Application and Assessment of Artificial Neural Networks for Biodiesel Iodine Value Prediction

Raquel M. de Sousa, Sofiane Labidi, Allan Kardec D. Barros, Alex O. Barradas Filho, Aldalea L. B. Marques

Abstract—Several parameters are established in order to measure biodiesel quality. One of them is the iodine value, which is an important parameter that measures the total unsaturation within a mixture of fatty acids. Limitation of unsaturated fatty acids is necessary since warming of higher quantity of these ones ends in either formation of deposits inside the motor or damage of lubricant. Determination of iodine value by official procedure tends to be very laborious, with high costs and toxicity of the reagents, this study uses artificial neural network (ANN) in order to predict the iodine value property as an alternative to these problems. The methodology of development of networks used 13 esters of fatty acids in the input with convergence algorithms of back propagation of back propagation type were optimized in order to get an architecture of prediction of iodine value. This study allowed us to demonstrate the neural networks’ ability to learn the correlation between biodiesel quality properties, in this case iodine value, and the molecular structures that make it up. The model developed in the study reached a correlation coefficient (R) of 0.99 for both network validation and network simulation, with Levenberg-Marquardt algorithm.

Keywords—Artificial Neural Networks, Biodiesel, Iodine Value, Prediction.

I. INTRODUCTION

The global search for clean energy alternatives, i.e., those generated from renewable resources that do not exacerbate the problems of greenhouse gases in the atmosphere, has made several countries initiate the search for alternatives and to conduct research directed to resource consumption sustainable [1]-[5]. Biodiesel, as a source of biomass derived energy, is able to contribute with sustainability to this world energy demand. Since it is an environmentally friendly [3]. There are many properties used to identify if the biodiesel is a quality fuel. Quality control of biodiesel is important since it can be influenced by many factors, including the quality of the raw material, the composition of the vegetable oil or animal fat fatty acids, production process, the use of other materials and parameters subsequent production [6]. Several international norms establish an appropriate standart, as European Standard (EN) 14214 in Europe, American Society for Testing and Materials (ASTM) D6751 in the USA [3]. In Brazil, resolution nº45 of the National Agency of Petroleum, Natural Gas and Biofuels (ANP) obligatorily controls the quality of biodiesel by employing the rules of the Brazilian Association of Technical Standards (ABNT), American Society for Testing and Materials (ASTM), the International Organization for Standardization (ISO) and the Comité Européen Normalisation (CEN).

Biodiesel has several properties that can be related to the raw material pollutants, the storage conditions and, mainly, the molecular structures that make it up. The inherent properties to the composition, such as cetane number, density, viscosity and iodine value, are directly related to percentage of each alkyl ester that makes up the raw material (vegetable oil or animal fat) [2], [5], [7], [8].

The iodine value (IV) is a measure of total unsaturation within a mixture of fatty acids, based on the formal addition of iodine in the double unions and determined by g of iodine/100g of sample. Diesel’s IV is virtually the same as the value of the oil used in its production [2], [6]. This value influences on several important properties of biodiesel, such as cold filter plugging point and oxidation stability. Oils and biodiesel with higher value of iodine tend to be less stable and more sensitive to oxidation due to the high degree of unsaturation [4], [7]. Use of IV is justified since this parameter indicates a tendency for both oil and fat to oxidation, in addition to indicating a tendency for both oil and fat to polymerization and formation of deposits inside the motor [6].

Determination of that property is carried out through official methods developed by the American Oil Chemists’ Society (AOCS) and the Association of Analytical Communities, although these methods tend to be laborious, long and involve several steps, using various solvents and toxic chemical products [7], [9]. Several free and non-invasive methods were developed in order to determine or predict the iodine value [4], [7], [9], [10]. Adewale et al. [7] and Baptista et al. [4] determined the iodine value by means of near-infrared spectroscopy techniques (FT-NIR). Nevertheless, these techniques can take a long time and be uncomfortable for the company. ANNs are presented here as a non-linear alternative for predicting quality parameters, when it comes to lower cost and less time.

There are several researches in relation to quality control measures of fuel and biofuel, by using ANNs as non-linear tool to predict. The study [5] used the neural network to determine the cetane number, which is one of the indicators of biodiesel quality related to the property of the esters. Cay et al. [11] used neural networks to predict the performance of motor and the exhaust gas emission for methanol and gasoline. Saldana et al. [12] used learning methods of machine that include support vector machines and artificial neural network in order to predict the density and viscosity of biofuel compounds. Jahirul et al. [1] and X. Meng et al. [3] used...
ANNs to predict the kinematic viscosity over influence of temperature. Besides there are studies such as [13] that use ANNs to predict the emission and performance of biodiesel. This work presents a method based on ANNs to predict the iodine value through training and testing of several algorithms of weight correction derived from backpropagation. The study is organized as follows: In Section II, we show the experimental data that was used and the development methodology of ANNs in the work. In Section III, the results and discussions and, finally, in section IV we show the final remarks.

