International Journal of Chemical, Materials and Biomolecular Sciences
ISSN: 2415-6620
Vol:5, No:10, 2011

Development of Gas Chromatography Model: Propylene Concentration Using Neural Network

Areej Babiker Idris Babiker, Rosdiazli Ibrahim

***Abstract*—**Gas chromatography (GC) is the most widely used technique in analytical chemistry. However, GC has high initial cost and requires frequent maintenance. This paper examines the feasibility and potential of using a neural network model as an alternative whenever GC is unvailable. It can also be part of system verification on the performance of GC for preventive maintenance activities. It shows the performance of MultiLayer Perceptron (MLP) with Backpropagation structure. Results demonstrate that neural network model when trained using this structure provides an adequate result and is suitable for this purpose. cm.

***Keywords*—**Analyzer, Levenberg-Marquardt, Gas chromatography, Neural network

I. Introduction

TODAY, Gas Chromatograph (GC) is the most widely used technique in analytical chemistry – a position it has held for over three decades. The popularity and applicability of the technique is principally due to its unchallenged resolving poser for closely related volatile compounds and because of the high sensitivity and selectivity offered by many of the detector systems. The technique is very accurate and precise when used in a routine laboratory [1]. On one hand, GC might be susceptible to damage and so needs frequent maintenance which is widely recognized as a significant contributory factor in the life cycle cost of a process plant. On the other hand, frequent maintenance is impractical as it require shutdown of system or facilities and slows down production process.

The demand for the use of Artificial Neural Networks to solve engineering problems is expected to increase significantly in the next ten years, mainly due to several breakthroughs in this field and also to the limitations of the existing conventional engineering problem solving techniques. Results to date have demonstrated the significant performance advantages of Artificial Neural Networks relative to currently available conventional methods [2]. It is also one of the black-box modelling approaches to do data-driven modelling. Neural network, with its remarkable ability to derive meaning from complicated or imprecise data, can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. It is consisting of many units i.e. processing unit's analogues to neutrons in the brain. Each node has a node function, associated with it which along with a set of local parameters determines the output of the node, given an input as in [3]. The neural network user gathers representative data, and then invokes training algorithms to automatically learn the structure of the data.

This paper focuses on developing a neural network model that can predict the output of GC which is Propylene in this process. Using this method would significantly lower maintenance costs besides achieving efficient process control. At the same time the model also serves as part of online verification tools to check the reliability of the GC or as an alternative solution to provide some consistent information whenever the GC is unavailable. The paper also examines and investigates performance of the neural network model based on training parameters and model structure. As stated, network architecture determines the number of connection weights and the way information flows through the network [4]. Thus, network architecture selection is an important set of tool for neural network modelling..

II. Methodology

Data from a chemical plant were gathered and analyzed. At this stage normalization and means of removal were performed to identify and justify the importance of each variable involved. Inputs and output are normalized to lie in between [-1 1]. The data is taken on 5-min basis for the duration of seven months. Data with timeout or behave irrelevant unlike the rest are filtered out. Data filtered is the data that is out of ranges or abnormal (timeout, zero reading). The total number of samples was 72,000 samples divided into seven months (January, February, March, April, May, June and July) with an average of 6,500 samples per month. As a typical approach the data will be divided into two sets: training and validation.

Neural network was modelled using MLP with back propagation structure. Besides being the most common approach in NN, this function is preferred because it uses the output of the first iteration to train the network along with the inputs. Neural network models were found useful especially when there is a complex relationship between the system inputs and outputs. Different error measurements have been used by different researchers. In this paper, Root Mean Square Error (RMSE) is used to evaluate the training functions. RMSE is given as following:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_{p_i} - y_i)^2} \qquad (1)$$

N represents the number of data while y_p represents the predicted output value and y is the current output for each

Babiker Idris Babiker with the Universiti Teknologi PETRONAS Bandar Seri Iskandar, 31720 Tronoh, Perak, Malaysia (phone: 60176063492, e-mail: areej5545@hotmail.com).

Rosdiazli Ibrahim is with the Universiti Teknologi PETRONAS Bandar Seri Iskandar, 31720 Tronoh, Perak, Malaysia (phone: 605-368-7821; fax: 605-365-7443; e-mail: rosdiazli@petronas.com.my).

training and validation set. The smaller the RMSE the better the prediction model is.

