Prediction of cetane number of biodiesel from its fatty acid ester composition using Artificial Neural Networks

R. Piloto[1], Y. Sanchez[1], L. Goyos[1] and S. Verhelst[2]

[1] Faculty of Mechanical Engineering, Technical University of Havana
Calle 127 s/n, Cujae, Marianao, Habana, 19390, Habana (Cuba)
Phone/Fax number: (+537) 7650976, e-mail: **rpiloto@ceter.cujae.edu.cu**

[2] Department of Flow, Heat and Combustion Mechanics, Faculty of Engineering, Ghent University
Sint-Pietersnieuwstraat 41, 9000 Gent (Belgium)
e-mail: sebastian.verhelst@ugent.be

Abstract. Models for estimation of the cetane number of biodiesel from their methyl ester composition using artificial neural networks were obtained in this work. An experimental data that covers 48 and 15 biodiesels in the modeling step and validation step respectively were taken. The selection of biodiesel samples took into account a wide range of ester compounds with different unsaturation characteristics and number of carbon atoms. A model to predict cetane number using artificial neural networks was obtained with better accuracy than 95 %. The best neural network for predict the cetane number was a backpropagation network (11:4:1) using a Conjugate Gradient Descend algorithm for the second training step and showing 96.3 % of correlation for the validation data and a mean absolute error of 1.5. The proposed network is useful for prediction of the cetane number of biodiesel in a wide range of composition but keeping the percent of total unsaturations lower than 80 %. The use of the artificial neural networks in this case let to study and understands the effect of individual fatty acids in the cetane number. Therefore they can be used for improving some biofuel properties related to the combustion process and its efficiency.

Key words

Cetane number, biodiesel, neural network, fatty acid, ester composition

1. Introduction

Several physical properties of biodiesel fuels depend on their fatty acid ester composition [1]-[3] Also related to the ester composition is the cetane number which is one of the most cited indicators of diesel fuel quality [3]-[5] and strongly influences the ignition delay phenomenon. It is generally dependent on the composition of the fuel and can also influence the engine stability, noise level, and exhaust emissions. While the ignition delay can be influenced by engine type and operation conditions, the cetane number mainly depends on the chemical composition of the fuel.

Equations for predicting the cetane numbers of diesel or biodiesel fuels have been published [4], [6]-[12], correlating this parameter with different input factors or using different mathematical methods. Yang [6] developed multiple linear correlation equations for predicting the Cetane Number (CN) for 12 hydrocarbons in order to compare with a model developed using Artificial Neural Networks (ANNs).

Determination of the CN by an experimental procedure at present is an expensive and time consuming process. Therefore, the obtaining of accurate models to predict the CN of biodiesel from its Fatty Acid Methyl Ester (FAME) composition in a wide range of feedstocks characteristics would be useful for the scientific community.

Most of the models published for cetane number prediction were developed with Multiple Linear Regression (MLR) techniques. That procedure requires the user to specify a priori a mathematical model to fit the data in order to obtain the empirical correlation. An alternative to avoid that problem is the use of artificial neural networks. Unlike the correlation techniques, the neural network can identify and learn the correlative patterns between the input and output data once a training set is provided.

The purpose of this work is to obtain models for the estimation of the cetane number of biodiesel from their FAME composition using ANNs, searching for the best suitable model to predict cetane number in the range of biofuels studied, covering biodiesels from 63 feedstocks.

2. Experimental set-up and procedures

48 different biodiesel fuels (including 10 pure fatty acids) were taken from references as input and output data for the implementation of ANNs for predicting the cetane number. The FAME main composition presented in biodiesel obtained from different feedstocks is covered by ten FAMEs selected [7], [9], [12]-[18]. The input data covers FAME composition and the output covers the cetane number. The validation of the models obtained was done using a separate data set selected from literature reports, which was not used for developing the models. The data selected for validation covers 15 samples.

