

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.

For more information visit www.intechopen.com



Applications of Artificial Neural Networks in Biofuels

Alex Oliveira Barradas Filho and
Isabelle Moraes Amorim Viegas

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/intechopen.70691>

Abstract

This chapter is focused on the application of artificial neural networks (ANNs) in the development of alternative methods for biofuel quality issues. At first, the advances and the proliferation of models and architectures of artificial neural networks are highlighted in the text by the characteristics of robustness and fault tolerance, learning capacity, uncertain information processing and parallelism, which allow the application in problems of complex nature. In this scenario, biofuels are contextualized and focused on issues of quality control and monitoring. Therefore, this chapter leads to a study of prediction and/or classification of biofuels quality parameters by the description of published works on the topic under discussion. Afterwards, a case study is performed to demonstrate, in a practical way, the steps and procedures to build alternative models for predicting the oxidative stability of biodiesel. The procedure goes from the processing of the data obtained by the near infrared until the evaluation of the alternative method developed by the neural network. In addition, some evaluation parameters are described for the assessment of the alternative method built. As a result, the feasibility and practicality of the application of neural networks to the quality of biofuels are proven.

Keywords: artificial neural network, biofuel, calibration, classification, quality parameters, oxidative stability

1. Introduction

In the last decades, artificial neural networks (ANNs) have undergone several transformations and improvements, which allowed their application in different areas of knowledge. Such an approach appreciated by the academic community, ANNs are distributed parallel systems, also known as connectionist systems, inspired by the biology of the human brain [1].

In this context, ANNs are simplified models of the human brain that consist of a large number of processing units (neurons) connected to each other. These units usually calculate mathematical

functions (non-linear and/or linear) and form a large network of communication, which allows solving high complexity problems [2].

The architectures that implement the connectionist approach are usually conditioned by a training and learning process rather than being explicitly programmed. In this way, the choice of the architecture has an extremely important character for the solution of certain problems [3].

Among the different tasks appropriate to the application of ANNs are:

- Classification and pattern recognition: process by which a received signal (input) is assigned to a particular group or category;
- Categorization: discovery of well-defined categories or classes in the input data. Unlike classification, classes are not previously known;
- Prediction: estimation of a numerical response based on input values, also called calibration;
- Optimization: characterized by the minimization or maximization of a cost function;
- Noise filtering: extraction of information about a certain response of interest from a noisy dataset.

The variety of ANN applications provide a stimulating scenario for contributions in the field of biofuels, which are defined as renewable fuels derived from biomass that can be used in internal combustion engines or for other types of energy generation. The aim of using biofuels is to reduce external dependence on oil (partial or total replacement of fossil fuels), minimize environmental impacts and develop agricultural production.

The main liquid biofuels produced in the world are ethanol and biodiesel. Ethanol, also known as ethyl alcohol, is a chemical with the molecular formula C_2H_6O , produced by the fermentation of sugars. Under normal conditions, it is a colorless and volatile liquid with an ethereal odor and pungent taste, miscible in water and in different organic solvents.

According to the U.S. Energy Information Administration (EIA), in 2014, the largest ethanol producers on the worldwide are the United States and Brazil [4]. In the USA, the main raw material used for the production of ethanol is corn, while sugarcane is more prominent in Brazil.

Biodiesel is a fuel composed of alkyl esters of long-chain carboxylic acids, produced by the transesterification and/or esterification of fatty material, fats of vegetable or animal origin, with a short-chain alcohol, such as methanol or ethanol [5]. For the production of biodiesel, a variety of raw materials has been used, including edible and non-edible oils, crude oils, fried oils and animal fats. The main raw materials used are soy, palm, cotton, rapeseed, jatropha and sunflower oils and bovine tallow, although it is possible to use all vegetable oils classified as fixed oils or triglycerides, and animal fats [6–8].

Unlike fuel ethanol, the EIA shows that most of biodiesel production in 2014 is not restricted to America alone but also to the continents of Europe, Asia and Oceania [4].

In general, the two biofuels (ethanol and biodiesel) have attracted international attention and, consequently, have had their production increased in comparison to previous years. Some topics studied and related to both ethanol and biodiesel are:

- Production of raw materials;
- Identification of alternative raw materials and other production routes for biofuels;
- Maximization of the production of biofuels;
- Contribution of biofuels to reduce greenhouse gas emissions and environmental impacts;
- Quality control;
- Forms of storage, transportation and distribution of biofuels.

However, despite the diversity of topics and works published in the scientific literature, the present research is targeted to the study of the application of ANNs in the quality control of biofuels [9–15]. Typically, studies related to the quality control of biofuels have the goal to search efficient methods that monitor the fuels produced and commercialized avoiding damages to the environment and consumer injury [9, 16].

It is important to mention that the quality of biofuels is ensured by technical resolutions or standards established by each country which set limit values for contaminants and other parameters [17].

2. State of art: ANN, quality parameters and biofuels

In this section, some papers published in scientific journals, which discuss applications of ANNs to the quality of fuel ethanol (pure or blends) and biodiesel, were selected and discussed. Articles were extracted from the Web of Science database. **Table 1** groups different articles by type of biofuel (ethanol or biodiesel).

In the first article of **Table 1**, Najafi et al., in the paper named “Performance and exhaust emissions of a gasoline engine with ethanol blended gasoline fuels using artificial neural network”, proposed an experimental analysis of the performance and pollutant emissions of a four-stroke SI engine operating with mixtures of ethanol and gasoline (0, 5, 10, 15 and 20%), with the aid of ANNs [18]. Analyzes of the fuel ethanol quality parameters were performed based on the standards of the American Society for Testing and Materials (ASTM). The authors showed that blends with ethanol and gasoline provided an increase in engine power and torque output. It was also found that for ethanol blends, specific brake fuel consumption decreases while thermal brake efficiency and volumetric efficiency increased [18].

Concerning to the use of ANNs, the work of Najafi et al. used the backpropagation algorithm and multilayer perceptron (MLP) architecture for non-linear mapping between the inputs (gasoline-ethanol mixtures and engine speed) and the output parameters (engine performance and exhaust emissions). The evaluation of the results was based on three criteria: correlation

Biofuel	Title of publication	Year
Etanol	Performance and exhaust emissions of a gasoline engine with ethanol blended gasoline fuels using artificial neural network	2009
	Ultrasonic determination of water concentration in ethanol fuel using artificial neural networks	2012
	Prediction of ethanol concentration in biofuel production using artificial neural networks	2013
	Application of GRNN for the prediction of performance and exhaust emissions in HCCI engine using ethanol	2016
	Artificial neural network prediction of diesel engine performance and emission fueled with diesel-kerosene-ethanol blends: a Fuzzy-based optimization	2017
Biodiesel	Application of artificial neural network to predict properties of diesel-biodiesel blends	2010
	Inference of the biodiesel cetane number by multivariate techniques	2013
	Neural network prediction of biodiesel kinematic viscosity at 313 K	2014
	Application of artificial neural networks to predict viscosity, iodine value and induction period of biodiesel focused on the study of oxidative stability	2015
	Attesting compliance of biodiesel quality using composition data and classification methods	2017

Table 1. ANN works applied to biofuel quality.

coefficient (r), root mean squared error (RMSE) and mean relative error (MRE). Thus, the work proves the feasibility of using the ANN approach to predict motor performance (brake power, torque output, brake thermal efficiency, volumetric efficiency, brake specific fuel consumption and equivalence fuel-air ratio) and the emissions (CO, CO₂, HC and NO_x) [18].

