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

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

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

Thermooxidative Properties of Biodiesels and Other Biological Fuels

Javier Tarrío-Saavedra[1], Salvador Naya[1], Jorge López-Beceiro[1],
Carlos Gracia-Fernández[2] and Ramón Artiaga[1]
[1]Escola Politécnica Superior, University of A Coruña, Ferrol
[2]TA Instruments, Madrid
Spain

1. Introduction

The aim of this chapter is to show how thermooxidative properties of biological fuels can be evaluated by pressure differential scanning calorimetry (PDSC) and used to correctly classify the fuels studied. The onset oxidation temperature (OOT) is an important parameter for estimating the oxidation stability that can be evaluated by the ASTM method. Nevertheless, in addition to the OOT, other meaningful information can be extracted from the PDSC tests. That additional information provides a better understanding of the thermooxidative process, allowing for identifying subtle differences between similar fuels. In fact, the following lines show that the features extracted from heat flow curves obtained by PDSC allow to characterize and to differentiate each type of fuel respect to the other ones if the adequate statistical tools are applied. Thus, the proposed statistical analysis of the PDSC curves allows to classify the different fuels types chosen for this study: two types of biodiesel, seven different classes of edible oils and two wood species. The statistical study consisted of the application of Analysis of Variance (ANOVA) procedures and the implementation of a simulation study, using parametric bootstrap and methods of multivariate supervised classification as Linear Discriminant Analysis (LDA), Logistic Regression and Naïve Bayes classifier.

Studying the thermooxidative properties of a fuel is important attending to various reasons. For example, vegetable oils are protected against oxidation thanks to antioxidants that precisely removed during the production process of biodiesel. For this reason, biodiesel is not stable, being susceptible to oxidation to a greater or lesser extent due to several factors including the presence of air, temperature, light, presence of hydroperoxides and antioxidants (Dunn, 2005; Knothe & Dunn, 2003; Knothe, 2007). The products resulting from the oxidation of biodiesel can damage internal combustion engines, it is therefore essential to study the oxidation stability of biodiesels. In the case of vegetable oils, they can produce significant changes in the salubriousness of food when the same oil is used repeatedly to fry due to the possible oxidation processes produced at the relatively high temperatures (Vorria et al., 2004). For this reason, the thermal stability to oxidation is an important parameter for oils. The study of thermooxidative characteristics of the species of wood is not as common as

in oils or biodiesel. However, this is justified as it would allow to estimate the resistance to combustion in an oxidizing atmosphere, under similar conditions to wildfire.

The thermal analysis techniques used to measure thermooxidative stability are: thermogravimetry (TG), differential scanning calorimetry (DSC) and PDSC. The oxidation stability can be obtained using the TG technique (a) measuring the increase in sample weight due to absorption of oxygen, (b) measuring the temperature corresponding to maximum weight and (c) the temperature at the beginning of oxidation (Van Aardt et al., 2004). DSC and PDSC techniques can be applied to study the exothermic oxidation process. The PDSC provides results in a shorter time than DSC, further reducing evaporation in the sample. It is also important to note that using PDSC we can estimate the oxidation stability under pressures similar to those operating in a diesel engine. The values that are determined to study the oxidation stability by DSC and PDSC are the oxidation induction time (OIT) and the the onset oxidation temperature (OOT). High values of both parameters are related to a high oxidative stability. The two methods have been used by several authors to study the oxidation stability of biodiesel (Knothe, 2007; Moser et al., 2007; Dunn, 2006; Xu et al., 2007, Polavka et al., 2005), findding correlations with other procedures (Dunn, 2005; Tan, 2002). The OOT parameter measures the degree of oxidative stability of a substance at a constant heating rate, both at high pressure and high temperature. It is a non isothermal dynamic method. The procedure for calculating the OOT is explained in ASTM E2009 (2008). Recent results concerning the characterization of thermooxidative fuels such as biodiesel or edible oils can be found in (Tarrío-Saavedra et al., 2010; Artiaga et al., 2010; López-Beceiro et al., 2011).

2. Materials

In the present chapter, three different types of fuels are tested:
1. Two types of biodiesel: obtained from the soybeam and from the palm.
2. Four classes of vegetable oils: soy, sunflower, corn and two olive oil spanish varieties named hojiblanca and picual.
3. Two species of comercial wood: *Pinus sylvestris* (Scots pine) and *Eucalyptus globulus*.

2.1 Biodiesel

Biodiesel is a liquid biofuel made from natural fats such as vegetable oils or animal fats through a process of esterification and transesterification. The resulting substance of these transformations can be applied as a partial replacement of petroleum products. The reaction of the base oils with a low molecular weight alcohol and a catalyst (usually sodium hydroxide), resulting in fatty acids formed by long chains of mono-alkyl esters which are very similar to "diesel " derived from petroleum. The commercial biodiesel used today are mixed with other fuels. In this paper we have studied two types of pure biodiesel, obtained from the soybeans and, on the other hand, from palm oil. They have been supplied by Entaban Biofuels Galicia, SA (Ferrol, Spain). See Table 1.

2.2 Vegetable oils

Table 2 shows the chemical composition retrieved from the USDA National Nutrient Database for Standard Reference-22 (USDA, 2009). This table is an indication to compare

Test	Soybean	Palm	Standard
Humidity	193 μg/g	216 μg/g	EN ISO 12937
Neutralization number	0.37 mg KOH/g	0.31 mg KOH/g	EN 14104
Ester content	99.3 %	97.3%	EN 14103
Methyl esters of linolenic acid	6.45 %	1.0%	
Iodine index	124.5 g I2/100g	66.1 g I2/100g	EN 14111
Free Glycerol	0.002 %	0.003 %	EN 14105 EN 14106
Total Glycerol	0.14 %	0.19 %	EN 14105
Monoglycerides	0.47 %	0.62 %	EN 14105
Diglycerides	0.10 %	0.15 %	EN 14105
Triglycerides	0.003 %	0.05 %	EN 14105
Stability to oxydation	6.76 h	12.5 h	EN 14112
Cold filter plugging point	-3 °C	9 °C	EN 116
Methanol	0.04 %	0.1 %	EN 14110
Sulphated ashes	0.003 %	0.002 %	ISO 3987
Density	877.7 kg/m³ at 23 °C	872 kg/m³ at 21 °C	------------
Density at 15°C	883.5 kg/m³	876.3 kg/m³	ISO 3575 EN ISO 12185
Kinematic viscosity at 40°C	4.1 mm²/s	4.4 mm²/s	EN ISO 3104
Na+K	< 1 μg/g	< 1 μg/g	EN 14538
Ca+Mg	< 1 μg/g	< 1 μg/g	EN 14538
Phosphorous	< 1 μg/g	< 1 μg/g	EN 14107

Table 1. Characteristics of soybean and palm based biodiesel studied.

different types of oils. While poliinsaturated acids like linoleic acid are abundant in corn and sunflower oils, the monoinsaturated oleic acid is the predominant fatty acid in olive oil. The levels of palmitic and estearic acids, which are saturated fatty acids in sunflower oil are lower than the others. Other components of vegetable oils are the acilglycerides, phospholipids, and non-glycerides compounds as vitamin E, vitamins D and A, sterols, carotenoids, methyl sterols and squalene.

