Journal of Applied Sciences Research, 8(3): 1672-1681, 2012
ISSN 1819-544X
This is a refereed journal and all articles are professionally screened and reviewed

ORIGINAL ARTICLES

Case Study of Artificial Neural Network Modeling on Catalyzed and Enzymatic Transesterification Process for Biodiesel Production

Liew Weng Hui

Department of Chemical & Process Engineering, Faculty of Engineering & Built Environment, The National University of Malaysia, 43600, Bangi, Selangor, Malaysia.

ABSTRACT

Biodiesel is reported as an important substitute for conventional fossil fuel to support the world demand. Both catalyzed and enzymatic transesterification process is known as the most widely used and potential process for biodiesel production. To enable the industrial process to be more efficient, understanding on the process characteristic of biodiesel production through modeling is crucial for process optimization, control and integration. In this study, artificial neural network (ANN) as advanced modeling tool is used to predict reaction data for catalyzed and enzymatic transesterification process. The ANN appears as feed-forward type with single hidden layer and coupled with Levenberg-Marquardt (LM) training algorithm. To illustrate the capability of ANN, two case studies with discrete and continuous data generated from literature models are presented. Unlike the available toolbox, the ANN program is self-developed where the training protocol and training parameters are modified. After optimization, the trained ANN wiith low number of hidden neurons has shown excellent prediction performance and robustness in modeling for both cases. The ANN performance is judged by mean squared error, correlation coefficient and number of epochs. The optimized ANN has manifested great potential for real application of biodiesel production process modeling.

Key words: mean squared error, correlation coefficient, optimization, Levenberg-Marquardt algorithm

Introduction

The world is facing the mutual issue of fossil fuel depletion, climate change and pollution problem. The searching of renewable and sustainable energy source is now critical to resolve the environmental problem and to improve the energy security. The World Energy Forum reveals that the fossil-based oil, coal and gas reserves are predicted to be exhausted in less than another ten decades (Sharma and Singh, 2008). However the combustion of fossil fuel has been elevated due to industrial activities, technology, transportation, energy generation and business development and improvement of living (Juan *et al*., 2010; Palligarnai and Boyi, 2010). As result, the heat-trapping greenhouse gases and other hazardous gases have significantly increased at atmosphere over the past decades. With all those described critical issues, the biodiesel is reported to be a promising source due to its nature as a renewable, energy-efficient and highly biodegradable fuel. In addition, it is environmental friendlier because it emits better quality of exhaust gases which contains less CO_2 and sulfur oxide (SOx), unburnt hydrocarbon, particulate matters and aromatics (Hideki *et al*., 2001; Palligarnai and Boyi, 2010). Transesterification, out of other primary biodiesel production methods like direct use and blending, micro-emulsions and thermal cracking (pyrolysis), is the most commonly adopted method of converting oils to biodiesel (Dennis *et al*., 2009). Transesterification takes place when triglycerides react chemically with acyl-acceptors such as straight chain aliphatic alcohols (e.g. methanol or ethanol) or acetate to form glycerol (trihydric alcohol) as by-products and esters of fatty acids, namely biodiesel, in the presence of catalyst or enzyme (Marchetti *et al*., 2007; Siddharth and Sharma, 2010). Chemically, one mol of triglyceride reacts with three mols of alcohols to produce three mols of esters as biodiesel and one mol of glycerol as by-products.

Among the transesterification reactions, catalyzed-transesterification which uses alkaline catalyst (such as sodium hydroxide, potassium hydroxide, sodium methoxide and potassium methoxide) is proven to generate high yield of biodiesel (>0.8 g/g oil) and by-product glycerol in relatively short reaction time (20 to 60 mins) and high concentration (nearly 100%) (Yee and Bhatia, 2010). However, transesterification process does pose the drawbacks such as high energy requirement, and difficulty in glycerol and catalyst recovery (Srivathsan *et al*., 2008). At current, the biodiesel production cost is still higher than petro-diesel due to higher cost of oil sources which take about 60 to 80% especially when facing competition on edible oils (Juan *et al*., 2010). As a result, the people may consider biodiesel as less financially profitable and less attractive for vast commercialization. Albeit its advantages of high yield and short reaction time, chemical-catalyzed

Corresponding Author: Liew Weng Hui, Department of Chemical & Process Engineering, Faculty of Engineering & Built Environment, The National University of Malaysia, 43600, Bangi, Selangor, Malaysia.
E-mail: wenghui25@hotmail.com, Tel: +6012-9113629

transesterification still portrays several issues such as pretreatment of feedstock (those with high content of FFA and water), recovery of glycerol, removal of catalyst and energy intensive nature such as high stirring speed and high temperature (Charpe, T.W., Rathod, V.K., 2010). Later, enzymatic transesterification with enzyme as biocatalyst emerges as another potential reaction without using base and acid catalysts. In usual, microbial lipase (glycerol ester hydrolase EC 3.1.1.3) which comes from variety of sources are used as the enzyme to accelerate the chemical reactions of biodiesel.

