Biodiesel characterization using electronic nose and artificial neural network
Proceedings of European Congress of Chemical Engineering (ECCE-6)
Copenhagen, 16-20 September 2007

Biodiesel characterization using electronic nose and artificial neural network

D. S. Giordani[*],[a] H. F. Castro,[a] P. C. Oliveira,[a] A. F. Siqueira,[a]

[a] *Engineering School of Lorena, University of São Paulo, Lorena, Brazil*
Estrada Municipal do Campinho, s/n – 12602-810 – Lorena – SP – Brazil
[*] *giordani@dequi.eel.usp.br*

Abstract

Biodiesel is an important new alternative transportation fuel. Biodiesel processing and quality are close related. The process used to refine the feedstock and to convert it to biodiesel determines whether the fuel will meet the applicable specifications. In this work, a new characterization method is proposed using an electronic nose coupled to artificial neural networks. Four samples of biodiesel from different sources and one petrodiesel were analysed and could be recognized by the e-nose. Furthermore, an innovative semi quantitative method is proposed entirely based on the smellprints correlated by feed-forward artificial neural network. The promising results of this work indicate that e-nose can be used to identify the biodiesel source and as indicative assay when expensive equipments are not available.

Keywords: biodiesel, electronic nose, artificial neural network, green chemistry

1. Introduction

Biodiesel are alkyl ester obtained from vegetal or animal oils by catalysed transesterification reaction, bases are the most common catalysts worldwide used, but enzymatic catalysis presents a growing importance as well. Many feedstocks can be used to produce the parent oils and this usually varies with geographic availability. It is obvious that the esters obtained vary with the parent oils, because they will contain the corresponding fat acids from the oil.
Many countries already have specific legislation to regulate the use of biodiesel and its blends to petrodiesel as a commercial fuel. In Brazil, the use of biodiesel as fuel has been legally authorized since 2005, allowing up to a volume fraction of 2% in petrodiesel, this blend is commonly named as B2. This content will become mandatory in 2008 and in 2013 the biodiesel content in petrodiesel will be increased to a volume fraction of 5%. However, the increase in biodiesel production in Brazil indicates that B5 blend will be available to be mandatory in 2009.

The advantages of the use of biodiesel compared with petrodiesel are related to its derivation from a renewable resource, reducing dependence on and preserving petroleum, biodegradability, reduction of most exhaust emissions (with the exception of nitrogen oxides, NOx), higher flash point, leading to safer handling and storage, excellent lubricity, a fact that is steadily gaining importance with the advent of low-sulfur petrodiesel fuels, which have greatly reduced lubricity. Adding biodiesel at low levels (1–2%) restores the lubricity (Khote, 2005).

Some problems associated with biodiesel are its higher price, which in many countries is offset by legislative and regulatory incentives or subsidies in the form of reduced excise taxes; cold flow properties that are especially relevant in cold countries; stability when exposed to air (oxidative stability) and slightly increased NOx exhaust emissions (Khote, 2005). However, the air quality effect of 100% market penetration of B20 into on-road heavy-duty fleets in several major urban areas was examined in a study that employed pollutant inventory and air quality models; the results suggest that the NOx increase does not have serious air quality implications (Morris et al., 2003).

The characterization of biodiesel on its various aspects has been largely studied and reported in the literature. Some authors have reported chromatographic and spectroscopic methods to classify biodiesel on its mono-, di-, or triacylglycerol composition (Freeman et al. 1986; Foglia et al., 1987; Freeman et al., 1984). Concerning the determination of the concentration of biodiesel in blends with petrodiesel, different methods have been used including ^{1}H NMR spectroscopy (Khote, 2001), chromatography (Foglia, 2005) and infrared spectroscopy (Pimentel, 2006). Recently, Felizardo et al., in 2007 used multivariate near infrared spectroscopy for predicting methanol and water content in biodiesel.