In 2012, the work titled “Ultrasonic determination of water concentration in ethanol fuel using artificial neural networks”, published by Liu and Koc, it was determined the concentration of water in ethanol by measurements of ultrasonic velocity and liquid temperature [19]. The aim of the research is to propose an alternative method to contribute to the inspection against the adulteration of fuels, which impairs the vehicle performance and can cause damages to the engine [19].

In the development of an alternative method, the authors Liu and Koc used an ANN based on the MLP architecture. A database was elaborated with 651 samples for the training and validation steps of ANNs. The activation functions, varied for each hidden layer, were the functions logistic sigmoid (logsig), tangent sigmoid (tansig) and linear (purelin), and the results were based on the mean square error (MSE) and on the determination coefficient (R^2). Thus, the research concluded that the results obtained by ANNs were far better when compared with other models [19].

In the paper “Prediction of ethanol concentration in biofuel production using artificial neural networks”, the authors Ahmadian-Moghadam et al. carried out, in 2013, an economic bioprocess to supply ethanol from sugar cane molasses. That research aims to contribute to the reduction of biofuel production prices and to have it as a more competitive resource in the market [20].

Ahmadian-Moghadam et al. applied ANNs to estimate the concentration of ethanol based on the sugar concentration and live and dead yeast cells. To do so, a database with 61 samples

was divided as follows: 60% for training, 15% for validation and 25% for test [20]. The performance of ANN models was evaluated by the mean absolute deviation (MAD), mean absolute percentage error (MAPE) and MSE. Authors concluded that the results showed the viability of the application of the ANN model to determine the final ethanol concentration in the biofuels production process in a large scale [20].

Bendu et al. pointed out, in the paper “Application of GRNN for the prediction of performance and exhaust emissions in HCCI engine using ethanol” published in 2016, the importance of the evaluation of performance and emission parameters of an ethanol-fueled homogeneous charge compression ignition (HCCI) engine. In addition, the authors identified the nature of the parameters as a non-linear problem, which indicated the need for more robust tools [21].

For this purpose, Bendu et al. used a generalized regression neural network (GRNN) consisting of four layers (input layer, radial layer, regression layer and output layer). The input parameters were the charge temperature and the engine load, while the performance and emission values were set as output parameters.

The engine performance parameters were brake thermal efficiency (BTE), exhaust gas temperature (EGT) and the exhaust emission parameters were NO, CO, smoke and unburned hydrocarbon emission (UHC). Summing up, the authors showed the viability of the method and pointed out that the GRNN model can also be used for the control and testing of the HCCI engine [21].

In 2017, Bhowmik et al. performed a study titled “Artificial neural network prediction of diesel engine performance and emission fueled with diesel–kerosene–ethanol blends: a fuzzy-based optimization” to explore the impact on performance and emission characteristics of a single cylinder indirect injection (IDI) engine fueled with blends of diesel and kerosene [22]. In this research, the authors indicated that the addition of ethanol to the mixtures of diesel and kerosene significantly improved the BTE, brake specific energy consumption (BSEC) and the emissions of NO_x, CO and total hydrocarbon (THC) of the engine [22].

Therefore, Bhowmik et al. built an ANN model to map the inputs (load, kerosene volume percentage and ethanol volume percentage) with respect to the outputs (BTE, BSEC, NO_x, THC and CO). The best topology found had a structure with five hidden neurons and presented satisfactory results for the problem addressed. The criteria for evaluation of the developed ANNs were based on MSE, MAPE and r [22].

In 2010, Kumar and Bansal published the paper “Application of artificial neural network to predict properties of diesel – biodiesel blends” whose aim was to evaluate tools for the determination of physical-chemical properties of diesel-biodiesel mixtures. Choosing an appropriate and efficient alternative method could help to avoid some overly time-consuming and costly experiments [23].

Also in the Kumar and Bansal paper, traditional linear regression (principle of least squares) and ANN were applied and compared. The ANNs optimization process was carried out by varying the architectures and training algorithms. The authors concluded that the best results were obtained by the ANN method [23].

In the work of Nadai et al., entitled “Inference of the biodiesel cetane number by multivariate techniques”, a method consisting of successive application of principal components analysis (PCA), fuzzy clustering and ANN in a dataset composed by structural information from proton nuclear magnetic resonance (^1H NMR) of biodiesel fatty esters was implemented [24]. The aim of that work was to obtain the cetane number of different types of complex mixtures from data of pure substances (esters). The authors pointed out two main characteristics that affect the cetane number values: the number of carbon-carbon double bonds and the structure of the alcohol moiety in each fatty ester [24].

In 2014, with the research “Neural network prediction of biodiesel kinematic viscosity at 313 K” Meng et al. performed the prediction of the kinematic viscosity property of biodiesel by artificial neural networks. The authors used 105 samples of biodiesel collected from the literature and 19 types of fatty acid methyl esters (FAMEs) were set as inputs. The results obtained suggested ANNs as an option in predicting kinematic viscosity with a correlation coefficient of 0.9774 [25].

In the paper “Application of artificial neural networks to predict viscosity, iodine value and induction period of biodiesel focused on the study of oxidative stability”, Barradas Filho and collaborators optimized ANN models to predict viscosity, iodine value and induction period (oxidative stability) of 98 biodiesel samples by its fatty esters composition [26].

Also in the work of Barradas Filho et al., the ANNs optimization occurred by varying the activation functions, the number of neurons in the hidden layers and the convergence methods. The evaluation criteria of the models were MSE, RMSE, MAPE and r and R^2 coefficients. After optimization, the ANN results were compared to other models and obtained the best performance [26].

In 2017, the work “Attesting compliance of biodiesel quality using composition data and classification methods” of Lopes et al. compared four classification methods (decision tree classifier, K-nearest neighbors, support vector machine and artificial neural networks) to evaluate the compliance of biodiesel samples concerning some quality parameters. This work aimed to obtain an alternative method with more accuracy when compared to other alternative methods [27].

3. Performance evaluation parameters

After adjustment of classification or calibration models, it is important to have parameters to evaluate the performance through the results obtained. Some of the most widely used parameters and a few explanations on each of them will be provided in this section. These parameters can also be employed, for instance, to aid comparing and deciding among different methods applied to the same problem addressed.

