Recent Developments on Statistical and Neural Network Tools Focusing on Biodiesel Quality

Delano B. Marques*, Alex O. Barradas Filho, Alexandre R. S. Romariz*, Isabelle M. A. Viegas, Djavania A. Luz, Allan K.D. Barros Filho, Sofiane Labidi, Antonio S. Ferraudo

1Department of Electrical Engineering, University of Brasilia, Brasilia-DF, Brazil
2Department of Electrical Engineering, Federal University of Maranhão, São Luis-MA, Brazil
3Department of Chemistry (LAPQAP/NEPE), Federal University of Maranhão, São Luis-MA, Brazil
4Department of Sciences, São Paulo State University Júlio de Mesquita Filho, Jabiocabal-SP, Brazil

Received 16 March 2014; Revised 4 May 2014; Accepted 9 May 2014; Published 8 August 2014
© 2014 Science and Engineering Publishing Company

Abstract
The performance of both the traditional linear regression and Artificial Neural Network (ANN) techniques has been compared to check the validity to predict the properties of biodiesel and mixtures of diesel and biodiesel. We present on this paper a review on statistical and ANN applications to the Biodiesel quality. A case study is also presented showing the prediction of oxidative stability of Biodiesel using, for the first time, other official quality parameters instead of the chemical composition as input data. In this sense, our hope is that this paper would complement a series of recent review papers and catalyze future research in this rapidly evolving area.

Keywords
Quality Assessment of Biodiesel; Prediction; Multivariate Analysis; Statistical; Neural Network

Introduction
Most standards adopted for biodiesel quality follows an international standard usually according to the EN 14214 (European) and ASTM D6751-11B (American). Table 1 shows different quality standards for biodiesel from different countries. Among the parameters established by all international norms, are those from the standardization of mineral diesel and those that were derived from analysis of vegetable oils, commonly used in the oleochemical industry (Monteiro et al., 2008).

Various studies have been published on analytical methods for biodiesel, and previous reviews directed for biodiesel analysis (Monteiro et al., 2008; Knothe, 2006; Zhang, 2012) showed the main analytical techniques developed until now. Among these analytical methods and other analytical needs stated earlier, the commonly used analytical methods for analyzing biodiesel are chromatography and spectroscopy (Zhang, 2012; Ghisi et al., 2011; Balabin et al., 2011).

Among the new methods involving the determination of inorganic contaminants, electroanalytical area is growing and showing many publications (Santos et al., 2012; Trindade et al., 2012; Tubino and Aricetti 2013; Martiniano et al., 2013).

The properties of pure biodiesel or diesel-biodiesel mixtures, when determined through instrumental analytical methods allow the gathering of much information about the biodiesel samples (Monteiro et al., 2008; Lôbo et al., 2009) but is a tedious process and as well as very time consuming (Kumar, 2010). Although there is a considerable number of articles on the application of ANN for Biodiesel quality, there are still many gaps about this subject and there are also several official parameters that have not yet been addressed. So the application of other tools in order to estimate these properties is considered of immense importance.

Chemometry based on mathematic statistical models has been widely used to assess the quality of complex samples such as biofuels, however, these tools are not
always satisfactory in terms of accuracy. This fact explained the expansion of the applications of ANN and the large increase in publications in recent years, including fuels and biofuels.

In the present work we organized a text that included concepts, informations and data from the literature, related to the application of prediction methods directed to the quality of biodiesel.

The literature shows that, in almost all studies about prediction of Biodiesel quality parameters, input data into the ANN typically involves the chemical composition of Biodiesel associated with at least one quality parameter, measured experimentally. This work proposes a new and alternative way, which is the use of other quality parameters (ester content, density, viscosity, acid value and iodine number) of biodiesel as input data, which are official parameters, described on official norm (EN 14112).

Our intention also highlights the quality aspects of biodiesel and present recent data on statistical and ANN tools, applied to biodiesel.

The need for these mathematic and computational tools consolidate the chemometrics, which is an area designed specifically for the application of Multivariate Analysis (MVA) in the chemistry area.