Few articles are found reporting the determination of the parent oils from which a biodiesel has been produced. There are some important reasons to characterize a biodiesel in terms of its parent oils; probably the most important is concerned to fiscal laws. Some countries apply different policies depending on the feedstock. Another important reason is related to the fact the each oil determines specific chemical and physical properties on the biodiesel, what are essential data to handle it in appropriate way. The quantitative analysis in terms of the quantity of biodiesel in blend with petrodiesel is also important due to the necessity of supervising the correct application of environmental laws.

Electronic noses measurements are based on a change in resistance in an array of chemical sensors when exposed to a chemical vapor. Its use has been reported in the literature to determine the origin of a variety of stuffs such as honey (Benedetti, 2004), cigarettes (Dehan, 2004), wine (Lozano et al., 2005), olive oil (Cosio et al., 2006), spoiled beef (Panigrahi, 2006), sesame oil and gasoline (Sobanski, 2006).

Artificial Neural Networks (ANN's) are computer programs designed to simulate the way in which the human brain processes information, they gather their knowledge by detecting relationships and patterns in data and hence they are able to learn from experience, differently from common programming. The major capabilities of the ANN's are recognition of patterns and data correlation when no mathematical models are available. Some authors have used successfully this tool in order to enhance the potentiality of e-nose analysis (Martín, 2001; Lozano et al., 2005; Onkal-Engin, 2005; Hai, 2006).

In this work, a 32 sensor-based e-nose was used to recognize four samples of biodiesel and one of commercial petrodiesel, since each sensor produces one set of data, 32 sets form the smellprint of the analyzed substance. Principle components analysis was used to reduce the quantity of data and to group them into a three dimensional space. Artificial neural networks resources were used to correlate the quantity of biodiesel in petrodiesel blends to the smellprints, in such a way to obtain a semi quantitative essay.

2. Materials and Methods

2.1. E-nose

A Smiths Detection Cyranose 320 e-nose was used in this work. Its measurement is based on a change in resistance of each chemical sensor in the 32-sensor array when exposed to a chemical vapor. This is a differential measurement with the sensor response measured as $(R_{max}-R_o)/ R_o$, with R_o being the resistance during a baseline gas flow and R_{max} being the maximum resistance during exposure to the sample vapor as shown in Figure 1.

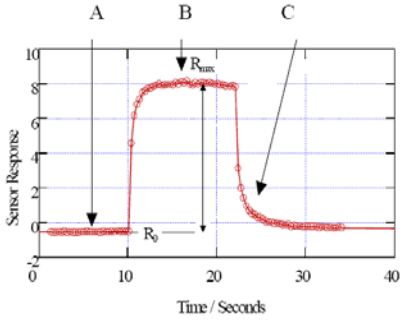

Figure 1: Typical sensor response through the (A) baseline purge, (B) sample exposure, and (C) sensor refresh.

The chemical sensors respond to the vapor headspace to which they are exposed. Across the array of unique sensors the responses are different and a response pattern is obtained that represents each particular headspace.

The sensor materials are thin films deposited across two electrical leads on an alumina substrate, creating the conducting chemiresistors. When the composite film is exposed to a vapor-phase analyte, the polymer matrix acts like a sponge and swells while absorbing the analyte. The increase in volume causes an increase in resistance because the conductive carbon-black pathways through the material are disrupted. When the analyte is removed the polymer releases the analyte and shrinks to its original size, restoring the conductive pathways. Each polymer used in the array is chemically unique and absorbs the analyte gases to a different degree, thus creating a pattern of differential response across the array. Figure 2 shows a typical sensor response to petrodiesel.

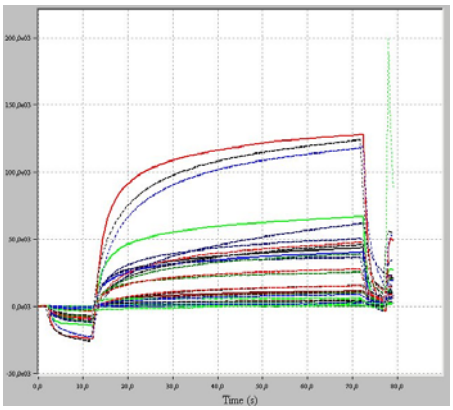

Figure 2: Response of 32-based sensor array when exposed to petrodiesel

2.2. Principal Component Analysis

As it can be seen, a great quantity of data is collected to each sample; occasionally it is not practical to treat all of them, because important information may be hidden inside the dataset, so it is necessary to use a technique to "clean" the dataset, reducing its size without loosing important information. The mathematical tool very often used to do that is Principal Component Analysis (PCA).

