International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 10 (2018) pp. 7529-7540

Predictive Capability Evaluation of RSM and ANN in Modeling and Optimization of Biodiesel Production from Palm *(Elaeisguineensis)* Oil

Dang Nguyen Thoai[1,2,*], Chakrit Tongurai[1], Kulchanat Prasertsit[1], Anil Kumar[3,4]

[1]*Department of Chemical Engineering, Faculty of Engineering, Prince of Songkla University, Songkhla 90112, Thailand.*
[2]*Department of Chemical Engineering, Faculty of Chemistry, Quy Nhon University, Binh Dinh 820000, Vietnam.*
[3]*Department of Mechanical Engineering, Faculty of Engineering, Prince of Songkla University, Songkhla 90112, Thailand.*
[4]*Department of Energy (Energy Center), Maulana Azad National Institute of Technology, Bhopal, India.*

[*]*Correspondence Author*

Abstract

In the present study, response surface methodology (RSM) and artificial neural network (ANN) are applied for biodiesel production via base-catalyzed transesterification. These models are also compared in order to optimize the methyl esters production process from edible oils. Methanol/oil molar ratio (3:1-9:1), sodium methoxide catalyst content (0.50-1.30 wt%), reaction temperature (45-65ºC) and time (30-70 min) were considered during process using Central Composite Design. RSM and ANN models show a high accuracy in terms of coefficient of determination (R^2> 0.99) and mean relative percent deviation (MRPD = 0.22-0.27%). Molar ratio and catalyst content are identified as two most important factors for base-catalyzed methanolysis. A high predicted output of FAME percentage of 98% was determined by the ANN model under optimum conditions; including MeOH/oil molar ratio of 5.88, catalyst content of 0.89 wt%, reaction temperature of 55 ºC in 50 min. Therefore, ANN model is a better solution over the RSM model and recommended for optimizing biodiesel production.

Keywords: Biodiesel; Response Surface Methodology; Artificial Neural Network; Transesterification.

Nomenclature

FAME	Fatty acid methyl ester
RSM	Response surface methodology
ANN	Artificial neural network
CCD	Central composite design
DOE	Design of experiment
EED	Essential experimental design
R	Correlation coefficient
R^2	Coefficient of determination
MSE	Mean square error
RMSE	Root mean square error
MAE	Mean absolute error
SEP	Standard error of prediction
MRPD	Mean relative percent deviation
GC	Gas chromatography

INTRODUCTION

The development of the world economy is depended on energy resources. As per the U.S. Energy Administration report of 2016, fossil fuels were the biggest source of energy (78%), whereas, the nuclear power and renewable energy were only 10% and 12%, respectively [1]. Many environmental complications like global warming, pollution, ozone layer depletion are due to fossil fuels [2]. Therefore, the great challenge is to produce energy from non-fossil and eco-friendly resources. Biodiesel has proven as a good replacement because of its renewability, biodegradability, non-toxic and high safety [3,4].

The most common method for biodiesel production is vegetable alcoholysis or transesterification [5,6]. In this method, vegetable oil reacts with alcohol in the presence of catalyst and create biodiesel and glycerol [5,7]. Base-catalyzed transesterification method has been studied among the catalyzed transesterification process by various researchers due to its high catalytic activity and, higher conversion of triglyceride to biodiesel [8,9].

Alcohol/oil molar ratio, catalyst content, time and reaction temperature are factors affect base-catalyzed transesterification process [2,3,8]. Process optimization is an important and notable issue and, requires to increase the biodiesel production efficiency and to reduce the production cost. The base-catalyzed transesterification process involved in these factors have been studied, modeled and optimized by RSM and ANN [8,10-15].

RSM is one of significant statistical method used in experimental design, modeling and optimization [16,17]. It gives relation between one or more responses with independent factors. It also determines the effect of independent factors, including single effects and interaction effects, on the whole process. Moreover, this method gives a mathematical relation for predicting the desire output. Thus, biodiesel can be modeled from RSM with minor estimation error in different conditions. Several researchers have been used this tool effectively for the efficiency evaluation of biodiesel production from base-catalyzed transesterification [6,18]. Thoai et al. have applied RSM in the optimization of the one-step methanolysis of refined palm oil (RPO) catalyzed by sodium methoxide as homogeneous base catalyst [8]. Moreover, Thoai et al. continued using RSM in the optimization of the novel two-step

transesterification [12]. RSM has also been used in based-catalyzed ethanolysis of sunflower oil [13,19].

ANN is the most popular artificial learning tool with a wide application range. It has been extensively accepted as an alternative technique to represent the complicated input and output relationship of the process [20]. It is able to use for prediction outputs of a new input data, if ring of data is successfully trained, validated and tested by ANN. It has been successfully used for several transesterification processes through based-catalyzed mechanism, including the one-step and two-step process [11-14]. Betiku et al. have modeled and optimize the two-step process for biodiesel synthesis from non-edible neem seed oil. The results demonstrated that the ANN model accurately represented the process [11]. In another study, Stamenkovíc et al. showed optimization capability of ANN in base-catalyzed one-step ethanolysis of sunflower oil [13]. RSM and ANN were also considered for developing and comparing their predictive and generalization abilities in the ethanolysis reaction of refined sunflower oil [13].

RSM and ANN have been applied from long time in order to model and optimize the alkyl esters production process from edible oils, however, their results for same study is hardly compared. The aim of present study is to combine the Central Composite Design (CCD) with both RSM and ANN performance for palm oil methanolysis process catalyzed by homogeneous base catalyst – sodium methoxide. This is first effort to study the predictive capability evaluation of RSM and ANN models of the said process.