<table>
<thead>
<tr>
<th>Property</th>
<th>Units</th>
<th>Brazil</th>
<th>United States</th>
<th>Europe</th>
<th>Germany</th>
<th>Australian</th>
<th>India</th>
<th>Japan</th>
<th>South Africa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Density</td>
<td>kg/m³</td>
<td>ANP</td>
<td>ASTM 14/2012</td>
<td>EN</td>
<td>DIN</td>
<td>(1) IS</td>
<td>JASO</td>
<td>SANS</td>
<td></td>
</tr>
<tr>
<td>Viscosity at 40°C</td>
<td>mm²/s</td>
<td>850 - 900</td>
<td>3.0 - 6.0</td>
<td>3.5 - 5.0</td>
<td>3.5 - 5.0</td>
<td>3.5 - 5.0</td>
<td>3.5 - 5.0</td>
<td>3.5 - 5.0</td>
<td>3.5 - 5.0</td>
</tr>
<tr>
<td>Water and sediment, max.</td>
<td>mg/kg</td>
<td>500</td>
<td>1.9 - 6.0</td>
<td>500</td>
<td>300</td>
<td>500</td>
<td>500</td>
<td>500</td>
<td>500</td>
</tr>
<tr>
<td>Total contamination, max.</td>
<td>24</td>
<td>-</td>
<td>24</td>
<td>20</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>Flash point, min.</td>
<td>°C</td>
<td>100</td>
<td>101</td>
<td>110</td>
<td>120</td>
<td>120</td>
<td>120</td>
<td>120</td>
<td>120</td>
</tr>
<tr>
<td>Ester content, min.</td>
<td>% mass</td>
<td>96.5</td>
<td>-</td>
<td>96.5</td>
<td>-</td>
<td>96.5</td>
<td>96.5</td>
<td>96.5</td>
<td>96.5</td>
</tr>
<tr>
<td>Carbon residue, max.</td>
<td>% mass</td>
<td>0.05 in</td>
<td>0.05 in</td>
<td>0.30 on the</td>
<td>0.03 in</td>
<td>0.05 in</td>
<td>0.05 in</td>
<td>0.30 on the</td>
<td>0.30 on the</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100% sample</td>
<td>100% sample</td>
<td>10% distillation residue of the sample</td>
<td>100% sample</td>
<td>100% sample</td>
<td>10% distillation residue of the sample</td>
<td>10% distillation residue of the sample</td>
<td></td>
</tr>
<tr>
<td>Sulfated ash, max.</td>
<td>% mass</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.03</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>Sulfur, max.</td>
<td>mg/kg</td>
<td>10</td>
<td>15</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>15 for S15 grade; 500 for S500 grade</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sodium and potassium, max.</td>
<td>mg/kg</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>Report</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Calcium and magnesium, max.</td>
<td>mg/kg</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>-</td>
<td>Report</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Phosphorous, max.</td>
<td>mg/kg</td>
<td>10</td>
<td>10</td>
<td>4</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Copper corrosion, 3h at 50°C, max</td>
<td>-</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Cold filter plugging point, max.</td>
<td>°C</td>
<td>(3)</td>
<td>Report</td>
<td>47</td>
<td>51</td>
<td>49</td>
<td>51</td>
<td>51</td>
<td>51</td>
</tr>
</tbody>
</table>

TABLE 1 - QUALITY STANDARDS FOR BIODIESEL FROM DIFFERENT COUNTRIES
There are many MVA tools applied to the diagnosis and prediction of properties. One suggestion for the classification of these studies in the literature could involve linear and nonlinear models (Nia et al., 2014; Morsy and Sun, 2013).

The main linear models are: Multi Linear Regression (MLR) (Balabin et al., 2007), Principal Component Analysis (PCA) (Abdi and Williams, 2010), Principal Component Regression (PCR) (Balabin et al., 2007; Dodge, 2003; Jolliffe, 1982) and Partial Least Squares Regression (PLS) (Balabin and Safieva, 2011).

PLS is a statistical (multivariate) method with a close relationship to PCR (Balabin et al., 2007). The main difference of the PLS method is that both the input (X) and output (Y) data are projected to new spaces. In the case of biodiesel analysis, X and Y are spectra and properties, respectively. Because of its simplicity and lower requirements for computational resources, PLS is widely used in the analysis of very different datasets. In addition, this approach can be regarded as a linear standard method for data analysis (Balabin et al., 2011).