From the original dataset, the method generates a new set of data, called set of principle components, that is linear combination of the original set and have the characteristic of being orthogonal one to each other, in such a way that no redundant information exists. The principal components form an orthogonal basis of the total space.

There are several ways to construct an orthogonal basis to a number of data columns. In PCA technique, the first set of principal components is a single axis in the space, when each observation is projected over this axis, the resulting values form new variables. The variance of these variables is the greatest among all the possible choices to the first axis. The second set is other axis in the space, perpendicular to the first, which creates a new variable whose variance is also the greatest. The total set of principal components is as large as the original set, but is known that the sum of the variances of just few initial data exceeds in 80% the sum of the variances of the entire group. In this way, it is possible to reduce the dataset to a size whose sum of variances represents the entire dataset; in such a way that the *generating forces* existing in the original set still exists in the new one.

2.3. Artificial Neural Networks

There are many types of neural networks but all of them have the same basic principle. Each neuron in the network is able to receive input signals, to process them and to generate an output signal (Haykin, 2002). Each neuron is connected with at least another neuron, and each connection is represented by a real number, called weight coefficient, which reflects the degree of importance of the given connection to the neural network.

The most popular ANN is the backpropagation feed-forward type, whose architecture

is based on an input layer containing all the entrance variables that feed the network and an output layer that contains the responses of the ANN to the desired problem. All the layers between input and output are called hidden layers. There is no limit to the number of hidden layers, but one hidden layer with an arbitrarily large number of processing elements (neurons) is generally enough to solve the majority of problems, although some rare functions require two hidden layers to be well modelled.

The output value of the i th neuron x_i is determined by Eqs. (1) and (2), which holds:

$$x_i = f(y_i) \qquad (1)$$

$$y_i = b_i + \sum_{j=1}^{n} w_j x_j \qquad (2)$$

where y_i is the potential of the i th neuron, b_i is the bias coefficient and can be understood as a weight coefficient of the connection formally added to the neuron, n is the number of input connections no the i th neuron, w_j is the weight coefficient of the connection between the input j and the i th neuron and x_j is the value of the input j. The function $f(y_i)$ is the so-called transfer function.

One of the most widely used transfer function is the sigmoidal, but there are others such as hyperbolic tangent and linear.

In the supervised training, i.e., that one whose targets are already known for a given input dataset, biases and weight coefficients are varied in order to minimize the sum of the squared differences between the computed and required output values (targets), what is done by minimization of an objective function E:

$$E = \sum_{j=1}^{n} \tfrac{1}{2}\left(\mathrm{x}_j - \hat{\mathrm{x}}_j\right)^2 \qquad (3)$$

where x_j and $\hat{\mathrm{x}}_j$ are vectors composed, respectively, of the computed and targets activities of the output neurons; n is the number of neurons.

Although backpropagation is the most popular training algorithm, it is known that sometimes it is slow to converge and tends to fall into local minima; there exist some variations of this algorithm developed to optimise convergence. The Levenberg-Marquardt algorithm was designed to approach second-order training speed. This algorithm appears to be the fastest method for training moderate-sized feed-forward neural networks (up to several hundred weights). It also has a very efficient MATLAB® implementation. The Levenberg-Marquardt algorithm has proven to be very efficient in works where it has been used (Laugier, 2003; Yu, 2003).

The main advantage of neural networks is that they are able to use some information a priori hidden in data. The process of capturing the unknown information is done during the training step of the ANN, when one may say that the ANN is learning how to output a satisfactory response for an input dataset. In mathematic language, the learning process is the adjustment of the set of weight coefficients in such a way that some conditions are fulfilled.

