

Comparison of Neural Networks Prediction and Regression Analysis (MLR and PCR) in Modeling Nonlinear System

Zainal Ahmad and Yong Fei Sean

School of Chemical Engineering Campus, Engineering Campus,
University Sains Malaysia, Seri Ampangan,
14300 Nibong Tebal, Seberang Perai Selatan,
Pulau Pinang, Malaysia
Tel: 604-5937788, Fax: 604-5941013
E-mail: chzahmad@eng.usm.my

Received Date: 20th July 2006 Accepted Date: 12th February 2008

ABSTRACT

Different methods for modeling nonlinear system are investigated in this paper. Neural network (NN) techniques, multiple linear regression (MLR) and principal component regression (PCR) are applied to two nonlinear systems which are sine function and distillation column. For the sake of studying these three distinctive methods, all the data is taken from simulation which is then be separated into training, testing and validation. Among those different approaches, the NN approach based on the nonlinear prediction technique gives a very good performance in for both case studies. It is also shown that MLR model suffers from glitches due to the collinearity of the input variables whereas PCR model shows good result in the prediction output. As a conclusion, the NN methods exhibit a consistent result with least sum square error (SSE) on the unseen data compared to the other two techniques.

Keywords: Artificial neural networks, multiple linear regression, principal component regression, principle component analysis, nonlinear process modeling

ABSTRAK

Pelbagai langkah pemodelan sistem tidak linear dibincangkan dalam artikel ini. Teknik rangkaian neural (NN), regresi linear pelbagai (MLR) dan regresi komponen prinsipal (PCR) diaplikasikan kepada dua sistem tidak linear iaitu fungsi sinus dan juga turus penyulingan. Untuk menjalankan kajian tentang tiga langkah yang berbeza ini, semua data dari simulasi yang kemudiannya dibahagikan kepada data latihan, ujian dan pengesahan. Antara tiga langkah ini, teknik rangkaian neural berdasarkan teknik jangkaan tidak linear memberikan keputusan yang paling baik bagi kedua-dua kajian kes. Kajian ini juga membuktikan bahawa model MLR memberikan keputusan yang kurang memuaskan disebabkan pembolehubah input yang bersifat linear sementara model PCR pula menunjukkan keputusan yang agak baik bagi jangkaan output.

Dapat disimpulkan bahawa teknik rangkaian neural menunjukkan keputusan yang konsisten dengan ralat jumlah kuasa dua terkecil (SSE) pada data tak terlihat berbanding dua teknik yang lainnya.

Kata kunci: Rangkaian neural buatan, regresi linear pelbagai, regresi komponen prinsipal, analisis komponen prinsipal, pemodelan proses tidak linear

INTRODUCTION

Chemical processes are highly nonlinear system exhibiting time dependent behavior. Modeling of these dynamics using first principle models is not always possible. As a result, alternative techniques either data-driven or knowledge-based have been devised (Cobourn et al. 2000). In this paper, modeling using conventional regression analysis or linear model and artificial neural networks (ANN) prediction or nonlinear model is introduced.

Artificial neural networks have been shown to be able to approximate any continuous nonlinear functions and have been used to build data based on empirical processes (Hertz et al. 1991; Haykin 1994; Turchetti 2004). Furthermore, the main advantage of neural network based process models is that they are easy to build. This feature is particularly useful when modeling complicated processes where detailed mechanistic models are difficult to develop Kumar (2004). However, a critical shortcoming of neural networks is that they often lack robustness unless a proper network training and validation technique is used. Robustness of the model can be defined as one of the baseline to judge the performance of the neural network models and it is strongly related to the learning or training classes.

Currently, applications of neural networks in process modeling and control are quite significant in industry especially in model based predictive control (MBPC) for example, by (Chen and Yea 2002; Xiong and Jutan 2002) and this is due to the ability of neural networks in modeling nonlinear processes for example, by (Shaw et al. 1997; Janczak 2005).

In process modeling, Aldrich and Slater (2001) modeled the fractional hold-up and drop size in the reactor using neural networks and it showed a very appealing result. Neural networks also have been used to predict the heat released by a chemical reactor as developed by Xiong and Jutan (2002) as well as Aziz et al. (2001). The prediction of heat released was compared to the actual heat released in the reactor and the single neural networks model performed

quite well and promoted to the significant performance on the model predictive control. Other research in chemical reaction in CSTR for instance was done by Shaw et al. (1997) where neural networks have been used to model the reactor temperature and the result was convincing.

Neural networks have also been used to model complex systems in bioprocess, for example by Lobanov et al. (2001) where neural networks were used as a biosensor to predict the glucose and ethanol in certain range of substrate and the accuracy of the estimation was good. Scheffer and Filho (2001) applied neural networks with the extended Kalman filter in the training to predict the production of the penicillin in a batch process. It was shown that the neural networks predictions were good even in the real data.

In the other hand, a conventional regression analysis which Multiple Linear Regression (MLR) and Principal Component Regression (PCR) are parts of statistical analyses that analyze the relationship between two or more independent quantitative variables. It is to predict the performance or response of a system when input variables vary. Principal component regression (PCR) and multiple linear regressions (MLR) are multivariate statistical techniques that have been applied to different sciences to obtain calibration models as an alternative to linear regressions. These statistical methods have provided good predictive models for the simultaneous analysis of multi-mixtures in pharmaceutical formulations (Dinc et al. 2002).

MLR is one of the most widely used statistical tools. It has been used in modeling and monitoring of the waste water treatment plant as shown by Teppola et al. (1999). MLR also being used in chemometry field as a multivariable calibrator for spectroscopic analysis Walmsley (1997) as well as in quality control (Andrade et al. 1997).

However, most often the data are plagued by quasi-multicollinearity (high intercorrelation among the predictors) which has a severe impact on the quality of the model of MLR. Several

alternative methods to MLR have been proposed so far in order to tackle this problem and among these methods is PCR. This involves transforming the original data matrix X into a matrix of new variables or 'principal components'. These principal components span the multidimensional space of X but are uncorrelated. The $m \times n$ data matrix X is transformed thus:

$$X = TP^T + E \quad (1)$$

where P are the 'loadings' describing the projections of a unit vector along the principal component, T are the 'scores' representing the co-ordinates of the data points on the respective principal component line, and E is the matrix of residual errors. PCR succeeds in solving the problem of collinearity by guaranteeing an invertible matrix for the calculation of the regression coefficients. In addition, the ability of PCR to be computed in to account for the greatest variability in the direction of loading, therefore PCR possesses the ability to omit the principal component that corresponding to the small eigenvalues or less variability. These will allow, only component that have a big eigenvalue is selected and the rest which have a high noise will be rejected.