For transesterification, the application of lipase enzyme portrays several important advantages which give great impacts to biodiesel development (Palligarnai and Boyi, 2010). This includes (1) capability in converting free fatty acids to biodiesel by esterification reaction, as well as glyceride stock by transesterification simultaneously at moderately low temperature (40 to 70°C). Enzymatic transesterification provides alternative to handle high free fatty acids feedstock and promotes low energy consumption due to low reaction temperature, (2) no soap formation during reaction which eliminates the cost and effort in product purification, (3) characteristic of biodegradable on lipase has greatly reduced any environmental contamination issue, (4) lipase is not corrosive to the production equipment, (5) using immobilized lipase, it can be easily separated and reused for next reaction with high catalytic activity, and (6) much less methanol is required and its amount is nearly generic as stoichiometric amount (compared with at least 5 to 6 molar ratio in catalyzed transesterification). Over the constraints, merits and demerits posed by biodiesel, it is thought that the high efficiency on transesterification process could contribute to lower down the production cost. In this case, the modeling and process control tool which functions to effectively capture the reaction data is crucial for achieving highest efficiency and lowest production cost.

Process modeling is important to express the mathematical representation of the process for further understanding of its characteristics. The result of modeling is useful for process control and improvement. Therefore, the quality of modeling could directly affect the plant operation. Conventionally, the process models for catalyzed transesterification were usually developed in the form of ordinary differential equation (ODE). The ODE relates the change of reaction component concentrations in time to the function involves reaction components concentration as well as the reaction rate constants. For instance, Hanny *et al*. (2010) had developed the models where the rate constants were found using curve fitting method by software KINTECTUS (version 3.961). Alternatively, the model can be formed using response surface methodology (RSM) based on central composite rotatable design (CCRD). The model can then be regressed using analysis of variance (ANOVA) by software Design Expert, as conducted by researchers Jibrail *et al*. (2008), Hameed *et al*. (2009) and Machavaram *et al*. (2008). For modeling on enzymatic transesterification, researchers commonly adapted the conventional kinetics model and Ping Pong Bi Bi model. The studies using conventional kinetics models have been successfully conducted by Selmi and Thomas (1998), Hernandez-Mantin and Otero (2008) and Benjamas Cheirsilp *et al*. (2008). Besides, Ping Pong Bi Bi model had been applied by researchers such as Fernando *et al*. (2009), Sulaiman *et al*. (2008) and Abdul Halim and Kamaruddin (2008). In overall, the conventional modeling method employs more effort in developing the model and determining the model coefficients. When the number of models and coefficients are high, this could make the modeling work becomes more complicated and tedious. In addition, extra computational effort will also be incurred when the structure of the model is complex.

In this study, artificial neural network (ANN) as an advanced modeling tool, is used to predict the dynamics of each chemical component involved in both base-catalyzed and enzymatic transesterification process. ANN performs black box modeling by rapidly linking the input to the output without the need to determine the coefficients in the original kinetics models. By resembling the neuron architecture and the nature of information processing in human brain, ANN works with the combination of a number of closely interconnected nodes for information distribution and processing (Baughman and Liu, 1995). By having the hidden layer processes with activation function like hyperbola tangent sigmoid, the ANN is capable to capture or model the highly complex or even illness-contained data. In this study, the input-output data was generated from the ODE models with the coefficients developed by Hanny *et al*. (2010). Using software Matlab® version 7.0.4 (Stephen, 2010), the ANN program is developed in m-file as single hidden layer, feed-forward type and coupled with Levenberg-Marquardt (LM) training algorithm.

Upon understanding the importance of finding a modeling tool for transesterification process, this study is aimed to perform modeling on biodiesel transesterification process by using self-developed ANN program with LM algorithm (Levenberg, 1994; Marquard, 1963). Unlike the existing ANN toolbox, this ANN program contains all the adjustable network parameters (such as number of epochs, initial damping factor, etc.) which could be changed from the m-file for performance optimization. Apart from that, ANN training sequence is self-designed with unique protocol to ensure high prediction accuracy and robustness of the trained ANN. During simulation, the ANN is trained to capture the dynamic pattern of concentration change for the chemical components in transesterification process. After simulation, the ANN performance is analyzed and discussed based on statistical parameters such as mean squared error (MSE) and correlation coefficient (R^2) which are able to intimate the best fit of the simulated curve in the case of least square regression (Steven and Raymond, 2003).

J. Appl. Sci. Res., 8(3): 1672-1681, 2012

Material and Methods

The research methodology in this study is designed to assess the ANN capability in modeling the biodiesel reaction concentration data which appears in both discrete and continuous form. In particular, this study is conducted sequentially on three major tasks, which are (1) input-output data generation, (2) development of computational program, and (3) network performance analysis and discussion, as described in the following section.

Input-Output Data Generation:

There are two case studies of ANN modeling which involves discrete data in catalyzed transesterification and continuous data in enzymatic transesterification. Therefore, there are two different models obtained from literature are used for data generation. For catalyzed transesterification, the model is obtained from literature by Jibrail *et al*. (2008) where it is developed using RSM based on CCD. For enzymatic transesterification, the model is obtained from the study by Fernando *et al*. (2009) where it is based on Ping Pong Bi Bi model. The software used for the data generation is known as Polymath (version 5.1, Prentice Hall).

