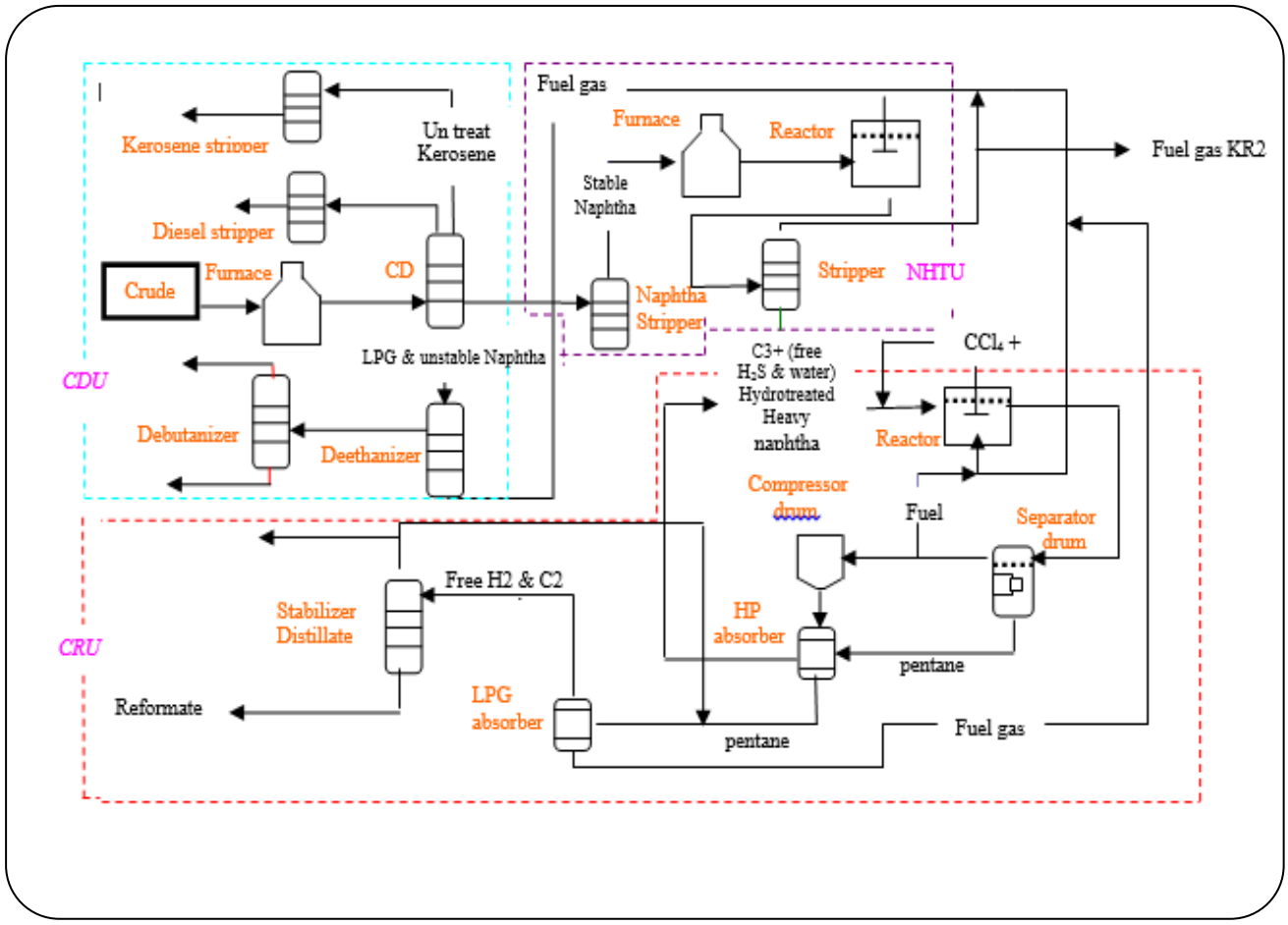

Fig. 1: Flow chart for the refinery process.

Table 3: Controller setting and set point.

	K_c	T_i	T_d	Set point
Flow 1	0.5	30	0	44.64 m^3/hr
Reflux flow	0.285	50	0	24 m^3/hr
Temp 5	0.4	80	20	135.7 °C

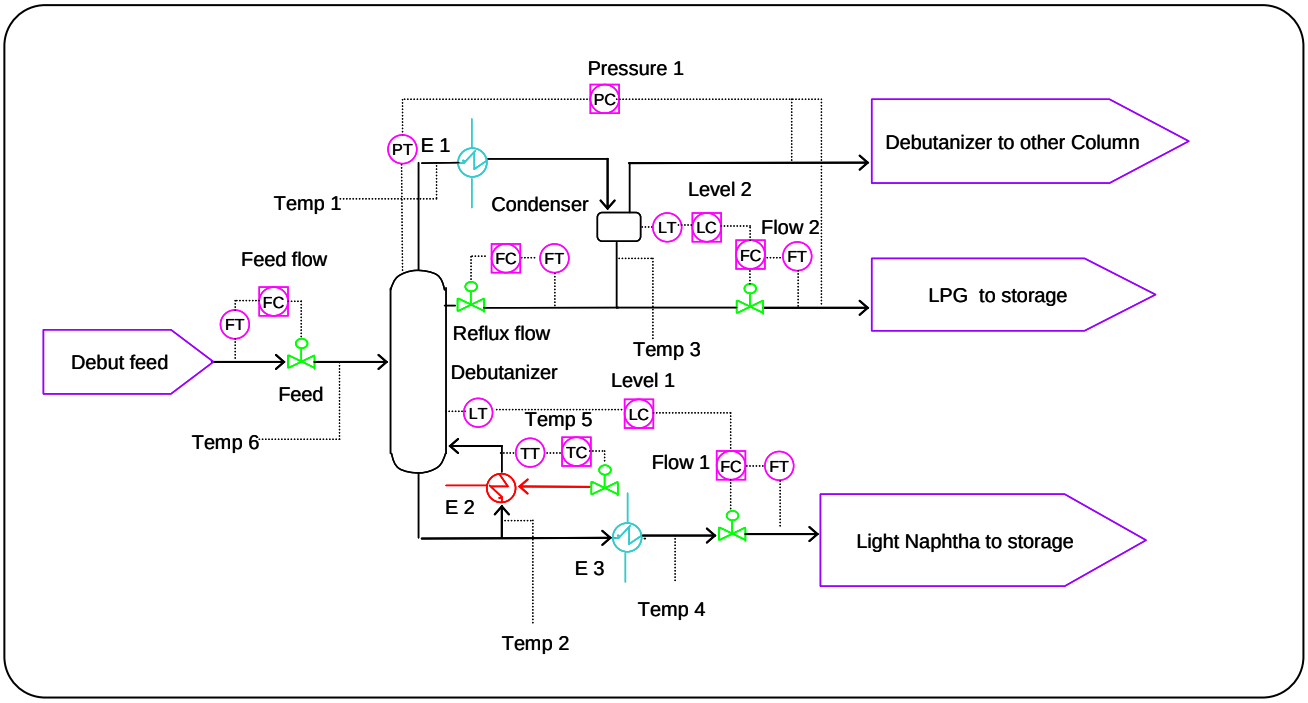

Fig. 2: Debutanizer column configuration.

to obtain the unavailable open loop data sets from the plant. The online and simulated close loop response of the compositions of n-butane at the top and bottom of the column were also established to obtain the unmeasured data. The simulation environment consists of a worksheet and Process Flow Diagram (PFD). The worksheet contains information on each flow and heat stream involved in the simulation. Comparison between the close loop responses in simulation to the actual plant from laboratory sample is performed to evaluate the deviation between the simulated and actual compositions, to ascertain that the simulation data available closely resemble the actual online industrial data.

Online close loop data for top and bottom compositions n-butane are available for the column which are obtained from laboratory sample (gas chromatographs) while the close loop data could be extracted to open loop since the controller setting and set point data are available. Table 3 shows the set point and controller settings for Flow 1, reflux flow and Temp 5.