Different networks were developed using five basic topologies, between (11:3:1) and (11:7:1). The ANNs used were the multilayer Perceptrons, with one hidden layer and between 3-7 units. The inputs of the network were ten, representing the chemical composition of 10 FAMEs and one input representing the total amount of the other FAMEs found in the biodiesel sample. The CN was the unique variable output of the network. A sample of one of tested network topologies is shown in Fig.1.
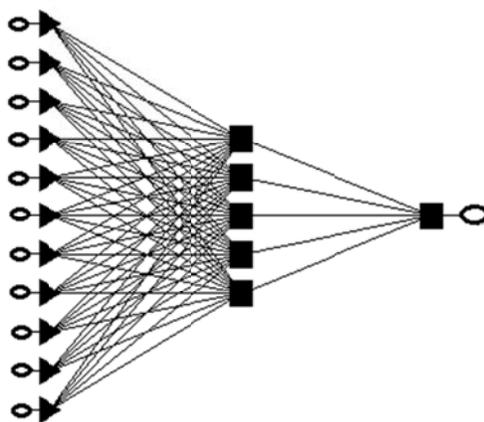
Fig.1 Network (11:5:1) for the prediction of the cetane number

The chemical formula and the structure of the FAMEs on which this research is focused are shown in Table I. The ten FAMEs listed represent the inputs for the CN modelling. The basic structural description for the input FAMEs used in this work (XX:X) covers the information about the number of carbon atoms (XX) and the number on the right (X) represents the number of unsaturations in the molecule.

In the training step two phases were implemented, keeping constant the phase 1 (back propagation) for all the ANNs evaluated. The training was developed for 10000 epochs with a learning rate of 0.01. Linear and logistic functions in the range of 0.9 were used as the output functions in different networks variants.

24 different ANNs were tested for the prediction of the CN using two phases. The phase one was a backpropagation (BP) and the second phase was varied among different possibilities: back propagation, conjugate gradient descend (CGD), Levenberg-Marquardt (LM), quick propagation (QP), quasi-Newton (QN) and Delta-bar-Delta (DBD). The experimental data used for the training step for the ANNs is shown in [19].

3. Results and Discussion

The search for the optimal network for predicting the CN was based in the absolute error as the objective function in the process, but combining it with adequate values of the correlation coefficient found. The results from the whole group of ANNs in order to find the better CN prediction capabilities based on the FAMEs as input parameters are shown in Table II and Table III.

The results shown in Table II are corresponding to the trained ANNs changing the phase two among six algorithms, varying the nodes number between 3 and 7 and using the linear output function.

Table III shows the results for the same method but corresponding to a logistic output function. As is shown in Table II, the prediction is better using the ANN (11:4:1) with a CGD algorithm. This selection is also based on the sum of squares corresponding to the residual values on the cetane number estimation. The lowest sum of squares was also obtained for the CGD with a total sum of 150 for 190 using the QN algorithm. The comparison between both ANNs is shown in Fig.2. The line represents the ideal values to obtain in the fit relating the actual values of cetane number and the predicted values by the ANNs selected. The selection of the best ANN for CN prediction is not based only in this plot, but also in the correlation coefficients obtained and in their behaviour in the validation step related to the networks prediction capability when the validation data is processed.

Table I. - Chemical structure and name of the FAMEs

fatty acid	structure	formula
lauric	12:0	$C_{12}H_{24}O_2$
myristic	14:0	$C_{14}H_{28}O_2$
palmitic	16:0	$C_{16}H_{32}O_2$
palmitoleic	16:1	$C_{16}H_{30}O_2$
stearic	18:0	$C_{18}H_{36}O_2$
oleic	18:1	$C_{18}H_{34}O_2$
linoleic	18:2	$C_{18}H_{32}O_2$
linolenic	18:3	$C_{18}H_{30}O_2$
eicosanoic	20:1	$C_{20}H_{38}O_2$
erucic	22:1	$C_{22}H_{42}O_2$

Table II. - Absolute errors and correlation coefficients (%) for the CN using a linear output function

nodes	BP	CGD	DBD	LM	QN	QP
3	2.7 (94.7)	2.4 (93.5)	2.4 (94.9)	1.9 (94.6)	1.9 (94.5)	2.1 (95.9)
4	2.9 (93.4)	1.6 (96.3)	2.3 (94.3)	2.4 (93.0)	1.5 (95.7)	2.4 (94.6)
5	2.3 (93.9)	2.6 (94.9)	2.0 (95.3)	2.3 (95.2)	2.1 (95.9)	2.1 (95.0)
6	2.3 (94.9)	2.5 (94.3)	2.0 (93.5)	1.9 (95.0)	2.4 (92.4)	1.9 (95.7)
7	2.7 (94.0)	2.3 (94.9)	2.6 (92.6)	2.0 (95.8)	1.8 (95.4)	2.1 (94.3)

In the comparison between models, even when there are some points with the same predicted values for both, a general analysis of outlier points shows that the best algorithm for CN prediction is the CGD.

