CHAPTER IV

SYSTEM MODELING APPLICATION

The use of NN approach to model the behavior of a BD column which has inherently complex and nonlinear behavior is presented in this chapter. The organization of this chapter is divided into two sections; the first section is the actual process used the mathematical model, and the other section is the use of NN approach to represent the actual process.

4.1 Batch Distillation Process

In this work, a rigorous mathematical model described by Distefano (1968) is employed to represent the actual behavior of a BD column. That model is formulated on a dynamic model using a plate–by–plate modeling approach. In the simulation running, two methods are employed to solve the model equations of the system. There are the explicit Euler method and the Bubble–Point method. The explicit Euler method with a step size of 0.0002 h is used for solving the differential equations which each one of plates consists of the state variables as follows: flow rates, enthalpies, and compositions in both liquid and vapor phase. Note that the temperature is not in a form of state variable, it can be obtained from vapor pressure's equation using the bubble–point calculation procedure. The vapor boil–up rate is set to the maximum value and the vapor flow rate at the top plate is maintained at a constant value. In addition, the details of the mathematical model and the steps for solving the model equations are presented in the Appendix A.

4.1.1 Problem Definition and Simulation

The separation of a ternary mixture in a conventional BD column is considered in this work. An initial mixture of cyclohexane 40.7 %, n-heptane 39.4 %, and toluene 19.9 % are charged to the still, which the amount of the mixture is equal to 2.93 kmol. The holdups of plate and condenser are considered the constant values kept as 3.093×10^{-3} kmol and 35.160×10^{-3} kmol respectively. The BD column is operated using a strategy of constant condenser vapor load kept constant as 2.75 kmol/hr. In addition, all specification of the column summarized in Table 4.1 was reported by Nad and Spiegel (1987).

Table 4.1 The specification of the BD column

• **Column conditions**		
Number of plates (N_P)	:	20 plates
Column pressure (P)	:	761.654 mmHg
Internal plate holdup (H_n)	:	3.093×10^{-3} kmol
Condenser holdup (H_C)	:	35.160×10^{-3} kmol
• **System**		
Components [1 2 3]	:	[Cyclohexane n-Heptane Toluene]
Composition (z)	:	[0.407 0.394 0.199]
Feed (F)	:	2.93 kmol
Vapor load to the condenser (V)	:	2.75 kmol/hr

The purpose of this section is to represent the dynamic behavior of a BD column during production period. As mentioned in chapter III, the initial conditions of production period are the results of start–up period. Therefore, the simulation of start–up period is necessary. The simulation of start–up period can be operated under total reflux ($r=1$) until the concentration of the light component reaches steady state. Then, production period is begun with constant reflux operation ($r=0.8$).