III. Results and Discussion

This section presents the results obtained and describes the methods used to obtain these results. The section is divided into two main subsections: Data Analysis and Neural Network.

A. Data Analysis

Model development was performed using five inputs: Hydrogen flow, Nitrogen flow, Propylene flow, Reactor Pressure and Reactor Temperature and one output: Propylene concentration. The data set experienced a number of processes such as normalization, correlation and means removal in order to have reliable and meaningful information to represent the right behaviour. The objective of data normalization is usually to allow meaningful comparisons of samples within the dataset [5]. All inputs are normalized to lie in between [-1, 1]. This is to overcome the circumstances that inputs and output variables span in different ranges. Correlation was performed in order to determine the input-output relationship. Input sensitivity is said to provide significant performance to the prediction result. The higher the correlation between the input and output will give higher sensitivity relationship between the two. A way to ascertain which factors are not most (and how well the network was trained) is also through the input sensitivity test as in [6]. Based on correlation coefficients the data in July it was found to have the strongest inputs-output relationship as indicated in Table 1. The weakest correlation is in January where most of the correlation coefficients values are less than 10%.

Table I
Data correlation coefficients

Month	H Flow	N Flow	Propylene Flow	Reactor Pressure	Reactor Temperature
Jan.	0.074	0.009	-0.077	0.119	0.017
Feb.	-0.08	-0.189	-0.152	-0.172	-0.155
Mar.	0.196	-0.419	-0.712	-0.899	-0.706
April	0.104	0.39	0.126	0.175	0.224
May	0.565	0.437	0.427	0.047	0.676
June	0.046	0.436	0.675	-0.565	0.427
July	-0.33	0.451	0.7206	0.715	-0.931

Thus, based on Table 1 the July data was chosen to develop the neural network model since it has the strongest inputs-output relationship. It has 8639 samples which have been divided into training set with 6000 samples and 2639 samples for model validation or verification activities. The training aims to minimize the error of the network output with regard to the input-output pattern of the training data set. As stated in the literature, during the training, the network recognizes the input-output pattern, computing the gradient and updating the network weights and bias [7]. At later stages, the rest of the month's data was used to test and check the reliability and robustness of the developed model. There are a few parameters investigated in developing the model and were considered in order to optimise the performance.

B. Neural Network

The type of problem amenable to solution by a neural networks is defined by the way the networks work and the way those networks are trained. Neural networks work by feeding in some input variables, and producing some output variables. Neural networks can therefore be used where one have some known information, and would like to infer some unknown information [8], [9].

Neural network modeling was used here to get the most accurate prediction result possible. Fig.1 below shows the structure of the network used whose inputs are reactor temperature, reactor pressure, N_2 flow, H_2 flow and Propylene flow and the output (target) is Propylene concentration. The network has three main layers, input layer with 5 neurons, output or target layer with one neuron and hidden layer with 7 neurons. This will be further discussed later.

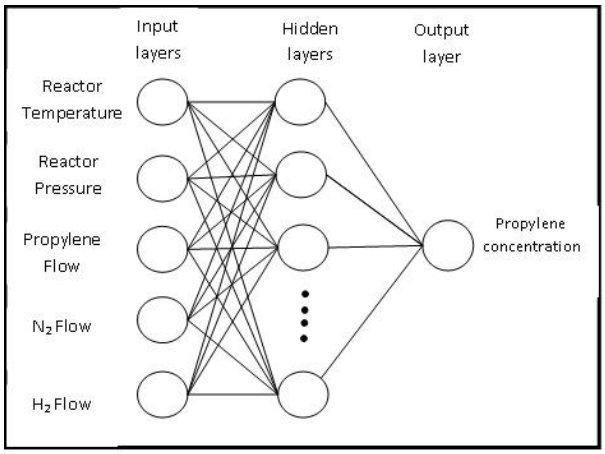

Fig.1 Neural Network model

In artificial neural network (ANN), the fundamentals of neural network concepts in developing a neural network model which will determine the reliability and robustness of the system is the most vital part. These parameter selections will actually affect the neural network presentation and improve the performance. The idea of this paper is to investigate and determine some parameters such as learning or training algorithm, the activation function, the training and the number of neurons.