2.3 Wood species

The wood is mainly composed by three components that conditon the degradation of wood in an inert atmosphere (Alén et al., 1996). These are hemicellulose, cellulose and lignin (Yang et al., 1999; Alén et al., 1996; Grønli et al., 2010). Cellulose represents about 40 and 60% in the overall weight of dry wood, while lignin represents the 23–33% in softwoods and the 16–25% in hardwoods; finally, the hemicellulose represents the 25–35% (Miller et al., 1999; Grønli et al., 2010). The hemicellulose decomposes at 200-260 °C, the cellulose at 240-350 °C and the ligning in a temperature range between 280-500 °C (Alén, 1996; Wang, 2009; Mohan, 2006). Therefore, changes in the thermooxidative stabiliy are expected due to percentage differences in these 3 componens. Percentages that characterize the different wood species.

	Common name	Sunflower mid-oleic	Soybean	Corn
Total fat/g		100	100	100
Saturated fat/g		9	15.6	12.9
14:00	Myristic acid	57	0	24
16:00	Palmitic acid	4219	10455	10580
17:00	Margaric acid	37	34	67
18:00	Stearic acid	3564	4436	1848
20:00	Arachidic acid	297	361	431
22:00	Behenic acid	836	366	0
Monounsaturated fat/g		57.3	22.8	27.6
16:1 undifferentiated	Palmitoleic acid	95	0	114
16:1 c		95	N.A.	114
17:01		N.A.	0	N.A.
18 undifferentiated	Oleic acid	57024	22550	27335
18:1 c		57024	22550	27335
20:01	Gadoleic acid	211	233	129
Polyunsaturated fat/g		29	57.7	54.7
18:2 undifferentiated	Linoleic acid	28925	50960	53510
18:2 n-6 cc		28703	50422	53510
18:2 tt		N.A.	533	N.A.
18:2 i		219	N.A.	286
18:03	Linolenic acid	37	6789	1161
18:3 n-3 ccc		37	6789	1161
Total trans fatty acids/g		0.2	0.5	0.3
Total trans-polyenoic fatty acids/g		0.2	0.5	0.3
Total omega-3 fatty acids		37	6789	1161
Total omega-6 fatty acids		28925	50422	53510
Tocopherols		41.08	94.64	14.3

Table 2. Chemical composition of sunflower, soybean, corn and olive oil, retrieved from the USDA National Nutrient Database for Standard Reference-22 (USDA, 2009).

3. Data collecting

A design of experiments consisting of 1 factor (type of fuel) at 9 different levels (soy biodiesel, palm biodiesel, sunflower oil, soy oil, corn oil, hojiblanca olive oil, picual olive oil, eucalyptus wood and Scots pine) was done. Three samples per each fuel type were considered, capturing the existing variability. In fact, this sampling process seeks to obtain a compromise between capturing the existing variability and the minimization of the time of the experimental test. The tests are carried out by PDSC to study the oxidation stability of the fuels and to compare these materials according to this concept. The PDSC tests were performed in a TA Instruments pressure cell mounted on a Q2000 modulated DSC. The experimental conditions were the following: T-zero open aluminum pan, a heating rate of 10 °C min^{-1} from room temperature to 300 °C -taking into account the recommendations to obtain a better oxidation peak (Riesen & Schawe, 2006)-, sample mass in the 3-3.30 mg

range, and an oxygen pressure of 3.5 MPa, applying a flow rate of 50 mL min^{-1} according to the ASTME2009 method. The experiments were manually stopped once the end of the exotherm was reached.

The Universal Analysis software supplied by the company TA was utilized to calculate the OOT using the standard E2009. The standard determines that the OOT corresponds to the temperature assigned to the crossover point between the tangent to the curve of heat flow at the point of maximum slope and tangent to the curve just before the occurrence of the peak corresponding to oxidation (which coincides with the baseline).

4. Analysis of variance (ANOVA)

The analysis of variance is a statistical tool performed to study the dependence of a quantitative variable with respect to one or more qualitative variables. In this chapter, an experimental design consisting of an nine-level factor was performed. The quantitative variable is the parameter OOT, an indicator of oxidation stability, while the factor is the type of fuel. The F test allows testing whether the mean OOT values for each fuel type are statistically equal or, conversely, there are at least one mean different (Maxwell, 2004).

H_0: $m_1 = m_2 = ... = m_9 = \mu$

H_1: at least one $m_i \neq \mu$, where μ is the global mean

If the before mentioned test is significant, Tukey's test can answer the question of which means are really different. Tukey's test applied to this case, provides information about what levels or types of fuel present OOT values statistically different (Maxwell, 2004). The significance level used in this work is 0.05.

5. Classification methods

The process of assigning a p-dimensional observation to one of several groups predetermined is called supervised classification. The principal aim is to obtain a discriminant function that summarizes the information corresponding to the different p variables that define a sample according to an indicator, with which each observation can be correctly classified as belonging to a group. In the statistical literature can be found several methods developed to address the classification problem.

5.1 Linear discriminant analysis

One of the most popular techniques in classification was proposed by Fisher (Fisher, 1936), this approach is called linear discriminant analysis (LDA) and basically divides the sample space into subspaces through the use of hyperplanes that allow to better separating the groups studied. The assumptions for the use of LDA are: multivariate normality and equal covariance matrices between groups. Under these assumptions, the LDA is based on finding a linear combination of features that describe or separate two or more classes of objects or events. The resulting combination can be used as a linear classifier, or more commonly, to reduce the dimension of the problem before a subsequent classification.