One problem that happens frequently to ANN is called overfitting; it happens when the network is much more trained than it would be enough to reach generalization of the phenomenon. In fact, the ANN seems to memorize the training examples, presenting good results only to that training data and diverging to any others. To overcome this problem, the most used technique is early stopping. The training

dataset is divided in such a way that the first part is used to train the network and the second part is used as a test dataset. The error calculated by Eq. 3 is usually reduced as the training process goes on, but the test error increases when ANN begins to overfit. Therefore, the training process is stopped and the weights and biases are reverted to the values they had just before overfitting.

2.4. Bio and petrodiesel samples

Four different samples of biodiesel and one sample of petrodiesel were used; the viscosities for all samples are shown in table 1, viscosities were determined in Brookfield Model DV-II+pro viscometer at 40ºC and may furnish a reasonable idea whether the biodiesel contain parent oil remaining from uncompleted transesterification. The palm and the babassu samples were obtained by enzymatic catalysis (Moreira, 2006); chicken grease sample was obtained by basic catalysis (Amaral, 2007), all in laboratory scale. Beef tallow sample (BIOMAX®) was obtained industrially and gently furnished by Fertibom Ltd. and used as received.

Parent oil	Viscosity (mm^2/s)
Palm	7,50
Babassu	4,50
Chicken grease	6,90
Beef tallow	5,38
Petrodiesel	3,74

Table 1: viscosities of bio and petrodiesel samples used in this work.

Five 1mL samples of each pure bio or petrodiesel studied were put into 40mL headspace vials and enough time was left to attain liquid-vapor equilibrium at room temperature, the headspace of each vial was then exposed to the e-nose in order to form a set of data large enough for statistical treatment. In order to form a dataset containing ten exposures, each vial was exposed twice to the e-nose. E-nose was adjusted to run 10s purging air to form the baseline and 30 seconds aspirating the headspace, this time was enough to the majority of the sensors to reach steady state. Purge and aspiration were at a sampling speed of 120mL/min. Six sensors were disabled because their resistance did not attain steady state.

Additionally, four B20 blends (20% of biodiesel with 80% of petrodiesel) were prepared with each biodiesel sample. Ten 1mL samples were prepared to each B20 blend and exposed to the e-nose once. The reason not to prepare five samples and exposing them twice such as in the former study was to obtain a homogeneous set of data, given that when working with a blend, where one of the components is much more volatile than the other one, the time to attain headspace equilibrium is more important and could influence the e-nose readings. Blends exposure were regulated to have 10s of air purging to the baseline at 120ml/min and 60s of headspace aspiration at 180ml/min to attain steady state.

The responses of the 26 sensors used were normalized using a simple weighting method type $(\Delta R/Ro)_i = (\Delta R/Ro)_i / \Sigma |\Delta R/Ro|_j$ to remove the effects of response size on

the smellprint pattern.
Finally, B2, B5, B10 and B20 blends with babassu biodiesel were prepared, ten samples of each were exposed to the e-nose. The baseline of the smellprints had to be conditioned with the pure petrodiesel, i.e. the base line was formed not by pure air purging, but by pure petrodiesel headspace, in order to enhance the sensitiveness of the sensors regarding to biodiesel presence in the mixture. No normalization techniques were used here to keep the original resistances registered by sensors, because in this case the response sizes of the smellprints are used to be correlated to concentration.

3. Results and Discussion

3.1. Pure Biodiesel

The e-nose was capable to recognize the smellprints of each diesel and to group them based on PCA techniques, as can be seen on Figure 3. Based on this result, it is possible to say that the e-nose is suitable equipment to recognize with efficiency and rapidity any of the presented samples. The results indicate that other biodiesel samples could be well recognized because the distance among the groups is large enough to hold other groups.

