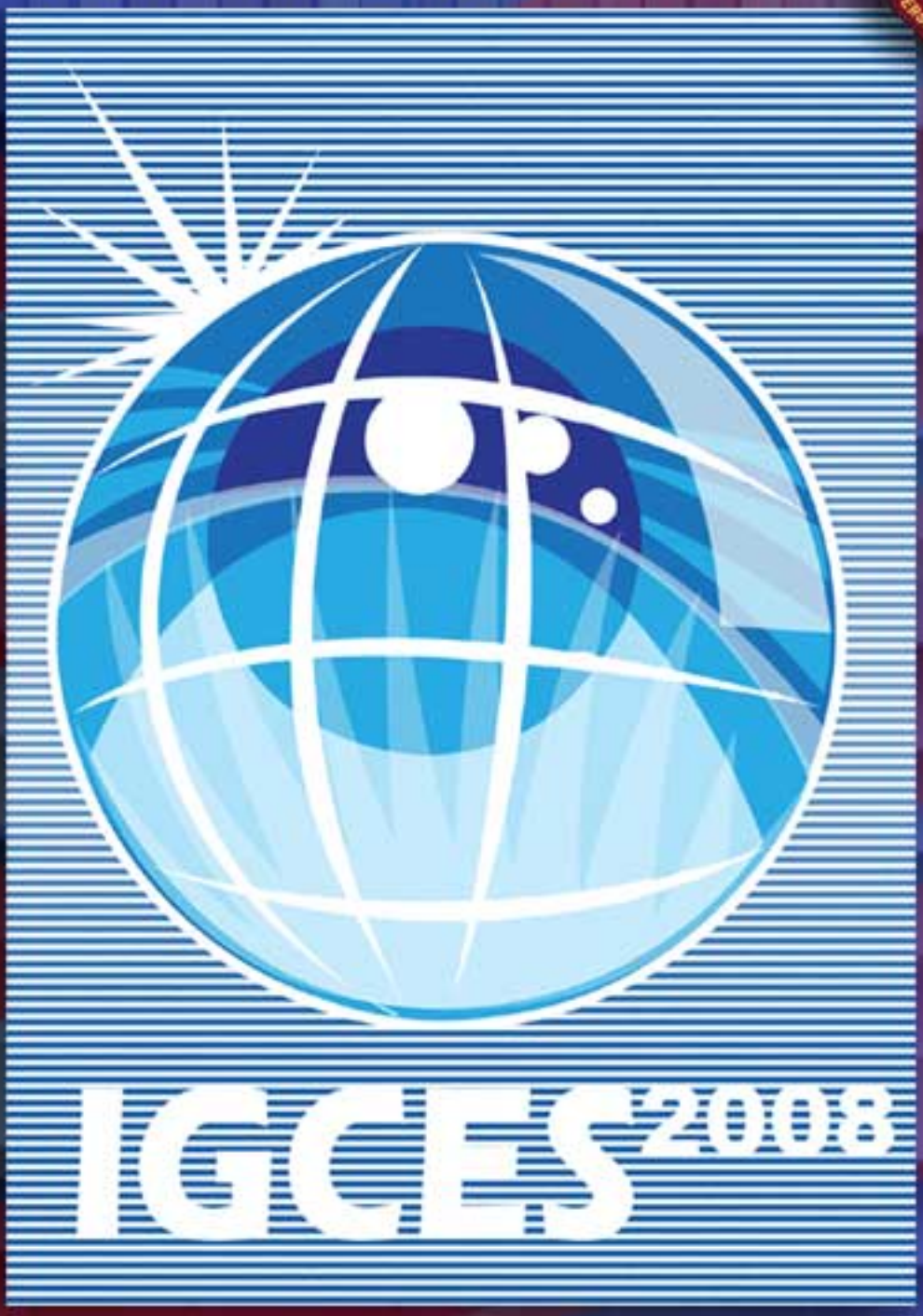
SPS-PGSS
UNIVERSITI TEKNOLOGI MALAYSIA



e-proceeding

ISSN 1823-3287

ANN Models for Determining Bio-Gasoline's Octane Number based on the Input Data Conditions