Input-Output Data for Study of Catalyzed Transesterification:

There are totally four input data in discrete form, which are reaction temperature, reaction period, ratio of oil-to-methanol and amount of catalyst. As tabulated in Table 1, each input parameter is varied into five data, for instance, reaction temperature is ranged from 50°C to 190°C with interval of 35°C.

Table 1: List of Input Parameters in 5 Intervals of Actual Values

Parameter	Unit	Data 1	Data 2	Data 3	Data 4	Data 5
Reaction temperature, x_1	°C	50	85	120	155	190
Reaction period, x_2	min	60	120	180	240	300
Ratio of oil/methanol, x_3	mol mol^{-1}	1:4	1:6	1:8	1:10	1:12
Amount of catalyst, x_4	wt%	1	2	3	4	5

From the literature, all four input variables relate to the product yield and the mathematical equation is generated after the regression analysis with RSM based on four variables CCD. Excluding the insignificant terms based on analysis of variance (ANOVA), the developed equation is as follow:

$$Y = 58.42 + 19.70x_1 + 12.23x_2 + 7.95x_4 - 5.39x_1^2 - 5.48x_2^2 - 3.74x_3^2 - 2.82x_4^2 + 7.13x_1x_2 + 12.04x_1x_4 - 13.26x_2x_4 \quad (1)$$

The Eq (1) consists of coefficient Y as predicted yield of palm oil FAME (mol mol^{-1}), x_i and x_j as input variables, b_o is the offset term, b_j is the linear effect, b_{ij} is the first order interaction effect and b_{jj} is the squared effect. Using this mathematical model by varying four variables in five different values, there are totally 321 sets of training data, 41 sets of testing data and 23 sets of validation data generated. In this study, the ANN models the single output as product yield based on four discrete input variables where its intrinsic relationship is deemed to be erratic. The ANN structure is having four units of input neurons and single output neuron, with single hidden layer.

Input-Output Data for Study of Enzymatic Transesterification

This modeling study is based on the enzymatic transesterification on sunflower oil using lipase Novozym® 435 as enzyme and solvent hexadecane as described by Fernando *et al*. (2009). Fernando *et al*. (2009) had modified the Ping Pong Bi Bi model by adding the mixture volume balance equation for the ethanol being introduced to the system, as shown in Eq (2).

$$\frac{dN_{oil}}{dt} = \frac{-m_{cat.}V.V_m.[oil][alc]}{Km_{oil}.[alc]\left(1+\frac{[alc]}{K_i}\right)+Km_{alc}[oil]+[oil][alc]} \quad (2)$$

In batch system, the change of alcohol concentration is dependent to the conversion of oil over the changed volume of mixture due to addition flow rate of alcohol (v_0). Subsequently, the model for the molar change of ester (biodiesel) and glycerol is expressed in Eq (3) and (4).

$$\frac{dN_{alc}}{t} = 3\frac{dN_{oil}}{t} + \frac{v_0 N_{oil0}\theta_{alc}}{V_0} \quad (3)$$

$$\frac{dN_{oil}}{t} = -\frac{1}{3}\frac{dN_{ester}}{t} = -\frac{dN_{glyc}}{t} \quad (4)$$

where,

V_m = Maximum initial rate of reaction
m_{cat} = Mass of enzyme
V = Mixture volume
[oil] = Oil concentration, N_{oil}/V
[alc] = Alcohol concentration, N_{alc}/V
Km_{oil} = Binding constant for oil
Km_{alc} = Binding constant for alcohol
K_i = Inhibition constant for alcohol
N_{oil0} = Initial molar amount of oil
V_0 = Addition flow rate of alcohol
θ_{alc} = Initial molar ratio of alcohol to oil

Using the available ODE and kinetic parameters as tabulated in Table 2 from Fernando *et al*. (2009) literature, the data for the change of molar amount of oil (N_{oil}), alcohol (N_{alc}), ester (N_{est}) (as biodiesel) and glycerol (N_{gly}) (as by-product) are generated. With the structure known as single input multiple output (SIMO) system, the time element is the input to the ANN while molar amounts for four reactants and products are computed as output data from output neurons. Finally, the generated data includes 80 sets of training data, 13 sets of testing data and 7 sets of validation data.

Table 2: Kinetic Parameter for Sunflower Oil

Kinetic Parameter	Unit	Sunflower Oil
Maximum initial rate of reaction, Vm	$mol.hr^{-1}.mgcat^{-1}.L^{-1}$	0.000240000
Binding constant for oil, Km_{oil}	$mol.L^{-1}$	0.000006331
Binding constant for alcohol, Km_{alc}	$mol.L^{-1}$	0.040000000
Inhibition constant for alcohol, Ki	$mol.L^{-1}$	0.000287970

Development of Computational Program:

The computational program used for ANN simulation is Matlab® version 7.0.4. Without using the available ANN toolbox, this program is self-developed using m-file by applying Levenberg-Marquardt (LM) algorithm. In addition, the internal parameters in the ANN program, such as initial range of weights and biases, ANN training protocol, updating factor of LM algorithm, etc., can be adjusted for optimization. The ANN program is written as feed-forward single hidden layer network. The ANN modeling simulation comprises of three phrases which are training, testing and validation phase prior to asserting the generated ANN model as well-trained and robust.