LDA is closely related to other statistical techniques such as analysis of variance (ANOVA) and regression analysis, however, in these two techniques, the dependent variable is a number, while in LDA is a categorical variable (class labels). Other statistical procedures related to LDA are the Principal Component Analysis (PCA) and Factor Analysis (FA), used when you look for linear combinations of variables that better explain the data.

Although the terms LDA and Fisher linear discriminant analysis are commonly used to indicate the same procedure of supervised classification, in fact, the early work of Fisher (Fisher, 1936) does not imply the assumptions of normality and equal covariances, undertaken by LDA.

The LDA has been successfully applied in fields as diverse as engineering, economics, computing science, biology, etc. Recently, the LDA has been applied in some works related to the classification of weeds (Lopez-Granados et al, 2008) and the classification of different species of wood through the use of features extracted by image processing (Mallik et al., 2011).

5.2 Logistic regression

The Logistic model is currently applied to these cases where the explicative variables (or set of different features) do not have a multivariate normal distribution (McLachlan, 2004). Considering only two classes (*C1* and *C2*), the Logistic regression equation used to solve this classification problem is the following[1]:

$$p = \frac{e^{\alpha + \beta' x}}{1 + e^{\alpha + \beta' x}} \tag{1}$$

where $p = P(Y = C1 \mid x)$ is the posterior probability of Y equal to *C1* class, $log\ (p/(1-p)) = a + \beta' x$ is the logit transformation, x is the p-dimensional vector of features or explanatory variables, β is a vector of *p* parameters and $p/(1-p)$ the odds ratio. Nevertheless, in the present study is necessary to use a classification model that could be applied in the case of the existence of multiples classes. This can be solved using the logit model generalized to more than two populations, i.e. for qualitative response with more than two possible classes. If *G* population are supposed, then, defining *p* as the probability that the observation *i* belongs to the class *g*, it is possible to write[2]:

$$p_{ig} = \frac{e^{\beta_{0g} + \beta'_{1g} x_i}}{1 + \sum_{j=1}^{G-1} e^{-\beta_{0j} - \beta'_{1j} x_i}} \tag{2}$$

This equation[2] is an estimator of the posterior probabilities, i.e. the probabilities of belonging to a specific class, given the values of a vector of features (values of x). The $p_{ig,}$ or posterior probabilities, satisfy a multivariate logistic distribution. The following expression is used to do the different possible comparisons[3]:

$$\frac{p_{ig}}{p_{ij}} = \frac{e^{\beta_{0g} + \beta'_{1g} x_i}}{e^{-\beta_{0j} - \beta'_{1j} x_i}} \tag{3}$$

The logistic regression has been applied for classifying species of wood through the use of features extracted by image segmaentation (Mallik et al., 2011).

5.3 Bayes Naïve classifier

Naïve Bayes classifier is a supervised multivariate classification technique based on Bayes theorem, particularly suitable when the dimension of the vectors of features or inputs is considerably high. Calculating the posterior probability for an event among a group of

possible outputs, $X = \{x_1, x_2, ..., x_d\}$, is intended. That is, using Bayes rule we intend to calculate the probability that a sample belongs to a particular class, C_j, from a group of possible classes $C = \{c_1, c_2, ..., c_k\}$, given some particular values corresponding to the characteristics that define the sample. Using Bayes rule, the probability that X belongs to C_j or posterior probability is[4]:

$$p(C_j \mid x_1, x_2, ..., x_d) \propto p(x_1, x_2, ..., x_d \mid C_j) p(C_j) \tag{4}$$

Using Bayes' rule, we estimate the class of the event or sample using the class corresponding to the largest posterior probability obtained. Since Naïve Bayes assumes that the conditional probabilities of the independent variables are statistically independent, the posterior probabilities can be rewrite as[5]:

$$p(C_j \mid X) \propto p(C_j) \prod_{k=1}^{\alpha} p(x_k \mid C_j) \tag{5}$$

In addition, due to the assumption that the predictor variables are statistically independent, we can reduce the size of the estimated density function using a kernel estimation consisting of one dimension.

The Naïve Bayes classifier can be modeled with normal, log-normal, Gamma and Poisson density functions.

Naïve Bayes method appears in the 80's and is the supervised classification method most popular based on the Bayes rule. Several variants and extensions of the Naïve Bayesian classifier have been developed, for example, Cestnik (Cestnik, 1990) developed the m-estimations of the posterior probabilities and Kononenko (Kononenko, 1991) designed a semi-naïve Bayesian classifier that goes beyond the "naive" and detects dependencies between attributes. The advantage of fuzzy discretization of continuous attributes in the Naïve Bayesian classifier is described in the work of Kononenko (Kononenko, 1992). Langley (Langley, 1993) studied a system that uses the Naïve Bayesian classifier at the nodes of decision trees. Other recent works are those for Webb et al. (Webb et al, 2005) and Mozina et al. (Mozina et al., 2004). This technique has been used successfully in classification problems of spam and in areas such as medicine (to resolve, among other tasks, medical diagnosis), acoustic (automatic classification of sound and voice), image classification (Kononenko, 2001; Tóth et al., 2005; Mallik et al., 2011).

5.4 *K* nearest neighbors (KNN)

K Nearest Neighbors (KNN) is a non-parametric supervised classification method, which has been used successfully in populations where the assumption of normality is not verified. This assumption is required by traditional techniques such as linear discriminant analysis. We can summarize the KNN operation in the following three points:

1. A distance is defined between samples (represented by feature vectors), usually the Euclidean or Mahalanobis distances.
2. The distances between the test sample, x_0, and the other samples are calculated.
3. The *k* nearest samples to those that we want to classify are selected. Then, the proportion of these *k* samples belonging to each of the studied populations is calculated. Finally, the sample x_0 is classified within the population corresponding to

the highest existing frequency. Among the different methods available for choosing the value of k, the minimization of error of cross validation is one of the most used.

KNN method was introduced by Fix and Hodges (Fix & Hodges, 1951). Later, he shown some of the formal properties of this procedure, for example, that the classification error rate is bounded by twice the Bayes error value when you have an infinite number of samples for classifying and k is equal to 1 (Cover & Hart, 1967). Once developed the formal properties of this classifier, he established a line of research that goes up today, highlighting the work of Hellman (Hellman, 1970), which show a new approach to rejection, Fukunaga and Hostetler (Fukunaga & Hostetler, 1975), which sets out refinements with respect to the Bayes error rate, or those developed by (Dudani, 1976) and Bailey and Jain (Bailey & Jain, 1978), in which new approaches were established to the use of weighted distances. Other interesting work on the subject is related to soft computing (Bermejo & Cabestany, 2000) and fuzzy methods (Jozwik, 1983, Keller et al., 1985). Recent interesting papers are those of Bremner et al. (Bremner et al., 2005), Nigsch et al. (Nigsch et al., 2006), Hall et al. (Hall et al., 2008) and Toussaint (Toussaint, 2005). They are also very interesting applications of this algorithm to the analysis of functional data (Ferraty & Vieu, 2006).