One of the main non-linear calibration models are neural networks (Balabin et al., 2007; Balabin et al., 2011). ANNs (Schalkoff, 1997) are a functional abstraction of the biological neural structures of the central nervous system (Anderson, 1983; Akkurt et al., 2003) and they can exhibit a surprising number of the characteristics of human brain, e.g., learning from experience and generalizing from previous examples to solve new problems (Öztas et al., 2006). Typically, a network consists of a set of sensing units (source nodes) that constitute the input layer, one or more hidden layers of computational nodes. The input signal is propagated forward through the network, layer by layer. Generally, neural networks are adjusted, or trained, in order to achieve a particular target for a give output.

The multilayer perceptrons (MLP) have been successfully applied to solve many difficult problems through their training in a controlled manner with very popular algorithm known as error back propagation algorithm (Rumelhart et al., 1986). An example of a network topology multilayer perceptron is shown in Figure 1 containing an initial layer with four input variables, two hidden layers with three neurons in each one and an output layer.

<table>
<thead>
<tr>
<th>Property</th>
<th>Units</th>
<th>Brazil</th>
<th>United States</th>
<th>Europe</th>
<th>Germany</th>
<th>Australia</th>
<th>India</th>
<th>Japan</th>
<th>South Africa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cloud point, max.</td>
<td>°C</td>
<td>ANP 14/2012</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Cold soak filter ability, max.</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Acid value, max.</td>
<td>mg KOh/g</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.80</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
</tr>
<tr>
<td>Free glycerin, max.</td>
<td>% mass</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>Total glycerin, max.</td>
<td>% mass</td>
<td>0.25</td>
<td>0.24</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>Monoglyceride content, max.</td>
<td>% mass</td>
<td>0.80</td>
<td>-</td>
<td>0.80</td>
<td>-</td>
<td>-</td>
<td>0.80</td>
<td>0.80</td>
<td>0.80</td>
</tr>
<tr>
<td>Diglyceride content, max.</td>
<td>% mass</td>
<td>0.20</td>
<td>-</td>
<td>0.20</td>
<td>0.40</td>
<td>-</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
</tr>
<tr>
<td>Triglyceride content, max.</td>
<td>% mass</td>
<td>0.20</td>
<td>-</td>
<td>0.20</td>
<td>0.40</td>
<td>-</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
</tr>
<tr>
<td>Methanol and/or ethanol content, max.</td>
<td>% mass</td>
<td>0.20</td>
<td>0.20 or Flash point &lt; 130°C</td>
<td>0.30</td>
<td>-</td>
<td>(4)</td>
<td>0.20</td>
<td>0.20</td>
<td>0.20</td>
</tr>
<tr>
<td>Iodine value, max.</td>
<td>gI2/100g</td>
<td>Report</td>
<td>-</td>
<td>120</td>
<td>115</td>
<td>Report</td>
<td>120</td>
<td>140</td>
<td>140</td>
</tr>
<tr>
<td>Oxidation stability at 110°C, min.</td>
<td>Hours</td>
<td>6</td>
<td>3</td>
<td>6</td>
<td>-</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>

(1) Barabás and Todorut (2011)
(2) Water and sediment maximum limit is 380 mg/kg;
(3) Cold filter plugging point varies according to area of commercialization of biofuel;
(4) Alcohol content maximum is 0.20% in mass for methanol or 0.20% in mass for ethanol.

(Continued from Table 1)
neural networks methods, nowadays, are due to the association between modern instrumental analytical methods and the efficient regression tools provided by modern mathematics. This association also makes possible the creation of accurate and robust methods for the prediction of object properties possible (Balabin and Lomakina, 2009; Balabin, 2009; Manly, 2005; Hobro and Lendl, 2009; Sun et al., 2010; Balabin et al., 2008). Therefore, progress in chemometrics has a direct effect on the field of analytical chemistry. A featured paper recently published (Nia et al., 2014) presents a comprehensive comparison of nonlinear methods (Kernel PLS (KPLS), support vector machines (SVM), least-squares SVM (LS-SVM), relevance vector machines (RVM), Gaussian process regression (GPR), ANN, and Bayesian ANN (BANN)) where a PLS regression is used as a linear benchmark.