The PID equation is used to determine the controller output (manipulated variable) derived from reference [18] as given below;

$$\Delta MV = K_c\left[e(t)+\frac{1}{\tau_i}\int_0^t e(t)+\tau_D\frac{de(t)}{dt}\right] \quad (1)$$

The tuning parameter is used to determine the process gain, K_p [19] as shown below;

$$K_p = 3K_c \quad (2)$$

Equations 1 and 2 were used to extract the close loop data to determine the process variable Temp 5 and manipulated variable of the reboiler and reflux

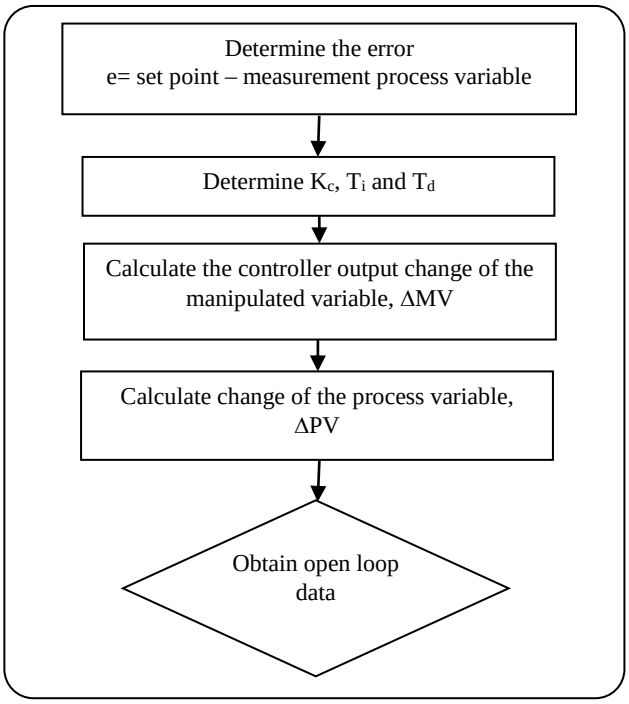

Fig. 3: Flow chart to extract close loop to obtain open loop data.

respectively are called MV2 and MV3. The manipulated variable reboiler flow rate could be obtained from Temp 5 which is applied to regulate the particular process variable. Equation 2 has been used to determine the process gain from the controller gain, K_c. Process variable for Flow 1, reflux flow and Temp 5 can be determined when the process gain, K_p and the change of the manipulated variable, ΔMV are calculated. Fig. 3 shows the procedure to extract the close loop data. There are three types of data. The first is the online open loop data. The second is the data extracted online in close loop and the last is the simulation data. These three data are combined together for training, validation and testing of the neural network model.

Principal component (PCA) and Partial Least Square (PLS) Analysis

PCA and PLS analysis are used prior to utilizing neural network to determine the important variables to be analyzed for composition prediction as this is crucial since there is a large number of variables surrounding the column. The PCA and PLS analyses will also determine the inputs to the neural network.

PCA is used to analyze all variables surrounding the column outlined in Table 4. If a correlation exists between the variables, small number of principal components will summarize a majority of the variation in X. To analyze the changes in the original data space, changes occurring within the principal components should be used. From the PCA, the important variables surrounding the column are determined. Partial Least Square (PLS) is used to relate the important variables from PCA with respect to the top and bottom compositions of n-butane. PLS regression is a method that generalizes and combines features from principal component analysis and multiple regressions. It is normally useful to predict a set of dependent variables (*Y*) from a large set of independent variables or predictors (*X*).

The data set which obtained from open loop online and simulation are combined together. SIMCA-P is used to perform PCA and PLS analysis for the debutanizer column. There are 2 important variables, which are the primary and observation variable. The primary variable consists of 23 variables surrounding the column while the observation variables are the top and bottom n-butane compositions. The observation variable is the number of observations established once the worksheet has been developed. Then PCA model is fitted to these data.

Table 4 shows the important variables involved in the PCA analysis for open loop response of the reboiler flow rate and reflux flow rate data set obtained from the plant and simulation. The simulated data are Temp 5, Pressure 1 and open loop compositions. The rest of the variables are obtained from the actual plant. The variables ‘a’ to ‘m’ are the important variables surrounding the debutanizer column where the variables ‘n_top’ to ‘r_bottom’ are the compositions at the top and bottom of the column.

From the PCA, variables which are not important are excluded while the important variables are analyzed again with respect to the top and bottom compositions n-butane using PLS. For PCA, component contribution plot are used to analyze all the important variables surrounding the column.

Table 5 shows the important variables involved in the PLS analysis where the manipulated variable is the reboiler flow rate. Table 6 shows the important variables involved in the PLS analysis where the manipulated variable is the reflux flow rate. Data is arranged according to Tables 5 and Table 6 for PLS analyses to further study the importance of the variables surrounding

Table 4: Important variables for PCA model.

Temp 1	a	Debutanizer top temperature
Temp 2	b	Debutanizer bottom temperature
Temp 3	c	Debutanizer receiver bottom temperature
Temp 4	d	Light Naphtha temperature after condenser E121
Temp 5	e	Reboiler outlet temperature to column
Temp 6	f	Debutanizer feed temperature
Level 1	g	Debutanizer level
Level 2	h	Debutanizer condenser level
Level 3	i	Debutanizer level indicator
Level 4	j	Condenser level indicator
Flow 1	k	Light Naphtha flow to storage
Flow 2	l	LPG flow to storage
Pressure 1	m	Debutanizer receiver overhead pressure
Component 1	n_top	Top composition propane
	n_bot	Bottom composition propane
Component 2	o_top	Top composition i-butane
	o_bot	Bottom composition i-butane
Component 3	p_top	Top composition n-butane
	p_bot	Bottom composition n-butane
Component 4	q_top	Top composition i-pentane
	q_bot	Bottom composition i-pentane
Component 5	r_top	Top composition n-pentane
	r_bot	Bottom composition n-pentane

Table 5: Important variables for PLS model (reboiler flow rate as manipulated variable).

Temp 1	a	Debutanizer top temperature
Temp 2	b	Debutanizer bottom temperature
Temp 5	e	Reboiler outlet temperature to column
Flow 1	k	Light Naphtha flow to storage
Component 3	p_top	Top composition of n-butane
	p_bot	Bottom composition of n-butane

Table 6: Important variables for PLS model (reflux flow rate as manipulated variable).

Temp 1	a	Debutanizer top temperature
Temp 2	b	Debutanizer bottom temperature
Temp 5	e	Reboiler outlet temperature to column
Flow 2	l	LPG flow to storage
Pressure 1	m	Debutanizer receiver overhead pressure
Component 3	p_top	Top composition of n-butane
	p_bot	Bottom composition of n-butane

the column. For PLS analysis, variable important plot are used to determine variables which are important with respect to the n-butane composition. From the PCA and PLS analyses, component contribution plot and variable important plot are used to identify the variables that are important to be selected the right inputs for neural network.

Neural network design

The online compositions at the top and bottom of the column in the refinery is currently measured using normal laboratory sampling. This is tedious and the results could not be obtained immediately therefore neural network are used as a benchmark because it is able to predict the composition faster with more accuracy and precision and could also handle non-linearities in the process variable surrounding the column as proposed in this study.

Open loop response of the reboiler and reflux data, which includes the composition of n-butane, are used to develop the dynamic neural network architecture. The selected input variables to the network including the composition of n-butane are time delayed while the outputs are the future predictions of n-butane. The type of dynamic network used for training, validation and testing the data set are nonlinear autoregressive network with exogenous inputs (NARX) with series-parallel architecture. The training algorithm used is the Levenberg-Marquardt method. Early stopping criteria are implemented to train the network while the performance function used is the mean square error.

These data sets are partitioned into three sets, which are the training, validation and test set. In the network, the number of layers used is 3 with only 1 hidden layer. The number of hidden layer is determined using statistical analysis and it is described in next section. The number of inputs to the network is 12 while the outputs are 2 (compositions of top and bottom). The transfer function to train the network is purelin (linear) for the entire layer and the networks are trained to predict simultaneously the top and bottom compositions of n-butane. Table 7 shows the important variables in the neural network where the data set are combined with the manipulated reboiler flow rate and reflux flow rate changes for n-butane after performing the PCA and PLS analyses.

Fig. 4 shows the neural network architecture for the n-butane composition prediction where the inputs for the neural network are from mv2 (k) to p_bot (k-1) while the outputs are the variable p_top (k+1) and p_bot (k+1). Fig. 5 shows the general procedure in developing the neural network architecture. Prior to implementing the neural network, the data are arranged by combining the open loop response from the simulation and online data. The data set are then trained until the network reaches its epoch and meet its performance criteria. The data set are also validated and tested after the network is trained. Since the extracted close loop data are available, the data are replaced as inputs to the neural network in the validation and test set by maintaining the actual architecture that are trained for the open loop response.

Model adequacy test for neural network to determine the hidden layer

To determine the number of neurons in the hidden layer, the following criteria is analyzed to determine the accuracy performance of the network using statistical analysis. The error is estimated using the Root Mean Square Error (RMSE) method given as,

Table 7: Important variables for neural network prediction.