Firstly, the program is initiated by loading the training, testing and validation data. This is followed by data scaling to obtain zero mean and unity standard deviation. Before scaling, the noise data might be added as random number to the original data for the purpose of studying the ability of ANN to handle noisy data. After scaling, weights and biases will be generated using random number in order to excite and initiate the network calculation. In training phase, the scaled training data are executed with fixed number of epochs. In every epoch, LM algorithm is used to calculate the adjustment rate of weights and biases by Jacobian matrix (error derivative over each weight and bias), error gradient, Hessian matrix and damping factor.

The training error is calculated instantaneously and compared with the previous error in iteration. If the newly calculated error is not lower than the previous error, the damping factor in LM algorithm will be multiplied with fixed factor of 10 in order to obtain the new damping factor. Error is then calculated and the process is repeated if the error is again not lower than the previous value. This iteration will continue until the error is less than the previously registered error. If this is successful, the iteration for that epoch will be stopped. If this error is lower than the error goal, the whole training process will be stopped and this whole calculation is counted as one epoch. Else, the second epoch will be initiated where the error calculated in first epoch will be registered as comparing error. In the second epoch, the calculated error should be lower than the comparing error (which calculated in first epoch). In this case, the error is ensured to be always maintained in negative gradient or decreasing trending over each epoch.

The testing phase is designated as cross-validation phase whereby the weights and biases obtained in training phase is applied to testing data set and the testing error is calculated. The calculated testing error will be compared with the initial testing error and the error goal which is similar value used in training phase. If the new testing error is larger than the previous testing error or the error goal, the network learning will return back to training phase and the training will be re-initiated using the lastly calculated weights and biases. If the calculated testing error is lower than the error goal, the testing phase will be stopped. Else, the testing phase will be ceased as well if the iteration limit in testing phase is reached. Either in each case, the computing will proceed to validation phase. In overall, the ANN simulation is executed from training phase to testing phase but it could be returned to training phase again if the result in testing phase is not satisfied.

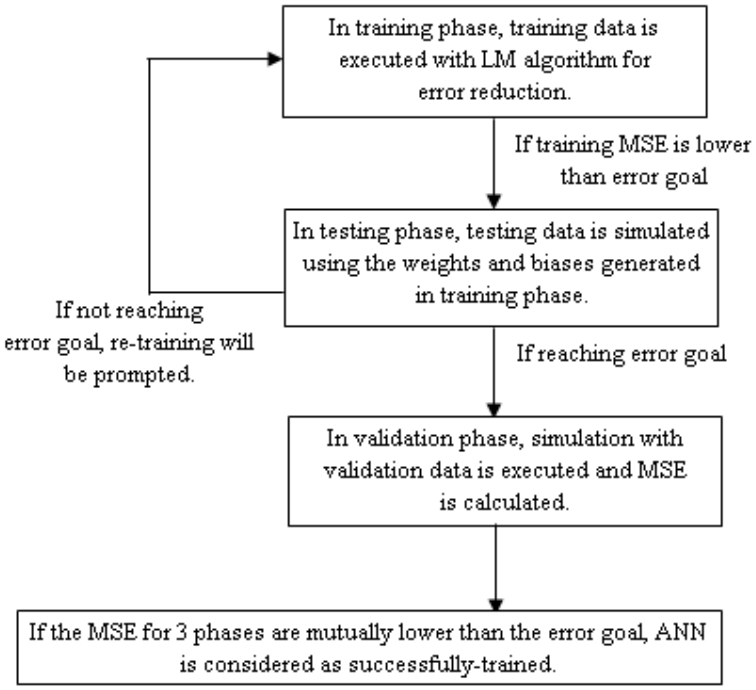

Fig. 1: Simplified Training Sequences for ANN

Finally, validation phase is introduced after testing phase. There is no any parameters adjustment in validation phase but to calculate the error in order to have final determination of the network performance. In case of the error in validation phase does not meet the error goal (even it has been achieved in both training and testing phase), the newly developed ANN is defined as not successful due to the incapability to tackle overfitting problem. Validation step is critical to provide final judgment to the ANN performance by using several data selected randomly over the whole range of input data. As final judgment, the MSE for all three phases are calculated to which must be mutually lower than the error goal in order to declare the network as successfully trained. Besides, the overall MSE which considers all three phases (instead of merely training phase) is calculated to represent the total error of the modeling.

Network Performance Analysis and Discussion:

The ANN performance could be optimized by varying the network training parameters easily through the self-developed m-file. Those parameters are inherent and commonly set as default values in program, such as the initial range of weights and biases, number of epochs, error goal, etc. In this simulation, several internal parameters have been pre-determined and this includes (1) initial range of weight of biases as -2 to +2, and (2) modification of updating formula of damping factor, λ in LM algorithm. Note that the default method to update λ is by multiplying and dividing the constant 10 in Eq (5). In this study, the updating formula is modified to be relating to training MSE and error goal (EG), as shown in Eq (6).

$$\text{New } \lambda = \text{Old } \lambda \times 10 \quad (5)$$

$$\text{New } \lambda = \text{Old } \lambda + \text{Old } \lambda \times [(\text{MSE} - \text{EG}) \times \text{Mul/Div}] \quad (6)$$

As summary, this self-developed ANN possesses the features in training protocol, includes (1) continuous error reduction during ANN training, (2) adjustment of network parameters and LM algorithm updating formula (Eq (6)) for performance optimization, (3) testing phase as cross-validation to avoid overfitting issue, (4) re-training after testing phase for continual reduction of error if the MSE cannot be lower than error goal in testing phase , and (5) final judgment by validation phase prior to declaring ANN as well-trained.