3.1. Evaluation parameters for classification

The first step to organize classification results for better visualization is to build a confusion matrix, like in the example from **Table 2**. The actual classes are represented in the columns, and the predicted classes, in the rows. The number of apple samples classified as apples is registered in cell AA; the number of apples classified as bananas is in cell AB and those

		Actual class		
		Apple (A)	Banana (B)	Coconut (C)
Predicted class	Apple (A)	9 (AA)	2 (BA)	1 (CA)
	Banana (B)	1 (AB)	5 (BB)	1 (CB)
	Coconut (C)	0 (AC)	2 (BC)	11 (CC)
	Sum	10	9	13

Table 2. 3 × 3 confusion matrix of results of fruits classification.

classified as coconuts are in cell AC. The same goes to the other fruit classes. The principal diagonal of the matrix represents the samples correctly classified (cells AA, BB and CC), and the other cells represent the misclassified ones. An ideal classifier would provide a confusion matrix in which all the cells out of the principal diagonal have zero value.

The evaluation parameters for classification methods are based on rates that can be obtained from the confusion matrix. These rates correspond to integer values as they are the numbers of samples classified and split according to some criteria, as will be explained below.

The example given in **Table 2**, and for banana class, the true positive rate (TP) corresponds to the number of bananas correctly classified as bananas (5 samples, cell BB) and the true negatives (TN) are the samples of the other classes (apple and coconut) classified in any class other than banana (21 samples, cells AA, AC, CA and CC). The false positive rate (FP) is the number of samples of other classes misclassified as bananas (2 samples, cells AB and CB) and the false negative rate (FN) corresponds to the banana samples not classified as bananas (4 samples, cells BA and BC).

For apple class, the TP, TN, FP and FN rates are 9, 19, 3 and 1, respectively, and for coconut class, these rates in the same sequence are 11, 17, 2 and 2. Once the TP, TN, FP and FN rates have been obtained for each class, their average values for all classes together can be used to calculate some global evaluation parameters within which the main ones will be briefly explained with the fruits example.

The accuracy (ACC), given by Eq. (1), reflects the global ability of correctly classification by the method. For the fruits example, ACC is 85.42%, which is the percentage of samples that were classified in its actual classes.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (1)$$

The sensitivity (SENS), also called “recall”, can be considered as a global TP rate. The SENS of the fruits classification is 78.13%, calculated by Eq. (2).

$$SENS = \frac{TP}{TP + FN} \times 100\% \quad (2)$$

The specificity (SPEC) can be calculated by Eq. (3) and it is a global TN rate. For the fruits example, SPEC is 89.06%.

$$\text{SPEC} = \frac{\text{TN}}{\text{TN} + \text{FP}} \times 100\% \quad (3)$$

The false positive rate (FPR) can be interpreted as a global rate of FP for all the classes combined and it is the inversely proportional to the SPEC. In the example discussed here, FPR is 10.94%, calculated by Eq. (4).

$$\text{FPR} = \frac{\text{FP}}{\text{TN} + \text{FP}} \times 100\% = 100\% - \text{SPEC} \quad (4)$$

Analogously, the false negative rate (FNR) is a global rate of FN (Eq. (5)). For the fruits classification, FNR is 21.87% and it is complementary to the SENS.

$$\text{FNR} = \frac{\text{FN}}{\text{TP} + \text{FN}} \times 100\% = 100\% - \text{SENS} \quad (5)$$

The ACC, SENS, SPEC, FPR and FNR are some of the main evaluation parameters for classification. Here an example of three classes was presented, giving a 3×3 confusion matrix and, therefore, the evaluation parameters should be calculated by the average TP, TN, FP and FN rates.

Problems with only two classes are simpler and more widespread in the literature, usually involving samples that “have” or “do not have” a specific characteristic and giving a 2×2 confusion matrix. In this case, TP, TN, FP and FN rates are obtained only for the “positive class” and the evaluation parameters are directly calculated by these rates instead of by the averages.

A two class example, already cited in Section 2, is the classification of biodiesel samples according to their compliance to some quality parameters. For each criteria, the samples were split in “compliant” and “non-compliant” [27].

3.2. Evaluation parameters for calibration

The evaluation parameters for calibration are quite different from those for classification. In calibration, these parameters are based on the difference between the actual response, that obtained experimentally by a reference method, and the predicted response, the one estimated by the calibration method.

The oxidative stability (h) of some biodiesel samples from the case study of Section 4 are show in **Table 3** with the actual (y) and predicted (y') responses given in hours. The residuals are the difference between the actual and the predicted responses. The other columns have values calculated to be used in equations of the evaluation parameters that will be explained.

The RMSE is an average deviation between the actual and the predicted values and it has the same unity from the responses. The example from **Table 3**, the RMSE calculated is 0.34 h, which means that the estimated responses differ, on average, in ± 0.34 h from the actual values. In papers, the RMSE is often abbreviated as RMSEC, RMSEP, RMSEV and RMSECV when

Actual response (h) (y)	Predicted response (h) (y')	Residual (h) ($y - y'$)	$(y - y')^2$	$\left(\frac{y_i - y'_i}{y_i}\right) \times 100\%$
19.36	19.20	0.16	0.0256	0.8264
8.93	9.01	-0.08	0.0064	0.8956
7.37	7.35	0.02	0.0004	0.2714
12.77	11.95	0.82	0.6724	6.4213
15.64	15.80	-0.16	0.0256	1.0230
6.60	6.45	0.15	0.0225	2.2727
5.53	5.75	-0.22	0.0484	3.9783
8.01	7.68	0.33	0.1089	4.1199

Table 3. Actual and predicted values of oxidative stability (h) of biodiesel samples.

calculated for the samples of calibration (training), prediction (test), validation and cross-validation, respectively, and it can be given by Eq. (6), in which n is the number of samples ($n = 8$, in this example).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y'_i)^2} \quad (6)$$

Another important error metric is the MAPE, which is a relative measure of the prediction accuracy, calculated by Eq. (7). From the example of oxidative stabilities of biodiesel, the MAPE is 2.48%, that is, on average, the predicted values deviate in 2.48% from their actual values.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - y'_i}{y_i} \right| \times 100\% \quad (7)$$

The Pearson correlation coefficient (r , Eq. (8)) is a measure of the linear relationship between two variables and it is expressed in values from 0 to $|1|$. The closer to $|1|$, the more linearly correlated the variables are. In cases of calibration methods, r coefficient is used to compare the actual and the predicted values. Since y and y' are expected to be equal, this represents a direct relationship and, then, the ideal r coefficient is +1.

$$r(y, y') = \frac{n \sum_{i=1}^n y_i y'_i - \left(\sum_{i=1}^n y_i \right) \left(\sum_{i=1}^n y'_i \right)}{\sqrt{\left[n \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i \right)^2 \right] \left[n \sum_{i=1}^n y'^2_i - \left(\sum_{i=1}^n y'_i \right)^2 \right]}} \quad (8)$$

For the oxidative stabilities, for example, r is 0.9977, which represents a high correlation between the actual and the predicted responses. However, it is important to perform the graphical analysis of the correlation by a scatter plot of the actual (y in x -axis) and the predicted (y' in y -axis) values, because not all samples with a perfect correlation coefficient are well distributed along the line of the expected identity function for $y = y'$.