Table 2 shows the number of learning algorithms used and their respective RMSE values. Based on the result obtained, Levenberg-Marquardt (LM) algorithm which has the least RMSE values of training and validation is chosen to train and validate the model.

TABLE II
TRAINING FUNCTIONS RMSE

Training Function/Algorithm	RMSE	
	Training	Validation
Levenberg-Marquardt	**3.46**	**1.778**
Batch Updates	6.14	13.65
BFGS Quasi-Newton	28.2	32.58
Bayersian Regularization	2.31	22.50
Conjugate Gradient Backpropagation	3.08	28.12
Conjugate Gradient Backpropagation with Polak-Ribiére	8.15	27.21
Gradient Descent	18.20	16.07
Gradient Descent and Adaptive Learning	29.78	25.67

To improve the output result further, the number of hidden layer neurons is to be decided. Usually, neural network will have 3 main layers. It can have more hidden layers depending on the complexity of the problem. The proposed network in this paper has three layers as shown earlier in Fig.1: input layer with five numbers of neurons representing the model inputs, output layer with one neuron representing the output and hidden layer with 7 neurons. The number of hidden layer neurons to be determined is based on trial and error. Table 3 below shows the different RMSE values for different number of neurons in the hidden layer.

TABLE III
SELECTION OF NEURONS NUMBER

Number of Hidden layer Neurons	2	3	6	**7**	11	15	20
Training	1.596	1.531	1.520	**1.460**	1.560	1.531	1.531
Validation	1.798	1.299	1.289	**0.768**	0.791	1.299	1.299

Table III also shows that as the number of neurons increases, the RMSE value for training and validation data set decreases. After it reaches the optimum number of neurons, the RMSE value starts to increase again. From the result it can be said that the hidden layer with 7 neurons is adequate for the model to predict the output clearly.

Moreover, one needs to choose the activation functions to calculate a layer's output from its net input. Log-sigmoid transfer function *Logsig* e.g. would give a range of 0 to 1 while hyperbolic tangent sigmoid transfer function *tansig* range is between -1 to 1. Tansig was chosen here so that output range would be within -1 to 1. This is besides its more adequate result. Fig. 2 and Fig. 3 show the performance of the trained network using July Month data.

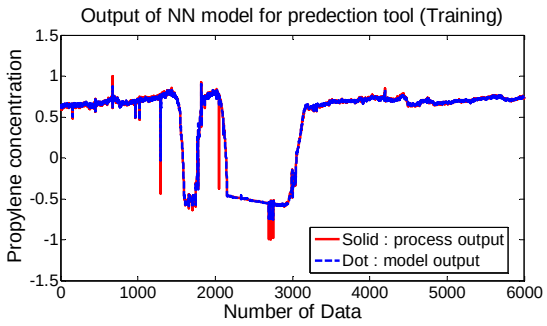

Fig. 2 Training data with RMSE = 1.46

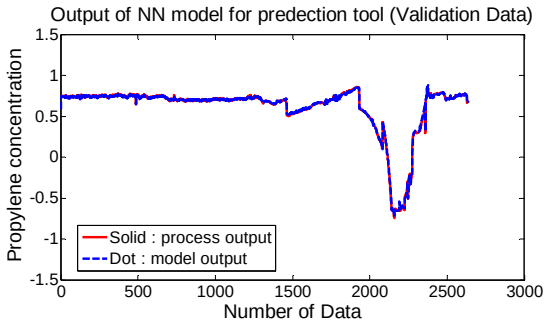

Fig. 3 Validation set with RMSE = 0.768

To ensure good performance of the developed model, January month's data with 6000 samples was used to test the model. As discussed previously, based on correlation coefficients January month data was found to have the weakest relationship between input and output. Thus, the result should point out whether or not the developed model is vigorous enough to handle such variation. Fig. 4 represents the network performance and Table 4 shows RMSE values of neural network model. As indicated the RMSE value for this exercise is approximately three times higher than the RMSE value for validation. However it is believe that the value was within an acceptable range considering that the relation between input and outputs was unconvincing. Fig. 4 also managed to highlight on how accurate the model predicts in terms of direction of the deviation between predicted and actual responses.