Theoretically, first principal component retains the most variance based on the Eigen value and covariance calculation; if the feature vectors are projected onto the first principal component, more variance will be retained than if the vectors are projected onto any other principal component. The second component retains the next highest residual variance, and so on. A smaller eigen value contributes much less weight to the total variance, hence if the feature vectors are projected onto a subset of principal components, omission of later components tends to introduce less classification error than if earlier components are omitted. In many cases, the first few components can retain nearly all of the variance, enabling satisfactory classification.

However, since it is a two-step method, PCR runs the risk of discarding useful predictive information, and that some noise will remain in the components eventually used for regression. Moreover, the transformation of data matrix X is performed separately without any regard to the output variable(s) (Warnes et al. 1996). In order to achieve more accurate prediction of output

variables, it is often appropriate to perform the Partial Least-Squares (PLS) regression. But in this paper, only MLR and PCR are highlighted.

In recent literatures (Kourtis 1995; Broderick 1995), PCR techniques have been found to be efficient in process monitoring applications since the high dimensional strongly correlated data can be reduced to a much smaller, thus, the collinearity problems are solved. The effect of noise can also be often greatly reduced by PCR techniques due to the modeling of the systematic correlation structures among different variables. Also, many types of sample anomalies and process problems can be automatically detected as outliers. PCR has outperformed the MLR especially when the data or variable is highly correlated (Andrade et al. 1997). PCR also being used in modeling of the fermentation process where it modeled the biomass concentration and the result showed that PCR performed quite well compared to MLR and slightly below the nonlinear model.

There are a wide number of techniques available to formulate such models, each varying in complexity and ease of development. The objective of this project is to observe the performance of nonlinear process modeling through three different approaches. The models will be applied to two nonlinear case studies and their suitability assessed from their performances on a common validation data set. Ideally, the best model will combine enough complexity to capture the process characteristics with sufficient simplicity to allow the model to be easily initiated and understood.

The rest of the paper is organized as follows. Modeling techniques are covered in Section 2 followed by case studies in Section 3. Section 4 covered the results and discussion and lastly, the conclusion is in Section 5.

MODELING TECHNIQUES

This paper compares three modeling techniques with fixed individual network structures for neural networks; the rest is regression techniques which is multiple linear regression (MLR) and principal component regression (PCR). The objective of carrying this comparison is to alleviate the development of the chemical processes modeling technique that are normally difficult to model. Many processes are so complex that significant engineering time

and effort is required to develop and validate detailed theoretical dynamic models which often ends up with a multi variable complicated mathematical model. As an alternative, simpler empirical nonlinear model is often developed using experimental data which is normally used linear model even though the system is nonlinear. One of the area that benefit from this are in modeling highly nonlinear process and implementing nonlinear model control (NMPC). NMPC basically requires an accurate model of the nonlinear process to get better predictions of the controlled variable over a prediction horizon in order to get good control performance.

Therefore it is essential to compare the performance of linear and nonlinear modeling technique to support that statement as well as improving the model based control approach. Furthermore, neural networks or artificial neural networks (ANN) are attractive whenever it is necessary to model complex or little understood processes with large input-output data sets, as well as to replace models that are too complicated to solve in real time.

Neural Networks

In this paper, feedforward neural network (FANN) structure is chosen as shown in Figure 1. The neural networks were trained by the Levenberg-Marquardt optimisation algorithm with regularisation and "early stopping". Regularisation is achieved by modifying the networks training objective to include a term to penalise unnecessarily large network weights as follows:

$$J = \frac{1}{N} \sum_{i=1}^N (\hat{y}(i) - y(i))^2 + \rho \|W\|^2 \quad (2)$$

where N is the number of data points, \hat{y} is the networks prediction, y is the target value, W is a vector of network weights and ρ is the regularisation parameter, which was set to 0.001 in this study. The individual networks are single hidden layer networks where the hidden layer neurons use the sigmoidal activation function while the output layer neurons use a linear activation function.

All weights and biases were selected randomly in the range of -0.1 to 0.1. Sum of squared errors (SSE) on the unseen validation data is the performance criterion for all the comparisons. To

cope with different magnitudes in the input and output data, all the measurements or the analysis data were scaled to zero mean and unit standard deviation.

Multiple Linear Regressions (MLR)

MLR is also known as Ordinary Least Square (OLS) and it show that vector of regression parameters, θ can be used to obtain a linear input-output model for a given data set through a correlation of the following form:

$$y = x_1\theta_1 + x_2\theta_2 + \dots + x_n\theta_n + e \quad (3)$$

By using MLR, the model parameter, θ can be estimated through the pseudo inverse with the following equation

$$\hat{\theta} = (X^T X)^{-1} X^T Y \quad (4)$$

where X is the matrix of input variable data and Y is a matrix (or a vector in a single output case) output variable data. In this chapter, for MLR analysis X will be the prediction output of the individual networks on the training and testing data and Y is the actual process output on the training and testing data. Once the model parameter is estimated, it can be applied to the unseen validation data as shown below:

$$\hat{Y} = \hat{y}_1\hat{\theta}_1 + \hat{y}_2\hat{\theta}_2 + \dots + \hat{y}_n\hat{\theta}_n \quad (5)$$

where \hat{y} is the final prediction on unseen validation data while \hat{y} is the prediction on unseen validation data and n is the number of input.