The development of computer tools in recent years and the creation of the information society have led to that the technique KNN be used successfully in such diverse fields as chemistry, biology, medicine, computer science, genetics and materials science (Tarrío-Saavedra et al., 2011; Mallik et al, 2011).

5.5 Validation procedure: Leave-one-out cross validation

When we want to classify samples using supervised classification methods, working with training and testing data, extracted from the observed instances, is necessary. Each instance in the training set consists of the corresponding class label and a vector of several sample features. The aim of the classification methods applied is to produce a model, using the training sample, to estimate the class labels corresponding to each data instances corresponding to the testing set for which we only know the features. Leave-one-out cross-validation is the procedure used to obtain the probabilities of correct classification for each test sample and, therefore for comparing the different classification methods proposed. This is a technique widely used for the validation of an empirical model, especially suitable for working with small samples sizes. This procedure consists on the following steps:

1. One instance is leaving out: the testing sample.
2. Then, a model is obtained using the remaining samples (the training sample).
3. Finally, the developed model is used for classifying the left out instance.

This sequence is repeated until all the instances are left out once. The percentages (measured as per one) of correct classification are obtained using this procedure.

All the classification methods have been implemented using R statistical package

6. Parametric bootstrap resampling

In the case of the parametric bootstrap, the model from which data was generated of the original sample is known or assumed, ie the type of distribution is known. Therefore, successive resamplings are obtained by substituting the parameters of the distribution of probability corresponding to the studied variables by the maximum likelihood estimators, calculated from the original sample. In the present chapter the normal distribution of the

features is assumed. In addition, we suppose a model where the observations (the chosen features) are independent, i.e. a diagonal covariance matrix is assumed. Taking into account these assumptions, knowing the mechanism that generates the data, generating new data from the parameters of the original sample is possible using the sample means and variances. This allows to do a simulation study to evaluate the discrimination power of the heat flow PDSC curves and their extracted features.

7. Results

The PDSC curves obtained using ASTME2009 are shown in Fig. 1. They represent the heat flow vs. temperature signals corresponding to the 9 different fuels, obtained using a heating ramp.
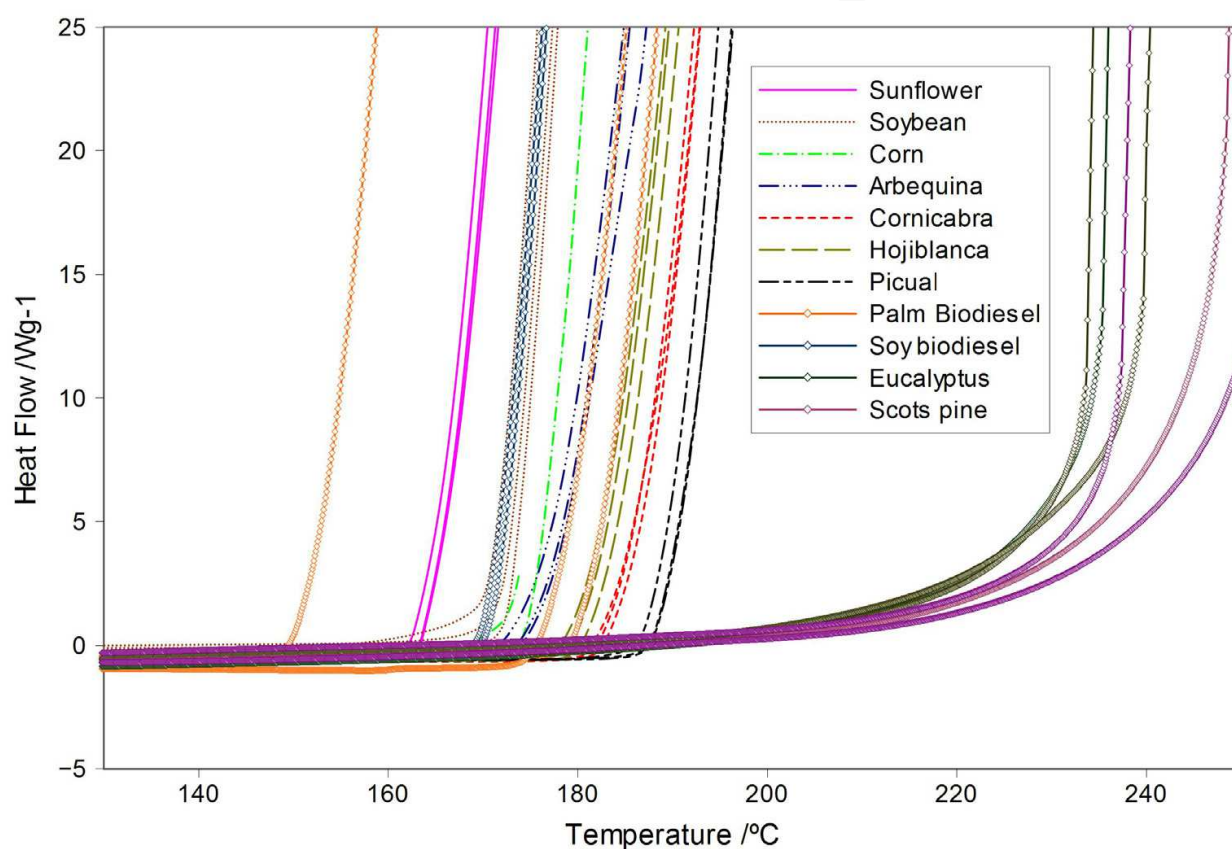
Fig. 1. PDSC curves corresponding to the studied fuel samples.