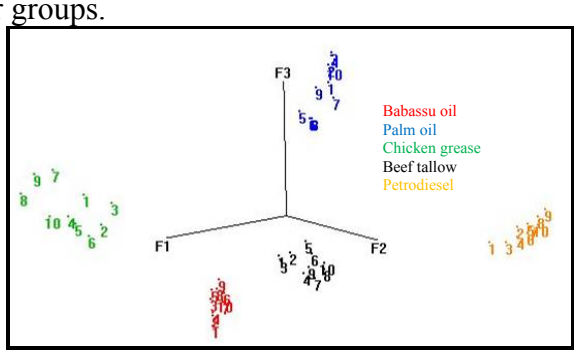

Figure 3: PCA projecting plot of four different samples of biodiesels and petrodiesel

Validation studies to this analysis method were done by repeating exhaustively recognition tests using the same headspace technique used in training step, i.e, at normal operating conditions. The test attained 100% of success. Afterwards, a qualification test was carried out by testing samples at the boundaries of operating conditions, which was accomplished doing the sample headspace aspiration with open vials, to reproduce possible field conditions. This test attained approximately 93% of success. The main purpose of method validation and qualification is to insure the repeatability, reproducibility and robustness of the method and model when using the e-nose to identify samples.

3.2. B20 blends

In a first essay, samples of the four blends were exposed to the e-nose according to the procedure in section 2.4. As it can be seen in Figure 4a, beef tallow and chicken

grease B20 blends had almost the same behavior and were grouped by PCA technique almost together, what turns impossible their recognition one from another in only one essay. The recognition could be done at least in terms of animal origin biodiesel and vegetal origin biodiesel. Nevertheless, in such situations it is recommended to perform the essay in two steps, what is commonly called tiered analysis. Then, a new exposure to the e-nose was done, this time with only chicken grease and beef tallow B20 blends, as a result both samples could be distinguished by the e-nose as shown in Figure 4b.

Thus, the strategy recommended to this set of B20 blends is tiered methodology. It is likely that this kind of occurrence may happen with other sets of blends when the same tactic should be used.

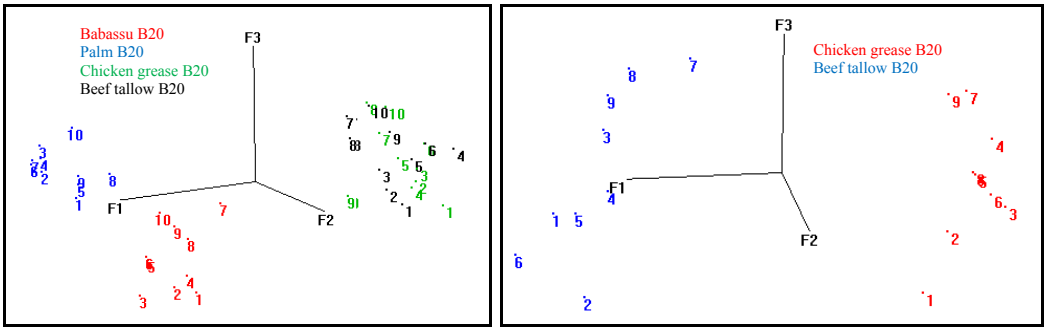

Figure 4: PCA projecting plot of (a) four different B20 blends of biodiesel
(b) only beef tallow and chicken grease B20 blends

Validation of this method was carried out as well as in the former one; results indicate that 88% of the essays were considered correct. Qualification of the method, with aspiration of open vials headspace achieve only 60% of correct determination, what indicates that, to B20 blends, essays should be done with the strict methodology of closed vials. In fact this is not surprising, because even though the groups are well separated one from another, the real distances among them are smaller than in pure biodiesel case.

3.3. Semi quantitative analysis

When a set of pure petrodiesel, B2, B5, B10 and B20 blends of babassu biodiesel were presented to the e-nose, the results of PCA projecting plot was not exciting as can be seen in Figure 5, where no group separation can be found, due to the great proximity of the resistances of the sensors when exposed to the vapors. A cross validation test using leave-one-out method showed a 68% of right predictions, which was not considered reasonable. However, a meticulous analysis of the smellprints and of the response intensities of the sensors, when no data regularization was applied, has shown two important facts: there were slight differences among them and that these differences could be greater if the baseline would be formed not by air purging, but by pure petrodiesel headspace. If compared to the human nose, it could be said that the e-nose is more sensible to few quantities of babassu biodiesel in the

blend when it is previously "adapted" to the petrodiesel smell.