MATERIALS AND METHODS

Materials

Refined palm oil (RPO) was bought from Morakot Industry Public Co. Ltd. (Thailand). Methanol (CH_3OH, 99%) and sodium methoxide (CH_3ONa, 96%) were supplied by Labscan Asia Co. Ltd. (Thailand) and Dezhou Long Teng Chemical Co. Ltd. (China), respectively. Sodium hydroxide (NaOH) was obtained from Merck (Germany), while sodium periodate ($NaIO_4$) was acquired from Fisher Chemical (UK) and bromothymol blue was provided by Ajax Finechem (Australia).

Methods

One-step biodiesel production

The RPO used for this study had the low free fatty acid (FFA) content (0.11 wt%) which is suitable feedstock oil for one-step biodiesel production. The catalyzed methanolysis reaction was carried out in a 0.5 L three-necked flask with magnetic stirring of 600 rpm at atmospheric pressure, and refluxed by water at 20 ^oC to condense the methanol vapor. RPO was preheated until attain the required temperature. Later, the mixture of methanol and catalyst was added in RPO. The beginning time for the reaction was recorded at the moment when all of methanol and catalyst were entered to the reactor. The product mixture was poured into separatory funnel to separate into two layers of ester and glycerol after finish this reaction. The settling time was 60 min. Glycerol was taken out of separating funnel and the ester phase was washed by hot water (80 ^oC) for three-three times without and with shaking. The washed methyl ester was dried by the heating at 110 ^oC for 90 min. Finally, the biodiesel product (FAME) was checked for the ester content.

All the experiments were performed three times to estimate its errors. Experiments were designed at various conditions; including MeOH/Oil molar ratio of 3/1-9/1, CH_3ONa catalyst content of 0.50-1.30 wt%, reaction temperature of 45-65 ^oC and reaction time of 30-70 min.

Procedure of the ester content determination in biodiesel

Ester content analysis using Gas Chromatography (GC)

Methyl ester content was analyzed by following the standard method on B-100 biodiesel specified by the Department of Energy Business, Ministry of Energy, Thailand [21]. This method is based on European Standard (EN 14103) and was carried out at Scientific Equipment Center, Prince of Songkla University, Thailand. The methyl esters were quantified directly in gas Chromatography (GC) equipped with flame ionization detector (GC-FID). The column selected for biodiesel have length 30 m, 0.32 mm internal diameter, film thickness 0.25 μm with helium as the carrier gas at a flow rate of 1.0 mL/min and split ratio of 50:1. The inlet temperature was kept at 290 ^oC and the initial temperature was hold at 210^oC (for 12 minutes) followed by increasing at a rate of 20 ^oC/min till 250 ^oC, hold for 8 minutes. The detector temperature was kept at 300^oC and the injection volume of 1μl was used for analysis. Methyl heptadecanoate was used as the standard of this analytical method. FAME content, C_{FAME} (%) is calculated from integration results for a particular determination according to Eq. (1).

$$C_{FAME} = \frac{(\sum A) - A_{EI}}{A_{EI}} \times \frac{C_{EI} \times V_{EI}}{m} \times 100\% \qquad (1)$$

where, ΣA is the sum of all methyl ester peaks from C12 to C24:1, A_{EI} is peak area for methyl heptadecanoate (internal standard), C_{EI} is concentration (mg/ml) of the methyl heptadecanoate solution (10 mg/ml), V_{EI} is volume (ml) of the methyl heptadecanoate solution used (5 ml) and 'm' is exact weight (mg) of the FAME.

Novel chemical method in determining of ester content

In the study of Thoai et al., based on the mechanism of transesterification, one mole glycerol is produced from one mole glyceride in a large excess of alcohol and catalyst [22]. From glycerol titration process, the remaining content of glycerides can be proximate calculated from the content of total glycerol. Finally, ester content can be converted proximately by subtracting the content of remaining glycerides from 100 wt%.

Design of experiments

The CCD was applied to investigate the influences of the experimental variables on the FAME content and to find the optimum conditions for the requested FAME content. The CCD

incorporates five levels (coded–α, –1, 0, +1, +α) in which axial points (±α) for a factor and 0 for all other factors. In addition, the center points code 0 were used to estimate pure error. The four important factors: MeOH/RPO molar ratio (X_1), catalyst content (X_2), temperature (X_3) and time (X_4) were investigated as independent variables. The experimental limit and code levels of independent factors are shown in Table 1. A list of 30 experiments including 2^4 factorial runs, 8 runs for axial points and 6 runs for center points were carried out for CCD with 4 independent factors. These experimental FAME contents were used in the analysis of variance (ANOVA). The performances of RSM and ANN models are statistical tested by correlation coefficient (R), coefficient of determination (R^2), adjusted R^2, mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), standard error of prediction (SEP) and mean relative percent deviation (MRPD). These parameters are determined from Eqs. (2) to (9) [13,23,24]:

$$R = \frac{\sum_{i=1}^{n}(y_{p,i}-y_{p,ave})(y_{a,i}-y_{a,ave})}{\sqrt{\left[\sum_{i=1}^{n}(y_{p,i}-y_{p,ave})^2\right]\left[\sum_{i=1}^{n}(y_{a,i}-y_{a,ave})^2\right]}} \quad (2)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_{a,i}-y_{p,i})^2}{\sum_{i=1}^{n}(y_{p,i}-y_{a,ave})^2} \quad (3)$$

$$Adjusted\ R^2 = 1 - \left[(1-R^2)\times\frac{n-1}{n-k-1}\right] \quad (4)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_{p,i}-y_{a,i}\right)^2 \quad (5)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_{p,i}-y_{a,i}\right)^2} \quad (6)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\left(y_{a,i}-y_{p,i}\right)\right| \quad (7)$$

$$SEP = \frac{RMSE}{y_{a,ave}}\times 100 \quad (8)$$

$$MRPD = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_{a,i}-y_{p,i}}{y_{a,i}}\right| \quad (9)$$

where, n is the number of experiments, $y_{p,i}$ is the predicted outputs, $y_{a,i}$ is the experimental results, $y_{a,ave}$ is the average experimental results, $y_{p,ave}$ is the average predicted output and k is the sum of input factors.