For ANN performance analysis, the number of hidden neurons will be varied from 1 to 15 units. Besides, this is followed by the study based on effect of noise. The noise is generated using random number according to the percentage of mean value for the output data. This means that the output data is added with noise and its relationship with the input data will be differed. In this study, the noise has been increased from 0 to 3.5%. All the simulation results are finalized and summarized in table and graphical form for better analysis and discussion. The ANN performance is examined based on the statistical data of MSE, R^2 and number of epochs.

Results and Discussion

Case Study 1: Modeling on Catalyzed Transesterification With Discrete Data:

There are four inputs of reaction temperature (50 to 190°C), reaction period (60 to 300 min), ratio of oil to methanol (1:4 to 1:12 mol mol^{-1}) and amount of catalyst (1 to 5 wt%) which generate the single output of biodiesel yield in unit of percent. The type of the input and output data is discrete data where it poses indirect relationship if compared with continuous data. It is aimed to judge the ANN capability in managing discrete data which might be a time-consuming exercise when using conventional modeling methods.

ANN Performance Based on Number of Hidden Neurons:

By varying the number of hidden neurons from 1 to 15, the ANN performance in MSE and R^2 is illustrated in Table 3. When the hidden neurons are low at 1 to 3 units, the ANN performs poorly with R^2 at 0.5894 only. When the number of hidden neurons is increased from 4 units, the R^2 has reached up to at least 0.90. With hidden neurons at 6 to 15 units, the R^2 is higher than 0.99. Exceptionally good performance with highest R^2 of 1 and MSE of 3.5116×10^{-3} has been achieved with 12 hidden neurons and it is selected as optimal number of hidden neurons for the best performance of ANN. From Figure 2, it is shown in correlation plot that the training, testing and validation data of biodiesel yield have widely covered the range of 0 to 100%. The result from the correlation plot is promising when the slope of linear curve are all shown as one for three plots. The y-intercept values are sufficiently low, with value of 0.0001, 0.03 and 0.0049 for training, testing and validation respectively.

The ANN model has correlated to the validation data accurately even with scattered coverage within the whole data range. For the training, there are totally 28 epochs and 2 testing phase undergone. This indicates that the first testing phase is unsuccessful to fulfill the error goal and the process returns to the training again. As shown in Figure 3, the reduction of training MSE has been steady and progressive where the MSE has continually decreased until it is below the target error. The ANN has shown a good learning and prediction capability over the discrete data as seen from the constant reduction of MSE. Apart from that, the ANN also portrays high robustness upon the fulfilling of stringent error target in testing and validation phase over the unseen data.

Table 3: ANN Performance under Variation of Number of Hidden Neurons for Case Study 1 and 2

No. of Hidden Neurons	Case Study 1			Case Study 2		
	Overall MSE	R^2	No. of Epochs	Overall MSE	R^2	No. of Epochs
1	5.299E+02	0.5894	3	1.136E-03	0.9715	5
2	5.278E+02	0.6079	2	2.837E-05	0.9998	10
3	5.548E+02	0.6334	3	1.993E-05	0.9998	9
4	4.662E+00	0.9969	18	2.210E-06	0.9998	10
5	9.636E+01	0.9333	10	9.400E-07	0.9999	21
6	6.160E+00	0.9960	23	3.710E-06	0.9998	4
7	3.698E+00	0.9976	18	9.800E-07	0.9999	11
8	7.545E-02	0.9999	25	9.800E-07	0.9999	15
9	7.099E-02	1.0000	20	1.050E-06	0.9999	8
10	3.710E+00	0.9978	10	9.400E-07	0.9999	20
11	1.389E-01	0.9999	27	5.140E-06	0.9997	6
12	3.512E-03	1.0000	28	2.490E-06	0.9994	8
13	2.682E+00	0.9985	13	9.700E-07	0.9999	17
14	6.235E-01	0.9998	20	9.500E-07	0.9999	14
15	5.836E+00	0.9964	24	1.270E-06	0.9999	85