The Fig. 1 shows that the curves are different at lesser or major extend depending on the class of fuel tested. At a first glance, it seems there are two main groups. The first group is corresponding to the different studied oils and biodiesel types, and the second one consists of the wood species. It is clear that the OOT values corresponding to the first group are significantly lower than the OOT ones of the second group. But there are differences in oxidation stability (measured by OOT parameter, according to the Fig. 2) within these main groups? For answering this question we have used well known statistical tools as the F test and Tukey test. Using F test we can confirm that at least one class of fuel presents different OOT mean value than the others with statistical significance (p-value ≈ 0 < 0.05). By means the Tukey test we can know which fuels are statistically different, observing the OOT

variable. Table 3 shows the result of Tukey test. Each column represents a group of fuels different from the others, on the basis of the OOT value. For example, in the group number 1 there are three species that present no different OOT values (p-value = 0.154 < 0.05). We can observe that there are not differences between soy and soy biodiesel OOT. However, they are different to all the olive oil varieties tested and to the wood fuels studied. In fact, olive varieties form an independent group. On the other hand, palm biodiesel OOT is statistically different from the sunflower, picual and wood species OOT. It is important to note that the high OOT values obtained for wood species may condition the results for the remaining fuel classes. Attending to the means, the following fuels are sorted from largest to smallest OOT: Scots Pine > Eucalyptus > Picual > Hojiblanca > Palm biodiesel > Corn > Soy > Soy biodiesel > Sunflower

Fuel class	N	Different Groups					
		1	2	3	4	5	6
Sunflower	3	165.8100					
Soy biodiesel	3	172.9450	172.9450				
Soy	3	173.0967	173.0967				
Corn	3		175.7933				
Palm biodiesel	3		181.6833	181.6833			
Hojiblanca	3			184.9933	184.9933		
Picual	3				192.2767		
Eucalyptus	3					235.6567	
Scots pine	3						245.1867
p-value		0.154	0.054	0.911	0.154	1.000	1.000

Table 3. Mean values of OOT in the homogeneous subsets for the fuel class factor. Different groups obtained by applying the Tukey test (with significance lever α = 0.05).

Foccusing to the soy and soy biodiesel OOT values, the OOT mean corresponding to soy oil is higher than soy biodiesel OOT, according to the theory. But when we want to compare an important quantity of fuels that presents a wide range of OOT values, the variance of the OOT measurements can prevent to distinguish the different fuels. The OOT is an important parameter that contains much information about the oxidation stability of a fuel. But, as we have observed, the OOT by itself is not enough to distinguish between all studied fuels. Obtaining more information about the PDSC curves is necessary to classify correctly among the different fuels. Therefore, additional features are chosen: the maximum slope of heat flow versus time (slope max, V) obtained in each case, the temperature at that point of maximum slope (T at max slope, H) and the slopes of the heat flow curves vs. temperature in the range from 5 to 10 Wg^{-1} (slope between 5 and 10, m). The Fig. 2 and 3 show the additional features extracted from the PDSC signals.

Moreover, having a large number of samples in a supervised classification problem is recommended. There are three samples of each fuel but, as shown in Fig. 1, the PDSC
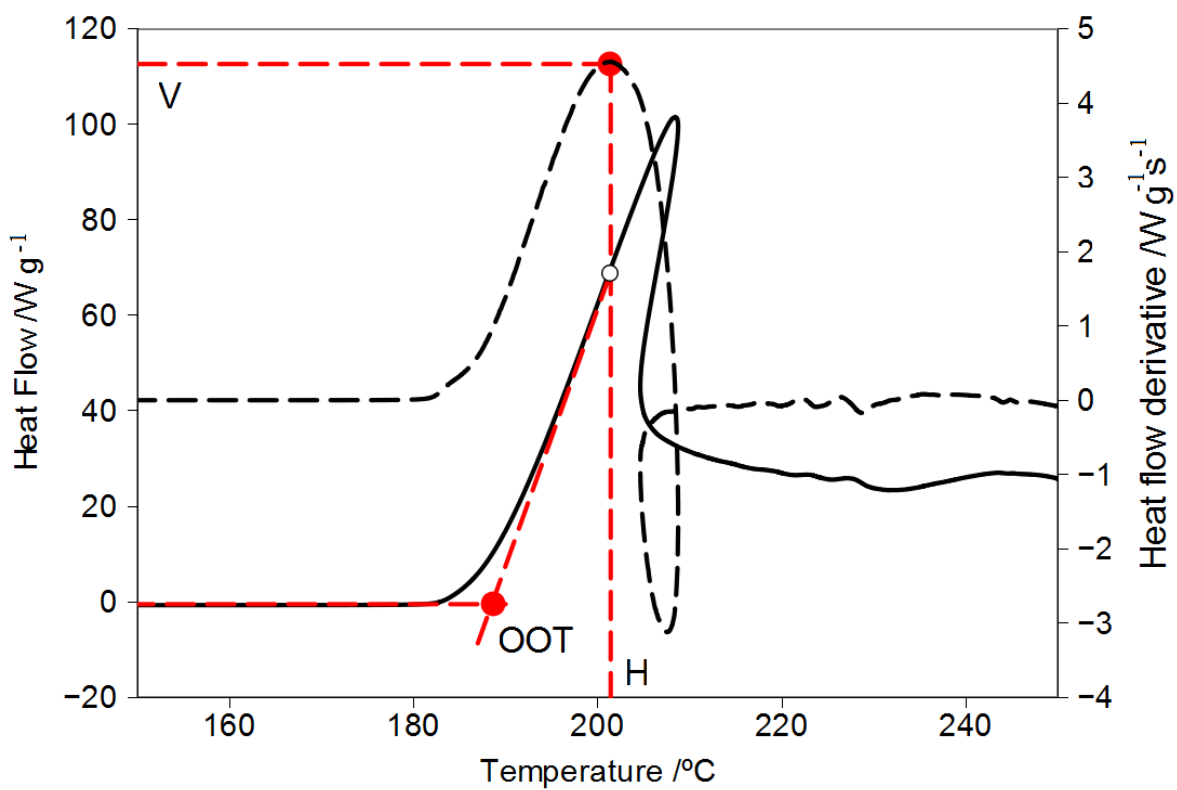
Fig. 2. OOT, H and V features extracted from the PDSC heat flow signal and its derivative.
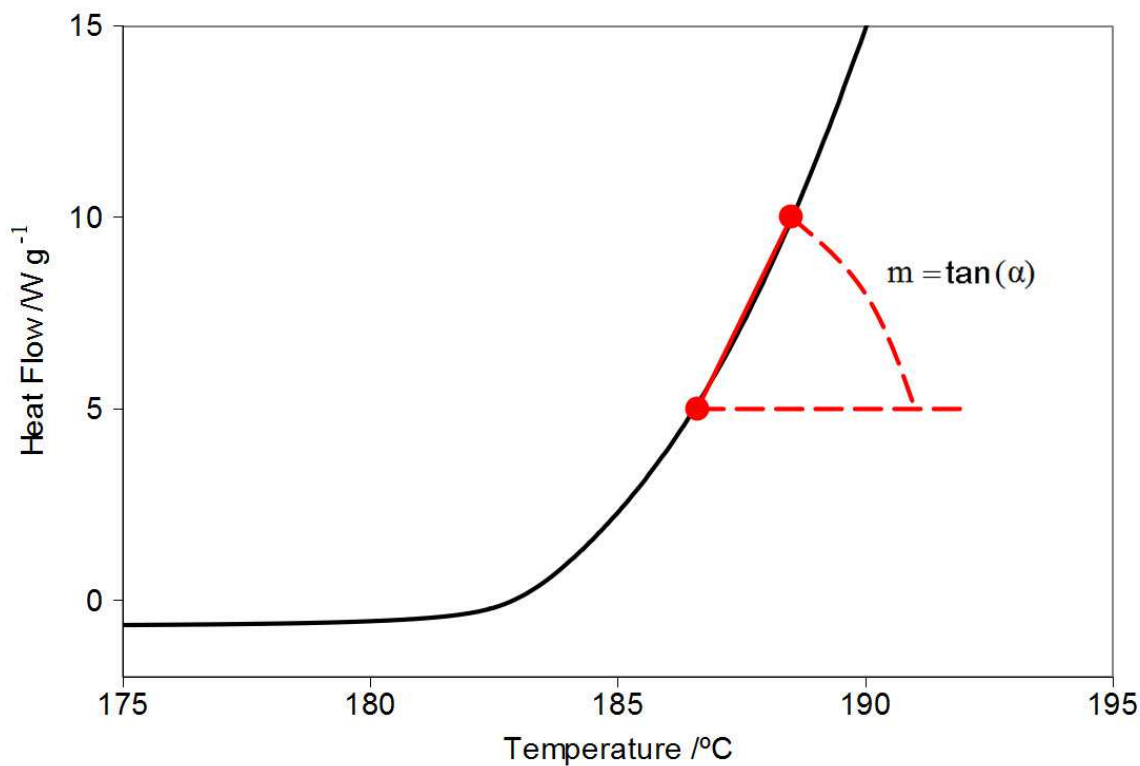