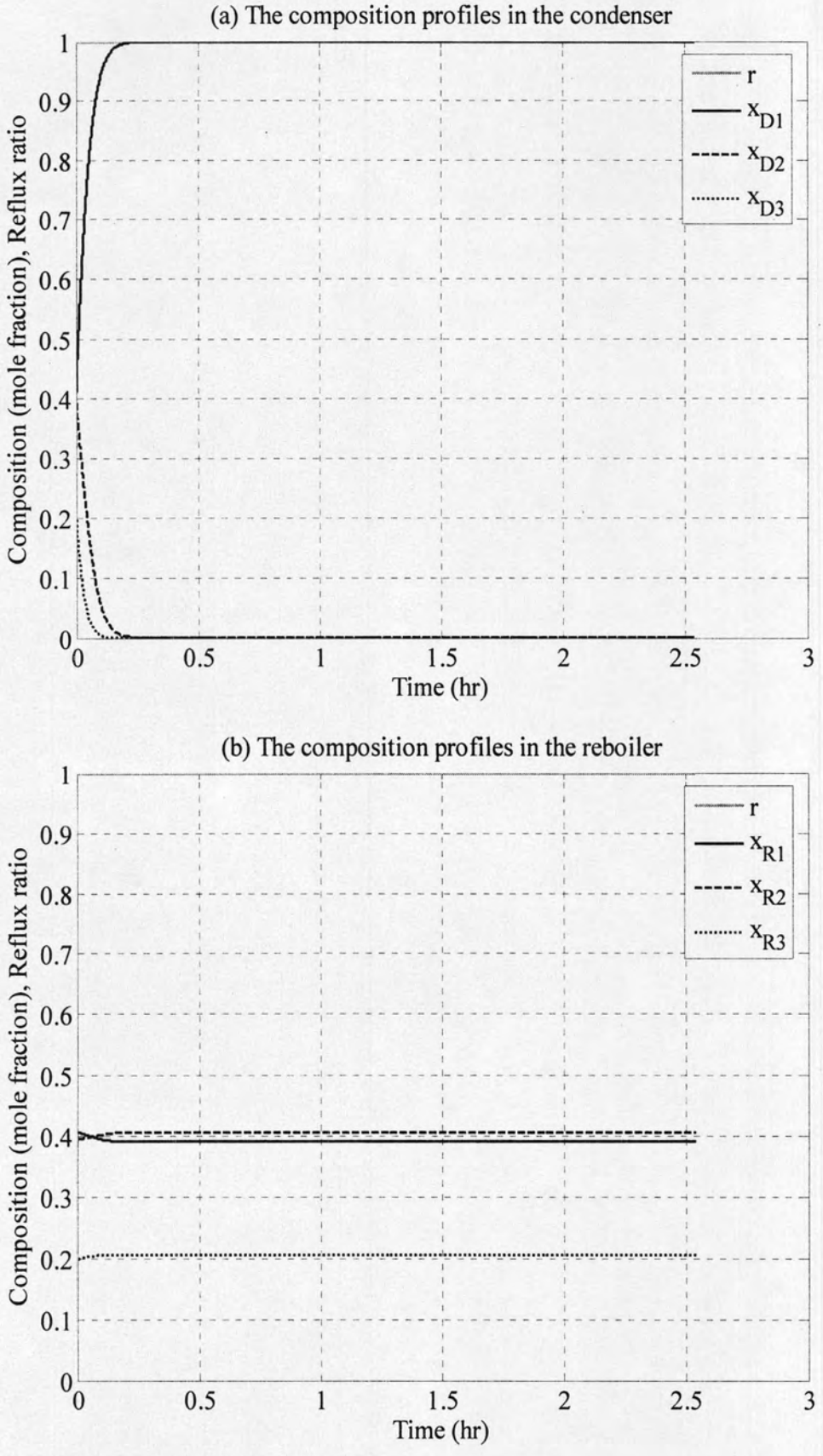


Figure 4.1 The composition profiles in start–up period with total reflux ratio (r = 1)