MV2	mv2 (k)	Manipulated reboiler flow rate
	mv2 (k-1)	Lag MV2
MV3	mv3 (k)	Manipulated reflux flow rate
	mv3 (k-1)	Lag MV3
Temp1	a (k)	Debutanizer top temperature
	a (k-1)	Lag Temp 1
Temp 5	e (k)	Reboiler outlet temperature to column
	e (k-1)	Lag Temp 5
Component 3	p_top (k)	Top composition n-butane
	p_top (k-1)	Lag top composition
	p_bot (k)	Bottom composition n-butane
	p_bot (k-1)	Lag bottom composition
	p_top (k+1)	Future predictions n-butane top composition
	p_bot (k+1)	Future predictions n-butane bottom composition

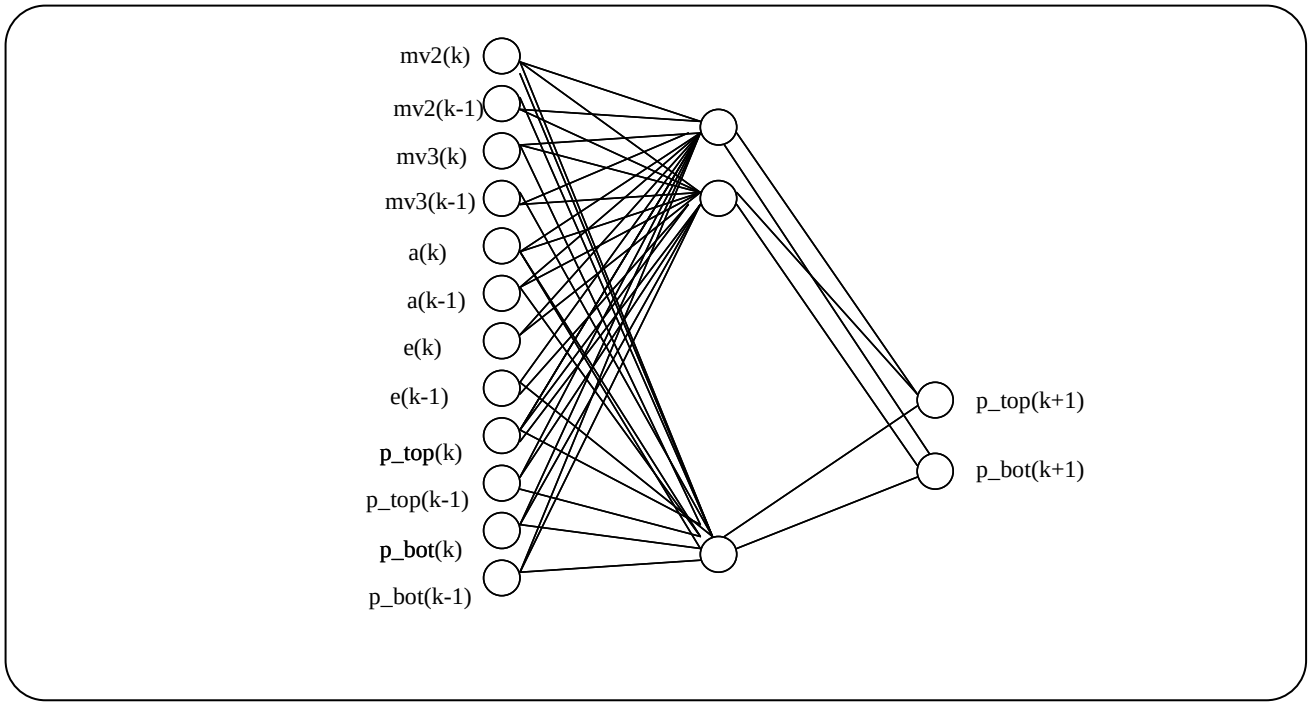

Fig. 4: Neural network architecture for n-butane.

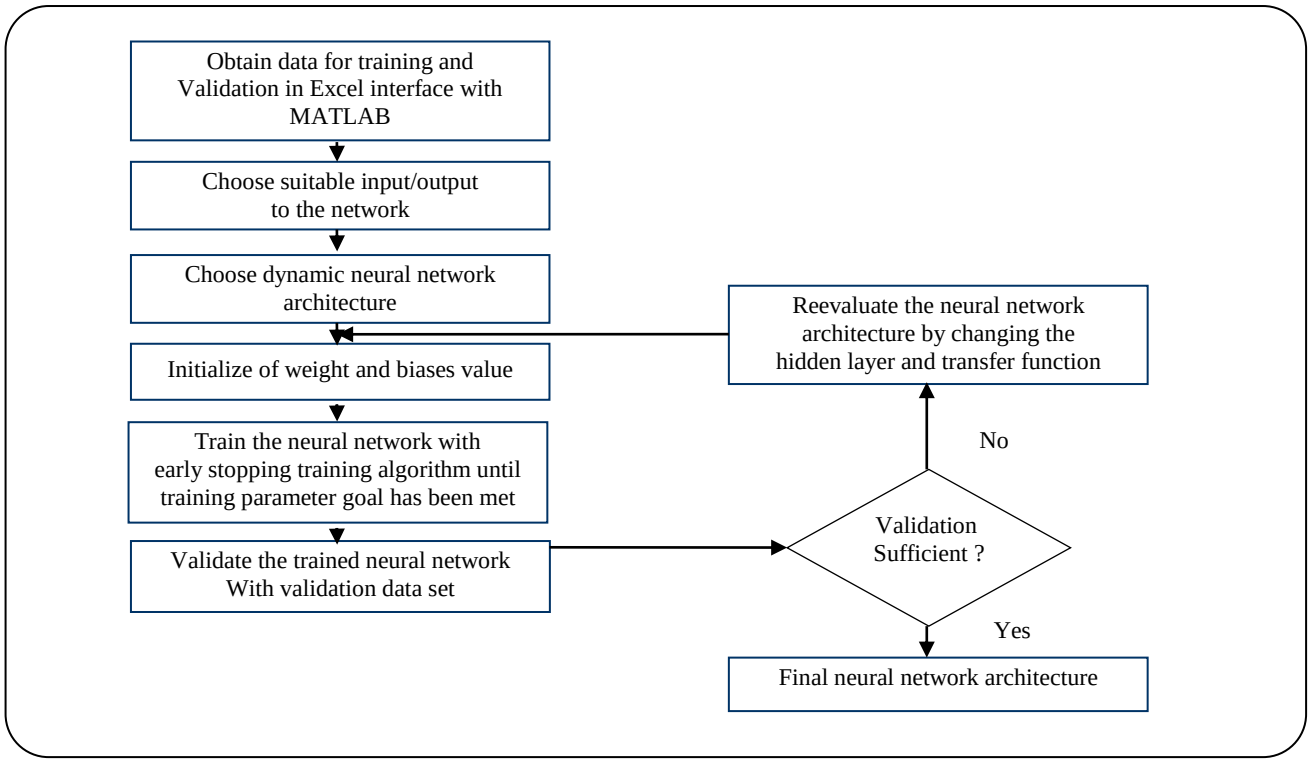

Fig. 5: Procedure of the neural network architecture.

$$RMSE=\sqrt{\frac{(x_{measured}-x_{predicted})^2}{N}} \quad (3)$$

Correct Directional Change (CDC) measures the capacity of a model to accurately predict the subsequent actual change of a predicted variable. The formula of CDC is defined as:

$$\frac{100}{N}\sum_{i}^{N}D_i \quad (4)$$

where formula of D_i is defined as:

$$y_i \times \overline{y_i}$$

The best known information criteria are the Akaike Information Criterion (AIC) and Bayesian information criteria (BIC):

$$AIC=MSE+\sigma^2\frac{2K}{T} \quad (5)$$

$$BIC=MSE+\frac{\log(N)\sigma^2 2K}{T} \quad (6)$$

The coefficient of determination is defined as:

$$R^2=1-\frac{\sum_{t=L}^{T}(y_t-f_t)^2}{\sum_{t=L}^{T}(y_t-\bar{y})^2} \quad (7)$$

Mean Absolute Percentage Error (MAPE) is the measure of accuracy in a fitted time series value given as:

$$MAPE=\frac{1}{N}\sum_{i=1}^{N}\frac{|F_t-A_t|}{A_t}\times 100\% \quad (8)$$

Pearson correlation coefficient (C_p), measures the goodness of fit of the regression given as:

$$C_p=\frac{\sum_{j=1}^{N_S}(E_{p,j}-\bar{E}_{p,j})(E_{a,j}-\bar{E}_{a,j})}{\sqrt{\sum_{j=1}^{N_S}(E_{p,j}-\bar{E}_{p,j})^2\sum_{j=1}^{N_S}(E_{a,j}-\bar{E}_{a,j})^2}} \quad (9)$$

The number of neurons in the hidden layer is determined from a range of 8 to 40. Fig. 5a shows the profile of the RMSE with respect to number of hidden nodes. Using the statistical analysis described above, with the following the deviation between actual and

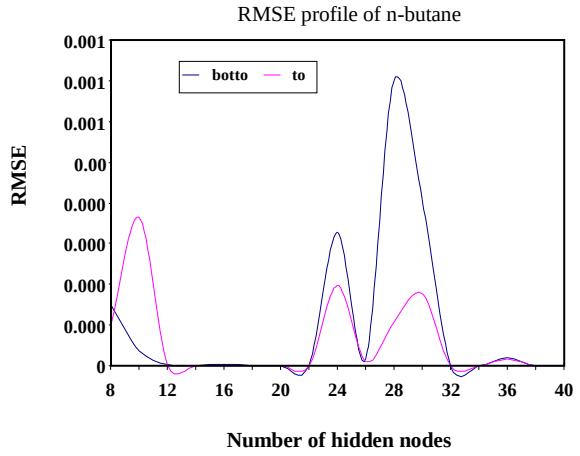

Fig. 5a: Profile of the RMSE with respect to number of hidden nodes.