Principal Component Regressions (PCR)

PCR is one way to deal with the problem of ill-conditioned matrices or when the predictor variables are quite highly correlated. Instead of applying regression process to the system properties on the original measured variables, the properties are regressed on the principle component scores of the measured variables (which are orthogonal and therefore well conditioned). The data matrix, $\hat{Y}_{(tr)}$ (which is the input based on the training data prediction output), will be decomposed into principal components as follows:

$$\hat{Y}_{(tr)} = t_1 p_1^T + t_2 p_2^T + \dots + t_m p_m^T = TP^T \quad (6)$$

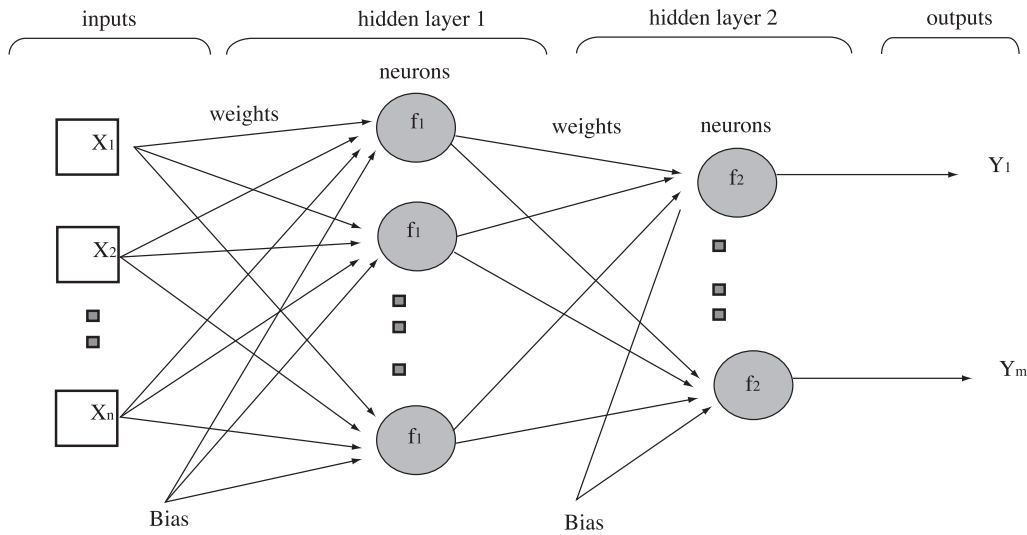


FIGURE 1. Typical feedforward artificial neural networks (FANN).

where t is the score vector (principal component) and p is the loading vector. The principal components are arranged in descending order of the data variations that they explain. PCR is a favourite tool for data compression and information extraction. It finds combination of variables or factors that describe major trends in a data set. The predictor matrix, $\hat{Y}_{(tr)}$ is replaced by its scores matrix, T which contains the *major* score vectors.

$$T = \hat{Y}_{(tr)} * P \tag{7}$$

Therefore,

$$Y_{(tr)} = TB + E \tag{8}$$

where B is the vector for model parameter. The least square estimation of B is

$$B = (T^T T)^{-1} T^T Y_{(tr)} \tag{9}$$

Therefore, the model parameter vector B can be changed into the model parameter vector, θ as follow:

$$\hat{\theta} = PB = P(T^T T)^{-1} T^T Y_{(tr)} \tag{10}$$

In this study principal components (PC) are calculated using singular value decomposition (SVD) even though there are several methods

that can be used to compute the PC. The main reason to using SVD is that it is easy to compute using Matlab™ and also easy to use. The SVD is presented below as Matlab™ command:

$$\begin{aligned} [U, s, V] &= \text{svd}(\hat{Y}_{(tr)}) \\ P &= V \\ T &= \hat{Y}_{(tr)} * V \end{aligned} \tag{11}$$

where $\hat{Y}_{(tr)}$ is the output prediction based on the training data, T is the principal component and P is the loading matrix.

Selections of the number of principal components used are determined through the cross validation technique. The PCR models are develop using the training data and tested on the testing data and the model giving the least model error on the testing data is selected. Evaluation of the PCR model will be based on the performance on the unseen validation data (Zhang, 1999a).

APPLICATION TO NONLINEAR PROCESS MODELING

Two nonlinear case studies were used to compare the performance of three different methods for modeling nonlinear processes. The performance criteria are based on the sum square error (SSE) from the validation data.

Case Study I: Sine Function

The static equation between input and output for sine wave function is

$$Y = \sin X \quad (12)$$

Testing and training data has to be generated randomly and with normally distributed within $(0; 2\pi)$. All the input and output data for training and testing has to be interrupted with noise where it is normally distributed as $N(0; 0.05)$. The output of validation data is clean but the input data will be disturbed with noise. This model will exhibit the following nonlinear trend as shown in Figure 2.

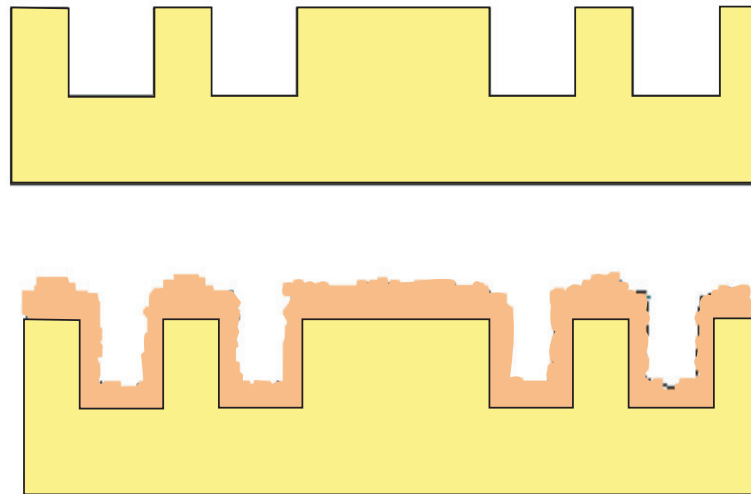


FIGURE 2. Sine wave function

The process of scaling to zero mean and standard deviation was not applied to this data due to the input, x and output; y which did not have any units. Thus, there would not be any unnecessary deviations presented as a result of different unit of input and output. The data were also divided into three set which is training data, testing data and validation data.

Case Study 2: Nonlinear System of Distillation Column

In this case study, the binary distillation column tray temperature was taken to predict the top and bottom composition as shown in Figure 3. There were 10 trays representing temperature measurements. In this case, the model for the distillation column between the trays temperature and the output composition for the bottom and top was developed. The proposed transfer function for this process model is :

$$yp_{(top)} = fn[x_1, x_2, x_3 \dots x_{10}] \quad (13)$$

$$yp_{(bottom)} = fn[x_1, x_2, x_3 \dots x_{10}] \quad (14)$$

where yp is the model prediction for top and bottom composition and x is the tray temperature from 1 to 10.