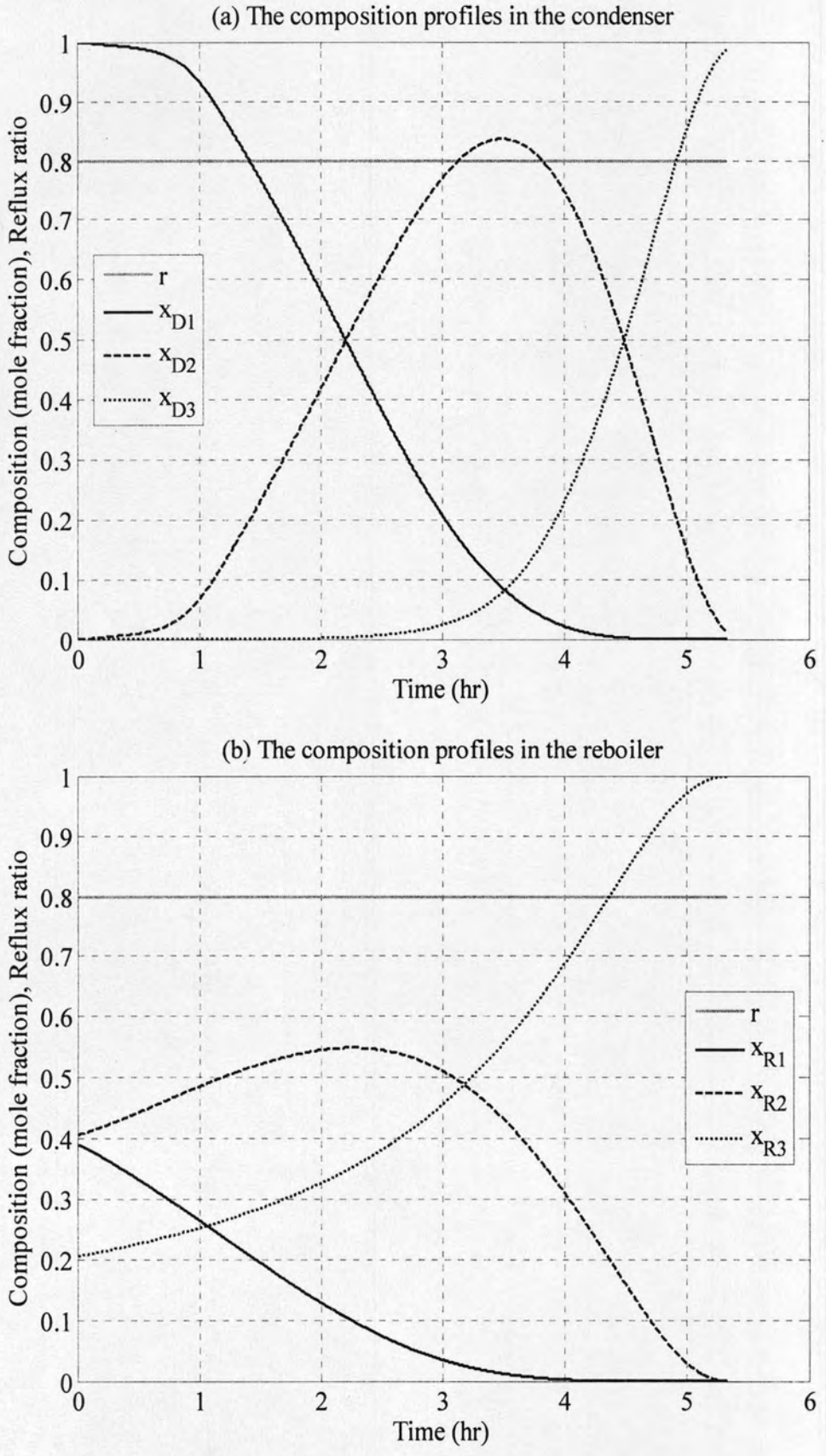


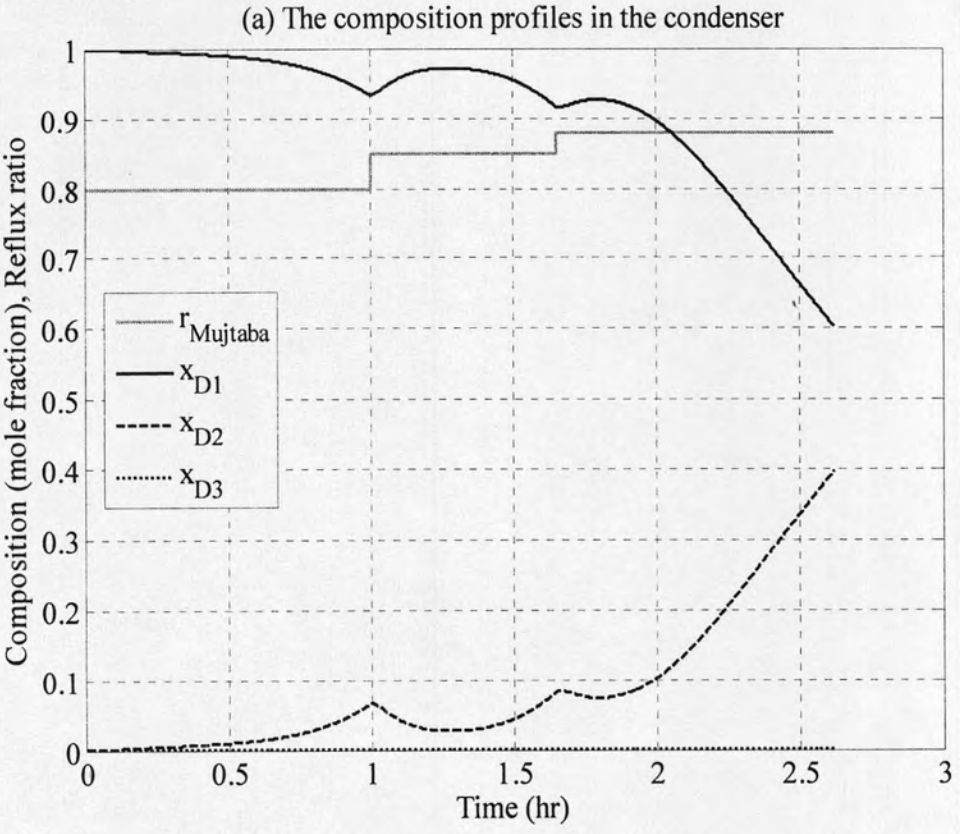
Figure 4.2 The composition profiles in production period with constant reflux ratio (r = 0.8)

4.1.2 Simulation Results

The simulation results of the BD column are presented in Figure 4.1 and Figure 4.2 for start–up and production periods respectively. The distillate and still composition profiles continuously vary with time in both start–up and production periods. During start–up period when the column is operated under total reflux operation, the composition of cyclohexane at the condenser (x_{D1}) increases while the compositions of n-heptane (x_{D2}) and toluene (x_{D3}) decreases until the system reach the steady state as shown in Figure 4.1 (a). Moreover, the composition profiles of cyclohexane (x_{R1}), n-heptane (x_{R2}), and toluene (x_{R3}) at the reboiler are shown in Figure 4.1 (b).

Then the product is withdrawn with a constant reflux ratio, the production period starts. The simulation results are shown in Figure 4.2., it can be seen that the compositions of a BD column are continually changing and the composition of distillate continuously gets heavier as the distillation proceeds and has the slow rate of change. For the first period, x_{D1} begins to decrease while x_{D2} increase and x_{D3} is close to zero. In the next period, x_{D2} and x_{D3} continue to increase for awhile and then decreases while x_{D1} continue decreases along batch time. Due to the amount of cyclohexane and n-heptane is less and less, the toluene will be distilled in the final period.

In addition, the base case which the reflux ratio profile is referred from Mujtaba (2004) is used for simulating the process to evaluate the accuracy of the neural network model. Note that, this work concern only first production period. The simulation results are shown in Figure 4.3.