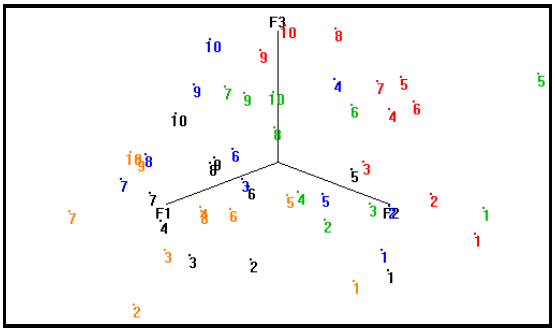

Figure 5: PCA projection plot of Babassu B0, B2, B5, B10 and B20 blend samples

Thus, since other authors have successfully used ANN resources to enhance the power of the e-nose (Martín, 2001; Lozano et al., 2005; Onkal-Engin, 2005; Hai, 2006), it was decided to apply this methodology.

The smellprint of each sample exposure to the e-nose was collected to form the training dataset of a feed-forward backpropagation ANN. Since there were ten samples of five blend concentration (B0, B2, B5, B10 and B20), a set of fifty essays was formed. Five essays were separated to form a testing group, so early stopping could be used to accelerate the search for global minima in the training process.

A 26:3:1 architecture was initially chosen to be tested, where 26 is quantity of neurons the input layer, corresponding to the number of sensors activated; 3 is the quantity of neurons in the hidden layer and 1 is number of neurons in the output layer, corresponding to the babassu biodiesel concentration in the blend. The network was trained using Levenberg-Marquardt algorithm and after some random weight initializations, the training process was ready.

The trained ANN was then submitted to a validation test, when a new set of blends was presented to the network. In order to test the capacity of the ANN to predict different values from those used in training, besides B0, B2, B5, B10 and B20, two new blends were added in the validation test: B7 and B15. As it can be seen in Figure 6, predictions were very close to the real values.

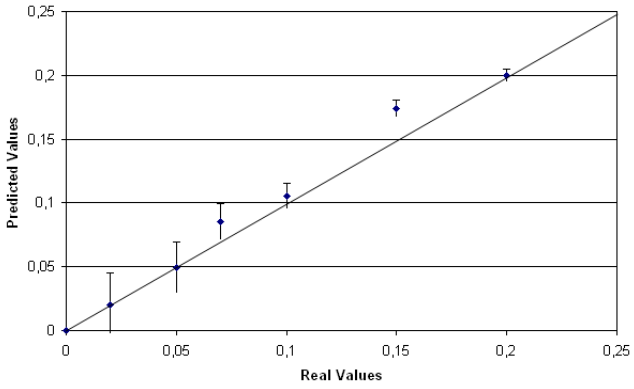

Figure 6: Correlation of real values and ANN predicted values

It can also be observed that to B7 and B15 blends predicted values were not as close as they were to the rest of the group, what indicates that the generalization capacity of the ANN need to be improved. The mean error to the validation was 1,71% if only B0, B2, B5, B10 and B20 samples are considered and 7,42% when B7 and B15 were included. It is obvious that better ANN's could be searched, an increase in the quantity of neurons in the hidden probably would have improved generalization, on the other hand the semi quantitative character of the method was already demonstrated.

Conclusions

The capacity of the e-nose to recognize pure biodiesel samples was shown even in conditions slightly different of those used to train it, what demonstrates that the equipment is perfectly suitable to perform field essays in a very rapid and practical way.

The electronic nose was also tested to recognize the origin of the biodiesel in B20 blends; the study revealed that it is also suitable to this task; however the essays must be always done in the same conditions of the training, i.e. using closed headspace vials. In some cases it is necessary to apply the tiered method to recognize some samples.