II. METHOD

A. Experimental Data

The data used for training, validation and testing of the network were collected from the literature [14]-[22] and the other one was ceded by the Laboratory of Analysis and Research on Petroleum Analytical Chemistry (LAPQAP/UFMA) that have related to the composition of samples and the physicochemical analysis. These data have 98 samples with 13 fatty acids esters of input, and in output they have the quality parameter iodine value. The distribution of each type of fatty acids ester in the biodiesel samples collected and the predominance of esters, such as oleic acid (C18:1), linoleic acid (C18:2), lauric acid (C12:0) and palmitic acid (C16:0) can be verified from Fig. 1. The description for the FAMEs in this study (XX: X) is about the information of the number of carbons (XX) and the number to the right represents the number of unsaturations of the molecule. Biodiesel samples were derived from various types of oils and oleaginosas, including, methyl and ethyl biodiesel from soybean and castor oil, palm oil, flaxseed, rapeseed, jatropha, moringa, babassu oil, poultry fat and algae.

B. Methodology of the Development of Artificial Neural Networks (ANNs)

The ANNs used in this study are of the type of feedforward backpropagation, which is one of the most popular and used in various fields and it uses the supervised learning paradigm [3]. The networks were configured with two hidden layers with the neurons varied in each training from two to 25 neurons in each layer. Several algorithms can be used to optimize the ANNs' architectures, which enable the convergence of the results quickly and efficiently. Thus, seven algorithms for training, validation and simulation of the network were selected which are gradient descent pattern, gradient descent with momentum and adaptive learning rate, BFGS quasi-newton, scaled conjugate gradient, resilient propagation, one step secant and Levenberg Marquardt. We can observe from Fig. 2, a neural network feed forward back propagation, showed through the 13 inputs of fatty acids esters, two hidden layers: the first layer with seven neurons and the second hidden layer with five neurons. In addition, the quality property of iodine value as output. This was one of the combinations tested during networks' development process.

The selected algorithms are divided into two categories of training strategies, methods of first and second order. The methods of the first order are those where only the local gradient determines the direction of minimization, which are known as direction methods of greater decrease (steepest descent or gradient descent). The second-order methods are considered the most efficient way to do the training of neural networks of MLP type. These algorithms resort to a mathematical rigor based on well-defined models of unconstrained nonlinear optimization, thus not presenting a natural bond with the biological inspiration initially proposed for ANNs. We can quote Newton’s method, which can be considered as the basic local method that uses second-order information, your application requires the calculation of the Hessian matrix [23]. By following the activation functions used in the present study were the logistic function and hyperbolic tangent function. Table I shows a brief summary of the scope of the study, the tested algorithms, the activation functions used, the stopping criteria and generalization of the network.

Altogether, 576 combinations of networks for each training algorithm were tested. The top ten network models of each training algorithm were saved through the k-fold cross-validation method [24]. In addition to the k-fold cross validation criterion with k = 11 as choice of the best network.
among the candidates, the training technique with early stopping was in order to detect when over fitting start during the training of a network [25]. Over fitting occurs while the error in the training set decreases. However, the validation error begins to increase sometime. If the training does not stop at this minimum point, the network will essentially learn noise contained in the training data after this point [3], [26], [25]. The error calculation used to measure the network performance was the Mean Squared Error (MSE), which it can be taken as the Mean Square Error, by the network and the desired value, \( y_j \), given by:

\[
MSE = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_j - y_j)^2
\]

Thus, k-10 parts were used for training and validation, and validation was used for testing. The inputs and outputs of data for training, validation and testing of the network were pre-processed between -1 and 1. The linear regression (R) and root mean square error (RMSE) are used in the network performance analysis. The R and RMSE value are an indication of the correlation between the outputs and targets of the network is given by:

\[
Y_j = \alpha + \beta X_j + \epsilon_j
\]

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{y}_j - y_j)^2}
\]

<table>
<thead>
<tr>
<th>Training parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Backpropagation algorithms</td>
<td>Gradient descent, Gradient descent with momentum, BFGS quasi-newton, Scaled conjugate gradient, Resilient Propagation, One-Step-Backpropagation, Levenberg-Marquardt</td>
</tr>
<tr>
<td>Activation function</td>
<td>tansig function, logsig function, purelin function</td>
</tr>
<tr>
<td>Number hidden layers</td>
<td>2</td>
</tr>
<tr>
<td>Performance function</td>
<td>MSE</td>
</tr>
<tr>
<td>Criteria generalization</td>
<td>Cross validation K-fold, Maximum validation error and early stopping</td>
</tr>
<tr>
<td>Maximum validation error iterations</td>
<td>20</td>
</tr>
<tr>
<td>Number of neurons by layer</td>
<td>2 - 25</td>
</tr>
</tbody>
</table>

### III. RESULTS AND DISCUSSIONS

In this study, the Matlab Version 2013b neural Networks Toolbox was used for the optimization process of ANNs. The experimental data of 98 samples with 13 esters of fatty acids in the input and the quality parameter of iodine value in the output were collected for training, validate and test of ANN method. The investigation the iodine value is very important to oxidation stability of biodiesel due to the number of unsaturation and because of warming of these unsaturated acids that can lead to formation of deposits or damage of lubricating oil [2], [6].