Abdul Wahid and Bambang Heru Susanto
Chemical Process System Laboratory, Dept. of Chemical Engineering, University of Indonesia
Email: wahid@che.ui.edu and bambanghs@che.ui.edu

Abstract

In this research, there are three models of Artificial Neural Network (ANN), multi-layer feed-forward (MLFF), generalized regression neural network (GRNN), and radial basis (RB), are used for the prediction of the octane number of bio-gasoline. The models provide a solution for any researcher on bio-gasoline in determining the bio-gasoline octane number easier and accurate.. ANN learns a relationship patterns of number of input and output data through training process, then predicts the output from a given new input. This research uses 49 data of bio-gasoline physical properties (density and 50% distillation temperature or TT50) as inputs and bio-gasoline octane numbers as outputs. The result shows that no model is generally proper for any input condition. However the models can be used for determining bio-gasoline's octane number according to the input data condition in each.

1. Introduction

Due to the shortage of petroleum products and its increasing cost, researchers are trying to develop alternative fuels especially for the full or partial replacement of the gasoline fuel [1]. In recent years, bio-gasoline from CPO has received considerable attention, both as a possible renewable alternative fuel and as an additive to the existing petroleum-based fuels.

Bio-gasoline exhibits several merits when compared to that of the existing petroleum fuels. It has fuel properties similar to that of gasoline and is produced easily and renewably from the crops. However, the constraint that often appears in the research on bio-gasoline is determining the octane number (ON) for predicting the quality of the bio-gasoline. It is not always possible to conduct engine test to determine ON because of the cost of the reference fuels and the more effort required. Hence, there have been many attempts either to calculate the ON itself or to derive an alternative parameter that could provide an indication of the ignition quality.

Determination of the octane number of biogasoline which is generally used with four methods: FT-IR method, ON-Tester equipment, ASTM D 2699 method [], and ON conversion from cetane index[3,4]. In ON-Tester equipment, ON is predicted by directly contacting sensor equipment to the bio-gasoline. In ASTM 2699 method, ON is predicted by VCRs (variable compression ratio).

The application of ANN in engineering field is interesting and increasing one. In this study, an attempt made to predict the ON of bio-gasoline using ANN modeling. They operate like a 'black box' model and do not require detailed information about the system being tested. Instead, they learn the relationship between the input parameters and the controlled and uncontrolled variables by studying previously recorded data, similar to the way a non-linear regression might perform.

Another advantage of using ANN is their ability to handle large and complex systems with many interrelated parameters. Moreover, it predicts quite well even if the sample size is very small

In this study, the various ANN models are tried for the prediction of the ON of the biodiesel. The comparison of predicted ON with actual ON is described in the following sections.

2. ANN Models for ON Prediction

In the previous study an ANN model is developed by Meusinger [5] to predict the ON of the fuel based on the chemical structure (fatty acid composition). He used the MLR (multi-layer regression). Murty [6] also use the MLR model for the same purpose.

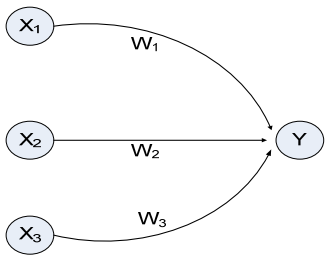


Figure 1. A simple artificial neural network

Figure 1 shows a simple ANN [7], with 3 input neurons (X1, X2, X3), 1 output neurons (Y) and the

relationship between the weights (W, W2, W3). Y receive input from the X1, X2 and X3 with the weight of the relationship is W1, W2 and W3, respectively. The summation of the three impulse neurons is:

$$net = x_1 w_1 + x_2 w_2 + x_3 w_3 \qquad (1)$$

The amount of impulse received by Y is a function of activation :

$$y = f(net) \qquad (2)$$

When the value of the activation function is quite strong, then the signal will be forwarded. Value of the activation function (network model output) can also be used as a basis for changing the weight.

