



Journal of Applied Sciences

ISSN 1812-5654

science
alert

ANSI*net*
an open access publisher
<http://ansinet.com>

Neural Network based Soft Sensor for Inferential Control of a Binary Distillation Column

Mohanad Osman and M. Ramasamy
Department Chemical Engineering, Universiti Teknologi Petronas,
Bandar Seri Iskandar, 31750 Tronoh, Perak, Malaysia

Abstract: Distillation column is one of the most widely used unit operations in process industries and the operation and control of this unit was always a challenging task due to the complexity of operational parameters interactions and the lack of a real time measurement of the products composition. Time delay involved with GC measurement of the product composition prevented the utilization of effective closed-loop control. To overcome this problem an inferential control scheme based on soft sensor estimation of the products composition is now widely adopted. In this work, a neural network based soft sensor is developed to be used in an inferential control scheme of a pilot-scale binary distillation column. Data were collected from the distillation column under different operating conditions with forced disturbances in number of operation variables. The collected data is pre-processed for the removal of outliers, normalized and segmented into training and test data subsets. The most important process variables in the model and their lags have been chosen systematically. Different neural networks have been trained using the preprocessed data. Trial and error method is used to find the optimum number of neuron in the hidden layer for each network. The performance of different networks is discussed. The developed soft sensor can be utilized in an inferential control scheme on the distillation column.

Key words: Soft sensors, neural networks, distillation

INTRODUCTION

Distillation is the most common unit operation in the chemical and petrochemical industries for separation and purification of components. Controlling distillation columns is the most challenging job for control engineers due to their highly nonlinear dynamic characteristics and complexity. The energy requirements for the distillation column might be reduced significantly by controlling the column operation at optimum conditions. Basically, there are five variables required to be controlled to achieve efficient operation, namely, composition of distillate stream, composition of bottom stream, liquid level of reflux drum, liquid level of base column and column pressure (De Canete *et al.*, 2008). In practice, on-line analyzer for composition is rarely used due to large measurement delay, the need for frequent maintenance as well as high capital and operating costs. In order to adapt to changing economic and market scenarios while maximizing profit, the demand for accurate inferential estimators (soft sensors) for controlling the product quality variable becomes paramount (Sit, 2005).

Soft-sensor is a mathematical model that utilizes the measured values of some secondary variables of a process in order to estimate the value of an immeasurable

primary variable (quality parameter) of particular importance (Zamproga *et al.*, 2005). Soft sensors have been widely reported to supplement online instrument measurements for process monitoring and control. Both First Principle Model (FPM)-based and data-driven soft sensors have been developed (Lin *et al.*, 2007). If a FPM describes the process sufficiently accurately, a FPM-based soft sensor can be derived. However, a soft sensor based on detailed FPM is computationally intensive for real-time applications (Zamproga *et al.*, 2005). Modern measurement techniques enable a large amount of operating data to be collected, stored and analyzed, thereby rendering data-driven soft sensor development a viable alternative.

Two main techniques are adopted in the development of data-driven soft sensors, Multiple Linear Regressions (MLR) such as Partial Least Squares (PLS) and Multiple Nonlinear Regressions (MNR). The main drawback of MLR techniques is their linear nature which hinders them from accurately describing processes with nonlinear nature like distillation. On the other hand, MNR techniques like Artificial Neural Network (ANN) and Neuro-Fuzzy can easily capture and accurately represent the nonlinear characteristics of processes and they become increasingly adopted (Hoskins *et al.*, 1988). In

this study, an attempt has been made to develop a neural network-based soft sensor for a pilot-scale binary distillation column.