Step	Parameter	MLP 4-3-1-1
Training	RMSEC (h)	1.31
	MAPE (%)	8.35
	R^2	0.9306
	r	0.9647
Validation	RMSEV (h)	0.43
	MAPE (%)	5.51
	R^2	0.9733
	r	0.9866
Test	RMSEP (h)	0.67
	MAPE (%)	6.89
	R^2	0.9544
	r	0.9769

Table 4. Evaluation parameters calculated from the results obtained by the ANN MLP 4-3-1-1 for the oxidative stability (h) of biodiesel samples.

Although the R^2 coefficient can be obtained by taking the square of the correlation coefficient, they have different meanings. The R^2 , calculated by Eq. (9) in which y_m is the average of the actual values, expresses how much the calibration model explains from the total variance and it can range from 0 to +1. For example, from **Table 4**, the R^2 obtained is 0.9954, which means that 99.54% of total data variance is explained by the regression, and the 0.46% remaining are attributed to residuals.

$$R^2 = r^2 = \frac{\sum_{i=1}^n (y'_i - y_m)^2}{\sum_{i=1}^n (y_i - y_m)^2} \quad (9)$$

Some of the main evaluation parameters for regression have been explained in this section. Besides the numerical parameters, it is also quite important to perform a graphical evaluation of the results by the correlation and residual plots. More details on this will be provided in the case study of the next section.

4. Case study

Biodiesel, like any other fuel, needs to meet some parameters specifications so it can be marketed with quality and safety. These quality parameters are established by standards in each country or region, such as the standards EN 14214 (Europe), ASTM D6751 (USA) and RANP 45/2014 (Brazil) [27].

Among the parameters of biodiesel quality, there are general parameters, which are also applied to petroleum diesel, and there is a special group of parameters related to the chemical composition and purity of the vegetable oils. These parameters can be grouped into four sets: contaminants from the raw material, parameters related to the evaluation of the production process, properties inherent to molecular structures and parameters related to the storage process [7].

One of the main problems assigned to the quality of biodiesel is the possibility of its oxidation caused by the presence of unsaturations in its ester molecules, which is one of the most relevant differences between biodiesel and mineral diesel composition. The main products formed by the oxidation of biodiesel can cause formation of insoluble gums in the engine, filter clogging, injector cocking and corrosion of the metal parts of the engine. Therefore, the evaluation of the oxidative stability of biodiesel is considered by many researchers in the literature to be a very important analysis that should be done because it is directly related to the deterioration capacity (oxidation) and to the time in which the biofuel can be stored (induction period) [26].

The oxidative stability of biodiesel is measured by the method EN 14112, also called Rancimat method, which consists of a system composed of a reaction vessel connected to a cell monitored by an electrode. The sample is placed in the vessel in a heating block at 110°C and a continuous stream of air is bubbled through. The increase in temperature and the presence of oxygen induce the accelerated oxidation of biodiesel. The primary products are formed, followed by secondary products of the oxidation among which are short chain volatile organic acids. These acids are carried to a cell containing deionized water and promote the increase in conductivity, which is measured by an electrode coupled to a device that records the conductivity as a function of time [28].

The induction period used to evaluate the oxidative stability of biodiesel is the time at which the conductivity curve increases rapidly, corresponding to the emergence of the secondary products of the oxidation. The standards EN 14214 and RANP 45/2014 state that the minimum oxidative stability of biodiesel should be 8 h at 110°C [29, 30], while ASTM D6751 specifies 3 h of oxidative stability [31].

Aiming to reduce the time, complexity and costs of analyzing biodiesel quality parameters, some papers in the literature report analytical methodologies alternative to official methods. In this context, the Rancimat method is a relevant case to be studied due to the long analysis time, since a sample of biodiesel that meets EN 14214 requirements will be under analysis for more than 8 h to obtain an oxidative stability result.

A case study of an application of ANN to predict oxidative stability of biodiesel will be presented below to better illustrate the main steps from data preprocessing and selection of sample sets (for training, validation and test) to the optimization of ANN configuration and application. Finally, some performance measures will be evaluated and discussed in the case study. All data handling, preprocessing, subset partitioning and ANN regression were carried out with software MATLAB® 2013a (MathWorks), PLS_Toolbox (Eigenvector) and an algorithm implemented in MATLAB [32].

4.1. Database acquisition and preprocessing

Biodiesels from soybean, corn, palm and babassu (an oleaginous abundant in the Northeast of Brazil) were synthesized via transesterification by methylic route and homogeneous alkaline catalysis and used to prepare 70 binary, ternary and quaternary mixtures (volumetric fractions) designed by simplex-lattice and centroid-simplex designs.

The oxidative stabilities of the samples were determined by the method EN 14112:2003 [33] using a Rancimat equipment Metrohm model 873. The average of two measurements for each sample was taken. The oxidative stabilities of the mixtures ranged from 4.81 to 25.47 h.

The spectra were acquired using a Fourier transform NIR spectrometer PerkinElmer model Frontier™ with a near infrared reflectance accessory (NIRA), equipped with a fast recovery deuterated triglycine sulfate (FR-DTGS) detector. All spectra were recorded with an average of 16 scans and spectral resolution of 2 cm^{-1} . The measured wavenumber range was $4000\text{--}12,000\text{ cm}^{-1}$, but the work range was restricted to $4000\text{--}6100\text{ cm}^{-1}$ because of non-informative signal (close to baseline) and increase of noise as wavenumber gets close to $12,000\text{ cm}^{-1}$.

The raw spectra (**Figure 1a**) showed bands characteristic of first overtone of C—H stretching ($5550\text{--}6100\text{ cm}^{-1}$) and of combination of C—H and C=O stretching modes ($4640\text{--}4700\text{ cm}^{-1}$) [34]. The bands around 4262 and 4334 cm^{-1} can be associated to the second overtone of C—H bending and to combination of C—H and C=C stretching modes, respectively [35].

For correction of spectra baseline deviations caused by systematic variations, the first derivative was calculated by the Savitzky-Golay filter [36] with a 15-point quadratic smoothing function. The window size of points to fit the polynomial function of Savitzky-Golay filter depends on how noisy the spectra are. In this case, a 15-point window was enough to smooth the spectral noise. The derivative NIR spectra of the full database can be seen in **Figure 1b**.

After applying Savitzky-Golay filter, the spectra were mean-centered and then used as input data (**X**-matrix) consisting of 1051 variables and, as output variable (response, **y**-vector), it was used the raw oxidative stabilities (h). From this point, only the preprocessed data was used.

4.2. Steps for construction of the ANN model and selection of sample sets

The construction and validation of models for multivariate classification or calibration go through three basic steps, and for each of them it is necessary a sample set. The first step, the training or calibration, consists of model adjustment with pairs of inputs and outputs (**X**, **y**) provided in the database. The coefficients or weights of the model are amended so the response calculated based on variations in **X** data is as closest to the real (experimental) response as possible. In the training, it is important to have representative samples concerning all the possible **X** and **y** variations that real samples can have.