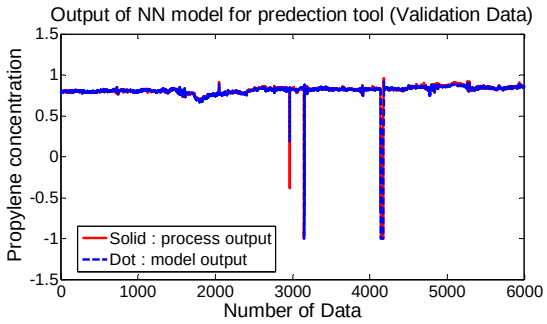

Fig. 4 Testing performance with RMSE = 2.478

TABLE IV
TRAINING FUNCTIONS RMSE

	RMSE		
	Training	***Validation***	***Testing***
Neural Network	1.46	0.768	2.478

In order to get an overall insight of the performance of the developed model, more simulation tests were performed utilizing the remaining data (February, March, April, May and June). Again the idea is to investigate the absolute fit of the model to observe the closeness of the actual data points to the model's predicted values.

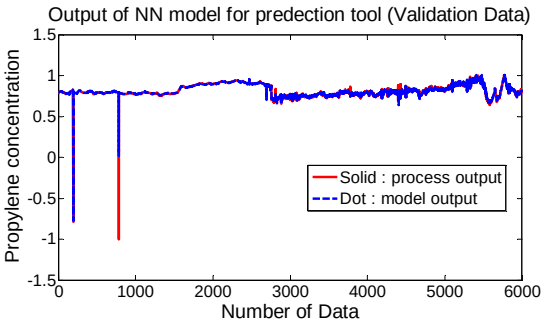

Fig. 5 February Month with RMSE = 1.253

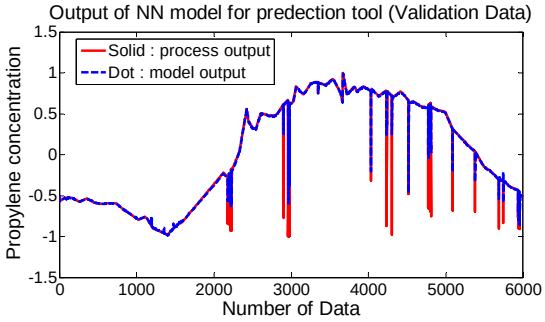

Fig. 6 March Month with RMSE = 0.0888

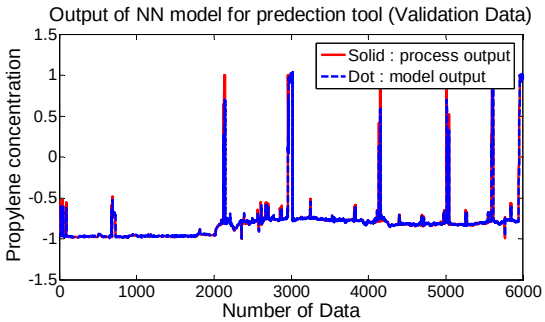

Fig. 7April Month with RMSE = 0.42096

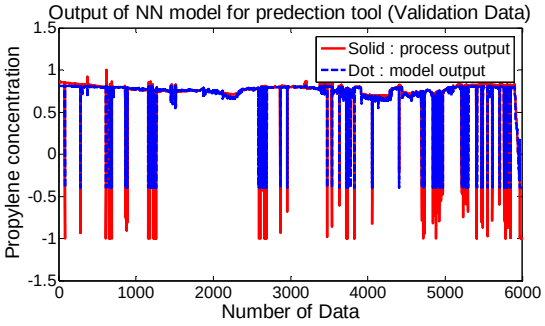

Fig. 8 May Month with RMSE = 2.831

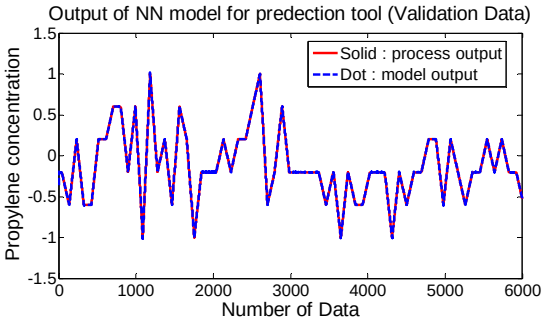

Fig. 9 June Month with RMSE = 0.0063

It can be noticed that RMSE of January month with the weakest correlation performs better than May month data, this is because of the nature of May data which has high variability compared to January data due to plant process errors. The model could maintain the accurate performance, though.