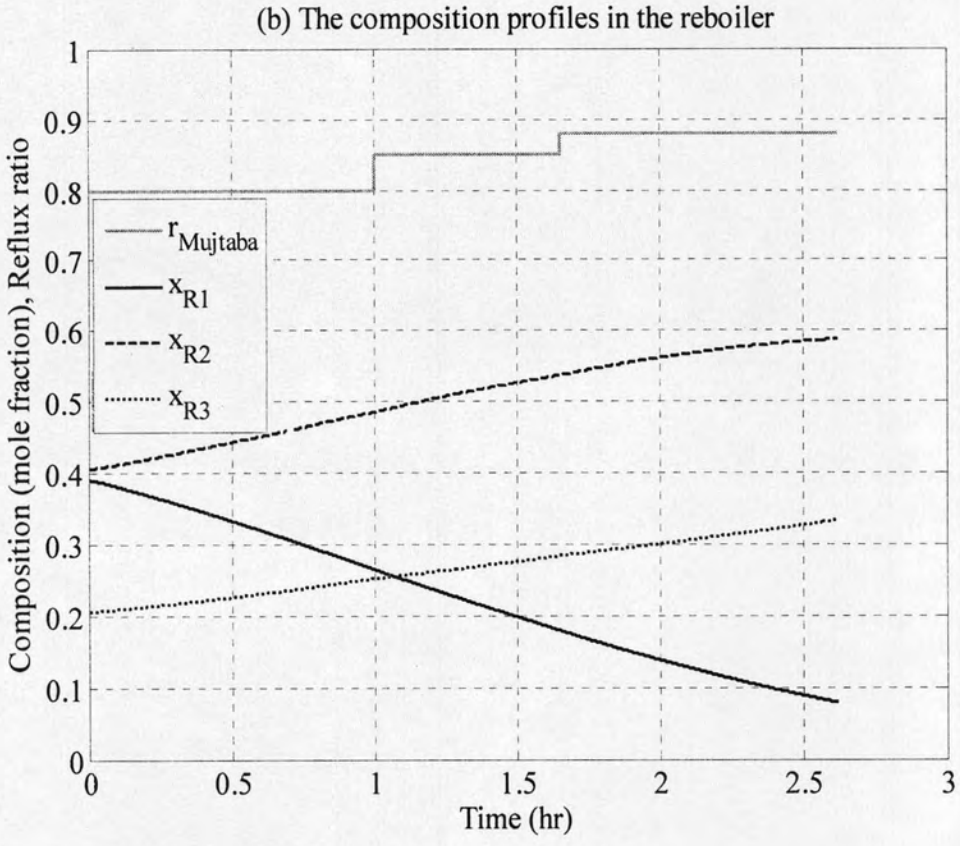
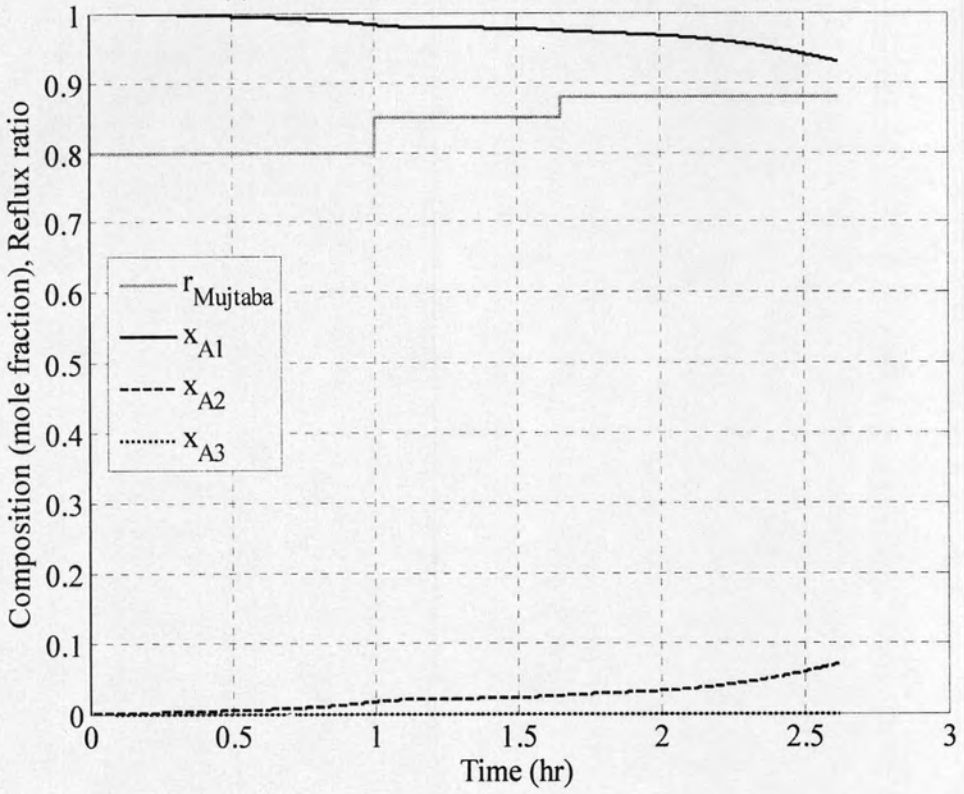


Figure 4.3 Simulation result of the base case

(c) The concentration profiles in the accumulator



(d) The temperature profiles of the plates in a column.