Fig. 3. The slopes corresponding to the heat flow curves vs. temperature in the range from 5 to 10 Wg^{-1} (slope between 5 and 10, m).

curves variability has been properly represented. So we can do a simulation study taking into account the sample parameters (mean and variance of the features measured in each fuel class). Then, a simulation study is presented to evaluate the power of classification of the chosen features. A parametric bootstrap resampling is chosen to increase the sample size until 100 items per fuel class. The parametric bootstrap is implemented for generating new values from OOT and the other chosen features, assuming that are independently distributed according a Gaussian distribution where the mean is the sample mean and the variance is the sample variance. The leave-one-out cross-validation method technique is used for the validation of the empirical model. It allow to estimate the probabilities of correct classification corresponding to the different classification methods. It works by leaving out one sample (represented by the features above mentioned); then a model is trained with the remaining parameter samples and, finally, the developed model is used for classifying the sample left out. This is repeated 900 times, until all the vectors have been left out once. Table 4 shows the probability of correct classification obtained by the above mentioned classification methods. These probabilities are very high, regardless of the method used. The best result corresponds to the use of logistic regression (99.7%) through almost all the samples are correctly classified. Table 5 shows the confusion matrices corresponding to the application of logistic regression, LDA, Bayes Naïve and KNN classification methods. The percentage of simulated samples correctly classified is shown in the diagonal of the matrices. The percentages of confusion obtained between the fuel types, two by two, are presented outside the diagonal. The little confussions existing between the two types of wood and between palm biodiesel and hojiblanca olive oil are solved using the logistic regresion method. According to these results, the OOT and the other characteristics are very useful parameters for classification purposes.

Classification method	Percentage of correct classification/ %
LDA	94.2
Logistic regression	99.7
Bayes Naïve Classifier	98.0
KNN	98.1

Table 4. Percentages of correct classification obtained by the three proposed methods. The best results are obtained by Logistic regression.

Method	Estimated	Actual								
		Corn	Eucal.	Hojib.	Palm Biod.	Picual	Scots p.	Soy	Soy Biod.	Sunfl.
Logistic Regress.	Corn	100	0	0	0	0	0	0	0	0
	Eucal.	0	100	0	0	0	0	0	0	0
	Hojib.	0	0	100	0	1	0	0	0	0
	PalmBiod.	0	0	0	99	0	0	0	0	0
	Picual	0	0	0	1	98	0	0	0	0
	Scots p.	0	0	0	0	0	100	0	0	0
	Soy	0	0	0	0	1	0	100	0	0
	Soy Biod.	0	0	0	0	0	0	0	100	0
	Sunfl.	0	0	0	0	0	0	0	0	100

LDA	Corn	75	0	0	0	0	0	0	0	0
	Eucal.	0	98	0	0	0	19	0	0	0
	Hojib.	0	0	100	4	2	0	0	0	0
	PalmBiod.	1	0	0	96	0	0	0	0	0
	Picual	0	0	0	0	98	0	0	0	0
	Scots p.	0	2	0	0	0	81	0	0	0
	Soy	7	0	0	0	0	0	100	0	0
	Soy Biod.	17	0	0	0	0	0	0	100	0
	Sunfl.	0	0	0	0	0	0	0	0	100
Bayes Naïve	Corn	99	1	0	0	0	0	0	0	0
	Eucal.	0	94	0	0	0	8	0	0	0
	Hojib.	0	0	100	3	0	0	0	0	0
	PalmBiod.	1	0	0	97	0	0	0	0	0
	Picual	0	0	0	0	100	0	0	0	0
	Scots p.	0	5	0	0	0	92	0	0	0
	Soy	0	0	0	0	0	0	100	0	0
	Soy Biod.	0	0	0	0	0	0	0	100	0
	Sunfl.	0	0	0	0	0	0	0	0	100
KNN	Corn	100	0	0	0	0	0	0	0	0
	Eucal.	0	96	0	0	0	10	0	0	0
	Hojib.	0	0	100	3	0	0	0	0	0
	PalmBiod.	0	0	0	97	0	0	0	0	0
	Picual	0	0	0	0	100	0	0	0	0
	Scots p.	0	4	0	0	0	90	0	0	0
	Soy	0	0	0	0	0	0	100	0	0
	Soy Biod.	0	0	0	0	0	0	0	100	0
	Sunfl.	0	0	0	0	0	0	0	0	100

Table 5. Confusion matrix or prediction percentages obtained by each classification method and leave-one-out cross-validation, using the features extracted from PDSC signals. The feature data set was tested with 9 classes or types of fuels. The results are shown as percentages.