The data was simulated and contains 200 samples set of steady state measurement from a distillation column. Then, the data were divided into three sets of data which were training data, testing data and validation data. The product

and bottom composition are measured in percentage, %, and the top and bottom temperature are measured in degree Celsius, °C. Since the product compositions are difficult to measure, it would be desirable to estimate them from tray temperature measurement.

RESULTS AND DISCUSSION

Case Study I: Sine Function

The first modeling is coming from the MLR. The basic equation for MLR is based on the Eqn. (3) and Eqn. (4). In Eqn. (3), θ is calculated from the Eqn. (4) which is from the input data X . In this case, X is coming from the training data set. This is due to the MLR itself which is not required the testing data. Then, the theta calculated from Eqn. (3) is used in the validation data to get a prediction. The predicted and actual validation data based on the MLR analysis are shown in Figure 4.

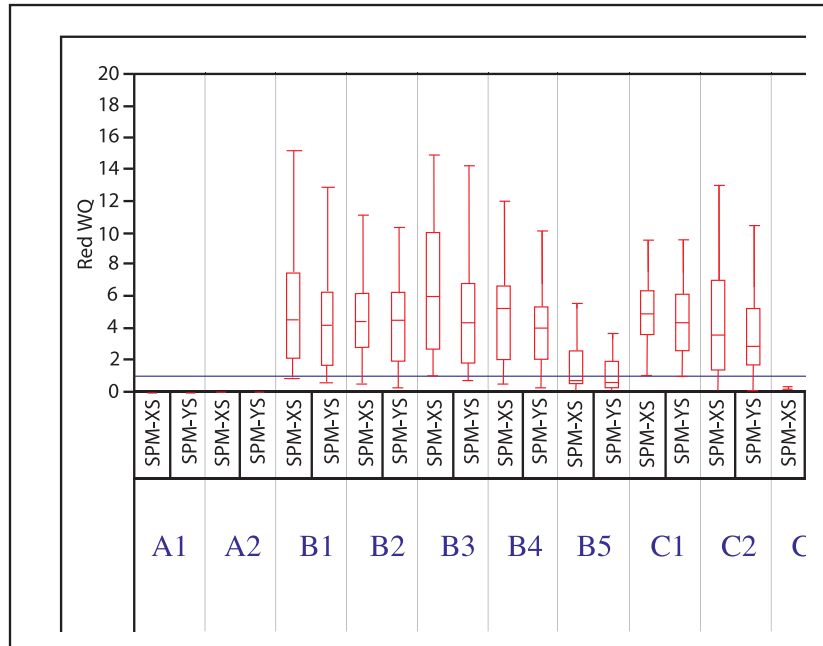


FIGURE 3. Distillation column for nonlinear system

The covariances for the training, testing and validation data are 3.3892 and 0.5121, 3.5407 and 0.4882, 3.3726 and 0.4983 for input and output respectively. The correlation coefficient for training, testing and validation data are - 0.7670, - 0.7486 and - 0.7462. \

It is shown that MLR performed quite badly to model the sine wave function. The SSE in the validation data were 7.97 and 7.99 by using theta calculated from training and testing data respectively. Therefore, the 'best' model in this case study is coming from the theta calculated

from the training data which 7.97 with the theta value of 1.0861 and - 0.3428. The final MLR model is:

$$Y = 1.0861 - 0.3428 X \tag{15}$$

where Y is the output and X is the input.

For the PCR analysis, the early stopping is based on the PCA which contributes the smallest SSE of both training and testing set. From Figure 5, second PCA is chosen as the SSE is the smallest.

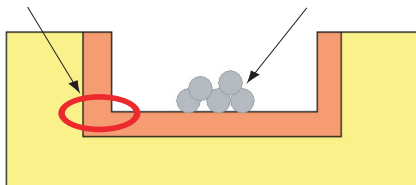


FIGURE 4. The relation of input and output (based on actual and predicted output) for validation set in MLR

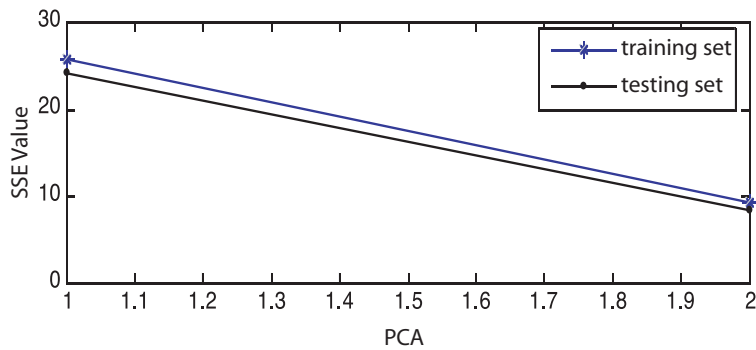


FIGURE 5. Sum Square Error (SSE) for PCR Analysis

Based on the SVD techniques, second PCA is chosen as the final prediction as shown earlier in Eqn. (11) and the final θ value is chosen as shown in Eqn. (10). The SSE for the validation data for this analysis is 7.81 which are slightly smaller than the MLR. The actual and predicted value of the output is shown in Figure 6.

Then, neural networks modeling technique is applied. First, the number of hidden neuron needs to be defined. Based on the literature, there is no specific method to determine the number of hidden nodes. Therefore in this case study, cross validation techniques is implemented. In order to determine the desired amount of hidden nodes in single hidden layer, a few samples of various

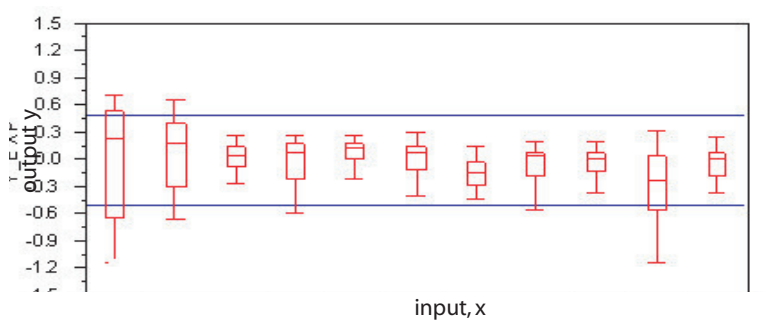


FIGURE 6. The relation of input and output (based on actual and predicted output) for validation set in PCR

From Figure 6, the predicted output is deviated quite far away the actual one. Consequently, the predicted output does not conform to the linearity of the actual output. Thus, model generated from this PCR is

$$Y = 1.1875 - 0.3799 * X \tag{16}$$

where Y is the output and X is the input.