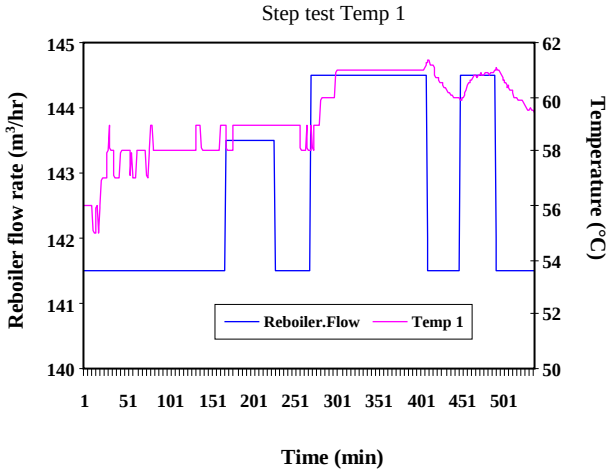

Fig. 6: Temp 1 Debutanizer top temperature.

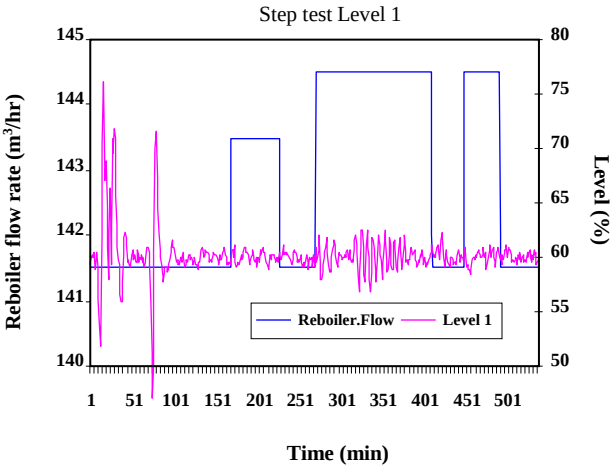

Fig. 7: Level 1 Debutanizer level.

composition prediction by neural network are determined with the following set of criteria; low RMSE, CDC approaching 100, small AIC and BIC, R^2 approaching 1, low MAPE and C_P approaching 1. Equations 3 – 9 are obtained from reference [20-22]. Based on the set criteria, the best neural network architecture is determined. The number of hidden nodes for the optimum neural network model has been found to be 22.

RESULTS AND DISCUSSION

Step test reboiler flow rate

To generate the input-output data for the neural network training, various step changes are applied to the inputs to obtain the corresponding outputs. The inputs for the system in this case is the reboiler flow rate and reflux flow rate. Figs. 6 to 9 show some of the step tests for the reboiler flow rate data sets. The step test of the reboiler flow rate, which is the manipulated variable, is generated by using multi amplitude rectangular pulse. The step test is important to determine the effect and fluctuations of the process variable when performing changes to the reboiler flow rate.

Step test reflux flow rate

Figs. 10 to 13 describe some of the step tests for the reflux flow rate data which demonstrates the effect and the fluctuations of the process variable when performing changes to the reflux flow rate. The fluctuations of the process variables will start to increase and decreases as the step test of the reflux flow rate changes. The step test of the reflux flow rate, which is the manipulated variable, is generated by using multi amplitude rectangular pulse.

Validation of online data with simulation data for composition

Figs. 14 and 15 represent the compositions at the top and bottom of n-butane from there Root Mean Square Error (RMSE) could be calculated for top and bottom compositions are 0.0251 and 0.008184 respectively. It indicates that there is small deviation between the online and simulation data. This implies that the simulation and close loop data agree well with each other including the variables that are not available from the open loop response for the plant. Once the close loop has been verified, then the open loop response for variable that is not available from plant can be obtained. The same step

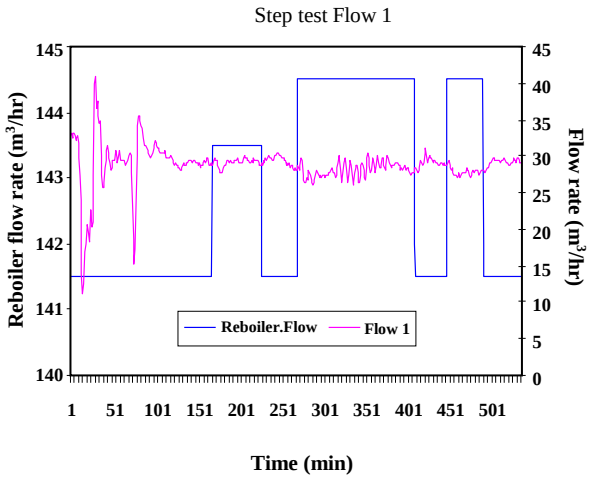

Fig. 8: Flow 1 Light Naphtha flow to storage.

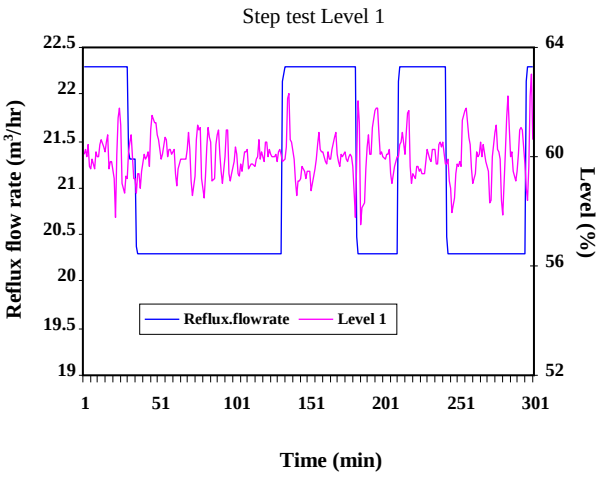

Fig. 11: Level 1 Debutanizer level.

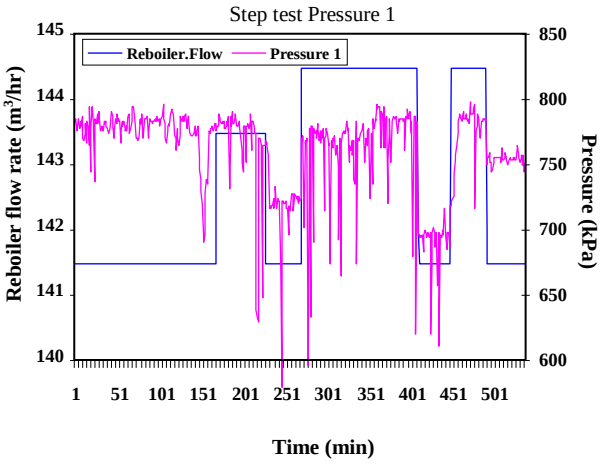

Fig. 9: Pressure 1 Debutanizer receiver overhead pressure.

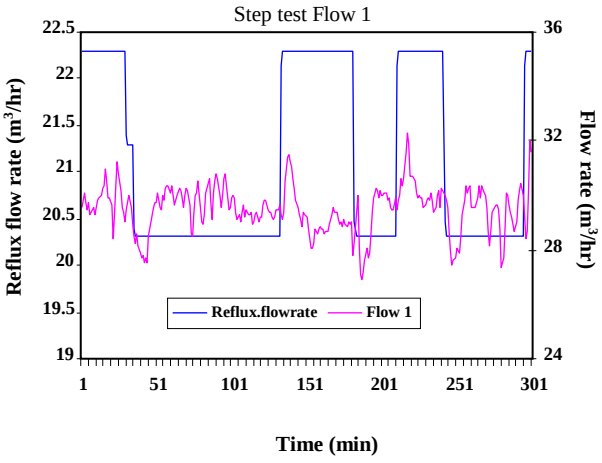

Fig. 12: Flow 1 Light Naphtha flow to storage.

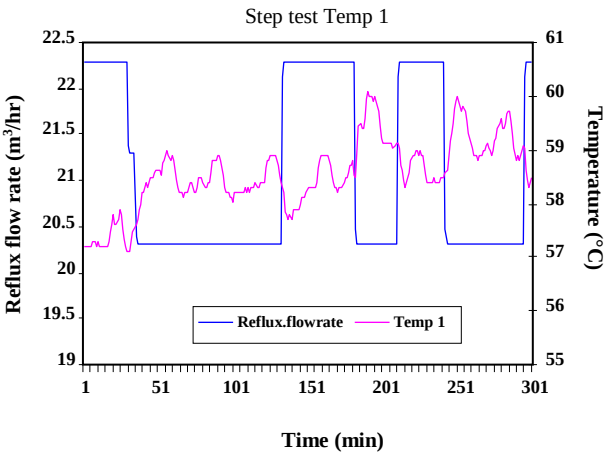

Fig. 10: T*emp 1 Debutanizer top temperature.*

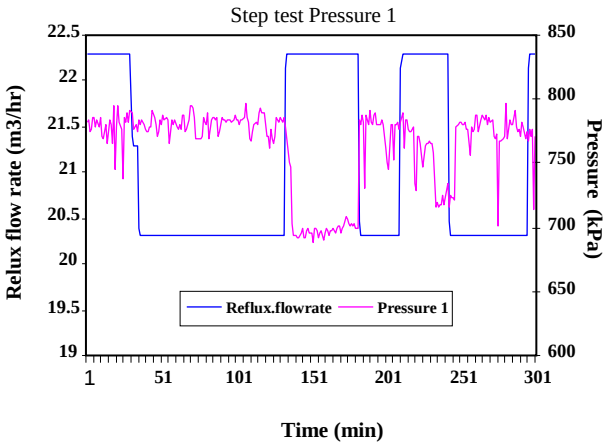

Fig. 13: Pressure 1 Debutanizer receiver overhead pressure.