Table 5 summarizes the RMSE values of February, March, April, May and June months

TABLE V
TRAINING FUNCTIONS RMSE

Month	**RMSE Values**
Feb	1.253
March	0.0888
April	0.42096
May	2.831
June	0.0063

From the obtained results, it can be seen that neural network developed model can predict the output of GC with high accuracy depending on the inputs of the sample. Moreover, the high variation of the inputs did not significantly affect the output of the model as shown in Fig. 5, 6, 7, 8 and 9. The model could show the concentration of the Propylene with maximum RMSE of 2.8 which it is believe to fall within the acceptable range of such operations.

IV. Conclusion

Based on the study above, the possibility of using neural network for predicting the output of Gas Chromatography was explored. More than 42,000 samples of GC analyzer data with high performance variation were used to validate the neural network model. This software model depends on inputs data of process to predict accurately the respective output within acceptable range of RMSE value +/- 3 for similar process. It is concluded that the developed reliable neural network model can be used as an alternative whenever GC is unavailable. It can also be part of the system verification on the performance of the GC for preventive maintenance. This model can improve system's efficiency and lower down the overall process cost.

Acknowledgment

Authors would like to thank UTP for the supportand PETRONAS for the project collaboration.

References

[1] Elizabeth Prichard, co-ordinating author, *Practical Laboratory Skills Training Guides*, *Gas Chromatography*, 2003.
[2] Mo-Yuen Chowp. *Methodologies of using neural network and fuzzy logic Tecnologies for Motor incipient fault detection*, P.1997
[3] S. N. Sivanandam, Sumathi & Deepa, *Introduction to neural network using MATLAB 6.0*, The Mc-Graw Hill companies.
[4] H.R. Maier and G.C. Dandy, "Neural network for the prediction and forecasting of water resources variables: a review of modelling issues and applications," *Environmental Modelling and Software*, vol. 15, 2000, pp. 101-204.
[5] Brown, R.H.; Matin, I.; , "Development of artificial neural network models to predict daily gas consumption," *Industrial Electronics, Control, and Instrumentation, 1995., Proceedings of the 1995 IEEE IECON 21st International Conference on* , vol.2, no., pp.1389-1394 vol.2, 6-10 Nov 1995
[6] Dejan Ivezic, "sort-term natural gas consumption forecast," vol.34,2006,pp.165-169.
[7] Fausett, L., *Fundamentals of Neural Networks.* New York: Prentice Hall, P (1994).
[8] Patterson, D., *Artificial Neural Networks.* Singapore: Prentice Hall, P.(1996).
[9] Ismail, M.J., Ibrahim, R. and Ismail, I. , "Adaptive neural network prediction model for energy consumption," *Computer Research and Development (ICCRD), 2011 3rd International Conference on* , vol.4, no., pp.109-113, 11-13 March 2011

Areej Babiker Idris Babiker is currently pursuing B.Hons in Electrical & Electronics Engineering majoring in Instrumentation and Control, Universiti Teknologi PETRONAS, Perak, Malaysia.

She is working on a research project entitled Development of Gas Chromatography Model Using Neural Network. The research project is a collaboration initiative between Group Technology Solutions (GTS) of Research & Technology Division and Universiti Teknologi PETRONAS (UTP). She has undergone 8 months industrial internship at White Nile Petroleum Operating Company in Khartoum, Sudan as an Electrical and Instruments Trainee Engineer.

Rosdiazli Ibrahim received the BEng. Electrical Engineering. from Univeristi Putra Malaysia in 1996. In 2000, he awarded a Master of Science in Automation and Control from the University of Newcastle Upon-Tyne, UK with Distinction. He was conferred a PhD degree in Electrical & Electronic Eng. from the University of Glasgow, UK in 2008.

He joined Universiti Teknologi PETRONAS in 1999 and currently is a Senior Lecturer in the Electrical & Electronic Engineering department. Formerly, he was an Instrument Engineer at PETRONAS Fertilizer (Kedah) Sdn. Bhd. He has presented and published several articles in the area of intelligent control and nonlinear multivariable process modeling which is very much in line with his research interest.