The ANNs were configured with two hidden layers and was set at two to 25 for each neuron layer of the neural network. The training and testing since studies were initially performed with ANN of only one hidden layer and low correlations were found, the predicted results for a layer were 0.76, 0.64, 0.91, for example, for the tested algorithms, while for networks with two hidden layers the coefficient reached 0.99. The decision taken to use two hidden layers in this work was justified by these results.

The coefficient of correlation results for choosing the early stopping criterion is reported in Table II. The use of cross validation by means of early stopping is an important step to avoid overfitting an underfitting. However, the choice of stopping point is an essential and critical criterion for network training. In this study, the definition of number of errors for early stopping was obtained through testing, with values of 6, 10, 15, 20 and 25. The Levenberg Marquardt was the convergence algorithm used for testing the values of errors, together with hyperbolic tangent activation function. Very small or very high values for validation with earlier stopping will not allow the network to learn enough to generalize well so that the classified incorrectly new data. As shown in Table II the results presented for error values equal to 6 in architecture network 8-15 with correlation coefficients of R = 0.98 for the training, R = 0.99 for validation and R = 0.35 for the testing. Already for error values equal to 25, the prediction results were also incorrect in the unknown patterns classification, we can mention the result of the network architecture 10-15 with R = 0.98 for training, R = 0.99 for validation and R = 0.54 for testing. By looking at the results of Table II, we can see that networks which generalized and obtained higher correlation coefficient for prediction were those having the value of 20 errors for early stopping, such as the network of architecture 12-5 with R = 0.97 for the training, R = 0.92 for the validation and R = 0.86 for the testing. Base on the results obtained with value of error equal to 20, this value was used with other algorithms in the following trainings.

From simulation of networks, correlation coefficient of prediction of higher value were showed by the algorithm Levenberg Marquardt, Resilient Propagation and One Step Backpropagation, both of them with correlation of more than 0.95 for both validation and testing. The network architecture obtained with correlation coefficient of 0.99 for training, validation and testing was a 9-7 network with algorithm Levenberg Marquardt and function of logsig activation. Table III shows the results obtained of each algorithm that was trained with their respective correlation coefficients and RMSEs. Related works [5], [1], [3] already demonstrated the ability of the Levenberg Marquardt in obtaining satisfactory results in predicting other properties of biodiesel quality as viscosity and cetane number. In this study, we confirm that ability also for the iodine value.
The results show the topology encountered and their corresponding results of correlation coefficient of training, validation, and testing.

As expected, the networks were tested with only Gradient Descent had inferior results compared to other algorithms tested due to they became trapped in minimum locals, with \( R = 0.82, R = 0.93 \) and 0.94 respectively for training, validation and testing into the network architecture 9-4. However, the results obtained by the algorithms Resilient Propagation and One Step Secant remained equivalent. Moreover, although the time of the training execution was not measured in this study, it is necessary to highlight that some algorithms have more time of training due calculate the hessian matrix such as the algorithms of BFGS and Levenberg Marquardt [23].

In relation to other methods for predicting the iodine value, the ANNs showed a superior results demonstrated by correlation coefficients compared to the work of [7], where the correlation results reach 0.97 and 0.98 in the validation tests, while the method of neural networks reach \( R = 0.99 \) for both validation or test. Although other methods present similar results in the determination of iodine value [4], [9], as [4] which has prediction results also \( R = 0.99 \), using ANNs is much simpler and faster than these methods.

Fig. 3 shows the correlation coefficient of the value obtained and the target value of prediction, obtained by Levenberg Marquardt, both of them have a value of 0.99, demonstrating the network capacity to generalize and simulate new ester data of fatty acid and predict the iodine value.

### IV. CONCLUSION

This study used 98 samples of ester fatty acids to predict the iodine value. Several networks were developed and tested, with several variations of backpropagation algorithm in order to get a better network architecture to predict the biodiesel quality property. This study researched the use of artificial neural networks to predict the quality parameter of iodine value in biodiesel. This is important since through the prediction of biodiesel’s property and from the composition of esters of the raw material is possible to examine the viability of the use of such raw material for the synthesis of biodiesel within the quality parameters. In addition, some methods of analysis of biodiesel’s properties can be long and/or complex. In this case, we achieved a better network architecture with the training algorithm Levenberg Marquardt with correlation index of 0.99 for both validation and testing. In this way, we can say that ANNs are a technology that can be used to predict the iodine value, as well as other properties related to the composition of esters of fatty acids.
REFERENCES