The step of validation (or internal validation) helps assess the progress of optimization and indicates when the model adjustment should be ended, so it occurs simultaneously to the training step. In the beginning of the training, the coefficients and weights are underfit and the errors are large. In the course of training, the errors decrease as the coefficients are adjusted

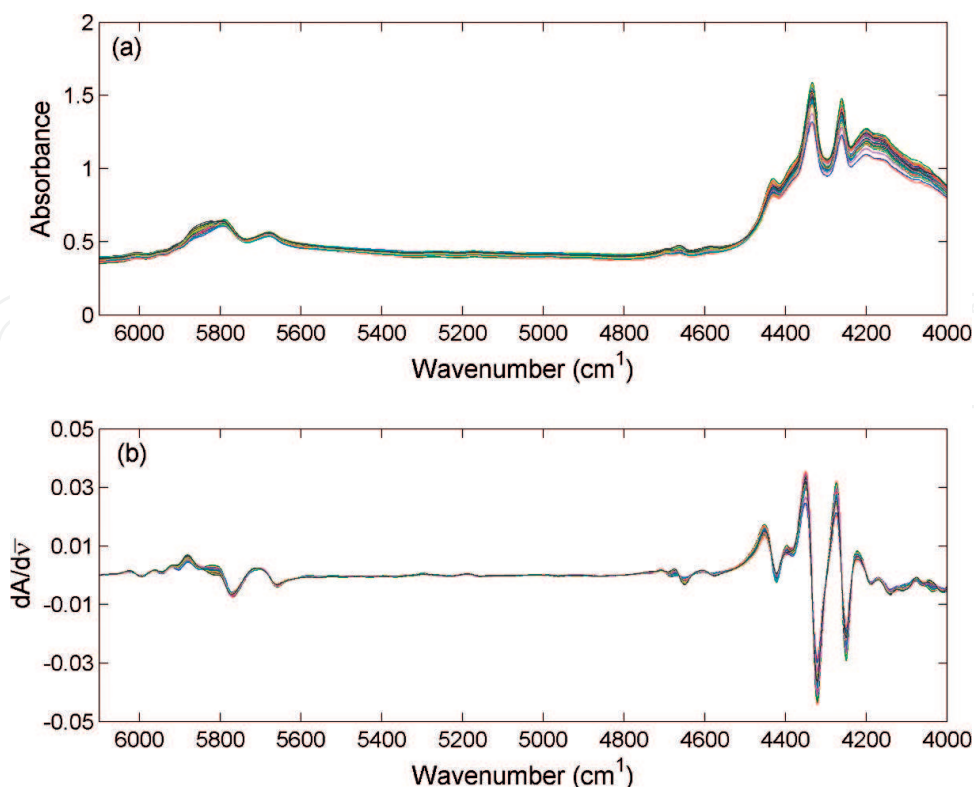


Figure 1. Raw (a) and derivative (b) NIR spectra of the 70 samples of biodiesel mixtures.

and begin to model even the natural noise coming from systematic errors of experimental measurements. At this stage, it occurs the so-called overfitting and the model will not be able to predict or classify external samples with accuracy, although the training and validation errors are small. Therefore, the aim of internal validation is to aid choosing the number of neurons and hidden layers in such a way to balance underfitting and overfitting.

The last step is the test (or external validation), in which samples that were not used during training or internal validation are estimated or classified by the optimized model, simulating a real application. The neural networks learn from the past (training samples) to estimate future cases (test samples).

There are some methods for selection of samples for training, validation and test. In this case study, it will be used the SPXY method (Sample set Partitioning based on joint x-y distances), which is based on the variability of both input (NIR spectra) and output (oxidative stability) variables [32]. The test set consisted of 30% of the database (21 samples), while the 49 remaining samples were split for training (39 samples) and validation (10 samples).

4.3. Dimensionality reduction and ANN configuration

As the **X**-matrix is composed of 1051 variables, it is necessary to apply a method for dimensionality reduction before the training of the neural networks. Otherwise, the modeling would consider too much noise and, because of the large number of input neurons, the ANNs would take too long to converge.

The partial least squares regression (PLS) was used for dimensionality reduction. The number of latent variables (LVs) was optimized based on full cross-validation method. Four LVs explained 99.15% of the \mathbf{X} -variance and 82.85% of \mathbf{y} -variance.

The feedforward MLP ANNs were trained using backpropagation algorithm with a fixed learning rate (0.125) as convergence method to minimize the RMSEC. The input layer is formed by four neurons receiving the four LVs, and the output layer is constituted of one neuron (oxidative stability). The number of neurons in the first hidden layer ranged from 1 to 20, and in the second hidden layer, from 1 to 10. It was also tested a topology with only one hidden layer. The hyperbolic tangent (\tanh , Eq. (10)) and purelin (Eq. (11)) functions were used as activation functions (or transfer functions) of the hidden and output layers, respectively.

$$f(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (10)$$

$$f(x) = x \quad (11)$$

4.4. Results and discussion

As the validation is the step that aids to assess and choose the best fit under varying conditions during optimization, the RMSEV was the criteria taken to choose the best number of hidden layers and neurons. The RMSEVs of the biodiesel oxidative stabilities (h) predicted by ANNs with different numbers of neurons in the hidden layers are represented in **Figure 2**, showing the dependence of the RMSEV on the ANN topology.