Table 1. Limit and code levels of independent factors for the modeling and optimization from RSM and ANN.

Factor			Limit and code level				
Independent variable	Symbol	Dimension	–α	–1	0	+1	+α
Molar ratio	X_1	mol/mol	3.00	4.50	6.00	7.50	9.00
Catalyst content	X_2	wt%	0.50	0.70	0.90	1.10	1.30
Temperature	X_3	°C	45	50	55	60	65
Time	X_4	min	30	40	50	60	70

RSM modeling

The four important factors: molar ratio (X_1), catalyst content (X_2), reaction temperature (X_3) and reaction time (X_4), are investigated as independent variables for modeling and optimization of FAME content (y). Multiple regressions were applied for the second-order polynomial regression model equation in order to find correlation between the response value and the independent variables. Eq. (10) shows the fitted quadratic response model.

$$y = \beta_0 + \sum_{i=1}^{4}\beta_i X_i + \sum_{i=1}^{3}\sum_{j=i+1}^{4}\beta_{ij}X_iX_j + \sum_{i=1}^{4}\beta_{ii}X_i^2 \quad (10)$$

where, y is the predicted response (FAME content); β_0, β_i, β_{ii}, β_{ij} are the regression coefficients (β_0 is referred to as the intercept term, β_i are linear terms, β_{ii} are quadratic terms and β_{ij} are interaction terms); X_i, X_j are coded as independent factors.

The statistical significance of the independent variables, their interactions and the quality of the fitted model are tested via F-value, P-value and ANOVA. ANOVA is also applied to predict the FAME content following the experimental variances.

Contour plots are formed via the multiple regression equation by keeping two independent terms at an average value while vary other two terms. Model gives the optimum conditions for achieving highest-FAME content from independent experimental factors.

RSM uses Essential Experimental Design (EED) software in MS Excel [25]. After loading EED, an additional menu option, DOE (Design of Experiment), is become available in the main menu of MS Excel (menu Add-Ins). ER (Essential Regression) software is used for multiple regression and polynomial regression of experimental data. Additionally, Minitab software, version 16.2.2 is used to check the accuracy of results.

ANN modeling

A feed forward, back-propagation multi-layer perception (MLP) neural network analysis is carried out through the Levenberg-Marquardt (LM) algorithm for modeling of the process parameters of the base-catalyzed methanolysis reaction. This is done by using the neural network toolbox of MATLAB 2015a (8.5.0.197613). Training parameters of the ANN are given in Table 2. The MLP network is well known and widely applied feed forward network analysis. The feed forward network is a straight forward network that requires outputs in order to train the model. The ANN operating ability is investigated by MSE. The selected ANN has three layers of neurons such as; an input layer, a hidden layer and an output layer. The hyperbolic tangent sigmoid transfer function (Tansig) and linear transfer function (Purelin) are chosen for input and output layers, respectively. The architecture of the ANN is shown in Figure 1. The sum of input layer neurons are four, correspond to MeOH/RPO molar ratio (X_1), catalyst content (X_2), temperature (X_3) and reaction time (X_4). The output layer is FAME content. The optimum hidden neurons number is found by a heuristic method. It also examines various numbers of neurons until the MSE of the output data is the lowest value.

Table 2. ANN parameters used for training, modeling and optimization of base-catalyzed methanolysis of RPO.

Property	Value/comment
Algorithm	Levenberg-Marquardt (LM) Back propagation (BP) MSE
Minimized error function	
Learning	Supervised
Input layer	No transfer function is used
Hidden layer	TANSIG
Output layer	PURELIN
Number of best interaction/Epoch	27
Number of input neurons	4
Number of hidden neurons	3
Number of output neurons	1

An effective ANN model can be developed if the design terms and its responses are normalized. The input factors and output value are normalized before training to eliminate the over fitting. The input values and output value are normalized as following equations:

$$x_{i,normalized} = \frac{X_{ij} - X_{i,ave}}{0.5(X_{i,max} - X_{i,min})} \quad (11)$$

$$y_{normalized} = \frac{Y_j - Y_{ave}}{0.5(Y_{max} - Y_{min})} \quad (12)$$

where, $x_{i,normalized}$: normalized input layer of input variable i; X_{ij}: the value of input variable i at experimental run j; $X_{i,ave}$: the average value of input variable i; $X_{i,max}$ and $X_{i,min}$: the maximum and minimum value of input variable i, respectively; $y_{normalized}$: normalized output variable; Y_j: the value of output variable at experimental run j; Y_{ave}: the average value of output variable; Y_{max} and Y_{min}: the maximum and minimum value of output variable, respectively.

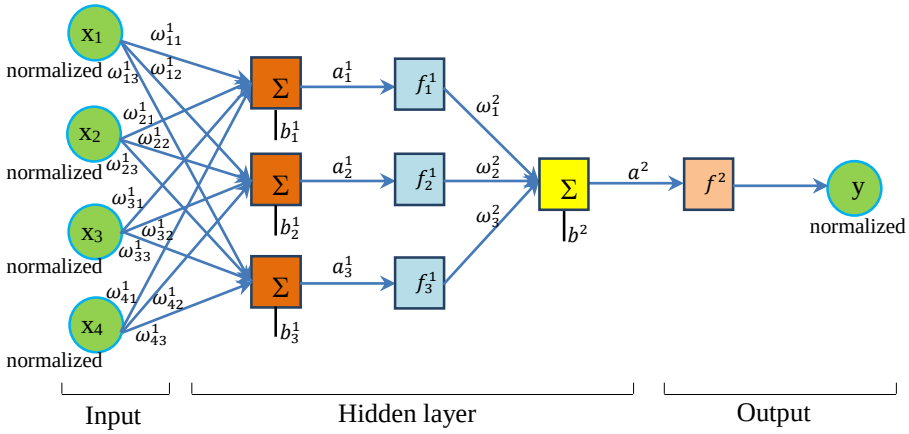

Figure 1 Structure of single hidden layer network of ANN in this study.