Artificial neural networks: Over the years, the application of ANN in process industries has been growing in acceptance. ANN is attractive due to its information processing characteristics such as nonlinearity, high parallelism, fault tolerance as well as capability to generalize and handle imprecise information (Hoskins *et al.*, 1988). Such characteristics have made ANN suitable for solving a variety of problems. This has been proven in various fields such as pattern recognition, system identification, prediction, signal processing, fault detection, soft sensors and others. In general, the development of a good ANN model depends on several factors. The first factor is related to the data being used. The model quality is strongly influenced by the quality of data used. The second factor is network architecture or model structure. Different network architecture results in different estimation performances. The third factor is the model size and complexity. A parsimonious model i.e., a model with a fewer number of parameters is required for effective on-line applications. A small network may not be able to represent the real situation due to its limited capability, but a large network may over fit noise in the training data and fail to provide good generalization ability. Finally, the quality of a process model also depends strongly on network training (Picton, 2000).

EQUIPMENT AND EXPERIMENTATION

The equipment used in this study is a pilot-scale binary distillation column shown in Fig. 1. The column is 5.5 m high with 0.156 m inner diameter, consisting of 15 bubble-cap trays and works under atmospheric pressure. The feed enters the column at tray number 7. The column is used to separate acetone as top product from isopropyl alcohol (IPA). Table 1 and 2 show the design specifications and the nominal operating conditions of the column.

The column was operated at different feed flow rates (0.5-0.9 L min⁻¹) and different feed compositions (0.01-0.1 mole fraction of acetone). Series of step changes in the reflux flow rate (0.2-0.9 L min⁻¹) and reboiler steam flow rate (0.017-0.5 kg min⁻¹) are introduced as excitation signals to produce a wide range of temperature profiles and top product concentrations.

Samples were collected from the top product stream after the reflux drum regularly at an interval of 6 minutes and the data on other variables such as flow rates, temperatures were acquired through the data acquisition system and stored at a sampling interval of 2 sec.

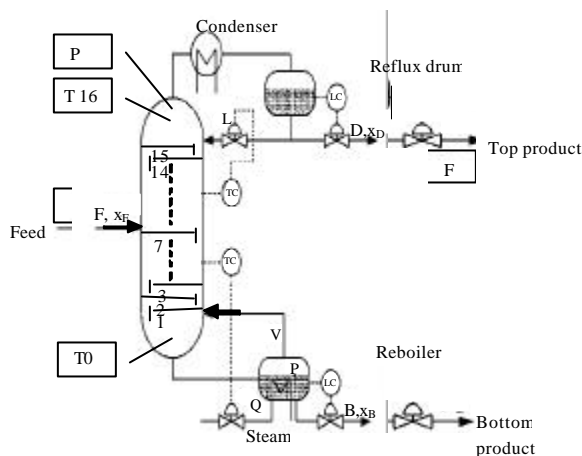


Fig. 1: Process flow diagram of pilot-scale binary distillation column

Table 1: Design specifications of the pilot-scale distillation column

Specifications	Description
Height	5.5 m
Inner diameter	0.156 m
Number of trays	15
Tray type	Bubble-cap
Feed tray	Tray No. 7
Operating pressure	1.013 bar
Feed mixture	Acetone and isopropyl alcohol mixture (industrial grade)

Table 2: Nominal operating conditions of the distillation column

Operating parameter	Normal operation value
Feed flow rate	0.7 L min ⁻¹
Reboiler steam flow rate	0.33 kg h ⁻¹
Reflux flow rate	0.7 L min ⁻¹
Top product flow rate	0.3 L min ⁻¹
Column bottom liquid level	0.4 m
Reflux drum liquid level	0.2 m

Table 3: Description of process variables

Tag	Description	Unit
F1	Reboiler steam flow rate	kg min ⁻¹
F4	Reflux flow rate	L min ⁻¹
P	Column top pressure	bar
T0	Column bottom temperature	°C
T _r	Tray temperatures along the column	°C

It may be noted that during the experiments, the column pressure was controlled by manipulating the cooling water flow rate, the reflux drum level by the top product flow rate and the column bottom level by bottom product flow rate. Gas Chromatographic analysis was done at a later time to find the mole fraction (X_A) of acetone in the top product from the collected samples. In total, 270 samples were collected with their corresponding operating conditions. The operating variables collected consist of reflux flow rate, reboiler steam flow rate, top pressure and temperatures of all trays in the column. Table 3 lists all the operating parameters collected as input variables along with their description and units of measurement.