8. Conclusion

The thermooxidative stability of 9 different types of fuels (including two types of biodiesel, soy and palm oil) has been measured using the OOT parameter. The use of the OOT parameter and ANOVA techniques allows to differentiate various groups of fuels: the varieties of olive oil, the two types of wood and finally the remaining fuels (although the sunflower oil is slightly different). But the OOT by itself is not enough to distinguish between all studied fuels with statistical significance.

The classification of the 9 fuels according to the thermooxidative properties has been possible using multivariate supervised classification method and additional features extracted from the PDSC curves as dataset: the maximum slope of heat flow versus time (slope max) obtained in each case, the temperature at that point of maximum slope (T at max slope) and the slopes of the heat flow curves vs. temperature in the range from 5 to 10 Wg^{-1} (slope between 5 and 10). That additional information provides a better understanding of the thermooxidative process, allowing for identifying subtle differences between similar fuels.

The evaluation of the discriminant power of the extracted thermooxidative features has been possible using parametric bootstrap resampling.

The overall percentages of correct classification are very high, in particular when Logistic regression classifier is used (99.7%); it seems to work better than LDA, Bayes Naïve and KNN.

9. Acknowledgment

10. References

Alén, R.; Kuoppala, E. & Pia, O. (1996). Formation of the main degradation compound groups from wood and its components during pyrolysis. *Journal of Analytical and Applied Pyrolysis*. Vol. 36, pp. 137–148.

Artiaga, R.; López-Beceiro, J.; Tarrío-Saavedra, J.; Mier, J. L.; Naya, S. & Gracia, C. (2010). Oxidation stability of soy and palm based biodiesels evaluated by Pressure Differential Scanning Calorimetry. *Journal of ASTM International*. Vol. 7, No. 4.

ASTM. Standard Test Method for Oxidation Onset Temperature of Hydrocarbons by Differential Scanning Calorimetry. E2009-08. *Annual Book of ASTM Standards*, ASTM Internacional, West Conshohocken, PA, 2008.

Bailey, T. & Jain, A. (1978). A note on distance-weighted k-nearest neighbor rules. *IEEE Transactions on Systems, Man, and Cybernetics*. Vol. 8, pp. 311-313.

Bermejo, S. & Cabestany, J. (2000). Adaptive soft k-nearest-neighbour classifiers. *Pattern Recognition*. Vol. 33, pp. 1999-2005.

Bremner, D.; Demaine, E.; Erickson, J.; Iacono, J.; Langerman, S.; Morin, P. & Toussaint, G. (2005). Output-sensitive algorithms for computing nearest-neighbor decision boundaries. *Discrete and Computational Geometry*. Vol. 33, pp. 593-604.

Cover, T.M. & Hart P.E. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*. Vol. 13, pp. 21-27.

Dudani, S.A. (1976) The distance-weighted k-nearest-neighbor rule. *IEEE Transactions on Systems, Man, and Cybernetics*. Vol. 6, pp. 325–327.

Dunn, R. O. (2005). Effect of antioxidants on the oxidative stability of methyl soyate (biodiesel). *Fuel Processing Technology*. Vol. 86, pp. 1071-1085.

Dunn, R. O. (2006). Oxidative stability of biodiesel by dynamic mode pressurized-differential scanning calorimetry (P-DSC). *Transactions of the ASABE*. Vol. 49, pp. 1633-1641.

Ferraty, F. & Vieu, P. (2006). Nonparametric Functional Data Analysis: Theory and Practice. Springer-Verlag, New York.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annual Eugenics*. Vol. 7, pp. 179-188.

Fix, E. & Hodges, J.L. (1951). Discriminatory analysis, nonparametric discrimination: Consistency properties. Technical Report 4, USAF School of Aviation Medicine, Randolph Field, Texas.

Fukunaga, K. & Hostetler, L. (1975). k-nearest-neighbor bayes-risk estimation. *IEEE Transactions on Information Theory*. Vol. 21, pp. 285-293.

Grønli, M. G.; Várhegyi, G. & Blasi, C. (2002). Thermogravimetric analysis and devolatilization kinetics of wood. *Industrial & Engineering Chemistry Research*. Vol. 41, pp. 4201–08.

Hall, P.; Park, B.U. & Samworth, R.J. (2008). Choice of neighbor order in nearest-neighbor classification. *Annals of Statistics*. Vol. 36, pp. 2135–2152.

Hellman, M.E. (1970). The nearest neighbor classification rule with a reject option. *IEEE Transactions on Systems, Man, and Cybernetics*. Vol. 3, pp. 179-185.

Jozwik, A. (1983). A learning scheme for a fuzzy k-nn rule. *Pattern Recognition Letters. Vol.* 1, pp. 287-289.

Keller, J.M.; Gray, M.R. & Givens, J.A. (1985). A fuzzy k-nn neighbor algorithm. *IEEE Transactions on Systems, Man, and Cybernetics.* Vol. 15, pp. 580-585.

Knothe, G. & Dunn, R. (2003). Dependence of oil stability index of fatty compounds on their structure and concentration and presence of metals. *J American Oil Chemists' Society.* Vol. 80, pp. 1021-1026.

Knothe, G. (2007). Some aspects of biodiesel oxidative stability. *Fuel Processing Technology.* Vol. 88, No.7, pp. 669-677.

Kononenko, I. (1991). Semi-naive Bayesian classifier, *Proceedings of the European Working Session on Learning-91*, pp. 206-219, Springer, Berlin.

Kononenko, I. (1992). Naive Bayesian classifier and continuous attributes. *Informatica*. Vol. 16, pp. 1–8.

Kononenko, I. (2001). Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in Medicine*. Vol. 23, pp. 89-109.

Langley, P. (1993). Induction of recursive Bayesian classifiers. *Proceedings of the European Conference on Machine Learning*, Vienna.