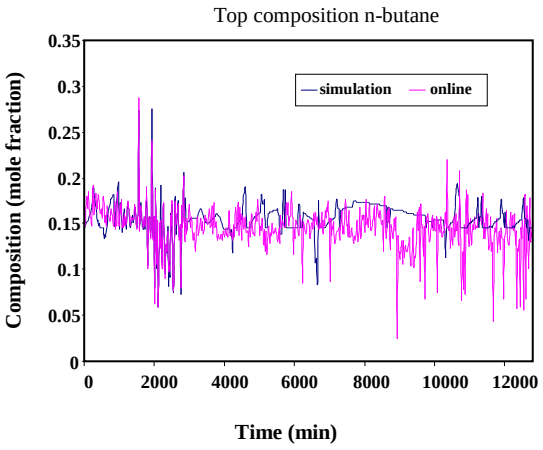

Fig. 14: Top composition n-butane.

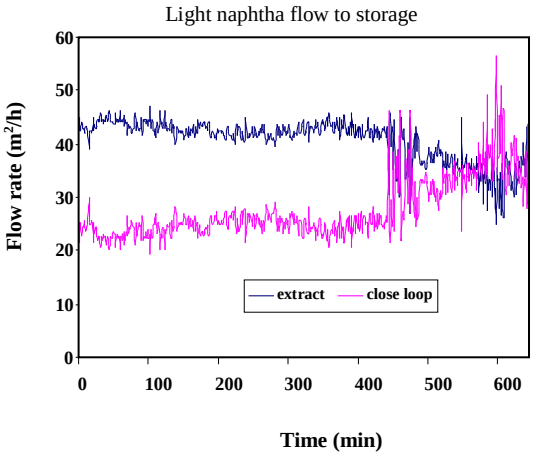

Fig. 16: Flow 1.

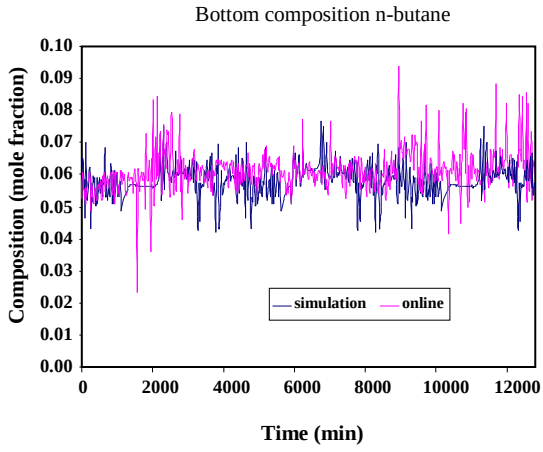

Fig. 15: Bottom composition n-butane.

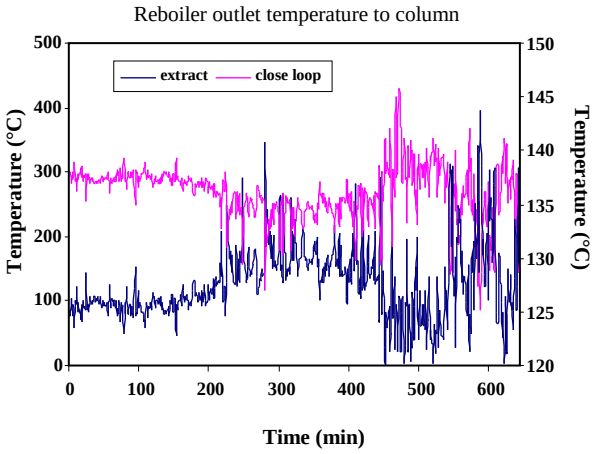

Fig. 17: Temp 5.

size for the manipulated variable for reboiler and reflux from plant are performed as inputs to obtain the fluctuation of the process variable as outputs in the simulation.

Extracted close loop data

Figs. 16 and 17 show the fluctuations of the two process variables that are in close loop response compared to the extracted closed loop. The fluctuations of the variables show a variation between close loop and data extracted from close loop. Flow 1 is controlled using PI and Temp 5 is controlled by using PID. The step change for the Manipulated Variable (MV) for Temp 5 is larger than Flow 1. This is because Temp 5 has small error with respect to time compared to Flow 1. Flow 1 has a large error because the variation between the fluctuations of the process variable and its set point is large compared to Temp 5. The controller setting for the plant affects the error to bring the process variables close to its set point.

PCA analysis

Step test reboiler flow rate

Component contribution plot shows all of the important components of the step test reboiler flow rate for the top liquefied petroleum gas product and bottom light naphtha product. The variables are outlined in Table 4. The values of R2 (variation explained) and Q2 (variation predicted according to cross validation) are also shown for each variable. The variables with high values of R2

have large loading values for the selected component. The Q2 values indicate the reliability of these R2 and loading values.

The component is considered less important if all of the variables have low values of R2 and Q2 in a component. This is applicable to variables c, d, f, g, h, i, j, l and m as depicted in Fig. 18. The highest value of R2 and Q2 for variable b is 0.48 and 0.44 respectively. From the plot, it could be concluded that variables a, b, e and k have high values for R2 and Q2 indicating that these components are important for composition prediction of n-butane with respect to the reboiler flow rate.

Step test reflux flow rate

Component contribution plot shows all of the important components of the step test reflux flow rate for the top liquefied petroleum gas product and bottom light naphtha product. The values of R2 (variation explained) and Q2 (variation predicted according to cross validation) are also shown for each variable. Variables with high values of R2 show large loading values for the selected component. Q2 value indicates the reliability of these R2 and loading values.

If all of the variables show low values of R2 and Q2 in a component, it indicates that the component is less important [23]. This is true for variables c, d, f, g, h, i, j and k as depicted in Fig. 19. The highest value of R2 and Q2 for variable e is 0.1 and 0.28 respectively. Form the plot it indicate that variables b, e, n_top and q_top have high values of R2 and Q2. It indicates that these components are important for composition prediction of n-butane with respect to the reflux flow rate.

PLS analysis for n-butane

From Fig. 20, the output variable of the PLS analysis is the n-butane composition and the input variable is the variable Temp 1 to Flow 1 outlined in Table 5. All of the variables for level have been excluded from the analysis, since level will not affect the fluctuations of the top and bottom compositions of n-butane.

From Fig. 20, the importance of these variables is determined by having the y axis value which is the Variable Important Plot (VIP) more than 0.5. If the value of the bar chart for the particular variable is less than 0.5, the variable is not important and it could be excluded from the analysis the variables i.e. k, b and e.

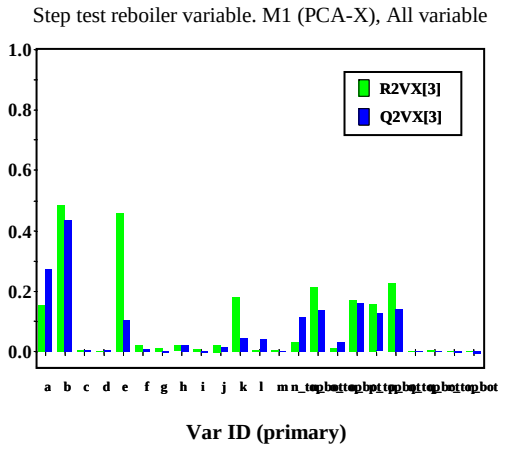

Fig. 18: Component contribution plot reboiler flow rate.

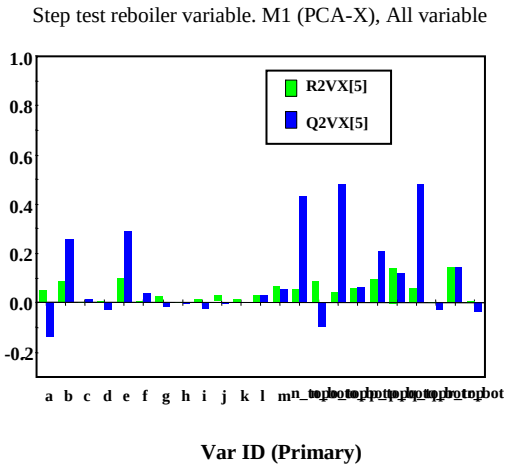

Fig. 19: Component contribution plot reflux flow rate.