A semi quantitative method was successfully developed to determine de quantity of biodiesel in blends of babassu biodiesel with petrodiesel using artificial neural networks to enhance the e-nose capacity. This technique is worth of further studies to improve generalization of the network predictions.

Electronic nose coupled to artificial neural networks is a very promising technique to biodiesel characterizations, especially taking into consideration the portability and the price of the equipment compared to the traditional techniques.

References

Amaral, T. B, Gonçalves, A. R., Almeida C.R.O. and Rocha, G.J.M. (2007) *29th Symposium on Biotechnology for Fuels and Chemicals*, Denver-CO, abstracts, p. 178.

Benedetti, S., Mannino, S., Sabatini, A. G., Marcazzan. L. G., (2004) *Apidologie,* 35, p. 1.

Cosio, M.S., Ballabio, D., Benedetti, S., Gigliotti, C., (2006) *Analytica Chimica Acta*, 567, p. 202.

Dehan L., Gholam Hosseini H., Stewart, J. R., (2004) *Sensors and Actuators B,* 99, p. 253.

Felizardo, P., Baptista, P., Menezes, J. C. and Correia, M. J. N., (2007) *Analytica Chimica Acta*, 595, p. 107.

Foglia, T.A., and K.C. Jones, (1997) *J. Liq. Chromatogr.Relat. Technol.* 20, p. 1829.

Foglia, T.A., Jones, K.C., Phillips, J.G., *Chromatographia,* (2005) 62, p. 115.

Freedman, B., E.H. Pryde, and W.F. Kwolek, (1984) *J. Am. Oil Chem. Soc.* 61, p.1215.

Freedman, B., W.F. Kwolek, and E.H. Pryde, (1986) *J. Am. Oil Chem. Soc.* 63, p. 1370.

Hai, Z., Wang, J., (2006) *Sensors and Actuators B,* 119, p. 449.
Haykin, S., N*eural networks – a comprehensive foundation,* (1999) 2nd Ed., Macmillan Col. Pub. Co., New York.
Khote, G. (2005). *Introduction, in: Khote et al. Biodiesel Handbook*, AOCS Press, Champaign, Illinois, p. 9.
Knothe, G., *J. Am. Oil Chem. Soc.,* (2001) 78, p. 1025.
Laugier, S. and Richon, D., (2003) *Fluid Phase Equilibria*, *210*, p. 247.
Lozano, J., Santos, J.P., Horrillo, M.C., (2005) *Talanta*, 67, 610–616.
Martín, Y. G., M. Oliveros, M. C. C. Pavon, J. L. P., Pinto, C. G., Cordero, B, M., (2001) *Analytica Chimica Acta,* 449, p. 69.
Moreira, A. B. R.; Pérez, V. H. ; Zanin, G. M.; Castro, H. F., (2006) *In: CHISA, Praga. Proceedings CHISA 2006*, 1. p. 1.
Morris, R.E., Polack, A.K., Mansell, G.E., Lindhjem, C., Jia, Y. and Wilson, G., (2003) *Summary Report. National Renewable Energy Laboratory*, NREL/SR-540-33793, Golden, CO.
Onkal-Engina, G., Demira, I., Engin, S. N., (2005) *Environmental Modelling & Software*, 20, p. 843.
Panigrahia, S., Balasubramaniana, S., Gua, H., Logueb, C., Marchelloc, M., (2006) *LWT*, 39, p. 135.
Pimentel,M. F., Ribeiro, G. M. G. S., Cruz, R. S., Stragevitch, L., Pacheco Filho, J. G. A., Teixeira, L. S. G., (2006) *Microchemical Journal*, 82, p. 201.
Sobanski, T., Szczurek, A., Nitsch, K., Licznerski, B.W., Radwan, W., (2006) *Sensors and Actuators B*, 116, p. 207.
Yu, D. L. and Gomm, J. B., (2003) *Control Engineering Practice*, 11, p. 1315.