According to the analysis above exposed, the best network for prediction of the cetane number is the implemented using the topology (11:4:1) of 11 inputs, 1 output variable and four nodes. The response surface obtained for this CGD (11:4:1) network selected for the prediction of the CN as a function of Myristic and Lauric ethyl esters percent is shown in Fig.3. In the same figure, in the right side, the plot of the Palmitoleic percent versus the CN

values is shown. It was the only two-dimensional relationship obtained with changes in behaviour of the CN influence depending of the magnitude (Palmitoleic percent). In the case of the other nine FAMEs, a constant trend was observed in the two-dimensional fits applied.

For the validation of models obtained by ANN, a data set not related to the modelling data was used. The validation data covers 15 samples from other references. The collected data includes the experimental evaluation of FAME composition and CN, covering a wide range of possible values of cetane number (between 41 and 69) taken from experiments using engine tests or an ignition quality tester (IQT). The prediction capability of the selected models for this external data were based in the comparison between the predicted cetane number by the ANN and the experimental value (actual value) trying to find less residuals for the whole tested data. The full data used can be consulted in [19].

Table III. - Absolute errors and correlation coefficients (%) for the CN using a logistic output function

nodes	BP	CGD	DBD	LM	QN	QP
3	2.2 (94.0)	2.1 (94.4)	2.6 (93.5)	2.2 (95.4)	2.4 (94.2)	2.3 (94.8)
4	2.8 (93.0)	2.6 (93.1)	2.4 (92.9)	2.3 (94.8)	2.5 (94.1)	2.4 (94.2)
5	2.2 (93.3)	2.5 (94.8)	2.0 (92.9)	2.4 (93.8)	2.2 (95.7)	2.1 (96.1)
6	2.1 (92.4)	2.3 (94.2)	2.0 (93.3)	2.7 (94.0)	2.0 (94.3)	2.4 (92.6)
7	1.9 (94.8)	2.3 (93.5)	2.2 (91.1)	2.5 (91.0)	2.5 (93.1)	2.3 (91.0)

Lack of accuracy was observed in certain cases of the validation, when the total percent of unsaturations in the FAME composition of the biodiesel reach certain levels.

As is observed in Fig.4, for higher values than 80 % in the total unsaturations, in some cases the uncertainty percent is increased, reaching values higher than 5 %. The zone in Fig.4 that covers the range 45-62 of cetane number corresponds to the common CN value of biodiesel from many feedstocks. As is observed, this is the zone that shows the lowest outlier points in both cases.
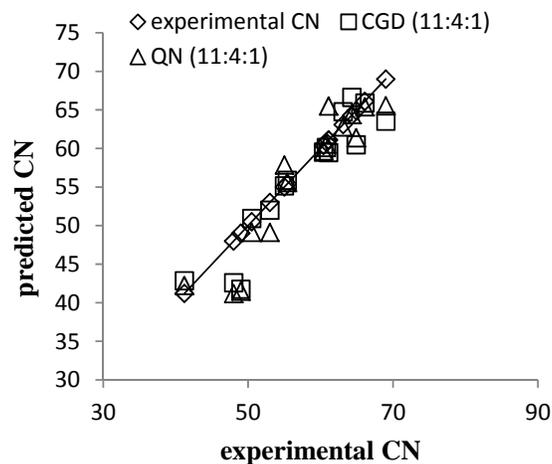
Fig. 2. Comparison between two ANNs for prediction of the CN
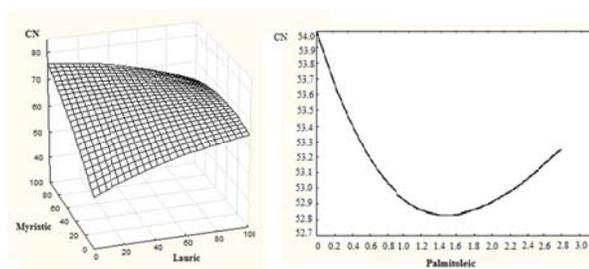
Fig. 3. Response surface and Palmitoleic percent influence in the cetane number estimation using a CGD (11:4:1).