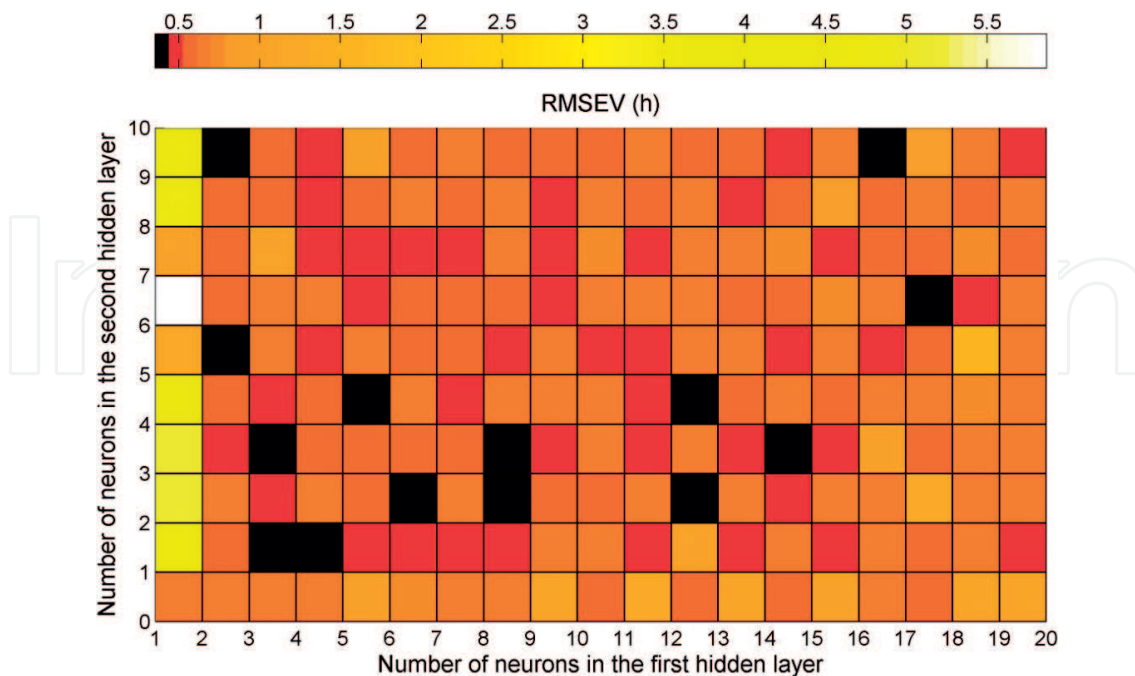


Figure 2. Dependence of the RMSEV of the biodiesel oxidative stability predicted by ANNs with different numbers of neurons in first and second hidden layers.

The RMSEVs when number of neurons in the second hidden layer is zero correspond to the topologies having only one hidden layer (the first one, varying from 1 to 20 neurons). These results presented high RMSEVs that did not vary much with the number of the hidden neurons, evidencing the convergence difficulty of the ANNs with only one hidden layer. Hence, the second layer was added to the optimization process.

Few neural network topologies presented RMSEV lower than 0.5 h, but the best ones are those represented by the black squares in **Figure 2**: MLP 4-2-5-1, MLP 4-2-9-1, MLP 4-3-1-1, MLP 4-3-3-1, MLP 4-4-1-1, MLP 4-5-4-1, MLP 4-6-2-1, MLP 4-8-2-1, MLP 4-8-3-1, MLP 4-12-2-1, MLP 4-12-4-1, MLP 4-14-3-1, MLP 4-16-9-1 and MLP 4-17-6-1. In the notation MLP A-B-C-D, A is the number of input neurons (four LVs), B and C are the number of neurons in the first and second hidden layers, respectively, and D is the number of output neurons (one, oxidative stability).

The 14 best topologies above mentioned had had RMSEV less than or equal to 0.43 h. For choosing among them, the smaller number of neurons is preferred (principle of parsimony: the simpler the better). Therefore, the topology MLP 4-3-1-1 was selected to expand results and predict the oxidative stability of the test samples, but the topologies MLP 4-3-3-1 and MLP 4-4-1-1 should provide similar results.

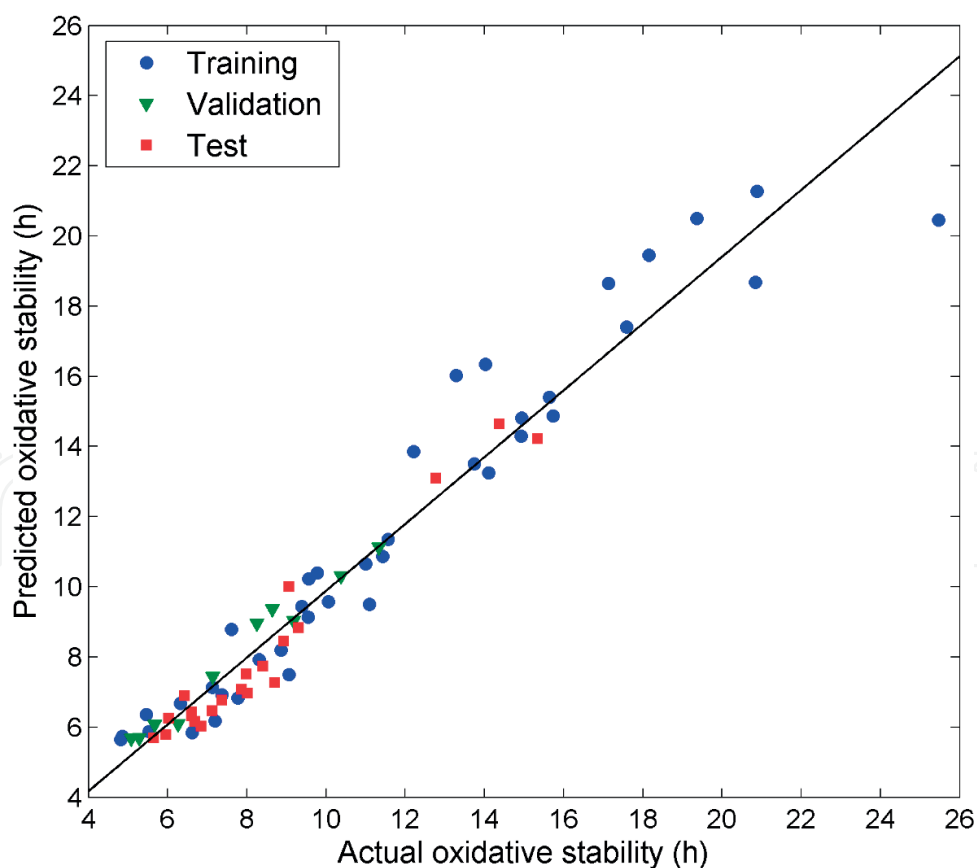


Figure 3. Scatter plot of the biodiesel oxidative stability values predicted by the ANN MLP 4-3-1-1 against the actual values measured by the method EN 14112.

The evaluation parameters calculated for the ANN MLP 4-3-1-1 can be verified in **Table 4**. These parameters can be interpreted as in Section 3.2.

The most important to evaluate the optimized model are the parameters obtained for the test dataset, since these samples simulate a real application with data not used to build nor optimize the model. The RMSEP was 0.67 and the MAPE for test samples was 6.89%, which means that the predicted oxidative stabilities for real samples differed in ± 0.67 h from the actual values and deviated about 6.89%, related to their actual values.

Still for the test samples, the correlation coefficient was 0.9769, indicating a high correlation between the actual and the predicted values of oxidative stability. The determination coefficient was also high, meaning that the ANN MLP 4-3-1-1 explained 95.44% of the total data variance, and the prediction errors represents 4.56% of the total variance.

The correlation plot for samples of all the three steps can be seen in **Figure 3**, in which the samples are well distributed along the line, especially the validation and test samples, leading to correlation coefficients higher than 0.96 for the three steps.

In residual plot (**Figure 4**), it is important to have approximately the same quantity of samples with positive and negative residuals, and the closer to the central line ($y = 0$) the smaller the RMSEs. In this case study, the majority of samples had residual lower than ± 1.5 h and they are well divided with positive and negative residuals. The higher residuals belong to the training samples, which indeed had the highest RMSE (1.31 h).