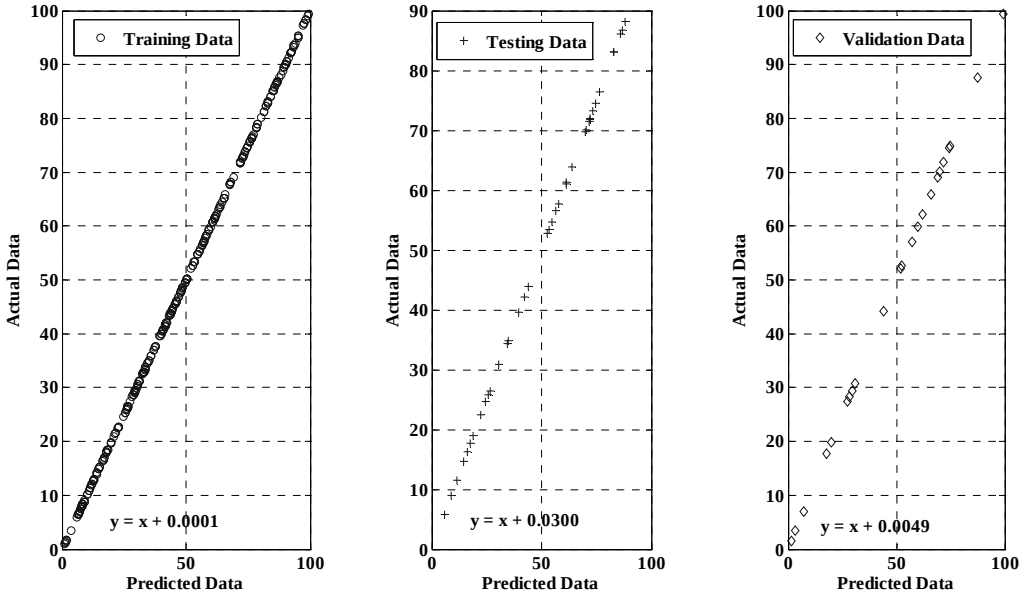

Fig. 2: Case Study 1: Correlation Plot for Predicted and Actual Discrete Data

ANN Performance Based on Effects of Noise:

Experimental data always contains noise which could be resulted by the inaccuracy or error of the measurement tools. This error brings some deviations (which may not be easily noticeable) from the original data and it could affect the modeling accuracy. Unlike the continuous data, the relationship between the input and output discrete data may still sporadic even with the addition of noise. It could be hard to distinguish if the discrete output data has already been introduced with noise. In this study, the noise is generated as random number and varied from 0.5 to 3.5% before being added to the output data. As displayed in Table 4, the overall trending for MSE and R^2 has changed in opposite pattern with each other. In particular, it is noted that the MSE increases while R^2 drops but the margin is insignificant and could be ignored. This is because the R^2 has still maintained nearly to unity value with noise changed from 0% to 3.5%. This result is interpreted that the ANN learns and simulates the data as usual protocol even it has been containing noise. Apart from that, there is no significant difference in terms of the number of epochs and this shows that the ANN can effectively reduce the prediction error during the training. As alternative to improve the prediction accuracy, it is recommended to maintain the measurement tools to always provide most reliable data for modeling.

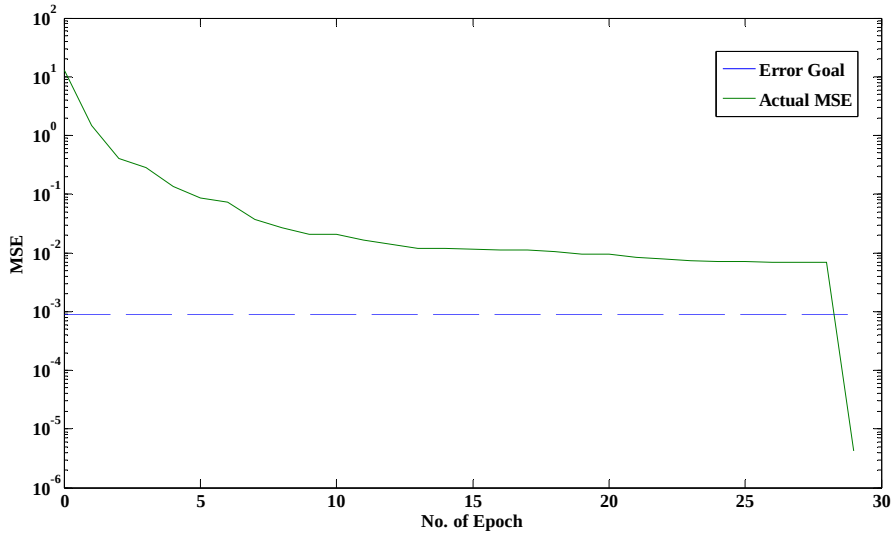

Fig. 3: Case Study 1: Trending for MSE Reduction During Training

J. Appl. Sci. Res., 8(3): 1672-1681, 2012

Table 4: ANN Performance Data under Variation of % Noise for Case Study 1 and 2

Noise %	Case Study 1			Case Study 2		
	Overall MSE	R^2	No. of Epochs	Overall MSE	R^2	No. of Epochs
0.0	3.512E-03	1.0000	28	9.400E-07	0.9999	21
0.5	2.051E-02	1.0000	26	1.250E-06	0.9998	24
1.0	7.038E-02	1.0000	46	2.000E-06	0.9998	28
1.5	4.574E-01	0.9997	20	3.780E-06	0.9997	18
2.0	2.688E-01	0.9998	22	5.190E-06	0.9996	26
2.5	4.500E-01	0.9997	22	6.600E-06	0.9994	17
3.0	6.275E-01	0.9996	24	8.890E-06	0.9992	19
3.5	8.325E-01	0.9994	24	1.139E-05	0.9990	39

Case Study 2: Modeling on Enzymatic Transesterification With Continuous Data:

The data used in this study is obtained from Ping Pong Bi Bi model indicating the change of molar amount for oil, alcohol, ester and glycerol with the time. Using the generated input-output data, ANN models the dynamic of the enzymatic transesterification reaction with input data of time. With single hidden layer, the ANN structure is having one input neuron and four output neurons. It is aimed to judge the ANN capability to model the continuous data with multiple outputs in single model.