The prediction output for MLR and PCR is very poor and it is due to the input and output correlation coefficient that are really small and the covariance matrix that quite big. Another reason for this is the PCR, it has one eigen-value and covariance plus the bias, therefore there is no omission of the principal component from the model input hence increased the classification error of the output.

hidden nodes have been done. SSE of training and testing data has taken into account while determines the number of hidden nodes in the models. The smallest SSE for both sets of data has been found in the sample using 5 hidden neurons, 0.0214 and 0.0368, respectively. Thus, 5 hidden nodes are the optimum values in this single hidden layer.

Then, after training and testing the neural networks model, the final weight and bias are applied to validation data. The predicted and actual data are shown in Figure 7. It is shown that the model is well suited the actual data where the SSE is very small which is 0.03.

In overall, neural networks model are far superior compare to conventional regression analysis in this case study as shown in Table 1.

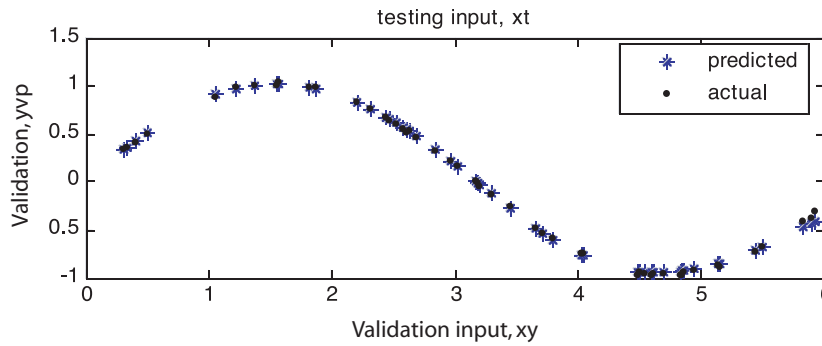


FIGURE 7. The relation of input and predicted output based on validation set for NN

TABLE 1. Overall result for case study 1

Modeling technique	SSE (Validation)
MLR	7.96
PCR	7.84
NN	0.03

Case Study 2: Nonlinear System of Distillation Column

In this case study the basic equation for MLR and PCR is quite similar where only θ value is different based on Eqn. (4) and Eqn. (10) respectively. In this case study, there are ten inputs ($x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9$ and x_{10}) and two outputs (y_t and y_b). The ten inputs are the ten tray's temperatures which x_1 represents the first tray's temperature. The unit is in °C. The two outputs are the top and bottom products' compositions, where y_t represents top and y_b represents bottom product's composition respectively. In this case study, there is no exact equation to represent this nonlinear system. However, it can be represented as Eqn. (13) and Eqn. (14) where θ is determined through MLR and PCR.

Figure 8 and Figure 9 show the predicted output in validation data using MLR analysis for top and bottom products, respectively. It is shown that the MLR predicted gives quite a bad performance at the beginning of the prediction where the predicted output deviates quite significantly compare to the actual value. However, this condition improved after some period of time.

The model generated from this MLR is shown below with an SSE of 53.46 and 53.25 for top and bottom product respectively.

a) Composition of top product (in %)

$$y_t = -0.0013 + 0.699 x_1 - 0.0128 x_2 + 0.0527 x_3 + 0.0364 x_4 + 0.1492 x_5 - 0.1340 x_6 - 0.2887 x_7 + 0.141 x_8 - 0.0332 x_9 - 0.0212 x_{10} \tag{17}$$

b) Composition of bottom product (in %)

$$y_b = 0.0116 - 0.8052 x_1 + 0.2611 x_2 - 0.041 x_3 - 0.0218 x_4 - 0.1627 x_5 - 0.0652 x_6 - 0.1679 x_7 + 0.0192 x_8 + 0.0184 x_9 - 0.0223 x_{10} \tag{18}$$

In PCR, again for early stopping method, the PCA is chosen which contribute the smallest SSE of both training and testing set. From Figure 10, the SSE for both training and testing data are not consistent, thus, the summation of both data and the smallest of the summation SSE should be taken. This falls into the tenth PCA. Therefore, the final θ is based on the tenth PCA and Figure 11 and Figure 12 show the predicted and actual output from the validation data. It is shown that PCR performed quite well which produce an SSE of 43.27 and 40.45 for top and bottom products, respectively.

Comparing to the MLR, PCR is better with the reduced SSE in validation data. The final model generated for PCVR is shown below:

a) Composition of top product (in %)

$$y_t = 0.0373 + 0.007 x_1 + 0.23 x_2 + 0.2983 x_3 - 0.2239 x_4 - 0.1915 x_5 + 0.0988 x_6 - 0.6667 x_7 + 0.1310 x_8 - 0.2489 x_9 - 0.1265 x_{10} \tag{19}$$