According to Fig.4, even when some values are well predicted by one or both ANNs, the general behaviour is to low accuracies when the unsaturation percent is higher than 80 %, therefore the selected best model can fail in this critical range of total unsaturations. Under this range,

the ANN (11:4:1) with CGD algorithm predicts the cetane number with equal or higher accuracy than 95 %. The model is not recommended for predicting cetane number of pure FAMEs different from the selected for this work.
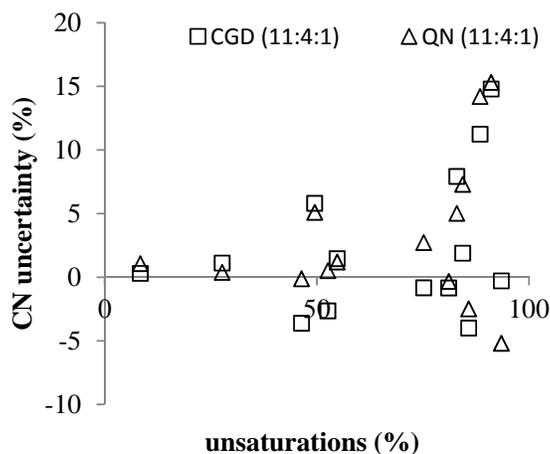
Fig. 4. Uncertainties percent vs. unsaturations percent for the selected ANNs

Yang [6] used a backpropagation neural network model with a training step and a validation step. Ramadhas [9] used an ANN to predict cetane number selecting four types of networks. Basu [20] also used backpropagation, Levenberg, quick propagation and delta-bar-delta as training algorithms in three layer (8:3:1) neural networks. For diesel fuels, he found correlation values for the network in the training step (R = 0.9539). Yang [6] obtained R = 0.8602 for the training step in a three layer backpropagation network with 2.1 for the mean absolute error but his network is only applied to diesel fuels.

Ramadhas [9] used four types of ANNs, not reporting the absolute error of the networks. The author used 5 inputs corresponding to 5 FAME while in this work it is extended to 11 inputs. Ramadhas used a data set that covers biofuels with cetane number between 22.7-75.6, similar to the range applied in the present work. Therefore the prediction capability of their ANNs can only be restricted to the composition of 5 FAMEs that is quite limited due to the amount of feedstocks, different in chemical composition that can be found in these biofuels.

4. Conclusions

A model to predict the cetane number based on the composition of ten FAMEs presented in biodiesel using an artificial neural network was obtained with better accuracy than 95 %. The best neural network for predicting the cetane number was a backpropagation network (11:4:1) using a Conjugate Gradient Descend algorithm for the second training step and showing 96.3 % of correlation for the validation data and a mean absolute error of 1.6. The proposed network is useful for prediction of the cetane number of biodiesel in a wide range of FAMEs composition but keeping the percent of total unsaturations lower than 80 %.

Acknowledgement

References

[1] W. Yuan, A.C. Hansen and Q. Zhang, "Vapor pressure and normal boiling point predictions for pure methyl esters and biodiesel fuels", in Fuel 2005, Vol. 84, pp. 943-950.
[2] K. Wadumesthrige, J.C. Smith, J.R. Wilson, S.O. Salley and K.Y. Simon, "Investigation of the Parameters Affecting the Cetane Number of Biodiesel", in J Am Oil Chem Soc. 2008, Vol. 85, pp. 1073-1081.
[3] L.F. Ramírez-Verduzco, J.E. Rodriguez-Rodriguez and A. Jaramillo-Jacob, "Predicting cetane number, kinematic viscosity, density and higher heating value of biodiesel from its fatty acid methyl ester composition", in Fuel 2012, Vol. 91, pp. 102-111.