ANN Performance Based on Number of Hidden Neurons:

Similar as Case Study 1, the ANN is simulated by varying the number of hidden neurons from 1 to 15 units and the performance results such as MSE and R^2 are tabulated in Table 3. It is noted that the MSE has reduced drastically when the hidden neuron is increased from 1 unit to 2 units and subsequently it has decreased progressively when it is further increased to 15 units. When the hidden neuron is 5 units, minimum MSE and highest R2 are achieved. There is no significant improvement when it is increased further to 15 units. With 5 units of optimal number of hidden neurons, the ANN with structure [1, 5, 4] achieves low MSE of 9.4×10^{-7} with R^2 of 0.9999. Referring to the Figure 4, the prediction model for molar amount of oil, alcohol, ester and glycerol have matched perfectly to the actual data throughout the trending. This proves the excellent ability of ANN to capture the dynamic trending for multiple output data which appears in different ranges and characteristics.

As compared with original Ping Pong Bi Bi equation and other reaction kinetics equations which are having four individual models for four species, ANN manages to model the trending for all species simultaneously in a simple model. ANN in this case provides advantage of reducing the effort in finding out the value of kinetic parameters as the original mechanistic model. Unfortunately, ANN poses the problem of not having any information on the process characteristics as available in mechanistic model. Rather, ANN is working on sufficient amount of data for learning and familiarization for almost every type of general or ill data. From linear correlation analysis between the predicted and actual data for molar amount of oil, the slope of the linear curve is 1 and the y-intercept value as low as 6.9×10^{-7}. The training of the ANN using 5 units of hidden neurons shows successful result with only 21 epochs totally to achieve the satisfactory performance. In training phase, the MSE is reduced in steady form without undue lagging along the epochs.

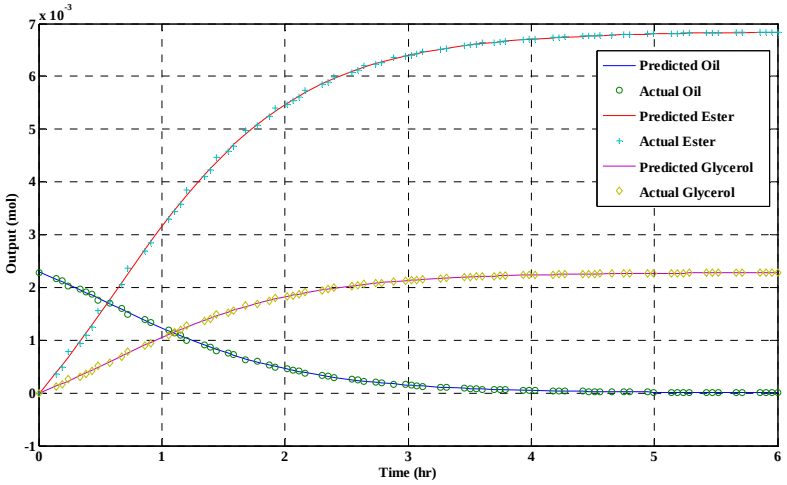

Fig. 4: Case Study 2: Actual and Predicted Training Data for Oil, Ester and Glycerol

ANN Performance Based on Effect of Noise:

With 5 units of ANN hidden neurons as optimized in previous section of Case Study 2, the noise is introduced from 0.5 to 3.5% to the original data. The objective of this study is to judge the ANN prediction and robustness in modeling the continuous data which contains noise. As shown in Table 4, when the noise is increased to 3.5%, the overall MSE has merely increased for 1.045×10^{-5}. Besides, the R^2 has also reduced insignificantly for 0.0009 compared with the result without noise. The ANN prediction performance has been maintained at satisfied level without appreciable deviation even with noise introduced to 3.5%. The prediction curve does not show unstable fluctuation over the noisy trend. Rather, it fairly captures the whole pattern in robust manner. ANN training has been completed rapidly with low number of epochs without showing symptoms of sluggish error convergence. The highest epochs is registered as 39 only when noise is 3.5%. It is noted that several simulation amongst the studies has undergone two testing phases. This suggests that the program sequence for re-training could improve the ANN performance by continuously reducing the error over the epochs and the phases, while eliminating the overfitting issue.