SOFT SENSOR DEVELOPMENT

Typically the development of a soft sensor can take the following steps:

- Collection of the data that will be used in the construction and validation of the soft sensor
- Pre-processing of the data, which includes outlier removal and normalization
- Soft sensor construction and validation

Data pre-processing: Sufficient data were collected from the operation of the distillation column as explained earlier. Before the collected data can be used for the development of a soft sensor, they must be pre-processed in three steps:

- Outliers detection and removal
- Data normalization
- Data segmentation into training and testing sets

Outliers can be defined as observations with abnormally extreme values (Eriksson *et al.*, 2001). Outlier detection is very critical for the soft sensor development because outliers have negative effect on the performance of the soft sensor model. Different reasons might be the source of outlier such as sensor failure, transmission problems or unusual operating conditions (Fortuna *et al.*, 2007). Outlier can be detected easily in the score plot of a PCA of the data (Eriksson *et al.*, 2001). Figure 2 shows the score plot of the first two principal components (t[1], t[2]) for all 270 observations collected. From this Fig. 2,

observations 38 through 42 can be identified as clear outliers and hence they have been removed from the bulk of data.

There is a significant difference in the numerical magnitude of the different data variables due to the difference in their units of measurement. Variables with larger magnitude will have an unjustified stronger influence in the development of the model (Fortuna *et al.*, 2007). Therefore, data normalization (or scaling) is required to assure an equal influence of the different variables in the model. Two types of normalization can be used, the first is the (min-max normalization) which is given by:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{1}$$

where, x is the measured variable, x' is the normalized variable, x_{\min} is the minimum value of the variables, x_{\max} is the maximum of the variables.

The second type is (z-score normalization). This normalization will make all the data to have a zero mean and unity variance and is given by:

$$x' = \frac{x - \bar{x}}{\sigma} \tag{2}$$

where, \bar{x} and σ are the mean and the standard deviation of the input variable, respectively. The z-score normalization was used in this study because it is less sensitive to the existence of outliers in the data (Fortuna *et al.*, 2007).

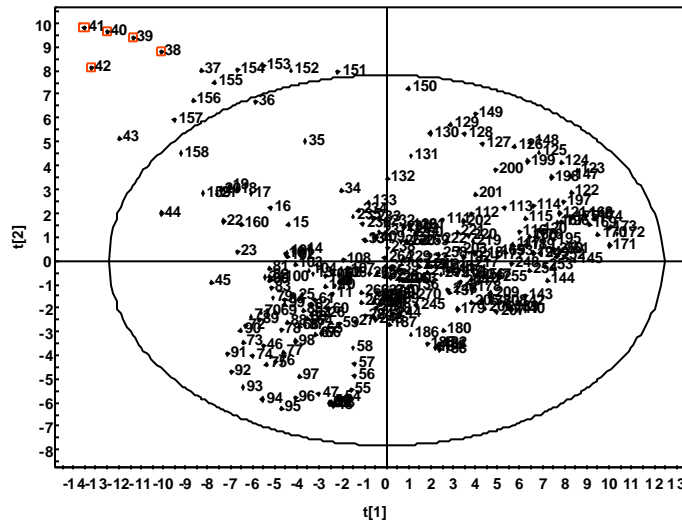


Fig. 2: Score plot of the first two principal components of a PCA study for 270 observations

After the removal of outliers and data normalization, the remaining 265 observations data are segmented into two subsets: a 215 observations training set and a 50 observations test set.

SOFT SENSOR CONSTRUCTION

The construction of a neural network (NN) based soft sensor can take the following steps:

- Selection of important process variables and their corresponding lags
- Selection of the NN model type and structure
- Model validation

Important variables selection: Usually data collected from industrial units suffer from co-linearity which makes many of them poorly informative in the prediction of the quality parameter (Wang *et al.*, 1996). Using redundant variables increases the complexity of the model and has bad effects on the performance of the soft sensor. One way to solve this problem is to select only part of the input variables that have the least co-linearity and the most correlation with the quality parameter.