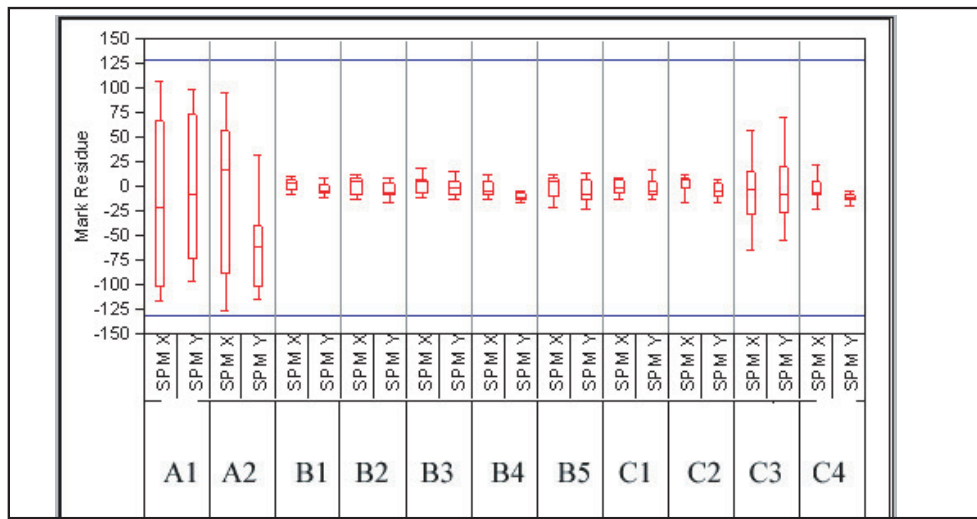


FIGURE 8. The actual and predicted top product's composition for validation set for MLR

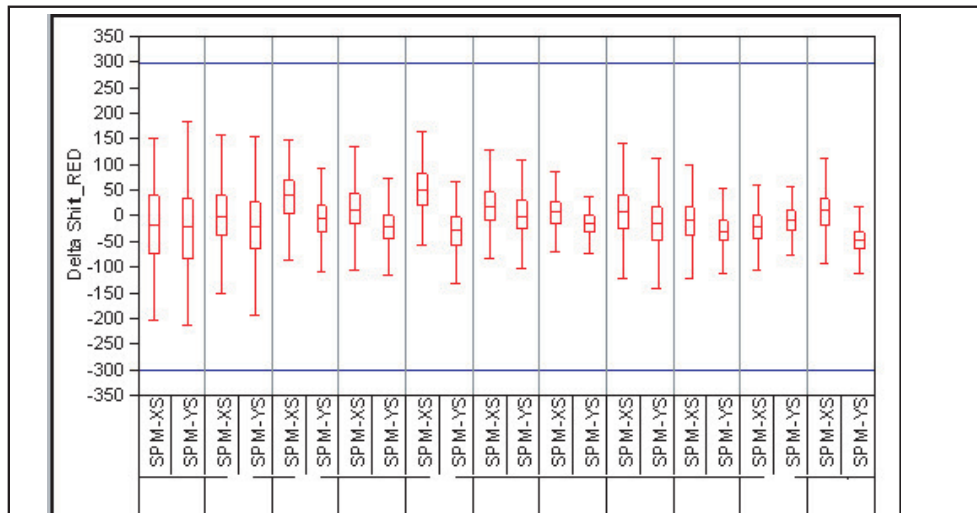


FIGURE 9. The actual and predicted bottom product's composition for validation set for MLR

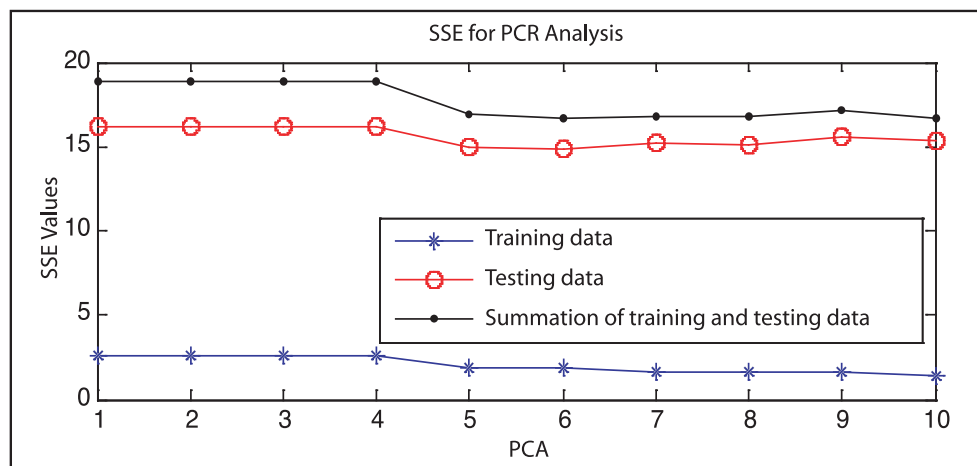


FIGURE 10. Sum Squared Error for PCA Analysis

b) Composition of bottom product (in %)

$$\begin{aligned}
 y_b = & 0.0254 - 0.0649 x_1 + 0.091 x_2 \\
 & + 0.1284 x_3 + 0.0956 x_4 - 0.0976 x_5 \\
 & - 0.2431 x_6 - 0.623 x_7 + 0.066 x_8 \\
 & - 0.1459 x_9 - 0.1095 x_{10}
 \end{aligned}
 \tag{20}$$

In neural networks modeling, the numbers of hidden nodes were determined using cross validation techniques for this case study which is 8 hidden nodes. Therefore, both neural network models for top and bottom product composition used the same number of hidden nodes. Subsequently, SSE for the validation data is 45.4397 and 46.3152 for top and bottom product composition respectively. Figure 13 and Figure 14 showed the predicted and actual output using neural network models. It is shown that the predicted composition for both top and bottom product have further deviation away from actual compare to MLR and PCR.

TABLE 2. Overall result for case study 2

SSE Validation		
Modeling techniques	Top	Bottom
MLR	53.46	53.25
PCR	43.27	40.45
NN	45.44	46.32

It is shown in Table 2 that MLR performed quite badly in this case study. It might be due to the high correlation of the input that causes the MLR performance deteriorates. PCR is the 'best' method in this case study while neural networks model is second. The PCR and NN model performance is quite the same but as expected that NN is better than PCR as shown in the first case study.