Numerous authors have compared performance of statistical and neural networks models on specific problems (Manjunatha et al., 2010; Sarle, 1994; Schumacher et al., 1996; Wu, 1992; Hruschka, 1993). Balabin and his collaborators have been among the leading researchers who have published on the relationship between statistical methods and neural networks comparing them in the application of these techniques in predicting parameters of Biodiesel quality. An important part of the matters presented and discussed here comes from their work. A larger and more detailed view on the subject can be found in their papers (Balabin et al., 2011; Balabin and Smirnov, 2011; Balabin et al., 2007; Balabin et al., 2008; Balabin, 2009; Syunyaev and Balabin, 2007; Syunyaev and Balabin, 2008; Balabin and Lomakina, 2011; Balabin and Smirnov, 2012).

Recent Prediction Methods Developed for Biodiesel Quality

Many of the well established instrumental methods, applied to biodiesel, combined with multivariate analysis, have been studied in order to establish calibration models, with many advantages including acceptable precision and accuracy, still being quick and not very laborious (Ghesti et al., 2006; Blanco et al., 2004; Brásio et al., 2013; Cheenkachorn, 2006).

Several multivariate calibration models are presented in the literature (Zhang, 2012; Lôbo et al., 2009; Balabin and Lomakina, 2009; Balabin, 2009; Balabin and Lomakina, 2011; Balabin and Smirnov, 2012), where it is shown that the data obtained by instrumental techniques are treated by multivariate analysis, the quantification of analytes being possible even in the
presence of interferents (Lôbo et al., 2009).

Cheenkachorn (2006) used statistical models and ANN tools to predict several properties of biodiesels using the fatty acid compositions of various vegetable oils. In this work the authors used data mainly collected from previous literature. The input variables of the models included different types of fatty acids such as palmitic acid, stearic acid acid, oleic acid, etc. and the predicted properties included viscosity, high-heating value (HHV), and cetane number. The results showed that both statistical models and the ANNs precisely predicted the properties of biodiesel derived from certain vegetable oils.

Balabin et al. (2011) compared the performance of linear and non-linear calibration techniques (MLR, PCR, PLS, polynomial PLS (Poly-PLS) and Spline-PLS versions, and ANN) for prediction of biodiesel properties from near infrared spectra. The model was created for four important biodiesel properties: density (at 15 °C), kinematic viscosity (at 40 °C), water content, and methanol content. They concluded that the ANN approach was superior to the linear (MLR, PCR, PLS) and “quasi”-non-linear (Poly-PLS, Spline-PLS) calibration methods.

Balabin and Smirnov (2011) compared the performance of 16 different feature selection methods for the prediction of properties of biodiesel fuel, including density, viscosity, methanol content, and water concentration. A comparison with a non-linear calibration model, (MLP-ANN) was provided. Among the feature selection algorithms tested were MLR, Interval PLS (iPLS), back-propagation ANN (BP-ANN) and others. Two linear techniques for calibration model building (MLR and PLS) were used for the evaluation of biofuel properties. The results of other spectroscopic techniques application, such as Raman, ultraviolet-visible (UV-Vis), or nuclear magnetic resonance (NMR) spectroscopies, could be greatly improved by an appropriate feature selection choice.

Balabin and Smirnov (2012) in their recent study made a rather general comparison of linear, such as partial least squares or projection to latent structures (PLS); quasi-nonlinear, such as the polynomial version of PLS (Poly-PLS); and intrinsically non-linear, such as artificial neural networks (ANNs), support vector regression (SVR), and least-squares support vector machines (LS-SVM/LSSVM), regression methods in terms of their robustness. Biodiesel was one of the systems chosen as representative examples of real-world samples in this study.