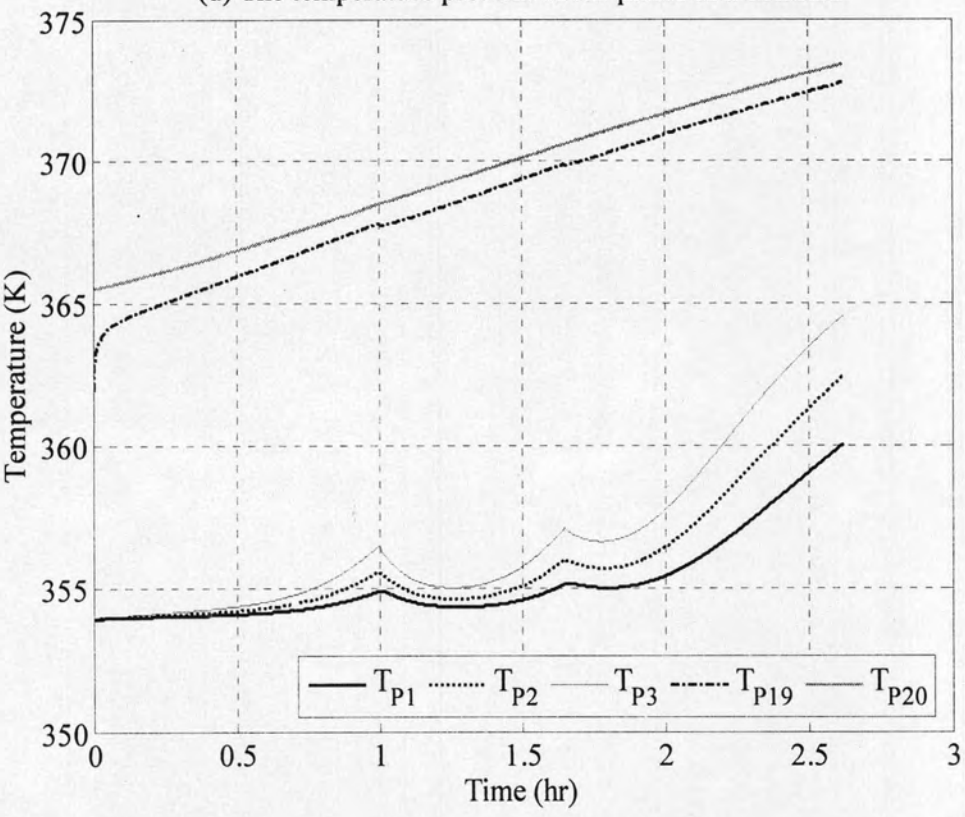


Figure 4.3 Simulation result of the base case (continued)

4.2 Neural Network Modeling

The neural network approach is employed to model the behavior of the BD column, which the objective of the model is to predict the composition at the condenser. Generally, the data for modeling the system must cover the entire system in order to represent accurately the system behavior. For the BD process, the basic process variables which consist of the composition, temperature, and reflux ratio are considered as the input data. Due to temperature measured to represent the process behavior can be measured in many locations, the practical consideration is necessary.

Quintero Marmol et al. (1991) suggested that $N_C + 2$ temperature measurements should be considered, where N_C the number of components is in the feed mixture. Therefore, five temperature measurements is decided due to ternary mixture is considered in this work. Moreover, Zamprogna et al. (2005) determined the optimal sensitive locations when using the temperature measurements as model input of the soft sensor. They also showed that a soft sensor using temperature measurements from those locations provides good estimates of the composition profiles. In addition, the optimal locations are two at the bottom and three at the top of the column. Note that, the number of plates is counted from the top to the bottom of the column.

As mentioned above, it can be concluded that the input variables consists of the composition, the temperature measured with five locations, and reflux ratio is used as the input data. To improve the accuracy of NN model, the data input should consist of the current and past values of the input variables. In addition, the data that represents all possible case is generated by solving the mathematical model mentioned in the previous section. The processes are ran 50 batches and 260 data per batch are collected.

For NN architecture, a multilayer feedforward network which consists of an input layer, one or two hidden layers, and an output layer is employed for modeling the process. The network architecture is shown in Figure 4.4.