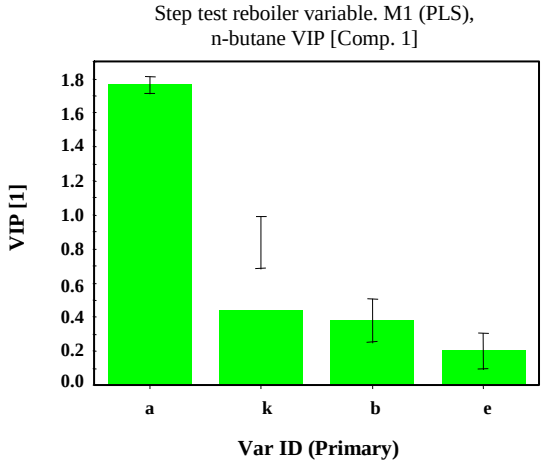

Fig. 20: Step test reboiler variable important plot.

From Fig. 21, the output variable of the PLS analysis is the n-butane composition and the input variable are the variables Temp 1 to Pressure 1 outlined in Table 6. All of the variables for the level have again been excluded from the analysis. The importance of a variable is considered important if its y axis value in the VIP bar chart is more than 0.5 [25]. If the value of bar chart of the particular variable is less than 0.5, that means the variable is not important and the variable could be excluded namely variables l, m and b.

Artificial Neural network modeling

Table 8 shows the summary of the neural network architecture for predicting the compositions of n-butane at the top and bottom of the column simultaneously. The data are partitioned according to the training; validation and test set as shown in Fig. 22. Fig. 22 a shows the training neural network by LM. Figs. 23 to 26 show the top and bottom compositions prediction of n-butane for validation and testing. The amount of data that are partitioned according to training is 65%, for validation is 18% and test is 17%. The results indicate the RMSE deviation between the open loop and the extracted close loop is 8.213×10^{-10} for the bottom composition validation and 0.16×10^{-9} for the top composition validation. The CDC value for bottom composition validation and testing are similar. The CDC value for top composition validation varies by 1.62 and for top composition testing varies by 0.32. The CDC value for bottom validation and testing is high calculated to be 100 for extract close loop and open loop response. High CDC value indicates that prediction by NN is very good. D_i is equal to 1 if $y_i \times \overline{y_i}$ is greater than zero. D_i is equal to zero if $y_i \times \overline{y_i}$ is negative. The high CDC value indicate D_i is equal to 1 based on the prediction is larger than D_i which is zero. The CDC for top validation and testing is low. This is because D_i which is zero is larger than D_i is equal to 1 and the subsequent actual change of the predicted variable is low. The regression value of R for top and bottom compositions test and validation is 1. Thus the prediction between the actual and simulated is similar. The AIC and BIC values for open loop for top composition validation are -572 and -564 respectively. The AIC and BIC values for open loop for bottom composition validation are -357 and -349 respectively. The values are larger compared to extract from close loop. The model for extract from close loop

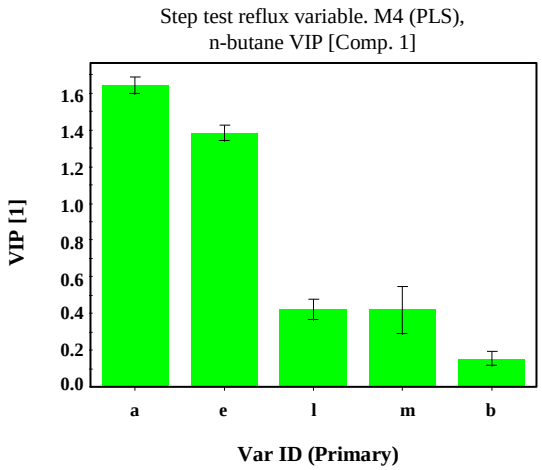

Fig. 21: Step test reflux variable important plot.

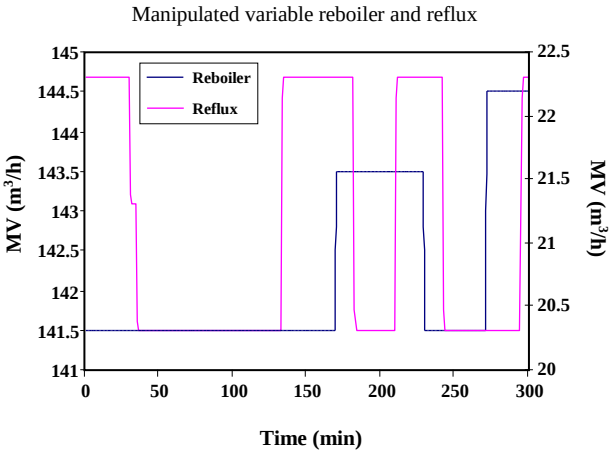

Fig. 22: Manipulated variable reboiler and reflux flow rate.

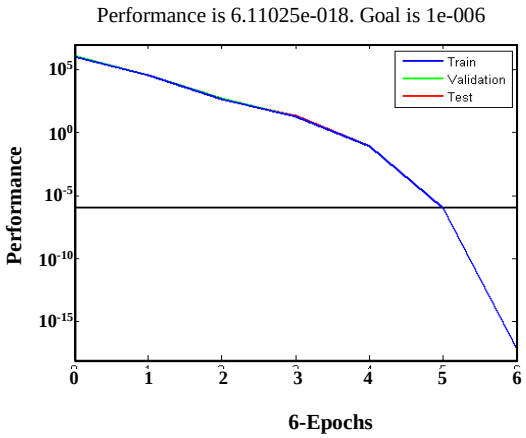

Fig 22a: Training neural network by LM.

Table 8: Neural network architecture.

Parameters	Description
Network	NARX series parallel network
Category	With partitioning
Training function	Levenberg-Marquardt
Adaptation learning function	Gradient descent
Performance function	MSE
Epochs	1000
Goal	1e-6
Number of layers	3
Layer 1: Number of Neuron Transfer function	12 Linear
Layer 2: Number of Neuron Transfer function	22 Linear
Layer 3: Number of Neuron Transfer function	2 Linear

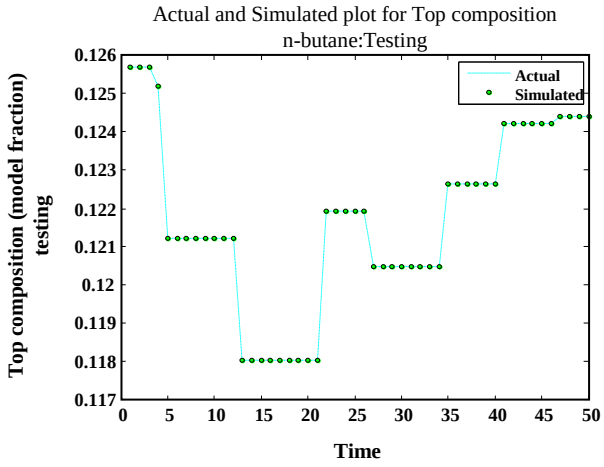

Fig. 23: Actual and simulated n-butane at the top composition for testing.

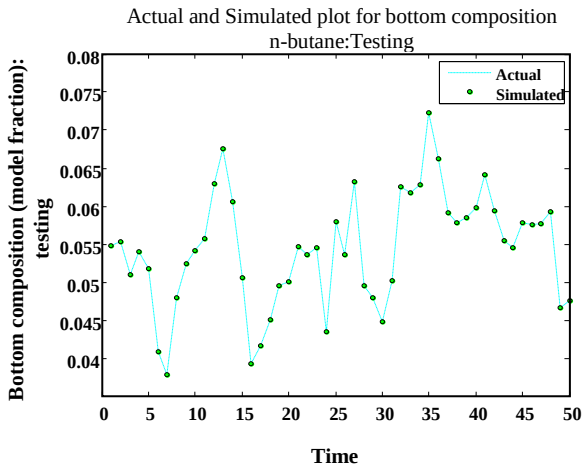

Fig. 24: Actual and simulated n-butane at the bottom composition for testing.

has low value of the information criteria gives the best performance. For the AIC and BIC, low value is preferred as it indicates better prediction. This is also applied to testing data set. The Akaike information criteria related to the square of residual to the number of free model parameters. The purpose is to weigh the error of the model against the number of parameters. The BIC is similar to AIC except that it is motivated by the Bayesian model selection principles. Cp value is close to 1. The Cp value for validation and test set for bottom and top compositions are calculated to be 1 and the MAPE should be close to 0. The MAPE values for top and bottom composition validation open loop are 1.04×10^{-6} and -6.4×10^{-6} respectively. The MAPE values are larger compared to extract from close loop. This is also applied to testing data set for MAPE values. When having a perfect fit, MAPE is zero. But in regards to upper level the MAPE has no restriction. The percentage error calculated for MAPE is to compare the error of fitted time series. Table 9 shows the statistical analysis

Table 9: Statistical analysis of n-butane open loop and extract close loop.