Response Surface Methodology (RSM) and ANN were employed to study the relationship between process variables and free fatty acid conversion and for predicting the optimal parameters (Talebian-Kiakalaieh et al., 2013). The highest conversion found was 88.6% at optimum condition being 14h, 65°C, 70:1 and 10wt% for reaction time, reaction temperature, methanol to oil molar ratio and catalyst loading, respectively. The RSM and ANN could accurately predict the experimental results, with R²=0.9987 and R²=0.985, respectively.

Studies involving statistical and neural network tools showed predictions on the biodiesel content in biodiesel/diesel blends. This theme has been one of the applications most studied involving chemometrics and ANN (Balabin and Lomakina, 2011; Ramadhas et al., 2006; Fimentel et al., 2006; Kalogeras et al., 2010; Fernandes et al., 2011) mainly including spectroscopic techniques.

Predictions involving engine emission characteristics, and engine performance was another themeexplored (Kumar et al., 2013). Good correlation were obtained recently (Kumar et al., 2013; Prasad et al., 2010; Kumar et al., 2012) between the ANN predicted values and the desired values for various engine performance values and the exhaust emissions.

Many other important properties of biodiesel and biodiesel/diesel blend also were objects of studies on predictions in the last years, such as flash point, fire point (Kalogeras et al., 2010; Saldana et. Al., 2011; Kumar and Bansal, 2007), methanol content (Balabin et al., 2011; Felizardo et al., 2007); cloud point, pour point, volatility at different temperatures 250 °C, 350 °C and 360 °C, cetane index, sulfur, flash point and cold filter plugging point (CFPP) (Kalogeras et al., 2010); water content (Balabin and Lomakina, 2011; Balabin et al., 2011; Felizardo et al., 2007); cold filter plugging point (Baptista et al., 2008); vapor pressures (Freitas et al., 2012); biodiesel feedstock (Zawadzki and Shrestha, 2009); oxidative stability (Canha et al., 2012); concentration of triglycerides (Oliveira et al., 2007); biodiesel production process modelling (Hui, 2012); engine torque, specific fuel consumption and exhaust gas components (Najafi et al., 2007) discrimination of methyl biodiesel from different sources (Flores et al., 2012).

Table 2 presents a selection of prediction methods applied to standard parameters (cetane number, viscosity, ester content, density, flash point, methanol and/or ethanol content, water content, oxidative stability, iodine value, cold filter plugging point, cloud
point, pour point and blend) related to the quality of biodiesel, available in the current literature. As we can see, there are a number of the standard parameters that still have not been addressed with this type of statistical or computational tools. On the other hand, many other studies involving biodiesel quality and applications of statistical and computational tools also have been found in the literature.

**TABLE 2—PREDICTION METHODS APPLIED TO STANDARD QUALITY OF BIODIESEL**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Official Method</th>
<th>Prediction Method</th>
<th>Tools</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Density</td>
<td>EN ISO 3675; EN ISO 12185 ASTM D1298; ASTM D 4052; ABNT NBR 7148; ABNT NBR 14065</td>
<td>Artificial Neural Network</td>
<td>batch gradient descent with momentum algorithm, levenberg–marquardt algorithm, scaled conjugate gradient algorithm</td>
<td>Kumar and Bansal (2010)</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>--------------------------</td>
<td>------------------------------------------------</td>
<td>--------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Methanol EN 14110; ABNT NBR 15343</td>
<td>Others</td>
<td>multiple linear regression, principal component regression, partial least squares regression, poly-partial least squares regression, spline-partial least squares regression, stepwise multiple linear regression, interval partial least square regression, backward interval partial least square regression, forward interval partial least square regression, moving window partial least squares regression, changeable size moving windows partial least squares, searching combination moving windows partial least squares, modified changeable size moving windows partial least squares, successive projections algorithm, uninformative variable elimination, including uninformative variable elimination, simulated annealing, genetic algorithms, genetic algorithm interval partial least squares.</td>
<td>Balabin et al. (2011)</td>
<td>Balabin and Smirnov (2011)</td>
</tr>
<tr>
<td>Cloud Point EN 23015; ASTM D2500</td>
<td>Artificial Neural Network</td>
<td>universal quasichemical model</td>
<td>Lopes et al. (2008)</td>
<td></td>
</tr>
<tr>
<td>Pour Point ASTM D-97; ASTM D-5949</td>
<td>Others</td>
<td>universal quasichemical model</td>
<td>Lopes et al. (2008)</td>
<td></td>
</tr>
</tbody>
</table>
Prediction of the Oxidative Stability of Biodiesel Using Ann Based on its Physicochemical Properties: A Case Study

Oxidative stability is an important parameter in evaluating the quality of biodiesel fuel. Its determination is usually arduous, expensive, and analysis is very time consuming, and the results obtained are not always accurate due to experimental errors. This work is aimed at predicting the oxidative stability of biodiesel based on its physicochemical properties.