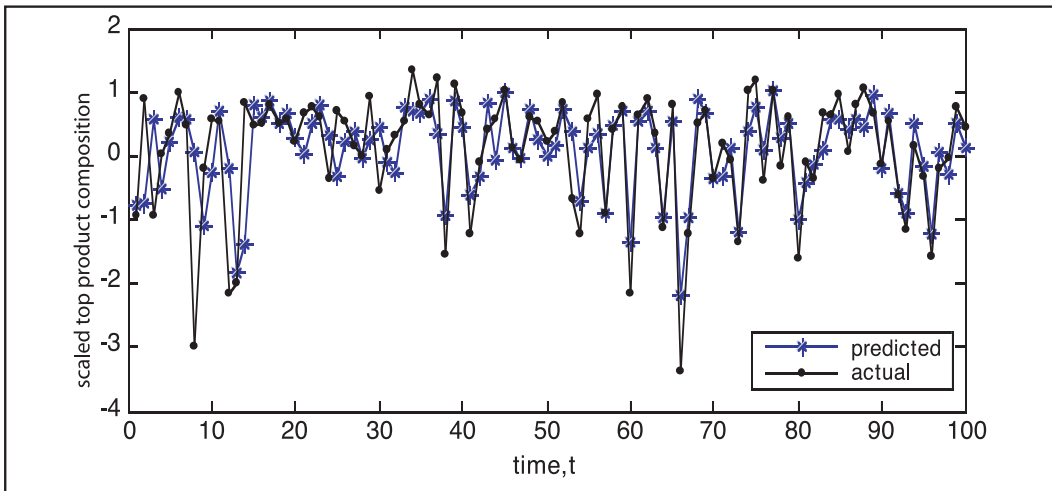


FIGURE 11. The actual and predicted top product's composition for validation set for PCR

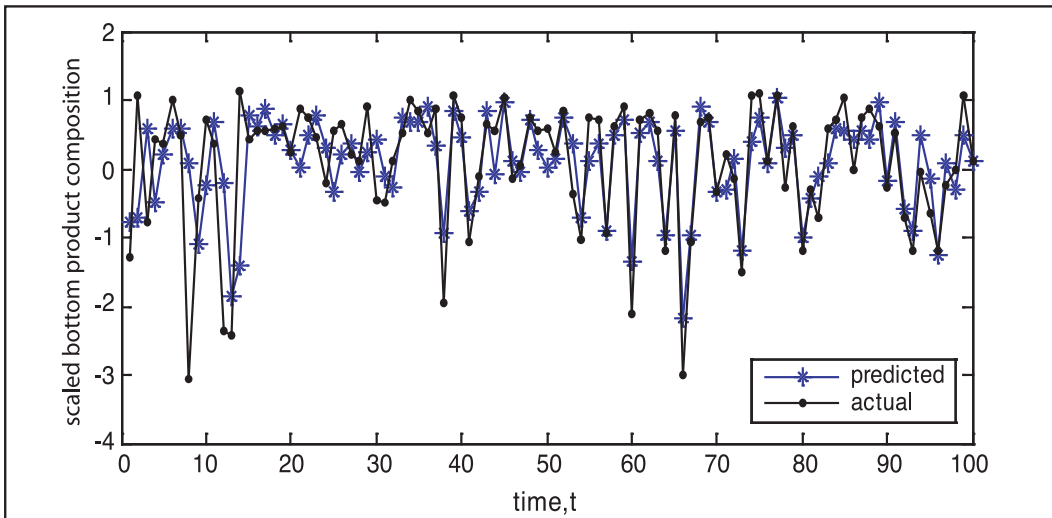


FIGURE 12. The actual and predicted bottom product's composition for validation set for PCR

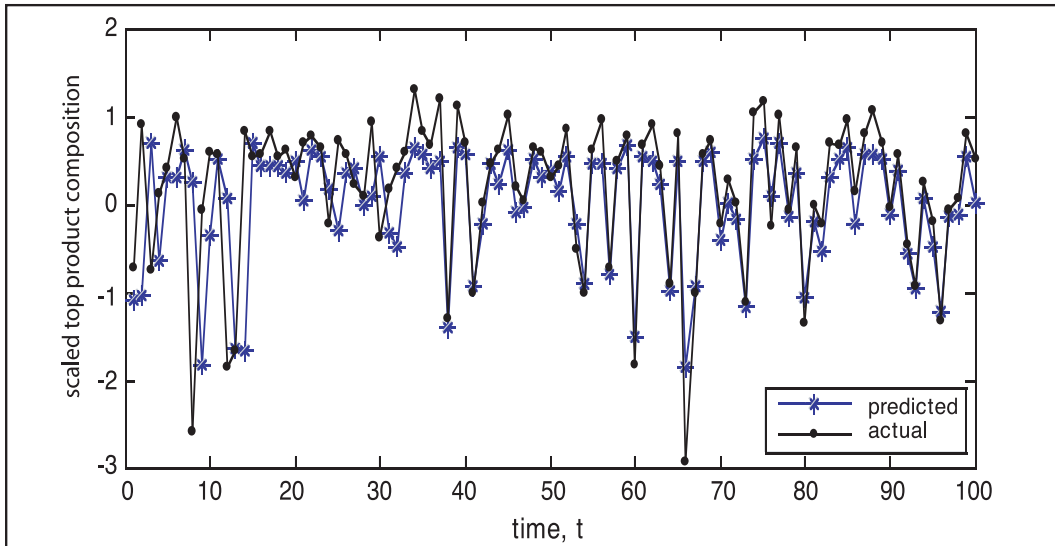


FIGURE 13. The actual and predicted top product's composition for validation set for NN

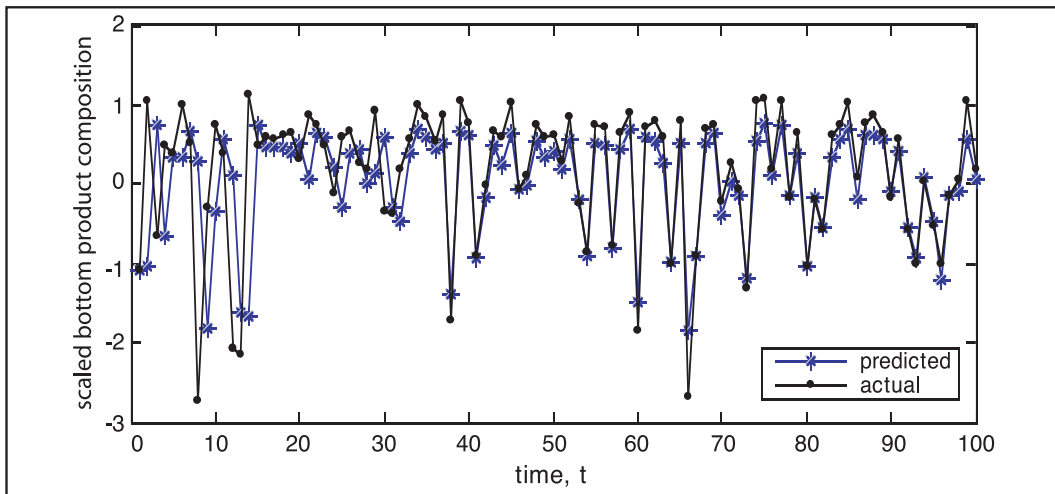


FIGURE 14. The actual and predicted bottom product's composition for validation set for NN