Conclusion:

The developed single hidden layer ANN has shown good prediction capability for both cases of discrete and continuous data in catalyzed and enzymatic transesterification process respectively. The ANN is self-developed where the training protocol and training parameters are modified. With the simplest structure of ANN, it generates sufficiently good modeling result as judged by MSE, R^2 and number of epochs. For both cases, low number of hidden neurons has generated optimal result. The predicted model has correlated perfectly to the original data without significant deviation. The ANN could effectively capture the relationship between the discrete input and output data without undergoing extensive training. ANN training is effective as note that the training MSE has been reduced progressively over the epochs until achieving the error goal. Testing and validation phase are designed for cross-validation to the trained ANN model for sake of tackling overfitting issue. It is deemed that the real application of ANN could reduce the modeling effort in order to facilitate the process optimization, control and integration.

References

Baughman, D.R. and Y.A. Liu, 1995. Neural Networks in Bioprocessing and Chemical Engineering. Academic Press, Inc.

BP Statistical Review of World Energy June 2010, See also: < http://www.bp.com/statisticalreview>

Charpe, T. W. and V. K. Rathod, 2010. Biodiesel Production Using Waste Frying Oil. Waste Management., 26: 487-494.

Dennis, Y.C.L., W. Xuan and M.K.H. Leung, 2009. A Review on Biodiesel Production using Catalyzed Transesterification. Applied Energy, 87: 1083-1095.

Fernando, L.P.P., P.M. Shayane and W.C.F. Pedro, 2009. Production of Biodiesel via Ethanolysis of the Sunflower and Soybean Oils: Modeling, Appl. Biochem Biotechnol, 161, 238-244, Humana Press.

Gui, M.M., K.T. Lee and S. Bhatia, 2008. Feasibility of Edible Oil vs. Non-edible Oil vs. Waste Edible Oil as Biodiesel Feedstock. Energy, 33: 1646-1653.

Hameed, B.H., L.F. Lai and L.H. Chin, 2009. Production of Biodiesel from Palm Oil (Elaeis guineensis) using Heterogeneous Catalyst: An Optimized Process. Fuel Processing Technology, 90: 606-610.

Hanny, J.B., M. Kayoko and T. Takayuki, 2010. Kinetic Study of Hydroxide-Catalyzed Methanolysis of Jatropha curcas-Waste Food Oil Mixture for Biodiesel Production, Fuel., 1: 1-7.

Hideki, F., K. Akihiko and N. Hideo, 2001. Biodiesel Fuel Production by Transesterification of Oils. Bioscience and Bioengineering, 92: 405-416.

Jibrail, K., K.T. Lee and S. Bhatia, 2008. Biodiesel Production from Palm Oil via Heterogeneous Transesterification. Biomass and Bioenergy, 33: 271-276.

Juan, J.C., D.A. Kartika, T.Y. Wu and Y.Y. Hin, 2010. Biodiesel Production from Jatropha Oil by Catalytic and Non-Catalytic Approaches: An Overview. Bioresource Technology, 102: 452-460.

Kansedo, J., K.T. Lee and S. Bhatia, 2009. Cerbera odollam (Sea Mango) Oil as A Promising Non-Edible Feedstock for Biodiesel Production. Fuel., 88: 1148-1150.

Levenberg, K., 1994. A Method For The Solution of Certain Problems in Least Squares. Quart. Appl. Math. 2: 164-168.

Machavaram, R., C.J. Prakash and R. Hifjur, 2008. Prediction of Optimized Pretreatment Process Parameters for Biodiesel Production Using ANN and GA. Fuel., 88: 868-875.

Marchetti, J.M., V.U. Miguel and A.F. Errazu, 2007. Possible Methods for Biodiesel Production. Renewable and Sustainable Energy Reviews., 11: 1300-1311.
Marquardt, D., 1963. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. SIAM J. Appl. Math., 11: 431-441.
Palligarnai, T.V. and F. Boyi, 2010. Environmentally Sustainable Biofuels: Advances in Biodiesel Research. Waste Biomass., 1: 47-63.
Sharma, Y.C. and B. Singh, 2008. Development of Biodiesel: Current Scenario. Renewable and Sustainable Energy Reviews., 13: 1646-1651.
Siddharth, J. and M.P. Sharma, 2010. Biodiesel Production from Jatropha curcas Oil. Renewable and Sustainable Energy Reviews., 14: 3140-3147.
Abdul Halim, S.F. and A.H. Kamaruddin, 2008. Catalytic Studies of Lipase on FAME Production from Waste Cooking Palm Oil in tert-butanol System. Process Biochemistry, 43: 1436-1439.
Srivathsan, V.R., L.N. Srinivasan and M. Karuppan, 2008. An Overview of Enzymatic Production of Biodiesel. Bioresource Technology, 99: 3975-3981.
Stephen, J.C., 2000. MATLAB, Programming for Engineers. Brooks/Cole Publishing Company.
Steven, C.C. and P.C. Raymond, 2003. Numerical Methods for Engineers. Fourth Edition, Mc-Graw Hill Companies, Inc.
Sulaiman A.Z., A. Dowaidar and H. Kamal, 2009. Dynamic Modeling of Biodiesel Production from Simulated Waste Cooking Oil using Immobilized Lipase, Biochemical Engineering Journal, 44: 256-262.
Yee, K.O. and S. Bhatia, 2010. The Current Status and Perspectives of Biofuel Production via Catalytic Cracking of Edible and Non-Edible Oils. Energy, 35: 111-119.