The output variable (FAME content) of the ANN model is determined and written as follows:

$$y_{normalized} = f^2(a^2) \quad (13)$$

$$a^2 = \left(\sum_{j=1}^{3} \omega_j^2 \times f^1\left(a_j^1\right)\right) + b^2 \quad (14)$$

$$a_j^1 = \left(\sum_{i=1}^{4} \omega_{ij}^1 \times X_i\right) + b_j^1 \quad (15)$$

where i: the sum of input terms (i=4); j: the sum of optimum neurons (j=3); a^1 and a^2: the linear combined outputs of the hidden layer and the output layer, respectively; b^1 and b^2: the bias of the hidden layer and the output layer, respectively; f^1 and f^2: the transfer function for the hidden layer and the output layer, respectively.

Finally, the output value is calculated or de-normalized to the original units by equation:

$$y = \left(y_{normalized} \times 0.5 \times (y_{max} - y_{min})\right) + y_{ave} \quad (16)$$

where y: output variable; $y_{normalized}$: normalized output variable; y_{max} and y_{min}: the maximum and minimum experimental output variables, respectively; y_{ave}: the average experimental output variable.

As mentioned above, total 30 experiments are completed with CCD design. Data are separated into three parts, including training (70% of total data points), testing (15% of total data points) and validation (15% of total data points) in ANN [24]. In the first, the training data are randomly chosen from the initial data. The weighted parameters of the interactions are calculated through a chain of repeats to get the minimum number of MSE between the calculated values and experimental FAME content. Subsequently, the testing data are applied to check the trained ANN. Finally, the validation data show the prediction of FAME content via the developed ANN modeling.

Evaluation ability of the RSM and ANN models

The developed models using RSM and ANN are investigated for predictive ability for the base-catalyzed methanolysis process. The coefficients of R, R^2, adjusted R^2, MSE, RMSE, MAE, SEP and MRPD are determined and employed for this purpose.

RESULTS AND DISCUSSIONS

The relationship between the four independent variables (MeOH/RPO molar ratio, catalyst content, reaction temperature and reaction time) and the FAME content is determined. The FAME content for each experimental run and from both RSM and ANN models are listed in Table 3.

RSM modeling

Analysis of variance (ANOVA)

Results of ANOVA in terms of the degree of freedom, the sum and means of squares, F-value and P-value are given in Table 4. The significance of the model, single terms, their squares and interactions is confirmed via their F-value and P-value. P-value less than 0.05 imply significant effects of these parameters on the FAME content.

Table 3. Designed independent factors and experimental results.

Run	Independent variables/ Input variables				Output variable/ FAME content (%)		
No.	X_1 (mol/mol)	X_2 (wt%)	X_3 (°C)	X_4 (min)	Experiment	RSM model	ANN model
1	4.50	0.70	50	40	83.51	83.29	83.29
2	7.50	0.70	50	40	94.48	94.45	94.56
3	4.50	1.10	50	40	90.60	90.32	90.64
4	7.50	1.10	50	40	98.64	98.59	98.02
5	4.50	0.70	60	40	86.87	86.39	86.80
6	7.50	0.70	60	40	95.31	95.11	96.13
7	4.50	1.10	60	40	93.42	93.21	92.99
8	7.50	1.10	60	40	98.83	99.05	98.61
9	4.50	0.70	50	60	85.76	85.55	85.78
10	7.50	0.70	50	60	96.28	96.20	96.27
11	4.50	1.10	50	60	91.56	91.47	91.82
12	7.50	1.10	50	60	98.74	99.23	98.68
13	4.50	0.70	60	60	87.92	87.67	88.05
14	7.50	0.70	60	60	95.59	95.88	95.59
15	4.50	1.10	60	60	93.34	93.38	93.80
16	7.50	1.10	60	60	98.78	98.70	98.80
17	3.00	0.90	55	50	80.84	81.55	81.06
18	9.00	0.90	55	50	98.45	98.03	98.37
19	6.00	0.50	55	50	87.07	87.52	87.02
20	6.00	1.30	55	50	97.53	97.37	97.57
21	6.00	0.90	45	50	93.45	93.54	93.50
22	6.00	0.90	65	50	95.92	96.12	96.23
23	6.00	0.90	55	30	93.61	94.10	94.14
24	6.00	0.90	55	70	96.21	96.01	96.24
25	6.00	0.90	55	50	94.87	94.95	94.82
26	6.00	0.90	55	50	95.32	94.95	94.82
27	6.00	0.90	55	50	95.19	94.95	94.82
28	6.00	0.90	55	50	94.79	94.95	94.82
29	6.00	0.90	55	50	94.47	94.95	94.82
30	6.00	0.90	55	50	95.04	94.95	94.82
					MSE	0.0879	0.0010
					R^2	0.9953	0.9958

From Table 4, molar ratio (X_1), catalyst content (X_2), reaction temperature (X_3), reaction time (X_4) and square terms of molar ratio and catalyst content (X_1^2, X_2^2) have positive effects on the FAME content. Moreover, two-way interaction of molar ratio

with catalyst content and reaction temperature (X_1X_2, X_1X_3), catalyst content and reaction time (X_2X_4), reaction temperature and time (X_3X_4) also have statistically significant effects on the FAME content. However, in other terms, X_3^2, X_4^2, X_1X_4 and X_2X_3 are observed to be insignificant on the FAME content.