In this study, the input data are the physical properties of biogasoline namely density and temperature on 50% destillation or TT50, while the output is octane number of biogasoline [8]. It tried to using new data for a octane number above 92. The range of data is 0.73 - 0.8865 kg/m^3 for density, 217 - 320^{o}C for TT50, and 72.53 to 109.07 for octane number.

There are three ANN models using in this study. These are multi-layer feed-forward (MLFF) which comes from the previous development of back-propagation model, radial basis (RB), and generalized regression neural network (GRNN).

The MLFF model has one layer with 2 input neurons (X1 and X2), one hidden layer with 4 neurons and one layer of output (Y). Network architecture produced the predicted optimum target [8]. The RB model has "sum-squared error goal" of 2 and "spreads meant" of 0.01 [8]. Meanwhile, in the GRNN model, which is the development of RB model, used default functions.

To determine which model is the best for all conditions, is conducted three following types of testing:

- using validation data in the range of training data (density and the TT-50)
- using validation data LESS THEN OUT of range of training data (density and the TT-50)
- using data validation GREATER THEN OUT of range of training data (density and the TT-50)

3. Results and Discussions

3.1 Tests using validation data in the range of training data (density and the TT-50)

In this test, the properties of bio-gasoline datas (density and the TT-50) should be in the range of data input in the training data. In this case the problem can round into :

- densities vary to close the value of the smallest, the middle, and the highest one.
- Temperature TT-50 varies to close the value of the smallest, in the middle, and highest one

The simulation results in Table 1 show that the GRNN model is the best in determining the ON when the density inputs are in the range. The MLFF model has smaller error (5.4%) than the RB model (6.2%). RB model also produces the same value for all variations of the density inputs. This is caused by the tendency of the RB model to choose ON slightly smaller than the octane number data in the database, if variations of input datas produces a small ON.

This mean that the GRNN model should be used if the data density is in the range of available data.

Table 1 Variations of error of the density and TT50 in the range of training data

No	Density Variations			TT50 Variations		
	GRNN	MLFF	RB	GRNN	MLFF	RB
1	10.2%	3.8%	20.6%	0.9%	1.8%	32.6%
2	4.4%	9.9%	4.0%	1.1%	10.5%	1.8%
3	3.4%	11.1%	3.1%	0.4%	1.7%	2.9%
4	0.2%	1.4%	2.1%	0.2%	0.2%	2.2%
5	0.5%	1.0%	1.1%	0.1%	1.2%	1.2%
	3.8%	5.4%	6.2%	0.5%	3.1%	8.1%

The same results also appear on the temperature variations of TT-50 in the range of data in the software. The GRNN model produces the best value, with the average error of 0.5%. Meanwhile, the MLFF model is in the second rank with an average error of 3.1%, and the RB model provide the biggest error, at 8.1%.

In this case the RB model also produces ON with the same value for all variations of TT-50. As mentioned before, if the input data produse small ON, the RB model will select only a number of smaller ON data in the database.

3.2 Tests using validation data LESS THEN OUT of range of training data (density and the TT-50)

In this test, the properties of the biogasoline datas (density and the TT-50) is smaller at outside the range of the input data in the database. In this case the problem can round into :

- smaller density is outside the range of data density in the database

- Temperature TT-50 is smaller at outside the range of TT-50 in the database.