Parameter	Extract	Open loop
rmse_bottom_validation	2.52E-09	3.35E-09
rmse_top_validation	1.15E-09	1.31E-09
CDC_bottom_validation	100	100
CDC_top_validation	17.30	15.68
R_bottom_validation	1	1
R_top_validation	1	1
AIC_bottom_validation	-352.80	-357.77
AIC_top_validation	-561.06	-572.12
BIC_bottom_validation	-344.99	-349.89
BIC_top_validation	-553.26	-564.24
MAPE_bottom_validation	-4.67E-06	-6.4E-06
MAPE_top_validation	9.23E-07	1.04E-06
Cp_bottom_validation	1	1
Cp_top_validation	1	1
rmse_bottom_testing	2.48E-09	3.33E-09
rmse_top_testing	1.15E-09	1.32E-09
CDC_bottom_testing	100	100
CDC_top_testing	16.32	16
R_bottom_testing	1	1
R_top_testing	1	1
AIC_bottom_testing	-344.941	-349.38
AIC_top_testing	-479.41	-487.47
BIC_bottom_testing	-337.29	-341.66
BIC_top_testing	-471.76	-479.74
MAPE_bottom_testing	-4.498E-06	-6.20E-06
MAPE_top_testing	9.46E-07	1.07E-06
Cp_bottom_testing	1	1
Cp_top_testing	1	1

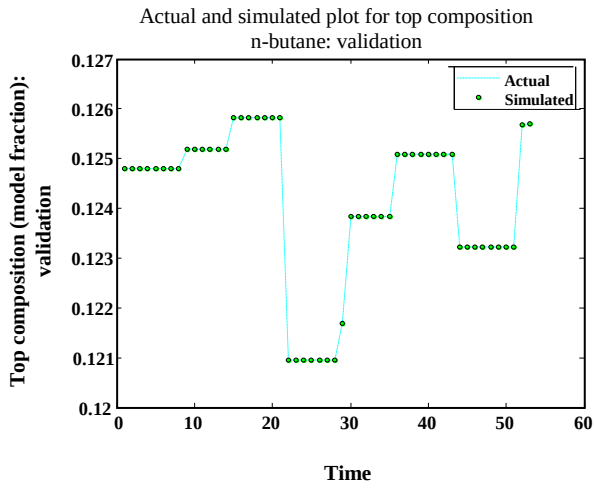

Fig. 25: Actual and simulated n-butane at the top composition for validation.

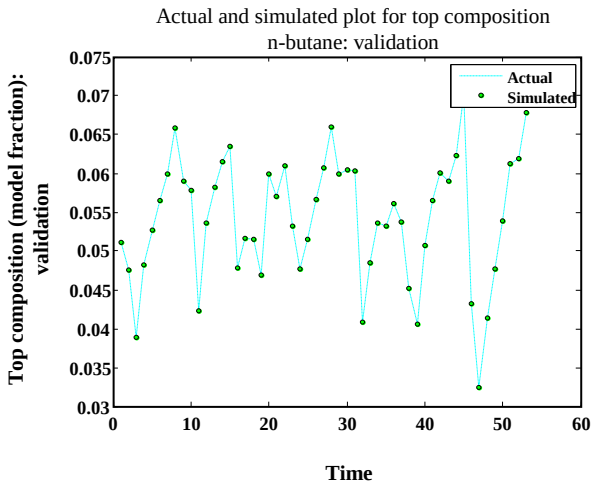

Fig. 26: Actual and simulated n-butane at the bottom composition for validation.

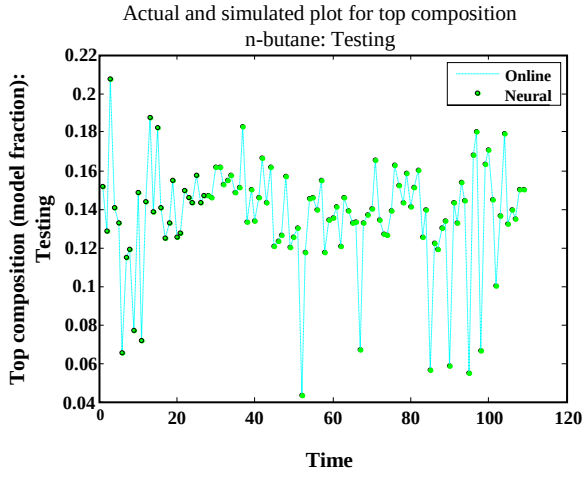

Fig. 27: Actual and simulated n-butane at the top composition closed loop for testing.

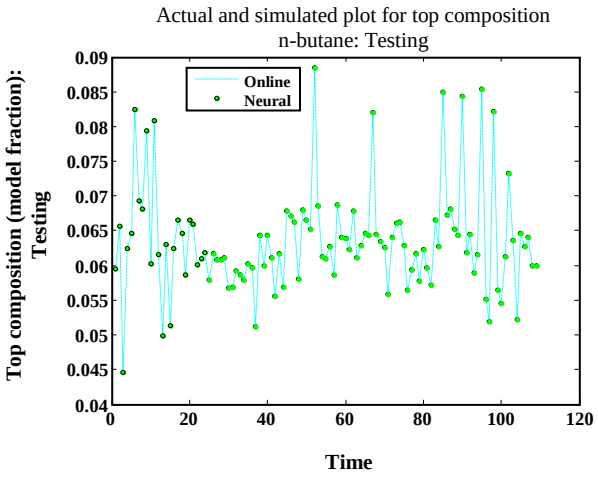

Fig. 28: Actual and simulated n-butane at the bottom composition close loop for testing.

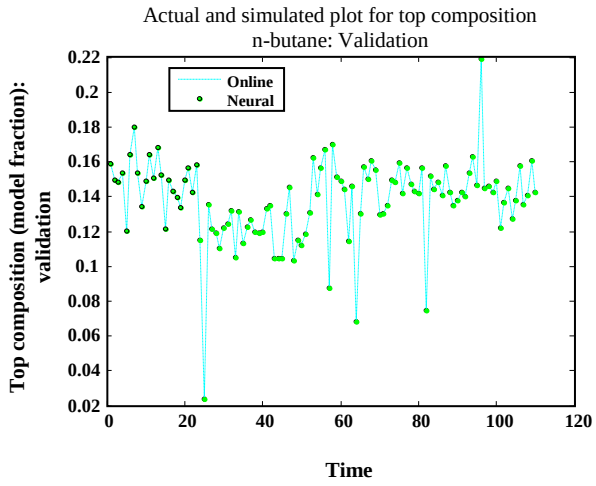

Fig. 29: Actual and simulated n-butane at the top composition close loop for validation.

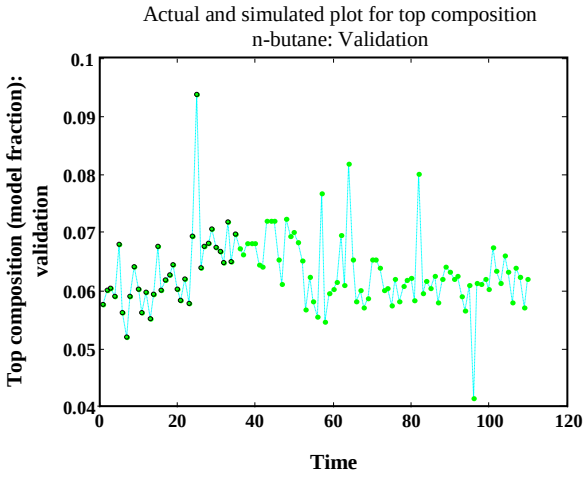

Fig. 30: Actual and simulated n-butane at the bottom composition close loop for validation.

of n-butane with partition open loop data and extracted close loop data. From the statistical analysis, the extracted close loop performs better than the open loop response because RMSE value is low, CDC is high, AIC and BIC is low and MAPE is low.

Validate online based on close loop data

Validate online prediction for the n-butane is implemented to estimate the composition of n-butane as the output for the column. The data of the composition prediction are partitioned to three sets, which are training, validation and testing. Only the validation and testing sets are shown in this paper. The neural network architecture for online validation is similar as shown in Table 8. The actual composition for n-butane is obtained from the laboratory measurement in the refinery plant. The required data collected surrounding the column consists of the closed loop data. The purpose for validate online is to monitor the composition for its precision and accuracy. The training data are trained until the epoch and the performance has achieved and then the network is validated and tested.

Figs. 27-30 show the top and bottom compositions prediction of n-butane for validation and testing. The results indicate that the RMSE is low at 3.73×10^{-9} for the bottom composition validation data set and 7.16×10^{-9}

Table 10: Statistical analysis for composition n-butane validation and testing based on closed loop data.