Due to the process dynamic behavior, different time lags are expected to exist between different variables and their corresponding effect in the primary output. Typically the choice of the important variables and their corresponding lags is done simultaneously. First a large set of variables and their different lags are considered as input candidates for the model. Number of methods can be used identify a subset of variables as the relevant model inputs.

In this study, a PLS method is used to determine the importance of the input variables and their most correlated lags with the primary output. The Variable Influence Plot (VIP) represents a Weighted Sum of Squares of the Weights (WSSW) assigned to the variables in the PLS model, which summarizes the influence and hence the importance of each variable in the prediction of the primary output (Fortuna *et al.*, 2006). For each of the input variables a sequence of 5 lags (k-1, k-2, ..., k-5) were obtained with one sample delay between each other, then the variables and their lags are used in the PLS model. Figure 3 shows the VIP of the variables and their lags. Two lags of each variable with the highest VIP value are shaded. These two lags for each variable are the ones that will be used later as input candidates for the soft sensor.

NN structure: Two neural networks are trained for the prediction of X_A , a Feed Forward (FF) network and a nonlinear autoregressive with exogenous input (NARX) network. Both networks are of three layers with an input layer, a hidden layer and an output layer. Levenberg-Marquardt backpropagation with momentum technique (Hagan *et al.*, 1996) is used as training algorithm for both types. The relationship between the inputs and the output of these networks can be given by the general equation:

$$\hat{y} = E_2 \left[\sum_{j=1}^{n_2} w_{ij}^2 E_1 \left(\sum_{i=1}^{n_1} w_{ij}^1 x_i + w_{0j}^1 \right) + w_{0j}^2 \right] \tag{3}$$

where, \hat{y} is the predicted output, x_j is the j th input from the input set x , w_{ij}^L is the weight of the i th input to the j th neuron of the L th layer, w_{0j}^L is the bias of the j th neuron of the L th layer, n_L is the number of neurons in the

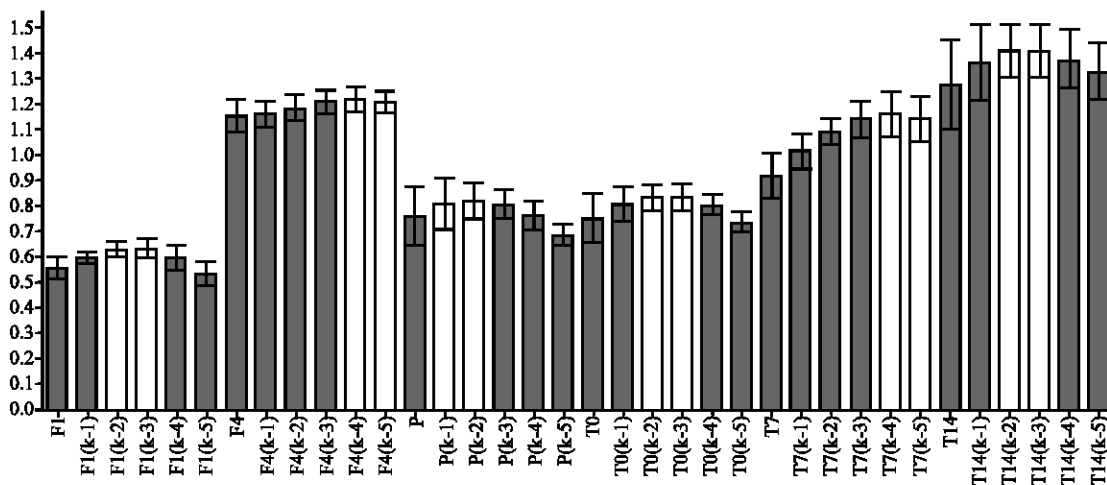


Fig. 3: VIP of the input variables and their lags

hidden layer and n_1 is the number of inputs to the network, F_1 is the activation function of the hidden layer, F_2 is the activation function of the output layer.