Table 2 Errors as variations of density and TT50 smaller at outside the range of data input

No	Density Variations			TT50 Variations		
	GRNN	MLFF	RB	GRNN	MLFF	RB
1	17.6%	5.7%	18.0%	100.0%	5.9%	33.8%
2	2.3%	2.3%	18.0%	100.0%	6.3%	33.5%
3	2.3%	2.3%	18.0%	100.0%	6.8%	32.9%
4	2.3%	3.1%	18.0%	100.0%	8.9%	31.3%
5	0.5%	1.9%	21.6%	0.9%	10.0%	30.5%
	5.0%	3.0%	18.7%	80.2%	7.6%	32.4%

The simulation results are shown by Table 2. From the table 2, the MLFF model is better than other models. The error of MLFF model is the smallest (3%), followed by the GRNN model (5%), and the RB model generates the biggest error, at 18.7%.

In this case, the RB model can not determine the large ON. ON produced by the RB model are have the same value for all variations of density, at 72.15. This ON is smaller than ON data in the database, at 72.53.

It means, if someone has data where the TT-50 data from bio-gasoline is at smaller outside TT-50 in the database, then it should choose the MLFF model for determining ON, because of the level of average error is the smallest (7.6%).

The GRNN model also can not determine ON at the four data input. The results of determining ON is 0 (zero). The neurons owned by GRNN model could not be able to determine the data input at outside the training data. Actually, when the TT-50 approaching the smallest temperature data owned by the training data (in the database), at 249^{o}C, the GRNN model could be able to determine ON accurately. The level of error is only 0.9%. But then, the TT-50 is farther from its smallest value will produce very bad result (fail).

The RB model, as mentioned before, still has produce the same value of ON for all the variations. It can not be able to determine the high ON. While the GRNN model generates ON of 0 (zero), the RB model preferably choose the smallest ON in the training data (in the databae) when failed to determine the appropiate ON.

These results also prove that the neurons developed by the MLFF model is able to capture the unknown data input (in the training data). Although the results still have errors high enough, but will be more reliable than the others.

4.3 Tests using data validation GREATER THEN OUT of range of training data (density and the TT-50)

In this test, if the properties of the biogasoline data (density and the TT-50) are higher at outside the range of input data in the database. In this case the problem can also round into :

- density is higher at outside the range of data density in the training data
- TT-50 is greater at outside the range of the TT-50 in the training data

Simulation results are indicated by Table 3. From the table, all of the ANN model have relatively small error, but the GRNN model is the best due to has smaller an average error than the others. At the high density, ON is small, in such a way that the RB model has a relatively small error. However, the results of its ON are all the same, at 72.15. The highest error occurred at the smallest TT-50 (307^{o}C). It means that the model can not be able to predict ON of bio-gasoline at small value of TT-50.

The MLFF model has the highest error in the highst density (0.8865 kg/m3) and the highest TT-50 (314 ^{o}C). This phenomenon did not happen in previous cases. It means that the MLFF model is not good enough for input data higher than those in the database.

Table 3 Variations of Error at input density at the higher of outside the training data

No	Density Variations			TT50 Variations		
	GRNN	MLFF	RB	GRNN	MLFF	RB
1	2.4%	4.0%	1.1%	100.0%	13.0%	2.7%
2	2.3%	3.6%	1.2%	100.0%	10.2%	0.7%
3	0.7%	2.1%	2.0%	0.8%	5.1%	0.5%
4	0.5%	1.4%	2.3%	0.7%	4.6%	0.4%
5	0.6%	1.9%	1.8%	1.5%	8.3%	1.9%
	1.3%	2.6%	1.7%	40.6%	8.3%	1.2%

The RB model is the best for the TT-50 condition at higher outside of the training data. The level of average error of the RB model is the smallest (1.2%). Followed by the MLFF model at 8.3%, and the GRNN model has the highest errors on average, at 40.6%.

The GRNN model can not be able to determine ON at the condition of the highest TT-50 (320^{o}C). This occurs when both density are 0.8225 kg/m^{3} and

0.8339 kg/m^3. Other results actually are very good with an average error below 1%, but failed in the two conditions, with the result that the average error of all five conditions become very high.