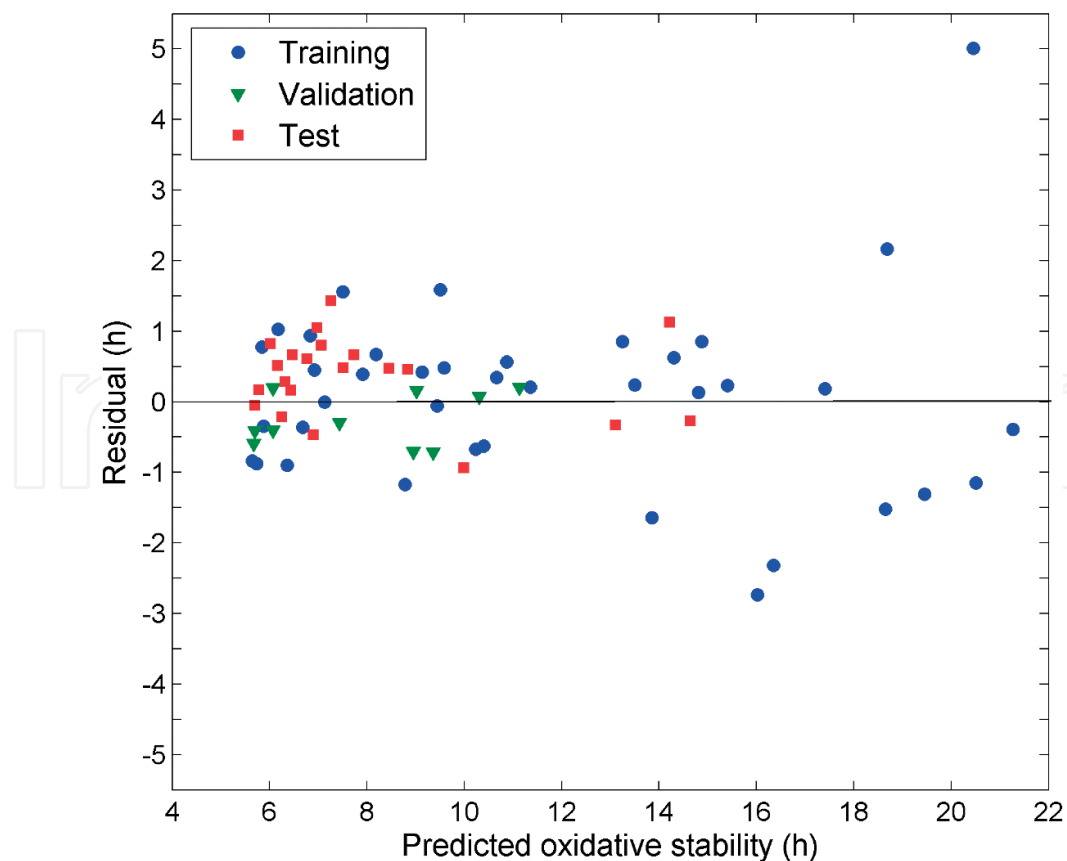


Figure 4. Residual plot the biodiesel oxidative stability values predicted by the ANN MLP 4-3-1-1.

5. Conclusion

The literature presents a variety of published works involving the feasibility of the application of artificial neural networks to biofuels. In this way, the increasing importance of the biofuels theme becomes more evident in the global energetic scenario.

The ANNs proved to be a promising tool in the development of more efficient and cost-effective alternative methods to control and monitor the quality of biofuels, when compared to official methods.

In addition, a presented case study allowed to understand, in practice, the procedures to be performed in the process of predicting a physical-chemical property of biodiesel, the oxidative stability, since data preprocessing, ANN setup and training and calculating and interpretation of the evaluation criteria.

Although the practical development was carried out by a regression approach, this work also explained about classifiers and procedures for both the construction and evaluation of models. Therefore, the present work can be helpful to instruct the basic procedures in the application of ANNs to the quality of biofuels.

Acknowledgements

The authors would like to thank the Foundation for Research and Scientific and Technological Development of Maranhão (FAPEMA) and the National Council for Scientific and Technological Development from Brazil (CNPq) for the financial support and fellowships received. We also thank the laboratories LAPQAP/NEPE from the Federal University of Maranhão and LAC from the Federal University of Pernambuco.

Author details

Alex Oliveira Barradas Filho^{1*} and Isabelle Moraes Amorim Viegas²

*Address all correspondence to: alex.barradas@ecp.ufma.br

1 Federal University of Maranhão, São Luís, MA, Brazil

2 Federal University of Pernambuco, Recife, PE, Brazil

References

- [1] Braga AP, Carvalho APLF, Ludermir TB. *Redes Neurais Artificiais. Teoria e Aplicações*. 2nd ed. Rio de Janeiro: LTC; 2007. 238 p. ISBN: 978-8521615644
- [2] Haykin S. *Neural Networks: A Comprehensive Foundation*. 2nd ed. Upper Saddle River: Prentice Hall, 1998. ISBN: 978-0132733502

- [3] Jain AK, Mao J, Mohiuddin KM. Artificial neural networks: A tutorial. *Computer*. 1996; **29**:31-44. DOI: 10.1109/2.485891
- [4] U.S. Energy Information Administration (EIA). International Energy Statistics, 2014. Available from: <https://www.eia.gov/tools/> [Accessed: 01-07-2017]
- [5] Verma P, Sharma MP. Review of process parameters for biodiesel production from different feedstocks. *Renewable & Sustainable Energy Reviews*. 2016;**62**:1063-1071. DOI: 10.1016/j.rser.2016.04.054
- [6] Moser BR. Influence of extended storage on fuel properties of methyl esters prepared from canola, palm, soybean and sunflower oils. *Renewable Energy*. 2011;**36**:1221-1226. DOI: 10.1016/j.renene.2010.10.009
- [7] Lôbo IP, Ferreira SLC, Da Cruz RS. Biodiesel: parâmetros de qualidade e métodos analíticos. *Química Nova*. 2009;**32**:1596-1608. DOI: 10.1590/S0100-40422009000600044
- [8] Schuchardt U, Sercheli R, Vargas RM. Transesterification of vegetable oils: A review. *Journal of the Brazilian Chemical Society*. 1998;**9**:199-210. DOI: 10.1590/S0103-50531998000300002
- [9] Sajjadi B, Raman AAA, Arandiyani H. A comprehensive review on properties of edible and non-edible vegetable oil-based biodiesel: Composition, specifications and prediction models. *Renewable & Sustainable Energy Reviews*. 2016;**63**:62-92. DOI: 10.1016/j.rser.2016.05.035
- [10] Saldana DA, Starck L, Pascal M, Rousseau B, Ferrando N, Creton B. Prediction of density and viscosity of biofuel compounds using machine learning methods. *Energy & Fuels*. 2012;**26**:2416-2426. DOI: 10.1021/ef3001339
- [11] Balabin RM, Safieva RZ. Near-infrared (NIR) spectroscopy for biodiesel analysis: Fractional composition, iodine value, and cold filter plugging point from one vibrational spectrum. *Energy & Fuels*. 2011;**25**:2373-2382. DOI: 10.1021/ef200356h
- [12] Balabin RM, Lomakina EI, Safieva RZ. Neural network (ANN) approach to biodiesel analysis: Analysis of biodiesel density, kinematic viscosity, methanol and water contents using near infrared (NIR) spectroscopy. *Fuel*. 2011;**90**:2007-2015. DOI: 10.1016/j.fuel.2010.11.038
- [13] Przybylski R, Zambiazzi RC. Predicting oxidative stability of vegetable oils using neural network system and endogenous oil components. *Journal of American Oil Chemistry Society*. 2000;**77**:925-932. DOI: 10.1007/s11746-000-0146-x
- [14] Ramadhas AS, Jayaraj S, Muraleedharan C, Padmakumari K. Artificial neural networks used for the prediction of the cetane number of biodiesel. *Renewable Energy*. 2006;**31**:2524-2533. DOI: 10.1016/j.renene.2006.01.009
- [15] Saeid B, Aroua MK, Raman AAA, Sulaiman NMN. Estimation of vegetable oil-based ethyl esters biodiesel densities using artificial neural networks. *Journal of Applied Sciences*. 2008;**8**:3005-3011. DOI: 10.3923/jas.2008.3005.3011
- [16] Marques DB, Barradas Filho AO, Romariz ARS, Viegas IMA, Luz DA, Barros Filho AKD, Labidi S, Ferraudo AS. Recent developments on statistical and neural network tools