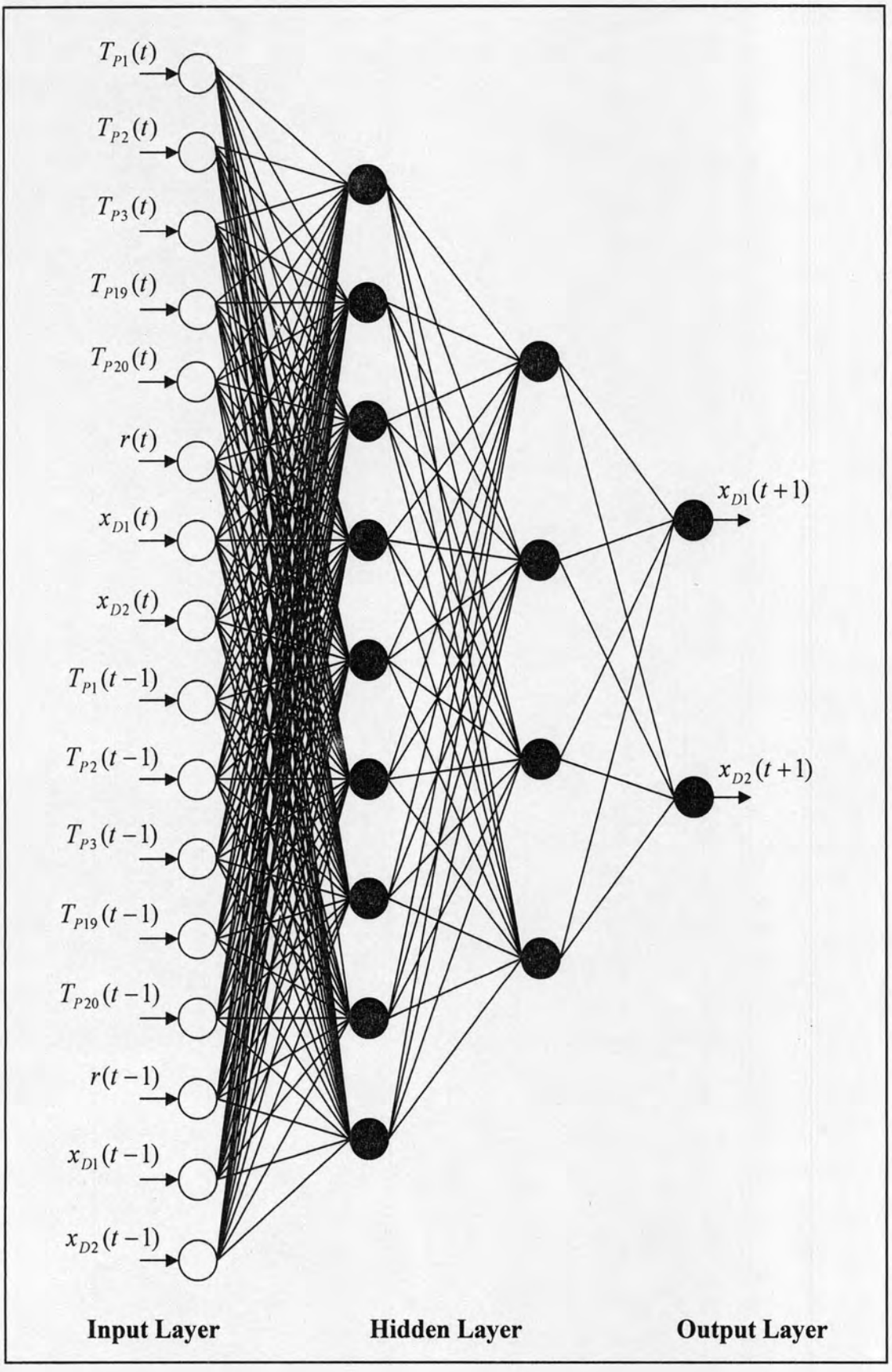


Figure 4.4 Neural network configuration for modeling the BD process

4.2.1 Neural Network Design

For the determination of the appropriate network, the data preparation and network design are concerned. For the data preparation, all data is normalized which it has mean of zero and standard deviation of one. For the network design, the algorithm for training the network is the backpropagation algorithm described in Appendix B which improves the rate and stability of learning by using the Levenberg–Marquardt optimization method. The used summation function which combines with the inputs and weights is a linear function. The used transfer functions are hyperbolic tangent sigmoid and linear functions in the hidden and output layers, respectively. The weights and biases are initialized by generating random numbers. The function which evaluates the performance of modeling is the mean square error (MSE). The network is trained off–line until convergence is reached. The convergence criteria for stopping the network training is the MSE value of the training set or the number of training cycle (epoch) to be kept as 1.0×10^{-8} or 1000, respectively. In addition, the technique of early stopping is used for improving the capability of the network generalization. In this technique, the data is partitioned into three sets as follow: 60% for a training set, 20% for a validation set, and 20% for a test set. The training set is used for updating the network weights and biases. The validation set is used for monitoring during training process, and the test set is used to compare different networks.

The network is trained off–line until convergence is reached. During network training, the early stopping technique (see Appendix B) is used to improve neural network generalization capability and to avoid overfitting of the training data.

The major issue in the configuration design is to determine the hidden layers size. Since there is no theoretic guidance available, the design has to be based on a heuristic approach. The appropriate configuration as well as the number of hidden nodes and layers is determined by comparative the error of validation set among several networks.

Table 4.2 The components of neural network design

Components		Design
Training algorithm	:	Backpropagation algorithm (BP)
Data normalization	:	Mean = 0 and Standard deviation = 1
Initial weights and biases	:	Random
Summation function ($\sum$)	:	Linear summation function
Transfer function (f)		
Hidden layer	:	Hyperbolic tangent sigmoid transfer function
Output layer	:	Linear transfer function
Performance criteria	:	Mean square error (MSE)
Convergence criteria	:	1.0×10^{-8} for training error, or 1000 epoch
Improve generalization	:	Early stopping

4.2.2 Performance evaluation

In order to test the appropriateness of neural network configuration, as the criteria to compare the results obtained from the mathematical model and neural network model, the mean relative absolute error (RAE) is used, which defined in the following.

$$\%RAE = \frac{1}{n}\sum_{n}\left(\left|\frac{y_{NN} - y_{ACTUAL}}{y_{ACTUAL}}\right|\right)\times 100 \qquad (4.1)$$

where y_{NN} is the output value from neural network model, y_{ACTUAL} is the actual value from the mathematical model, and n is the number of data.