The important operational variables molar ratio, catalyst content, temperature and reaction time and have F-values of 2318.83, 828.49, 56.45 and 31.41, respectively, and P-value <0.0001 (Table 4). Molar ratio and catalyst content have very high F-value as compared to other individual variables. This means that the molar ratio and catalyst content are the two most important factors in the present study. The increases the methoxide anion concentration speed up the FAME formation rate. This shows importance of the MeOH/RPO molar ratio in the enhancement the forward reaction rate. It shifts the reaction equilibrium toward the formation of product at higher concentration of methanol. The present results are similar to previous researches which shows the effects of molar ratio and catalyst content for base-catalyzed methanolysis [15,26].

Table 4. Results of ANOVA.

Source/ Term	Degree of freedom (DF)	Sum of squares (SS)	Mean square (MS)	F–value	P–value	Remarks
Model	14	639.242	45.660	259.84	<0.0001	Significant
Linear	4	568.491	142.123	808.79	<0.0001	Significant
X_1	1	407.468	407.468	2318.83	<0.0001	Significant
X_2	1	145.583	145.483	828.49	<0.0001	Significant
X_3	1	9.920	9.920	56.45	<0.0001	Significant
X_4	1	5.520	5.520	31.41	<0.0001	Significant
Square	4	54.002	13.500	76.83	<0.0001	Significant
X_1^2	1	45.592	45.592	259.46	<0.0001	Significant
X_2^2	1	10.732	10.732	61.07	<0.0001	Significant
X_3^2	1	0.024	0.024	0.13	0.720	Not significant
X_4^2	1	0.020	0.020	0.11	0.741	Not significant
2-Way interaction	6	16.749	2.792	15.89	<0.0001	Significant
X_1X_2	1	8.309	8.309	47.28	<0.0001	Significant
X_1X_3	1	5.941	5.941	33.81	<0.0001	Significant
X_1X_4	1	0.263	0.263	1.49	0.240	Not significant
X_2X_3	1	0.043	0.043	0.25	0.628	Not significant
X_2X_4	1	1.238	1.238	7.04	0.018	Significant
X_3X_4	1	0.956	0.956	5.44	0.034	Significant
Residual	15	2.636	0.176			
Lack of Fit (LOF)	10	2.171	0.217	2.33	0.181	Not significant
Pure Error	5	0.465	0.093			
Total	29	641.878				

R^2 = 0.9953, adjusted R^2 = 0.9929, R^2 for prediction = 0.9842

Prediction of FAME content by RSM

FAME content in the final biodiesel product is strongly influenced by four operational variables. From Table 3, the initial regression model is shown in Eq. (17):

$$Y = -42.86 + 16.68X_1 + 64.68X_2 + 1.036X_3 + 0.466X_4 - 0.573X_1^2 - 15.64X_2^2 - 0.00117X_3^2 + 0.000270X_4^2 - 2.402X_1X_2 - 0.08125X_1X_3 - 0.00854X_1X_4 - 0.05187X_2X_3 - 0.139X_2X_4 - 0.00489X_3X_4 \quad (17)$$

The fit of the designed model is checked as per F-value, P-value, lack of fit error (LOF), R^2, adjusted R^2 and R^2 for prediction [8,17,18]. The model's F-value 259.84 and the very low P-value (<0.0001) indicates that the corresponding model is noteworthy (Table 3). The LOF of 0.181 (much higher 0.05) implies that LOF is meaningful relative to the pure error [8]. Pointless LOF is good for predicted model. Additionally, in the evaluation the importance of the suggested model, large differences between R^2, adjusted R^2 and predicted R^2 also demonstrate the significance of the model [17,18]. These coefficients (R^2, adjusted R^2, predicted R^2) are very high (0.9953, 0.9929 and 0.9824, respectively) and prove the worth of the model (Table 4).

Correlation is linear and most of experimental points are located on the 45-degree line as depicted in Figure 2. Therefore, the suggested model is precise description of the process.

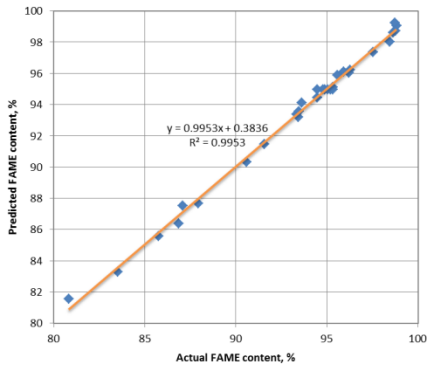

Figure 2 Comparison of predicted and experimental FAME content.

The final practical model based on the coded factor, ANOVA data and by eliminating the irrelevant model terms is given in Eq. (18):

$$Y = -34.86 + 16.25X_1 + 61.82X_2 + 0.86X_3 + 0.442X_4 - 0.573X_1^2 - 15.63X_2^2 - 2.402X_1X_2 - 0.08125X_1X_3 - 0.139X_2X_4 - 0.00489X_3X_4 \quad (18)$$

ANN modeling

Development of ANN

The FAME content is predicted based on the ANN with LM algorithm includes four input layer neurons and one output. The ANN model depends on the decisive optimum neuron numbers.

The influence of the sum of neurons in the hidden layer is investigated in order to determine the optimum neurons. This process consists of examine the chain of various neurons until the MSE are the lowest value. The number of neurons is varied from 1 to 25. Results for the ANN model are shown in Figure 3. The optimum sum of neurons for the ANN model is 3 neurons with the minimum MSE of 0.00097664 (Figure 4). Initially high MSE reduced rapidly to a smallest value. The MSE depends on the number of neurons of ANN model and are shown in Table A (Appendix). Table B in Appendix also shows weights, bias and the transfer functions for the ANN model with 2 layers and 3 optimum neurons.

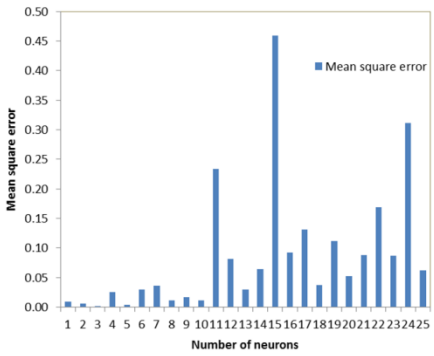

Figure 3 Validation MSE response for the ANN model.