The changes that occur due to the biodiesel oxidation, such as the formation of hydroperoxides, which may polymerise with other free radicals to form insoluble sediments and gums, can compromise fuel quality, which means, effects on physicochemical parameters such as ester content, viscosity, density, iodine number and acid value. These alterations as a result of oxidation reactions can also cause the corrosion of fuel system components, hardening of rubber components and fusion of moving components (Monyem, 2001; Ortech Corporation, 1995; EN 14112/2003).

Rancimat is an official test method (EN 14112, EN 15751) used in almost all countries to determine oxidative stability of Biodiesel. It is an accelerated oxidation test in which the biodiesel to be tested is run at elevated temperatures (110 °C) whilst exposing the sample to a stream of purified air (10 L/h) accelerating the oxidation process of the biodiesel. After passing through the biodiesel, the air is fed into a collection flask containing distilled water and a probe to measure conductivity.

As the biodiesel sample degrades, the volatile organic acids produced are carried to the collection flask, and the conductivity of the solution is recorded by the probe. Oxidation stability will be given by the induction period, defined as the time between the start of the test and the sudden conductivity increase of the solution in the collection flask. This results in auto-oxidation in a few hours, instead of months.

Materials and Methods

Selected data of pure biodiesels related to the physicochemical properties, ester content, iodine number, viscosity, density, acid value and oxidative stability, used in the present work, were obtained from the interlaboratory official program of the Brazilian government, coordinated by the Brazilian National Agency of Petroleum, Natural Gas and Biofuels (from Portuguese: Agência Nacional do Petróleo, Gás Natural e Biocombustíveis - ANP).

The data corresponded to the results from the analysis that were carried out, and provided through reports, by different testing laboratories, participating in this important program in Brazil. These laboratories followed the official ANP procedures (ASTM D6751-11B, EN 14.214:2010 and NBR RANP 14/2012) for Biodiesel analysis. The biodiesel used was provided by manufacturers approved by ANP, and was obtained via methylate route.

The automatic module of the artificial neural networks of the Statistics 10.0 software was used in this study.

The MLP used in this study, consisted of 5 input variables (ester content, density, viscosity, acid value and iodine number), one hidden layer, in which the number of neurons varied from 3 to 30 during the training phase, and an output layer (induction period). The learning algorithm employed was the BFGS (Broyden-Fletcher-Goldfarb-Shanno) method that is a variation of the backpropagation.

BFGS is a quasi-Newton method that uses the quadratic Taylor approximation of function objective, whose idea is to make an iterative approximation of the matrix Hessian inverse (H). The current approximation Hi is used in each iteration to define the next descending direction of the method. Ideally, the approximations converge to the inverse of the Hessian matrix and they are carried out using Equation 1, where qi and pi are vectors determined by other equations (Silva, 1998).

$$H_{i+1} = H_i + \frac{p_i p_i^T}{p_i^T q_i} \left[ 1 + \frac{q_i^T H_i q_i}{p_i^T q_i} \right] - \frac{H_i q_i^T p_i}{p_i^T q_i}$$

(Eq. 1)

In this work, the following activation functions were tested: identity, logistic, hyperbolic tangent, sine and exponential.

In the supervised learning paradigm, the output values generated by the network are compared to the desired values output using Equation 2, which is the squared error cost function to be minimized (Sum of Squares (SOS)), where d(1), d(2), ..., d(q) are the output values of the training examples (desired output) and y(1), y(2), ..., y(q) are the values calculated by the network. The training happens until the error is below a threshold predefined or until the number of iterations or cycles is reached (Jain, Mao and Mohiuddin, 1996).
\[ E = \frac{1}{2} \sum_{i=1}^{p} || y^{(i)} - a^{(i)} ||^2 \]  
(Eq. 2)

Results and Discussion

The experiments were done to determine the best number of neurons in the hidden layer and to evaluate the performance of ANNs in predicting induction period of biodiesel. Under supervised learning, the desired output for the input variables was given to the network.