4.2.3 Results and Discussions

Training, validation, and test performances of the neural network modeling are also illustrated in the Figure 4.5. The training stopped after 16 epochs because the network begins overfitting which is the validation error will begin to rise up while the training error still to decrease, this means that the network starts to memorize the training set. Moreover, the validation error decreases slower than that of the training set and reaches a minimum than the test error, this indicate that the data set is good partitioning. Note that the test error is not used during training but it is used to compare several models after training model.

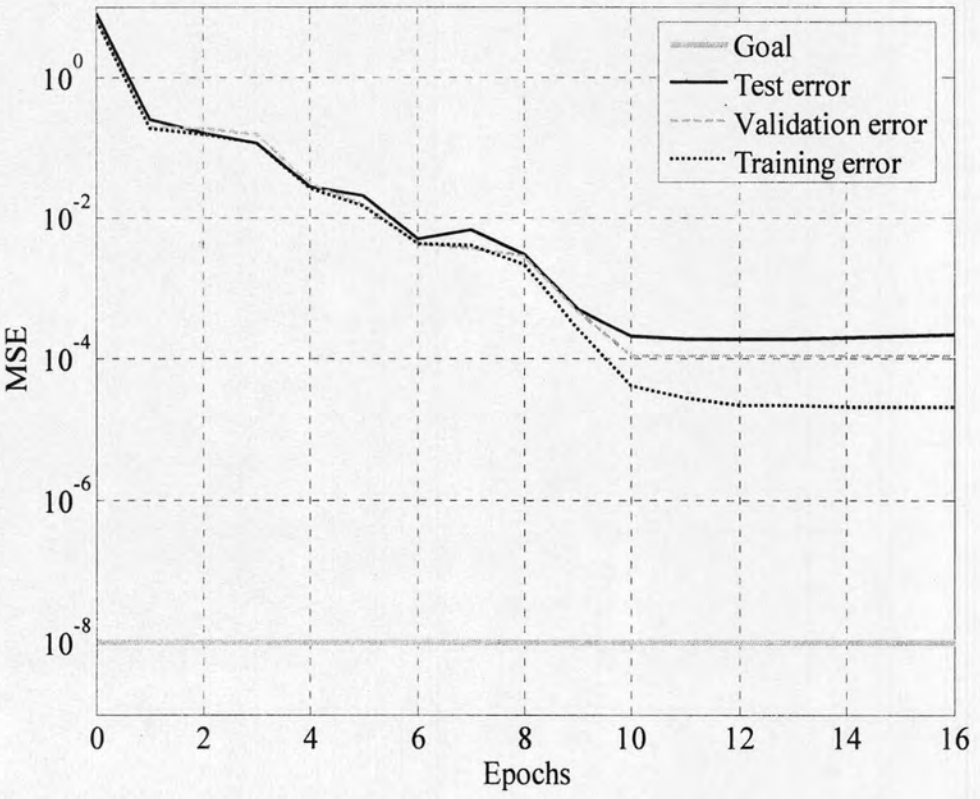


Figure 4.5 Training, validation, and test performances

From the both results mentioned above, the network configuration prefers the network with two hidden layers to the network with one hidden layer, which is considered the MSE value. Therefore, the configuration of $16-6-4-2$ network is selected as the appropriate network employed to represent the behavior of the BD process.

In addition, the appropriate network is also to test the network capability using the simulation under base case condition which the reflux ratio is 0.8. The results of the simulation are shown in Figure 4.6 which the percentages of RAE are shown in Table 4.5.

Table 4.3 Effect of number of hidden nodes in a hidden layer on the MSE obtained with the training, validation, and test sets.

Number of nodes S1	Training error $\times 10^{-5}$	Validation error $\times 10^{-5}$	Test error $\times 10^{-5}$
2	3.2516	2.7053	5.2924
4	2.9149	3.6410	5.8685
6	2.5216	2.7317	4.5925
8	**2.4039**	**2.4159**	**3.9064**
10	2.3381	2.8014	4.9127
20	2.2835	3.7449	4.6931
30	2.2319	4.2731	7.4962
40	2.1154	7.0291	15.366
50	2.0211	10.902	20.917

As mentioned in the previous section, the appropriate configuration of neural network model can be determined by varying the hidden layer size which is the number of hidden layers varied from one to two layers and the number of hidden nodes varied as follows: from two to fifty nodes for one hidden layer and from two to ten per layer for two hidden layers.

The results of the determination of the appropriate network with only one layer are shown in Table 4.3, it can be seen that the training error decreases as the number of hidden nodes increase to the number of hidden nodes is eight nodes which provide the MSE of the network training as follows: 2.4039×10^{-5} of the training set, 2.4159×10^{-5} of the validation set, and 3.9064×10^{-5} of the test set. Then the number is further increase, the validation and test errors also increase. It means that the early stopping arise at this point $(16-8-2)$ which is selected as the appropriate network for one hidden layer. Note that the test error of twenty hidden nodes does not have this trend due to the network stuck in a local minimum point. Therefore, it can be concluded that the network with too few hidden nodes is incapable of complex process representation. On the other hand, if the network has too many hidden nodes it will be poor generalization for untrained data.