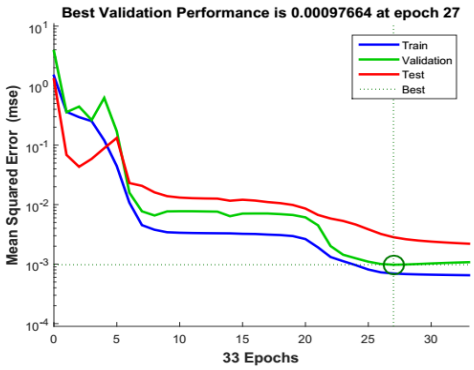

Figure 4 MSE values for training, validation and testing of the developed ANN model.

Prediction of FAME content by ANN

Fig. 5 compares the predicted and actual FAME content for training (R = 0.99881), validation (R = 0.9986), testing (R = 0.95991) and the overall regression (R = 0.99795) of the developed ANN model as per the 4-3-1 configuration (4 input variables, 3 neurons in hidden layer and 1 output variable). Most data points are distributed on the 45-degree line which shows a very good mutual relationship between the experimental data and predicted outputs. Results also confirm the developed ANN model is absolutely agreed to predict the output values of the validation and testing data.

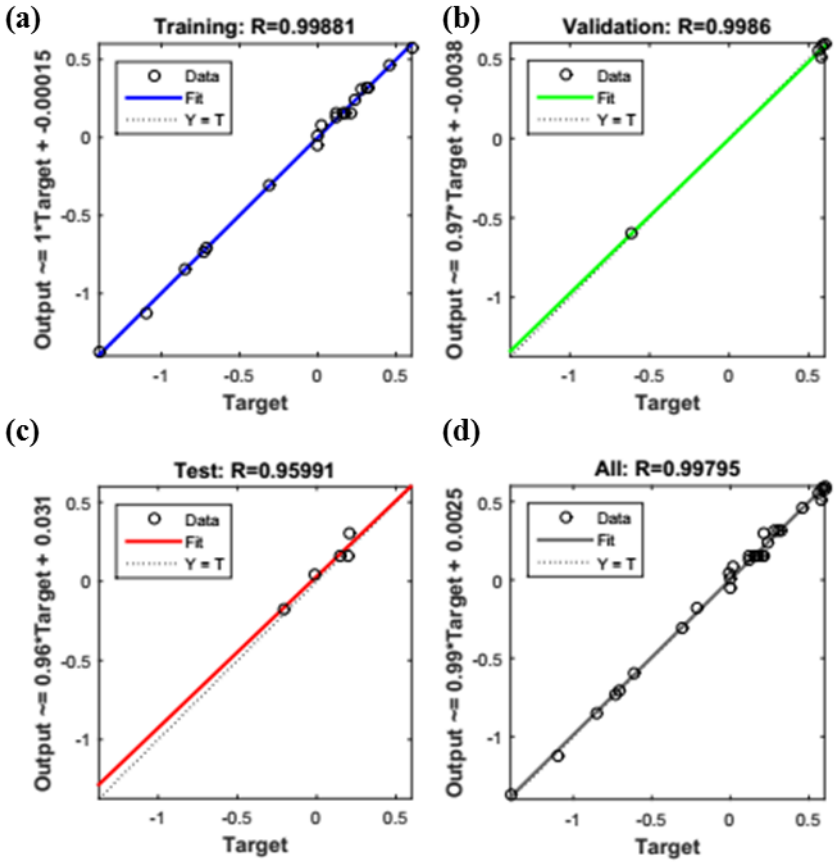

Figure 5 Comparisons of the predicted and experimental FAME content (output) for training (a), validation (b), testing (c) and the overall regression (d) for 3 neurons.

Predictive capability of RSM and ANN models

The capability of the developed RSM and ANN models in prediction of the FAME content in biodiesel is evaluated in terms of their R, R^2, adjusted R^2, mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), standard error of prediction (SEP) and mean relative percent deviation (MRPD). These results are presented in Table 5. If the value of the R is close to 1 then there is a good correlation between experimental and predicted values. The two models have very high values of R^2, demonstrate the authentic suitability of these models [24]. The adjusted R^2 is used in testing over fitting of R^2. These are also high for the two models which, the models. Moreover, the RMSE – the square root of the MSE – is also determined for both models. MSE value from ANN model (0.0010) is much lower as compared to RSM model (0.0879). The similar difference is also obtained for RMSE, with 0.0313 and 0.2964, respectively. Results confirm the ANN model is better than the RSM model (Table 5). MAE, SEP and MRPD check the significance and accuracy of the models [23,24,27]. The lower values of these statistical parameters, better the performance of the model.

Table 5. Performance evaluation of RSM, ANN models.

Parameter	RSM	ANN
R	0.9979	0.9980
R^2	0.9953	0.9958
Adjusted R^2	0.9921	0.9903
MSE	0.0879	0.0010
RMSE	0.2964	0.0313
MAE	0.2448	0.0233
SEP (%)	0.3173	0.0335
MRPD (%)	0.2667	0.2232

Several studies have shown that ANN is better than RSM model in prediction capability [13,23,24,27-29]. Further, these results have not proven the difference between MSE and RMSE for the RSM and ANN models. The present study has passed this difficulty and contributes a fully confirmation about the effectiveness of the developed ANN and compared to the RSM model.