The input set x for the FF network is defined as:

$$x = [u_1(k-d_1), u_1(k-d_2), \dots, u_n(k-d_1), u_n(k-d_2)] \quad (4)$$

where, u is an input variable, k is the sample number, d_1 and d_2 are two successive lags of the input variables.

While the input set x for the NARX is defined as:

$$x = [u_1(k-d_1), u_1(k-d_2), \dots, u_n(k-d_1), u_n(k-d_2), y(k-1), \dots, y(k-d_y)] \quad (5)$$

where, y is the actual output, d_y is the maximum lag (degree) of the recurrent output.

Two different activation functions of the hidden layer (F_1) are investigated for each network. In one case a purelin activation function is used, which simply means to use the weighted sum of variables directly as the activation signal from the hidden layer to the output layer, or:

$$F_1(z) = z \quad (6)$$

In the second case, a tan-sigmoid is used for F_1 , which is given by:

$$F_1(z) = \frac{1}{1 + e^{-z}} \quad (7)$$

Accordingly, there were four neural networks investigated for their prediction ability of the quality parameter, a FF network with a purelin hidden layer activation function (ff-lin), a FF network with a tan-sigmoid hidden layer activation function (ff-tsig), a NARX network with a purelin hidden layer activation function (narx-lin), a NARX network with a tan-sigmoid hidden layer activation function (narx-tsig). For all of the investigated networks a purelin function is used as the activation function for the output layer F_2 .

The next step of the soft sensor construction is to choose the input variables from the input candidates that will give the best network performance. The Root Mean Square Error (RMSE) between the predicted and the actual output in the test set of observations is used as a performance index for the neural network. RMSE can be described by:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N [\hat{y}_i(k) - y_i(k)]^2}{N}} \quad (8)$$

where, N is the number of observations in the test data set.

Table 4: Input Variables for Investigated Neural Networks

Network	Input variables set	No. of inputs
ff-lin	T0(k-2), T0(k-3), T7(k-4), T7(k-5), T14(k-2), T14(k-3), F4(k-4), F4(k-5).	8
ff-tsig	T0(k-2), T0(k-3), T7(k-4), T7(k-5), T14(k-2), T14(k-3), P(k-1), P(k-2).	8
narx-lin	T0(k-2), T0(k-3), T14(k-2), T14(k-3), $X_A(k-1)$.	5
narx-tsig	T0(k-2), T0(k-3), T14(k-2), T14(k-3), P(k-1), P(k-2), $X_A(k-1)$	7

The algorithm used to select the set of input variables x for each network, is to add the variables one by one (as a set of two lags of each) to the particular model in order of decreasing VIP values, the network is trained using the training set of observations, simulated for the test set and the performance is tested by calculating RMSE for the test set. The variable that improves the network performance (decreases RMSE) is kept and the one that increases RMSE is discarded and then the next input variable is tested and so on. The different networks investigated had different input structures that give the best performance as shown in Table 4. For the NARX networks different maximum lags (degrees) of the recurrent output variable $y(k-d_y)$ (which is represented here by ($X_A(k-d_y)$)) are investigated and it was found that only a first degree lag ($y(k-1)$) improves the network performance while the other lags increases the RMSE, that's why only $X_A(k-1)$ is included in the NARX networks model.

Different number of neurons in the hidden layer can significantly affect the network performance (Hagan *et al.*, 1996), hence the number of neurons in the hidden layer has to be optimized. To find the optimum number of neurons in the hidden layer, the four neural networks with input structures as described in Table 4 are trained with different number of neurons in the hidden layer. The performance of the networks with different number of neurons in the hidden layer is shown in Fig. 4. For networks with purelin hidden layer activation function the network performance was independent of the number of neurons in the hidden layer. Accordingly a hidden layer with only one neuron is used in the structure of ff-lin and narx-lin networks because of the decreased computation effort compared to a multi-neuron layer. On the other hand the performance of networks with tan-sigmoid function fluctuates with minimum RMSE value and hence best performance at 2 neurons for ff-tsig network and 5 neurons for.