From Table 4.4, it can be observed that the results of two hidden layers network remain to be the same as the one hidden layer network, which the training error decreases as the number of hidden nodes increase. Moreover, the training capability significantly increases when the number of nodes in the first hidden layer is increased. The hidden layer size indicates that the number of weights and biases optimized while training are a large or small number. It can be seen that the extension of hidden layer size causes to the problem of local minimum points. Therefore, the appropriate configuration for two hidden layer network done by trial and error is six nodes of the first hidden layer and four nodes of the second hidden layer $(16-6-4-2)$ which the MSE of training set, validation set, and test set are 2.4039×10^{-5}, 2.5643×10^{-5}, and 3.6109×10^{-5} respectively. Note that the determination of the appropriate network is determined by evaluating the mean square error of the test set

Table 4.4 Effect of number of hidden nodes in two hidden layer on the MSE obtained with the training, validation, and test sets.

Number of nodes		Training error $\times 10^{-5}$	Validation error $\times 10^{-5}$	Test error $\times 10^{-5}$
S1	S2			
2	2	3.4491	6.6454	11.582
	4	3.1488	3.0152	6.4505
	6	2.8810	3.5561	6.5541
	8	2.8564	3.1374	6.5313
	10	2.9756	3.3492	6.5070
4	2	2.9918	3.4159	5.9575
	4	2.4729	2.6139	4.4893
	6	2.3853	2.5476	3.9484
	8	2.4645	2.9304	4.8784
	10	2.5295	2.6826	4.6625
6	2	2.3845	2.8162	4.4872
	4	**2.3598**	**2.5643**	**3.6109**
	6	2.3676	2.8597	4.2433
	8	2.3720	2.8454	4.8634
	10	23549	2.6612	3.9521

Table 4.4 Effect of number of nodes in two hidden layer on the MSE obtained with the training, validation, and test sets (continued).

Number of nodes		Training error $\times 10^{-5}$	Validation error $\times 10^{-5}$	Test error $\times 10^{-5}$
S1	S2			
8	2	2.3726	2.6814	3.7762
	4	2.3994	2.4847	3.9219
	6	2.3525	2.7783	4.6092
	8	2.3343	2.9464	4.7283
	10	2.3539	2.8401	4.0334
10	2	2.3550	2.7106	4.0901
	4	2.3569	2.7213	4.1937
	6	2.3530	2.7490	4.6428
	8	2.3244	3.6531	6.4277
	10	2.3047	2.9010	5.2714

Table 4.5 The percentage of RAE of the simulation results of NN model with two hidden layers (16 – 6 – 4 – 2) compared with the base case results

Output	% RAE
Top composition for Cyclohexane (Xd1)	0.5753
Top composition for n–Heptane (Xd2)	82.5423

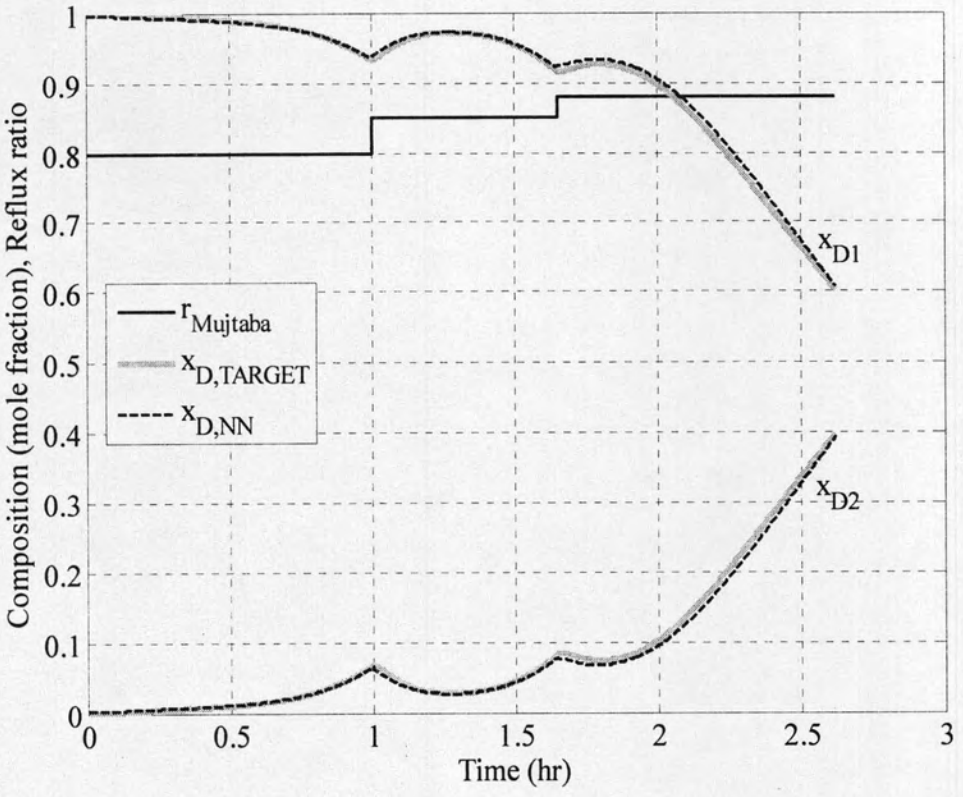


Figure 4.6 Comparison of output profiles calculated from the mathematical model and the NN model with two hidden layers (16 – 6 – 4 – 2)