453 samples and their respective physicochemical parameters were used to generate the evaluated database. Several samples of this database were incomplete with respect to the physicochemical analysis. Missing data were supplemented with the median of each parameter. Besides the size of database, the type of data is also extremely important for predictive studies.

Database was divided into subgroups: 80% of the data were used for training purposes and 20% for testing purpose and also to verify the generalizability of the trained network. This was because, in order to make the prediction, the network learnt a rule using the training examples. The network was fed with the five parameters established before, as the main parameters influenced by the biodiesel oxidation, seconded the literature (Canha et al., 2012; Hoekman et al., 2012).

Results for the performance of the ten best ANNs with different numbers of neurons are presented in Table 3 for the database.

As we can see all networks showed good performance to predict the oxidative stability of biodiesel, however, preference should be given for models with simpler topologies with minimum error and maximum value of the correlation coefficient.

<table>
<thead>
<tr>
<th>Network</th>
<th>Topology</th>
<th>Correlation Coefficient</th>
<th>Error</th>
<th>Activation function</th>
<th>Hidden layer</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Training</td>
<td>Test</td>
<td>Training</td>
<td>Test</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>5-26-1</td>
<td>0.778388</td>
<td>0.889690</td>
<td>0.002324</td>
<td>0.002254</td>
<td>Tanh</td>
</tr>
<tr>
<td>2</td>
<td>5-6-1</td>
<td>0.785254</td>
<td>0.884360</td>
<td>0.002262</td>
<td>0.002456</td>
<td>Tanh</td>
</tr>
<tr>
<td>3</td>
<td>5-14-1</td>
<td>0.733176</td>
<td>0.885739</td>
<td>0.002728</td>
<td>0.002345</td>
<td>Tanh</td>
</tr>
<tr>
<td>4</td>
<td>5-10-1</td>
<td>0.788032</td>
<td>0.890015</td>
<td>0.002238</td>
<td>0.002323</td>
<td>Tanh</td>
</tr>
<tr>
<td>5</td>
<td>5-16-1</td>
<td>0.775358</td>
<td>0.884420</td>
<td>0.002261</td>
<td>0.002383</td>
<td>Tanh</td>
</tr>
<tr>
<td>6</td>
<td>5-13-1</td>
<td>0.801687</td>
<td>0.891741</td>
<td>0.002107</td>
<td>0.002237</td>
<td>Tanh</td>
</tr>
<tr>
<td>7</td>
<td>5-28-1</td>
<td>0.800537</td>
<td>0.884446</td>
<td>0.002119</td>
<td>0.002409</td>
<td>Tanh</td>
</tr>
<tr>
<td>8</td>
<td>5-8-1</td>
<td>0.816125</td>
<td>0.890790</td>
<td>0.001970</td>
<td>0.002216</td>
<td>Tanh</td>
</tr>
<tr>
<td>9</td>
<td>5-6-1</td>
<td>0.772202</td>
<td>0.889273</td>
<td>0.002385</td>
<td>0.002477</td>
<td>Tanh</td>
</tr>
<tr>
<td>10</td>
<td>5-22-1</td>
<td>0.781152</td>
<td>0.884469</td>
<td>0.002303</td>
<td>0.002346</td>
<td>Tanh</td>
</tr>
</tbody>
</table>

FIGURE 2 - CORRELATIONS BETWEEN EXPERIMENTAL VALUES (TARGET) AND PREDICTED (OUTPUT) INDUCTION PERIOD BY ANN/BFGS METHOD.

This work confirmed the general dependence of oxidative stability property (induction period) on physicochemical properties, ester content, iodine number, viscosity, and density.

Among the 10 networks that presented the best performance, one containing 8 neurons in the hidden layer was highlighted due to the best coefficients presented, whose medium correlation coefficient was 0.85 to this network.