Optimization of FAME content by the RSM and ANN models

Actual FAME content under the experimental conditions are between 80% to 100% (Fig. 2). In order to evaluate the optimization capability of the RSM and ANN models, the FAME content of 96.5% and 98% were chosen as a desired target of base-catalyzed methanolysis. The optimum condition for temperature and reaction time are same for both models with regard to same desired target (Table 4). In a contrary, the optimum molar ratio and catalyst content from these models have a remarkable difference. The molar ratio and catalyst content are the two most important factors for base-catalyzed methanolysis in the present study as per ANOVA results (Table 4). Therefore, the evaluation for the RSM and ANN models is as per these two important factors. The values of molar ratio and catalyst content required for base-catalyzed FAME synthesis by ANN model are lower in comparison with RSM model (Table 6). Thus, ANN model is better in prediction capability as compare to RSM model.

Table 6. Optimization conditions and model validation.

Model	RSM		ANN	
	96.5 % FAME	98% FAME	96.5 % FAME	98% FAME
MeOH/RPO molar ratio	6.15	7.57	5.49	5.88
CH_3ONa content (wt%)	1.01	1.20	0.87	0.89
Reaction temperature (^oC)	55	55	54.7	54.9
Reaction time (min)	50	50	49.8	50

To check the validation of RSM and ANN models, triplicate experiments were repeated under the predicted optimum conditions from each model with the same target of the ester content (98%). Table 7 shows the experimental value of the ester content in final biodiesel production. The experimental ester content value obtained from ANN is in close agreement with the suggested target. As a result, ANN proved to be more effective than RSM in optimizing the biodiesel production in the present study.

Table 7. The validation of RSM and ANN models with the ester content of 98% as a target.

Model	Molar ratio (mol/mol)	Catalyst content (wt%)	Temperature (^oC)	Time (min)	Ester content from experiment (%)
RSM	7.57	1.20	55	50	99.64
ANN	5.88	0.89	54.9	50	97.95

CONCLUSION

RSM and ANN models were developed and compared for their predictive and generalization abilities in the methanolysis process of palm oil catalyzed by sodium methoxide in the present study. Following conclusions are drawn:

1. The predictive capability of the two models for sodium methoxide-catalyzed methanolysis was compared using the same experimental conditions from the CCD.
2. High values of R, R^2, predicted R^2 (> 0.99) clearly indicates high accuracy of both RSM and ANN models.
3. Both models have proven the important role of the molar ratio and catalyst content for base-catalyzed methanolysis.
4. Earlier, the difference of the predictive capability between RSM and ANN is only based on R^2, adjusted R^2 and R^2 for prediction, this may be the first study in evaluating in terms of MSE, RMSE, MAE, SEP, and MRPD.
5. Lower values of the ANN models demonstrated that the ANN model is a better choice compared to the RSM model by paying attention to parameters (MSE, RMSE, MAE, SEP, MRPD) and recommended for similar studies.

ACKNOWLEDGEMENTS

This research work was supported by Graduate School of Prince of Songkla University under Thailand's Education Hub for ASEAN Countries (Contract No.: TEH-AC 047/2014).

APPENDIX A

Table A. Mean square error (MSE) depends on the number of neurons of ANN model.

Number of neurons	MSE	Number of neurons	MSE	Number of neurons	MSE
1	0.0092157	10	0.0112460	18	0.0367090
2	0.0059002	11	0.2336700	19	0.1121600
3	0.0009766	12	0.0815890	20	0.0527680
4	0.0251790	13	0.0297210	21	0.0875460
5	0.0034538	14	0.0638450	22	0.1693700
6	0.0294600	15	0.4596200	23	0.0871460
7	0.0361070	16	0.0922340	24	0.3110600
8	0.0113680	17	0.1307600	25	0.0621300
9	0.0163470				

APPENDIX B

Table B. Weights, bias and transfer function of ANN model (optimum neurons: 3)

Weights 1st layer	Weight size for 1st layer	Bias	Size [3x1]	Transfer function
$w_{1,1}^1$	-1.3116	B_1^1	2.8172	$f_1^1 = tanh$
$w_{1,2}^1$	0.84862	B_2^1	1.0268	$f_2^1 = tanh$
$w_{1,3}^1$	0.45611	B_2^1	-3.4715	$f_3^1 = tanh$
$w_{2,1}^1$	2.1058			
$w_{2,2}^1$	0.39497			
$w_{2,3}^1$	-0.97206			
$w_{3,1}^1$	0.23611			
$w_{3,2}^1$	0.173670			
$w_{3,3}^1$	-1.5561			
$w_{4,1}^1$	-0.45005			
$w_{4,2}^1$	0.069986			
$w_{4,3}^1$	1.7798			
Weights 2nd layer	Weight size for 2ndlayer	Bias	Size [1x1]	Transfer function
w_1^2	0.88161	B^2	-0.77158	$f^2 = 1$
w_2^2	2.5727			
w_3^2	1.5377			

REFERENCES

[1] B.T. Fichman, Monthly Energy Review, U.S. Energy Information Administration. March 2016, 2016. https://www.eia.gov/totalenergy/data/monthly/pdf/mer.pdf.

[2] P. Verma, M.P. Sharma, G. Dwivedi, Impact of alcohol on biodiesel production and properties, Renewable and Sustainable Energy Reviews, 56 (2016) 319–333.

[3] Y.C. Sharma, B. Singh, S.N. Upadhyay, Advancements in development and characterization of biodiesel: A review, Fuel, 87 (2008) 2355–2373.

[4] A. Srivastava, R. Prasad, Triglycerides-based diesel fuels, Renewable & Sustainable Energy Reviews, 4 (2000) 111–133.

[5] G. Knothe, J. Van Gerpen, The Biodiesel Handbook Editors, 2005.

[6] A. Zarei, N.A.S. Amin, A. Talebian-Kiakalaieh, N.A.M. Zain, Immobilized lipase-catalyzed transesterification of Jatropha curcas oil: Optimization and modeling, Journal of the Taiwan Institute of Chemical Engineers, 45 (2014) 444–451.