[4] G. Knothe, "Dependence of biodiesel fuel properties on the structure of fatty acid alkyl esters", In Fuel Processing Technology 2005, Vol. 86, pp.1059-1070.
[5] G. Knothe, "Cetane numbers of branched and straight-chain fatty esters determined in an ignition quality tester", in Fuel 2003, Vol. 82, pp. 971-975.
[6] H. Yang, Z. Ring, Y. Briker, N. McLean, W. Friesen and C. Fairbridge, "Neural network prediction of cetane number and density of diesel fuel from its chemical composition determined by LC and GC-MS", in Fuel 2002, Vol. 81, pp. 65-74.
[7] A. Gopinath, S. Puhan and G. Nagarajan, "Relating the cetane number of biodiesel fuels to their fatty acid composition: a critical study", in Journal of Automobile Engineering 2009, Vol. 223, pp. 565-583.
[8] K. Krisnangkura, "A simple method for estimation of cetane index of vegetable oil methyl esters", in J Am Oil Chemists' Soc. 1980, Vol. 63, pp. 552-553.
[9] A.S. Ramadhas, S. Jayaraj, C. Muraleedharan and K. Padmakurami, "Artificial neural networks used for the prediction of the cetane number of biodiesel fuel", in Renewable Energy 2006, Vol. 31, pp. 2524-2533.
[10] M. Lapuerta, J. Rodriguez-Fernandez and O. Armas, "Correlation for the estimation of the density of fatty acid esters fuels and its implications. A proposed biodiesel cetane index", in Chemistry and Physics of Lipids 2010, Vol. 163, pp. 720-727.
[11] M. Lapuerta, J. Rodriguez-Fernandez and E. Font de Mora, "Correlation for the estimation of the cetane number of biodiesel fuels and implications on the iodine number", in Energy Policy 2009, Vol. 37, pp. 4337-4344.
[12] A.I. Bamgboye and A.C. Hansen, "Prediction of cetane number of biodiesel fuel from the fatty acid methyl ester (FAME) composition", in Int Agrophysics 2008, Vol. 22, pp. 21-29.
[13] M.J. Ramos, C.M. Fernández, A. Casas, L. Rodríguez and A. Pérez, "Influence of fatty acid composition of raw materials on biodiesel properties" in Bioresource Technology 2009, Vol. 100, pp. 261-268.
[14] A.M. Mohibbe, A. Waris and N.M. Nahar, "Prospects and potential of fatty acid methyl esters of some non-traditional seed oils for use as biodiesel in India", in Biomass & Bioenergy 2005, Vol. 29, pp. 293-302.
[15] D. Tong, C. Hu, K. Jiang and Y. Li, "Cetane Number Prediction of Biodiesel from the Composition of the Fatty Acid Methyl Esters" in J Am Oil Chem Soc. 2011, Vol. 88, pp. 415-423.
[16] J.A. Kinast, "Production of Biodiesels from Multiple Feedstocks and Properties of Biodiesels and Biodiesel/Diesel Blends", in NREL (Ed.). NREL/SR-510-31460, NREL, Illinois (2003).
[17] A.K. Agarwal, "Biofuels (alcohols and biodiesel) applications as fuels for Internal Combustion Engines", in Progress in Energy and Combustion Science 2007, Vol. 33, pp. 233-271.
[18] B.R. Moser, "Biodiesel production, properties, and feedstocks", in In Vitro Cell Dev Biol Plant 2009, Vol. 45, pp. 229-266.
[19] R. Piloto, Y. Sanchez, M. Lapuerta, L. Goyos and S. Verhelst, "Prediction of the cetane number of biodiesel using artificial neural networks and multiple linear regression", in Energy Conversion & Management 2012. Vol. 65, pp. 255-261.
[20] B. Basu, G.S. Kapur, A.S. Sarpal and R. Meusinger, "A Neural Network Approach to the Prediction of Cetane Number of Diesel Fuels Using Nuclear Magnetic Resonance (NMR) Spectroscopy", in Energy & Fuels 2003, Vol. 17, pp. 1570-1575.