Model validation: Table 5 summarizes the performance of the optimum structure of each type of the investigated neural network. It can be seen from this table that all of the networks had a very good performance with an RMSE less than 0.04, or an average error less than 7% of the total

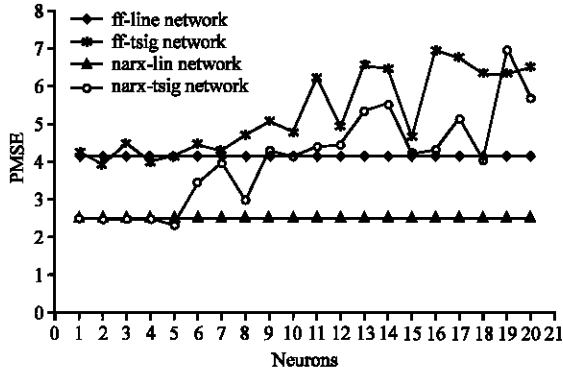


Fig. 4: Performance of Neural Networks as Function of Number of Neurons in the Hidden Layer: ff-lin network, ff-tsig network, narx-lin network, narx-tsig network

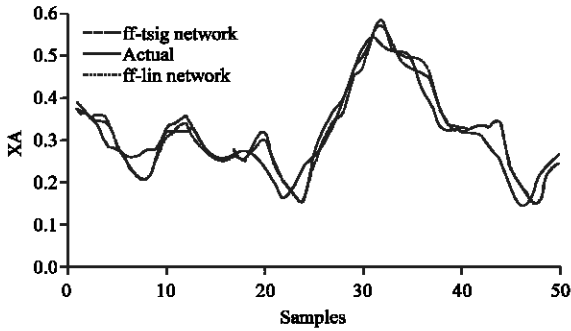


Fig. 5: Performance of FF neural networks

Table 5: Performance of the optimum structure for the investigated neural networks

Network	Network structure	RMSE	% error
ff-lin	8-1-1	0.0412	6.9
ff-tsig	8-2-1	0.0393	6.6
narx-lin	5-1-1	0.025	4.2
narx-tsig	7-5-1	0.0232	3.9

range of top product composition measured (0.023-0.64). For further validation of the prediction accuracy of the optimum NN structure based soft sensor, plots of the predicted and actual top product composition (X_A) are shown in Fig. 5 and 6. It is clear from these figures and Table 5 that networks with tan-sigmoid hidden activation functions have outperformed those with purelin functions of the same type. This indicates a high level of nonlinearity for the investigated system. In general, NARX networks of both types have outperformed FF networks and have given an excellent prediction performance as can be seen in Fig. 5 and 6.

Inferential control scheme: The soft sensor can be used in an inferential control scheme as shown in Fig. 7. Here,