[7] A. Talebian-Kiakalaieh, N.A.S. Amin, A. Zarei, I. Noshadi, Transesterification of waste cooking oil by heteropoly acid (HPA) catalyst: Optimization and kinetic model, Applied Energy. 102 (2013) 283–292.

[8] D.N. Thoai, C. Tongurai, K. Prasertsit, S. Photaworn, Optimization of the first step transesterification of refined palm oil using solid sodium methoxide catalyst, Vung Tau, Vietnam, 3 (2016) 137–144.

[9] G. Vicente, M. Martínez, J. Aracil, Integrated biodiesel production: A comparison of different homogeneous catalysts systems, Bioresource Technology, 92 (2004) 297–305.

[10] A. V. Veličković, O.S. Stamenković, Z.B. Todorović, V.B. Veljković, Application of the full factorial design to optimization of base-catalyzed sunflower oil ethanolysis, Fuel, 104 (2013) 433–442.

[11] E. Betiku, O.R. Omilakin, S.O. Ajala, A.A. Okeleye, A.E. Taiwo, B.O. Solomon, Mathematical modeling and process parameters optimization studies by artificial neural network and response surface methodology: A case of non-edible neem (Azadirachta indica) seed oil biodiesel synthesis, Energy, 72 (2014) 266–273.

[12] D.N. Thoai, C. Tongurai, K. Prasertsit, A. Kumar, A novel two-step transesterification process catalyzed by homogeneous base catalyst in the first step and heterogeneous acid catalyst in the second step, Fuel Processing Technology, 168 (2017) 97–104.

[13] O.S. Stamenković, K. Rajković, A. V. Veličković, P.S. Milić, V.B. Veljković, Optimization of base-catalyzed ethanolysis of sunflower oil by regression and artificial neural network models, Fuel Processing Technology, 114 (2013) 101–108.

[14] K.M. Rajković, J.M. Avramović, P.S. Milić, O.S. Stamenković, V.B. Veljković, Optimization of ultrasound-assisted base-catalyzed methanolysis of sunflower oil using response surface and artifical neural network methodologies, Chemical Engineering Journal, 215–216 (2013) 82–89.

[15] J.M. Avramović, O.S. Stamenković, Z.B. Todorović, M.L. Lazić, V.B. Veljković, The optimization of the ultrasound-assisted base-catalyzed sunflower oil methanolysis by a full factorial design, Fuel Processing Technology, 91 (2010) 1551–1557.

[16] G.R. Moradi, S. Dehghani, F. Khosravian, A. Arjmandzadeh, The optimized operational conditions for biodiesel production from soybean oil and application of artificial neural networks for estimation of the biodiesel yield, Renewable Energy, 50 (2013) 915–920.

[17] R.H. Myers, D.C. Montgomery, C.M. Anderson-Cook, Response surface methodology: Process and Product optimization using designed experiments, 3rd ed., John Wiley and Sons (2009).

[18] D. Montgomery C., Design and Analysis of Experiments, 5th ed. John Wiley and Sons (2001).

[19] M. Charoenchaitrakool, J. Thienmethangkoon, Statistical optimization for biodiesel production from waste frying oil through two-step catalyzed process, Fuel Processing Technology, 92 (2011) 112–118.

[20] N. Tehlah, P. Kaewpradit, I.M. Mujtaba, Artificial neural network based modeling and optimization of refined palm oil process, Neurocomputing, 216 (2016) 489–501.

[21] P. Suksri, Y. Moriizumi, H. Hondo, Y. Wake, An introduction of biodiesel to Thai economy, (2008) 1–24.

[22] D.N. Thoai, S. Photaworn, A. Kumal, K. Prasersit, C. Tongurai, A novel chemical method for determining ester content in biodiesel, Energy Procedia, 138 (2017) 536–543.

[23] E. Betiku, V.O. Odude, N.B. Ishola, A. Bamimore, A.S. Osunleke, A.A. Okeleye, Predictive capability evaluation of RSM, ANFIS and ANN: A case of reduction of high free fatty acid of palm kernel oil via esterification process, Energy Conversion and Management, 124 (2016) 219–230.

[24] A. Sarve, S.S. Sonawane, M.N. Varma, Ultrasound assisted biodiesel production from sesame (Sesamum indicum L.) oil using barium hydroxide as a heterogeneous catalyst: Comparative assessment of prediction abilities between response surface methodology (RSM) and artificial neural network (ANN), Ultrasonics Sonochemistry, 26 (2015) 218–228.

[25] D.D. Steppan, J. Werner, R.P. Yeater, Essential regression and experimental design for chemists and engineers, (1998).

[26] N.N. Mahamuni, Y.G. Adewuyi, Optimization of the synthesis of biodiesel via ultrasound-enhanced base-catalyzed transesterification of soybean oil using a multifrequency ultrasonic reactor, Energy and Fuels, 23 (2009) 2757–2766.

[27] J.M. Avramović, A. V. Veličković, O.S. Stamenković, K.M. Rajković, P.S. Milić, V.B. Veljković, Optimization of sunflower oil ethanolysis catalyzed by calcium oxide: RSM versus ANN-GA, Energy Conversion and Management, 105 (2015) 1149–1156.

[28] E. Betiku, S.S. Okunsolawo, S.O. Ajala, O.S. Odedele, Performance evaluation of artificial neural network coupled with generic algorithm and response surface methodology in modeling and optimization of biodiesel production process parameters from shea tree (Vitellaria paradoxa) nut butter, Renewable Energy, 76 (2015) 408–417.

[29] E. Betiku, S.O. Ajala, Modeling and optimization of Thevetia peruviana (yellow oleander) oil biodiesel synthesis via Musa paradisiacal (plantain) peels as heterogeneous base catalyst: A case of artificial neural network vs. response surface methodology, Industrial Crops and Products, 53 (2014) 314–322.