Parameter	Online
rmse_bottom_validation	3.73E-09
rmse_top_validation	7.16E-09
CDC_bottom_validation	98.16
CDC_top_validation	98.16
R_bottom_validation	1
R_top_validation	1
AIC_bottom_validation	-796.17
AIC_top_validation	-505.38
BIC_bottom_validation	-785.36
BIC_top_validation	-494.58
MAPE_bottom_validation	5.93E-06
MAPE_top_validation	5.53E-06
Cp_bottom_validation	1
Cp_top_validation	1
rmse_bottom_testing	4.20E-09
rmse_top_testing	8.06E-09
CDC_bottom_testing	98.14
CDC_top_testing	98.14
R_bottom_testing	1
R_top_testing	1
AIC_bottom_testing	-750.06
AIC_top_testing	-461.92
BIC_bottom_testing	-739.30
BIC_top_testing	-451.160
MAPE_bottom_testing	6.65E-06
MAPE_top_testing	6.23E-06
Cp_bottom_testing	1
Cp_top_testing	1

for the top composition validation data set. On the other hand the CDC is high for the top and bottom compositions for validation at 98.16. The high CDC values indicate D_i is equal to 1 based on the prediction is larger than D_i which is zero. The CDC is also high for testing data sets. The regression value of R is 1, thus the prediction between the actual and simulated is similar for validation and testing data sets. The AIC and BIC values for online top composition validation are -505 and -494 respectively. The AIC and BIC values for online bottom composition validation are -796 and -785 respectively. For the AIC and BIC, low value is preferred as it indicates better prediction also applies to testing data set. The MAPE values for top and bottom composition validation online are 5.53×10^{-6} and 5.93×10^{-6} respectively. The MAPE should be close to 0 and this also applies to testing data set. Cp values are close to 1. The Cp value for validation and test set for bottom and top composition are calculated to be 1. Table 10 shows the statistical analysis for the n-butane composition prediction validation and testing based on close loop data. Figs. 31-32 show the residual analysis of n-butane for testing and validation data set for top and bottom composition. From the analysis it can be concluded that the residual is very small and hence the prediction obtained for the neural network estimation is highly accurate.

CONCLUSIONS

This paper presents a case study in utilizing a neural network model to estimate the top and bottom compositions of a debutanizer column. The use of online close loop data, open loop data and simulation data makes the model robust and highly suitable for online use. PCA and PLS analyses have also been found to facilitate the correct and right inputs of the variables for the column since a large amount of data are available from the industry.

The neural network prediction of n-butane gives high accuracy and the error between the prediction and actual composition is small and this will ensure that neural network can also be used as an inferential estimator for composition estimation online. From the statistical analysis for top and bottom compositions n-butane, it indicates the RMSE is small for extract from close loop compared to open loop response. Therefore the extract from close loop perform is better than the open loop response.

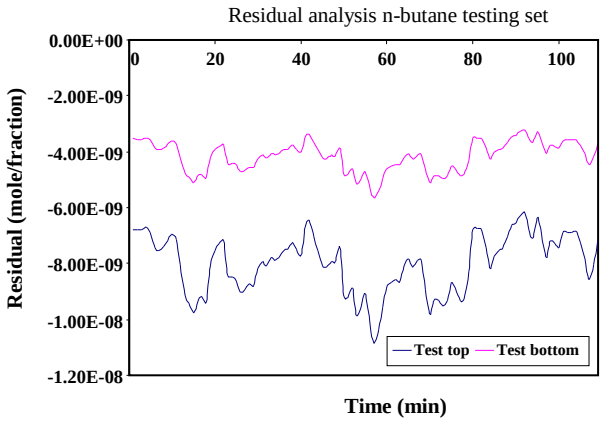

Fig. 31: Residual analysis n-butane test data set.

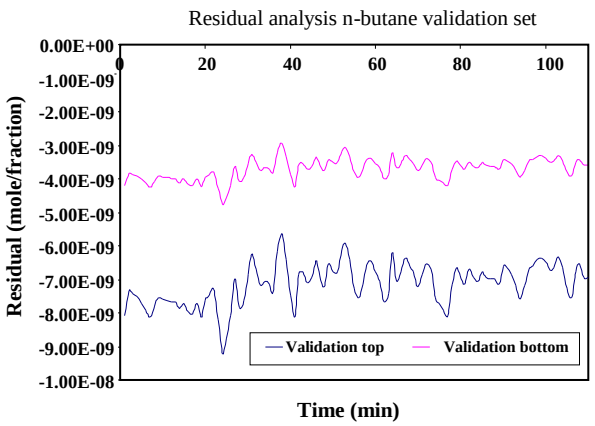

Fig. 32: Residual analysis n-butane validation data set.

Acknowledgment

The authors would like to acknowledge PETRONAS for providing the required data and information for the research and Universiti of Malaya for providing the research grant. (PS107/2010B)

Nomenclature

A_t	Actual value
$x_{meamsured}$	Measure value
C_p	Person correlation co-efficient
$x_{predicted}$	Predicted value
D_i	Product $y_i \times \overline{y_i}$
y_i	Difference actual and average actual
E_a	Actual value
E_p	Predicted value
$\overline{y_i}$	Difference predicted and average predicted
$\overline{E_a}$	Average actual value
$\overline{E_p}$	Average predicted value
σ^2	Variance
F_t	Predicted value
y_t	Data set certain value
K	Number of free model parameters
f_t	Associated model value
MSE	Mean square error
y	Mean of observed data
N	Number of observation
R^2	R squared

Received : Jun. 1, 2014 ; Accepted : Jun. 20, 2016

REFERENCES

[1] Seader J.D., Header E.J., "Separation Process Principle" John Wiley and Sons, 2nd edition (2006).

[2] Guptal S., Ray S., Samanta A.N., Nonlinear Control of Debutanizer Column Using Profile Position Observer, *Computers and Chemical Engineering*, **33**: 1-36 (2008).

[3] Hori E.S., Skogestad S., Selection of Control Structure and Temperature Location for Two Product Distillation Columns, *Chemical Engineering Research and Design*, **85**: 293-306 (2007).

[4] Peng B., Li Xi., Sheng M., Song S., Zhao G., Li Xi., Study of Dual Temperature Control Method on Cyclic Total Reflux Batch Distillation, *Chemical Engineering and Processing*, **46**: 769-772 (2007).

[5] Luyben W.L., Evaluation of Criteria for Selecting Temperature Control Trays in Distillation Columns, *Journal of Process Control*, **16**: 115–134 (2006).

[6] Enagandula S., Riggs James B., Distillation Control Configuration Selection Based on Product Variability Prediction *Control Engineering Practice*, **14**: 743-755 (2006).

[7] Liu X., Kruger U., Littler T., Xie L., Wang S., Moving Window Kernel PCA for Adaptive Monitoring of Nonlinear Processes, *Chemometrics and Intelligent Laboratory Systems*, **96**: 132-143 (2009).

[8] Zamprogna E., Barolo M., Seborg Dale E., Optimal Selection of Soft Sensor Inputs for Batch Distillation Columns Using Principal Component Analysis, *Journal of Process Control*, **15**: 39-52 (2005).

[9] Chen G., McAvoy Thomas J., Piovoso Michael J., A Multivariate Statistical Controller for Online Quality Improvement, *Journal of Process Control*, **8**: 139-149 (1998).

[10] L Smith Cecil, "Industrial Process Control" Proceedings of the workshop American Institute of Chemical Engineers, 70-75 (1979).

[11] Singh V., Gupta I., Gupta H.O., ANN Based Estimator for Distillation-Inferential Control, *Chemical Engineering and Processing*, **44**: 785–795 (2005).

[12] Fortuna L., Graziania S., Xibilia M.G. , Soft Sensors for Product Quality Monitoring Debutanizer Distillation Columns, *Control Engineering Practice*, **13**: 499-508(2005).

[13] Singh V., Gupta I., and Gupta H.O., ANN-Based Estimator for Distillation Using Levenberg-Marquardt Approach, *Engineering Applications of Artificial Intelligence*, **20**: 249–259(2007).

[14] Baratti R., Corti S., A Feedforward Control Strategy for Distillation Column, *Artificial Intelligence in Engineering*, **11**: 40-412 (1997).

[15] Zilochian A., Bawazir K., "Chapter 7: Application of Neural Network in Oil Refineries " *CRC Press* (2001).

[16] Mujtab I M, Hussain M A., Optimal Operation of Dynamic Processes Under Process-Model Mismatches: Application to Batch Distillation, *Computers Chemical Engineering*, **22**: 621-624 (1998).

[17] Greaves M A, Mujtaba I.M., Barolo M., Trotta A., Hussain M.A., Neural Network Approach to Dynamic Optimization of Batch Distillation : Application to Middle-vessel Column, *Trans IChemE*, **81**: 393-401 (2003).

[18] Seborg Dale E., Edgar Thomas F., Mellichamp Duncan A., "Process Dynamic and Control", John Wiley and Sons, 2nd edition (2003).

[19] Dwyer Aidan O., "Handbook PI and PID Controller Tuning Rules" *World Scientific* (2003).

[20] Mohamed Ramli Nasser., Hussain M. A., Mohamed Jan Badrul., Abdullah Bawadi., Composition Prediction of a Debutanizer Column using Equation Based Artificial Neural Network Model, *Neurocomputing*, **131**: 59-76 (2014).

[21] Wan J., Huang M., Ma Y., Guo W., Wang Y., Zhnag H., Li W., Sun X., Prediction of Effluent Quality of a Paper Mill Wastewater Treatment Using Adaptive Network-Based Fuzzy Inference System, *Applied Soft Computing*, **11**: 3238–3246 (2011).

[22] Sharma R., Singh K., Singhal D., Ghosh R., Neural Network Application for Detecting Process Faults in Packed Towers, *Chemical Engineering and Processing*, **43**: 841–847 (2004).

[23] Eriksson L., Johansson E., Kettaneh-Wold N., Trygg J., Wilstrom C., Wold S., "Multi and Megavariate Data Analysis Part I Basic Principles and Applications", (Umetrics Academy 2nd edition, (2006).