focusing on biodiesel quality. *International Journal of Computer Science and Applications*. 2014;**3**:97-110. DOI: 10.14355/ijcsa.2014.0303.01

- [17] Hoekman SK, Broch A, Robbins C, Cenicerros E, Natarajan M. Review of biodiesel composition, properties, and specifications. *Renewable and Sustainable Energy Reviews*. 2012;**16**:143-169. DOI: 10.1016/j.rser.2011.07.143
- [18] Najafi G, Ghobadian B, Tavakoli T, Buttsworth DR, Yusaf TF, Faizollahnejad M. Performance and exhaust emissions of a gasoline engine with ethanol blended gasoline fuels using artificial neural network. *Applied Energy*. 2009;**86**:630-639. DOI: 10.1016/j.apenergy.2008.09.017
- [19] Liu B, Koc AB. Ultrasonic determination of water concentration in ethanol fuel using artificial neural networks. *Transactions of the ASABE*. 2012;**55**:1865-1872. DOI: 10.13031/2013.42339
- [20] Ahmadian-Moghadam H, Elegado FB, Nayve R. Prediction of ethanol concentration in biofuel production using artificial neural networks. *American journal of Modeling and Optimization*. 2013;**1**:31-35. DOI: 10.12691/ajmo-1-3-2
- [21] Bendu H, Deepak BBV, Murugan S. Application of GRNN for the prediction of performance and exhaust emissions in HCCI engine using ethanol. *Energy Conversion and Management*. 2016;**122**:165-173. DOI: 10.1016/j.enconman.2016.05.061
- [22] Bhowmik S, Panua R, Debroy D, Paul A. Artificial neural network prediction of diesel engine performance and emission Fueled with diesel–kerosene–ethanol blends: A fuzzy-based optimization. *Journal of Energy Resources Technology*. 2017;**139**. DOI: 10.1115/1.4035886
- [23] Kumar J, Bansal A. Application of artificial neural network to predict properties of diesel-biodiesel blends. *Journal of Science, Engineering and Technology*. 2010;**6**:98-103. DOI: 10.3126/kuset.v6i2.4017
- [24] Nadai DV, Simões JB, Gatts CEN, Miranda PCML. Inference of the biodiesel cetane number by multivariate techniques. *Fuel*. 2013;**105**:325-330. DOI: 10.1016/j.fuel.2012.06.018
- [25] Meng X, Jia M, Wang T. Neural network prediction of biodiesel kinematic viscosity at 313 K. *Fuel*. 2014;**121**:133-140. DOI: 10.1016/j.fuel.2013.12.029
- [26] Barradas Filho AO, Barros AKD, Labidi S, Viegas IMA, Marques DB, Romariz ARS, de Souza RM, Marques AL, Marques EP. Application of artificial neural networks to predict viscosity, iodine value and induction period of biodiesel focused on the study of oxidative stability. *Fuel* 2015;**145**:127-135. DOI: 10.1016/j.fuel.2014.12.016
- [27] Lopes MV, Barradas Filho AO, Barros AK, Viegas IMA, Silva LCO, Marques EP, Marques ALB. Attesting compliance of biodiesel quality using composition data and classification methods. *Neural Computing and Applications*. 2017. DOI: 10.1007/s00521-017-3087-4
- [28] Yaakob Z, Narayanan BN, Padikkaparambil S, Unni KS, Akbar PM. A review on the oxidation stability of biodiesel. *Renewable and Sustainable Energy Reviews*. 2014;**35**: 136-153. DOI: 10.1016/j.rser.2014.03.055

- [29] EN 14214. Liquid Petroleum Products—Fatty Acid Methyl Esters (FAME) for Use in Diesel Engines and Heating Applications—Requirements and Test Methods. Comité Européen de Normalisation. 2012.
- [30] RANP 45/2014. Resolução ANP n° 45, de 25 de agosto de 2014 - DOU 26.08.2014. Agência Nacional do Petróleo, Gás Natural e Biocombustíveis. p. 2014
- [31] ASTM D6751. Standard Specification for Biodiesel Fuel Blend Stock (B100) for Middle Distillate Fuels. American Society for Testing and Materials. 2012.
- [32] Galvão RKH, Araujo MCU, José GE, Pontes MJC, Silva EC, Saldanha TCB. A method for calibration and validation subset partitioning. *Talanta*. 2005;**67**:736-740. DOI: 10.1016/j.talanta.2005.03.025
- [33] EN 14112. Fat and Oil Derivatives—Fatty Acid Methyl Esters (FAME) - Determination of Oxidation Stability (Accelerated Oxidation Test). Comité Européen de Normalisation. 2003.
- [34] Oliveira JS, Montalvão R, Daher L, Suarez PAZ, Rubim JC. Determination of methyl ester contents in biodiesel blends by FTIR-ATR and FTNIR spectroscopies. *Talanta*. 2006;**69**: 1278-1284. DOI: 10.1016/j.talanta.2006.01.002
- [35] De Lira LFB, de Albuquerque MS, Pacheco JGA; Fonseca TM, Cavalcanti EHS, Stragevitch L, Pimentel MF. Infrared spectroscopy and multivariate calibration to monitor stability quality parameters of biodiesel. *Microchemical Journal* 2010;**96**:126-131. DOI: 10.1016/j.microc.2010.02.014
- [36] Savitzky A, Golay MJE. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*. 1964;**36**:1627-1639. DOI: 10.1021/ac60214a047