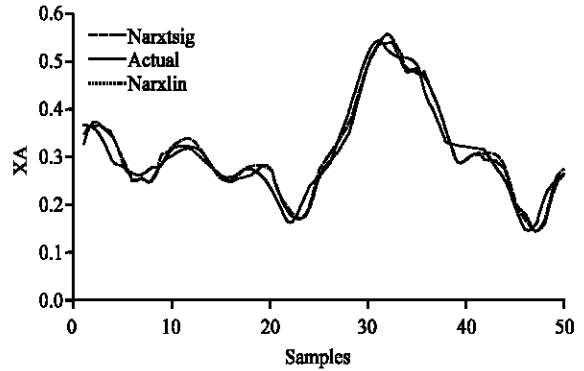


Fig. 6: Performance of NARX neural networks

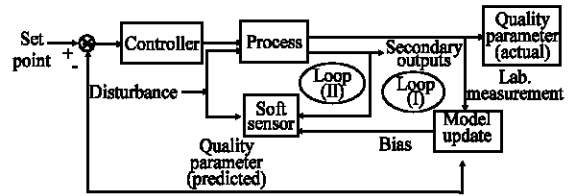


Fig. 7: Inferential control scheme

Loop (II) is updated each few seconds and the secondary outputs (network input variables) are used to predict the value of the quality parameter (top product composition) which is used by the controller to control the process. Loop (I) is updated every 30 to 40 min when the actual values of the top product composition are available and this value is used to update the soft sensor model by calculating the error between the actual and the predicted values which is then added to the model as a bias.

CONCLUSIONS

Two types on neural networks each with two different hidden layer activation function studied have shown a good prediction ability for the top product composition of the pilot distillation column. NARX networks had a better prediction performance with both types of hidden layer activation function than FF network. Using tan-sigmoid as hidden layer activation function had a better performance than purelin activation function. Input variables and number of neurons in the hidden layer are found to depend significantly on the type of the network and its activation function. The constructed soft sensor with its optimized input variables set and neural network can be used in an inferential closed loop control scheme for the pilot-scale distillation column.

ACKNOWLEDGMENTS

Authors gratefully acknowledge the financial support under the Science Project 03-02-02-SF0003 by the Ministry of Science, Technology and Innovation, Malaysia and Universiti Teknologi PETRONAS.

REFERENCES

- De Canete, J.F., S. Gonzalez-Perez and P. Saz-Orozco, 2008. Artificial neural networks for identification and control of a lab-scale distillation column using LABVIEW. *Int. J. Intell. Syst. Technol.*, 3: 111-112.
- Eriksson, L., E. Johansson, N. Kettaneh-Wold and S. Wold, 2001. *Multi and Megavariate Data Analysis Principles and Applications*. Umetrics Academy, Sweden, ISBN-10: 919737301X.
- Fortuna, L., S. Graziani, A. Rizzo and M.G. Xibilia, 2007. *Soft Sensors for Monitoring and Control of Industrial Processes*. Springer, London, ISBN: 978-1-84628-479-3, pp: 270.
- Fortuna, L., S. Graziani, M.G. Xibilia and G. Napoli, 2006. Comparing regressors selection methods for the Soft Sensor design of a Sulfur Recovery Unit. *Proceedings of the 14th Mediterranean Conference on Control and Automation, (MCCA'06)*, IEEE Explore, London, pp: 1-6.
- Hagan, M.T., H.B. Demuth and M.H. Beale, 1996. *Neural Network Design*. 1st Edn., PWS Publishing Co., Boston, MA, USA., ISBN: 0-53494332-2.
- Hoskins, J.C. and D.M. Himmelblau, 1988. Artificial neural network models of knowledge presentation in chemical engineering. *Comp. Chem. Eng.*, 12: 881-890.
- Lin, B., B. Recke, J.K.H. Knudsen and S.B. Jørgensen, 2007. A systematic approach for soft sensor development. *Comput. Chem. Eng.*, 31: 419-425.
- Picton, P.D., 2000. *Neural Networks (Grassroots)*. 2nd Edn., Palgrave Macmillan, New York, ISBN-13: 978-0333802878.
- Sit, C.W., 2005. Application of artificial neural network-genetic algorithm in inferential estimation and control of a distillation column. M.Sc. Thesis, Universiti Teknologi Malaysia.
- Wang, X., R. Luo and H. Shao, 1996. Designing a soft sensor for a distillation column with the fuzzy distributed radial basis function neural network. *Proceedings of the 35th Conference on Decision and Control*, Dec. 11-13, IEEE, Kobe, Japan, pp: 1714-1719.
- Zamprogna, E., M. Barolo and D.E. Seborg, 2005. Optimal selection of soft sensor inputs for batch distillation columns using principal component analysis. *J. Process Control*, 15: 39-52.