As in the previous explanation, the MLFF model has the largest error in the higher temperature TT-50. Meanwhile, the RB model is to be the best for data input at the higher of the training data. Nevertheless, the results remain the same for all conditions.

4.4 Test using ”New Data”

"New Data" are data that is developed from the density of bio-gasoline which has the same ON, but the TT-50 are different. Table 4 shows the results of producing new data from old data.

These new data are in the range of training data. Based on the results of which have been described previously, the GRNN model is the best for this condition. Table 5 shows evidence that GRNN is the best model to predict when ON of data inputs are in the range of training data. No errors made by the GRNN models, or in other words, the rate of error is 0 (zero). The MLFF model is the secon best one with the average error of 1.7%. While the RB model is slightly above the error of the MLFF model, at 1.8%.

Tabel 4 ”New Data” from old data

Data Asal			*Data Baru*		
Densitas	TT 50	Bilangan Oktana	Densitas	TT 50	Bilangan Oktana
0.7400	251.6	92.00	0.74	252.5	92
0.7400	249.8	92.00			
0.7400	253.4	92.00			
0.7400	255.2	92.00			
0.7400	264.2	90.00	0.74	263.75	90
0.7400	262.4	90.00			
0.7400	262.4	90.00			
0.7400	266	90.00			
0.7300	276.8	88.00	0.73	271.85	88
0.7300	269.6	88.00			
0.7300	269.6	88.00			
0.7300	271.4	88.00			

Tabel 5 Determination of the ON using the "new data"

Data of Biogasoline			Octane numbers			Errors		
Densitas	TT-50	ON	GRNN	MLFF	RB	GRNN	MLFF	RB
0.74	252.5	92	92	93.957	94.03	0.0%	2.1%	2.2%
0.74	263.75	90	90	89.375	90.362	0.0%	0.7%	0.4%
0.73	271.85	88	88	85.972	85.5	0.0%	2.3%	2.8%
			Average			0.0%	1.7%	1.8%

5. Conclussions

1. None of the three ANN models tested, generally can be used for all the data input
2. There are condition in application of the three ANN models for determining ON of bio-gasoline :
    a. Using the MLFF model if the data of density and temperature TT-50 is smaller at outside the range of training data
    b. Using the RB model if the temperature TT-50 is greater at outside the range of training data
    c. Using the GRNN model if the data density and the TT-50 are in the range of training data, and the density is higher at outside the range of training data as well.

6. References

[1] Kurtubi, "Indonesia: Net Oil Importer", Center for Petroleum and Energy Economics Sudies (CPEES), Jakarta 2004

[2] RISTEK, "The Master Plan of Rusnas Activities in Oil Palm Industry", Jakarta 2004

[3] Nasikin, M., et al. (2004). "The influence of acid strength of acid catalysis reactions that occur in the Triglyceride from palm oil" *Research Report of Department of Gas dan Petrochemical FTUI*

[4] Lob, A., et al. (1995). "Prediction of reformate RON by FT-IR". *Fuel*. 74(2):.227-231

[5] Meusinger, R., R. Moros. (2001). "Determination of octane numbers of gasoline compounds from their chemical structure by ^{13}C NMR spectroscopy and neural networks". *Fuel*. 80: 613- 621

[6] Murty, B. S. N., R. N. Rao. (2004). "Global optimization for prediction of blend composition of gasolines of desired octane number and properties". *Fuel Processing Technology*. 85: 1595– 1602

[7] Siang, Jong Jek. (2005). *Artificial Neural Network and Its Programming using Matlab* – Ed. I., Yogyakarta: Andi

[8] Wahid, A., et al. (2007). "On Determination of Bio-Gasoline Octane Number using Artificial Neural Network". *Proceeding's The 10th "Quality in Research (QIR)" International Conference*, Depok - Indonesia. *4 – 6th Dec 2007: ICT-06*