Figure 2 shows the results obtained after application of the ANN chosen to predict oxidative stability. The performance of the model was carried out simultaneously for both sets, calibration and validation.

From this figure it is possible to observe that although some samples present anomalous behavior (outliers), in the general the predicted values are fairly uniform and well distributed along the straight line of correlation, with a higher number of samples presenting an oxidative stability lower than 10 h and a medium induction period value of 6.16 h. These results correspond to region and value expected for
the most of the Biodiesel biofuel samples. Predicted values for all samples range from 0.00 to 35.5.

Furthermore, in spite of its large calibration range, the model allows obtaining an excellent agreement between the values predicted by the model and those determined by the reference method. Therefore, these results confirm that this model can be used to predict the oxidative stability of biodiesel with values similar to the Rancimat ones.

Conclusions

The aim of this work was to present an update of the literature about the application of ANN to predict quality parameters of biodiesel. The work also presented a case study on the prediction of oxidative stability.

A first important conclusion is clear, namely that these computational tools are very important to solve complex problems in analytical chemistry. The published results show that both statistical models and the ANNs precisely can predict the properties of biodiesel.

The number of publications presented here is neither complete and can appear exhaustive but this is merely a sample of publications that demonstrate, on the one hand, the usefulness of applications of ANNs and statistical tools currently and, on the other hand also show how large the interest in biodiesel is and is growing increasingly.

The existence of many standards that have not yet been addressed by these tools and, on the other hand, the efficiency and importance observed of the application of statistical and computational tools directed to the quality parameters of Biodiesel indicated clearly the importance of continuity of these studies.

A case study showed the prediction of oxidative stability of Biodiesel. The method used 4 other official quality parameters as input data which avoided the need of additional experiments to determine the mass fractions of FAMEs by chromatography, usually required by the prediction methods for biodiesel quality parameters.

Results indicated that the proposed ANN method was able to predict the induction period with the medium error of 0.002, considered an adequate and small value. In terms of accuracy, the prediction method presented a good value whose medium correlation coefficient was 0.85 for the network number 8. The application of the model presented a medium value of IP of 6.6h. Good correlation was observed between the experimental values (target) and the predicted induction period (output) for all biodiesel samples by the method proposed.

Based on the results presented here, both the data from the recent literature as the case study on the prediction of oxidative stability of Biodiesel, indicated that ANNs offered a very important alternative method for Biodiesel quality, which should not be under-estimated, mainly considering that ANN was extremely versatile for mapping complex and nonlinear relationships.

ACKNOWLEDGMENT

The authors thank Brazilian National Agency of Petroleum, Natural Gas and Biofuels (ANP/CPT) from Brazil by the supplying the database on biodiesel analysis for this research and UFMA-LAPQAP/NEPE for the support and collaborations during the development of this research. The authors also thank the Brazilian agencies FAPDF, FINEP, and CAPES for financial support and provision of research fellowships and scholarships.

REFERENCES


Balabin RM, Lomakina EL. Support vector machine regression (SVR/LS-SVM) - an alternative to neural


Ferrão MF, Viera MS, Pazos REP, Fachini D, Gerbase AE, Marder L. Simultaneous determination of quality parameters of biodiesel/diesel blends using HATR-FTIR
spectra and PLS, iPLS or siPLS regressions. Fuel 2011,90(2):701-6.
Mahamuni NN, Adewuyi YG. Fourier transform infrared
spectroscopy (FTIR) method to monitor soy biodiesel and soybean oil in transesterification reactions, petrodiesel-biodiesel blends, and blend adulteration with soy oil. Energy Fuels 2009, 23(7):3773-82.


Martiniano LC; Gonçalves VRA; Yotsumoto Neto S; Marques EP; Fonseca TCO; Paim L; Gouveia AS; Stradiotto NR; Aucelio RQ; Cavalcante GHR; Marques EP; Marques ALB. Direct simultaneous determination of Pb(II) and Cu(II) in biodiesel by anodic stripping voltammetry at a mercury-film electrode using microemulsions. Fuel 2013, 103: 1164-67.


& Fuels 2011, 25(9): 3900-08.