CONCLUSION

This study has investigated three different modeling techniques for the modeling of two nonlinear process system. These three techniques have varied in their success. The linear models which are conventional linear regression MLR and PCR have not been able to sufficiently capture the inherent non-linear process characteristics especially in case study 1. Such defect can be proved by the larger prediction errors and unsatisfactory model fits obtained. MLR and PCR did not perform adequately consistent to model the nonlinear process especially MLR which suffer from the collinearity of the variables input. The non-linear modeling techniques which is a neural networks modeling technique have been much more successful where the performances for the two case studies were quite consistent with the smallest SSE in validation data.

NN require much less specification than the regression method and achieve similar and often better performance. Although training the NN can sometimes take a substantial length of time, rich mappings can be obtained which perform well when presented with new unseen data. The ability of NN to learn process characteristics with little prior knowledge is a desirable trait that eases their implementation and heightens their modeling potential. Combining some user-knowledge of the process with this learning ability makes NN powerful and flexible tools that are well-suited to model a nonlinear process.

ACKNOWLEDGEMENTS

We wish to thank the Universiti Sains Malaysia (USM) for providing support for our research through short term grant 6035182.

REFERENCES

- Aldrich, A. & Slater, M.J. 2001. Simulation of Liquid-Liquid Extraction Data with Artificial Neural Networks. I. M. Mujtaba and M. A. Hussain (Ed). In *Application of Neural Networks and Other Learning Technologies in Process Engineering*. London, Imperial College Press. pp: 3-22
- Andrade, J. & M, Muniategui, S., Lopez-Mahia, P and Prada, D. 1997. Use of Multivariate Techniques in Quality Control of Kerosene Production. *Fuel*. 76 (1): 51-59.
- Aziz, N., Mujtaba, I.M & Hussain, M.A. 2001. Set Point Tracking in Batch Reactors: Use of PID and Generic Model Control with Neural Networks. I.M. Mujtaba and M. A. Hussain (Ed). In *Application of Neural Networks and Other Learning Technologies in Process Engineering*. London, Imperial College Press. pp: 217-241
- Broderick, G., Paris, J & Valade, J.L. 1995. A Composite Representation of Pulp Quality. *Chemom. Intell. Lab. Syst.* 29: 19-28.
- Chen, J. & Yea, Y. 2002. Neural Network-Based Predictive Control for Multivariable Processes. *Chemical Engineering and Communication*. 189: 866-894.
- Cobourn, W. G., Dolcine, L., French, M & Hubbard, M.C. 2000. A Comparison of Nonlinear Regression and Neural Network Models for Ground-level Ozone Forecasting. *Journal of the Air and Waste Management Association*. pp: 1999 - 2009.
- Dinc, E., Yucesoy, C & Onur, F. 2002. Simultaneous Spectrophotometric Determination of Mefenamic Acid and Paracetamol in a Pharmaceutical Preparation using Ratio Spectra Derivative Spectrophotometry and Chemometric Methods. *Journal of Pharmaceutical and Biomedical Analysis*. 28: 1091-1100.
- Haykin S. 1994. *Neural Networks*, Macmillan College Publication Company.
- Hertz, J., Krogh, A & Palmer, R.G. 1991. *Introduction to the Theory of Neural Computation*. Redwood City, CA, Addison-Wesley.
- Janczak, A. 2005. *Identification of Nonlinear Systems using Neural Networks and Polynomial Models : A Block-oriented Approach*, Springer, New York.
- Kourti, T & MacGregor, J.F. 1995. Process Analysis, Monitoring and Diagnosis using Multivariate Projection Methods. *Chemom. Intell. Lab. Syst.* 28: 3-21.
- Kumar, S. 2004. *Neural Networks: A Classroom Approach*. McGraw-Hill, New Delhi.
- Lobanov, A. V., Borisov, I.A, Gordon, S.H, Greene, R.V., Leathers, T.D & Reshetilov, A.N. 2001. Analysis of Ethanol-Glucose Mixtures by Two Microbial Sensors: Applications of Chemometrics and Artificial Neural networks for data Processing. *Biosensor and Bioelectronics*. 16: 1001-1007.
- Scheffer, R. & Filho, M. 2001. Process Identification of a Fed-batch Penicilin Production Process-Training with the Extended Kalman Filter. I. M. Mujtaba and M. A. Hussain (Ed). In *Application of Neural Networks and Other Learning Technologies in Process Engineering*. London, Imperial College Press. pp: 99-118

- Shaw, A.M., Doyle III, F.J. & Schwaber, J.S. 1997. A Dynamic Neural Networks Approach to Nonlinear Process Modelling. *Computer and Chemical Engineering*. 21 (4): 371-385.
- Teppola, P., Mujunen, S & Minkkinen, P. 1999. Kalman Filter for Updating the Coefficients of Regression Models. A Case Study from an Activated Sludge Waste-water Treatment Plant. *Chemometrics and Intelligent Laboratory Systems*. 45: 371-384.
- Turchetti, C. 2004. Stochastic Models of Neural Networks. IOS Press, Amsterdam,
- Walmsley, A. D. 1997. Improved Variable Selection Procedure for Multivariate Linear Regression. *Analytica Chimica Acta*. 354: 225-232.
- Warnes, M. R., Glassey, J., Montague, G.A & Kara, B. 1996. On Data-Based Modelling Techniques for Fermentation Processes. *Process Biochemistry*. 31 (2): 147-155.
- Xiong, Q. & Jutan, A. 2002. Grey-Modelling and Control of Chemical Processes. *Chemical Engineering Science*. 57: 1027-1039.