López-Granados, F.; Peña-Barragán, J. M.; Jurado-Expósito, M.; Francisco-Fernández, M.; Cao, R.; Alonso-Betanzos, A. & Fontenla-Romero, O. (2008). Multispectral discrimination of grass weeds and wheat (*Triticum durum*) crop using linear and nonparametric functional discriminant analysis and neural networks. *Weed Research*. Vol. 48, pp. 28-37.

López-Beceiro, J.; Pascual-Cosp, J.; Artiaga, R.; Tarrío-Saavedra, J. & Naya, S. (2011). Comparison of olive, corn, soybean and sunflower oils by PDSC. *Journal of Thermal Analysis and Calorimetry*. Vol. 104, pp. 169-175.

Mallik, A.; Tarrío-Saavedra, J.; Francisco-Fernández, M & Naya, S. (2011). Classification of Wood Micrographs by Image Segmentation. *Chemometrics and intelligent laboratory systems*. Vol. 107, pp. 351-362.

Maxwell, S. E. & Delaney, H. D. (2004). Designing experiments and analyzing data. A model comparison perspective, Lawrence Erlbaum Associates, New Jersey.

McLachlan, G.L. (2004) Discriminant Analysis and Statistical Pattern Recognition, Wiley-Interscience, Hoboken, New Jersey.

Miller, R. B. (1999). Structure of wood. In: Wood handbook: *Wood as an engineering material*, Woodhead Publishing Limited, Department of Agriculture, Forest Service, Forest Products Laboratory, Madison, WI.

Mohan, D.; Pittman, J. C. U. & Steele, P. H. (2006). Pyrolysis of wood/biomassfor bio-oil: a critical review. *Energy Fuel.*Vol. 20, pp. 848–89.

Moser, B. R.; Haas, M. J.; Winkler, J. K.; Jacksona, M. A.; Erhana, S. Z.; List, G. R.. (2007). Evaluation of partially hydrogenated methyl esters of soybean oil as biodiesel. *European Journal of Lipid Science and Technology*. Vol. 109, pp. 17-24.

Mozina, M.; Demsar, J.; Kattan, M. & Zupan, B. (2004). Nomograms for Visualization of Naive Bayesian Classifier, *Proccedings of PKDD-2004*, pp. 337-348.

Nigsch, F.; Bender, A.; Van Buuren, B.; Tissen, J.; Nigsch, E. & Mitchell, J.B.O. (2006). Melting Point Prediction Employing k-nearest Neighbor Algorithms and Genetic Parameter Optimization. *Journal of Chemical Information and Modeling*. Vol. 46, pp. 2412-2422.

Polavka, J.; Paligová, J.; Cvengros J, et al. (2005). Oxidation stability of methyl esters studied by differential thermal analysis and rancimat. *American Oil Chemists' Society*. Vol. 82, pp. 519-524.

Riesen, R. & Schawe, E. K. (2006). Description of the O_2 pressure influence on the oxidation time of motor oils. *J Thermal Analysis and Calorimetry*. Vol. 83, pp. 157-161.

Tan, C. P.; Che Man, Y. B.; Selamat, J. & Yusoff, M.S.A. (2002). Comparative studies of oxidative stability of edible oils by differential scanning calorimetry and oxidative stability index methods. *Food Chemistry*. 2002, Vol. 76, p. 385– 389.

Tarrío-Saavedra, J. ; López-Beceiro, J.; Naya, S. ; Gracia, C. ; Mier, J. L. & Artiaga, R. (2010). Factores influyentes en la estabilidad a la oxidación del biodiesel. Estudio estadístico. *Dyna*. Vol. 85, No. 4, pp. 341-350.

Tóth, L.; Kocsor; A. & Csirik, J. (2005). On naive bayes in speech recognition. *International Journal of Applied Mathematics and Computer Science*. Vol. 2, pp. 287-294.

Toussaint, G.T. (2005). Geometric proximity graphs for improving nearest neighbor methods in instance-based learning and data mining. *International Journal of Computational Geometry and Applications*. Vol. 15, pp. 101–150.

U.S. Department of Agriculture, Agricultural Research Service (2009). USDA National Nutrient Database for Standard Reference, Release 22. In: *Nutrient Data Laboratory*, 11.09.2010, Avaliable from
http://www.nal.usda.gov/fnic/foodcomp/Data/SR15/dnload/pk260w32.exe.

Van Aardt, M.; Duncan, S. E.; Long, T. E.; O'Keefe, S. F.; Marcy, J. E. & Sims, S.R. (2004). Effect of antioxidants on oxidative stability of edible fats and oils: thermogravimetric analysis. *Journal of Agricultural and Food Chemistry*. Vol. 52, pp. 587-591. 13.

Vorria, E.; Giannou, V. & Tzia, C. (2004). Hazard analysis and critical control point of frying-safety assurance of fried foods. *European Journal of Lipid Science and Technology*. Vol.106, pp. 759–65, ISSN 1438-9312

Wang, S.; Wang, K.; Liu, Q.; Gu, Y; Luo, Z; Cen, K & Fransson, T. (2009). Comparison of the pyrolysis behavior of lignins from different tree species. *Biotechnology Advances*. Vol. 27, pp. 562–7.

Webb, G.I.; Boughton, J. & Wang, Z. (2005). Not So Naive Bayes: Aggregating One-Dependence Estimators. *Machine Learning*. Vol. 58, pp. 5-24.

Xu, Y. X.; Hanna, M. A. & Josiah S J. (2007). Hybrid hazelnut oil characteristics and its potential oleochemical application. *Industrial Crops and Products*. Vol. 26, pp. 69–76.

Yang, H.; Lewis, I. R. & Griffiths, P. R. (1999). Raman spectrometry and neural networks for the classification of wood types. 2. Kohonen self-organizing maps. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*. Vol. 55, pp. 2783–2791.
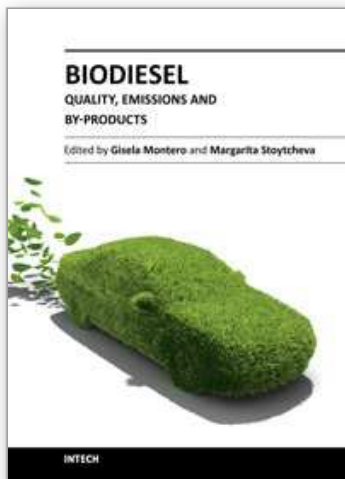
Biodiesel- Quality, Emissions and By-Products

Edited by Dr. Gisela Montero

This book entitled "Biodiesel: Quality, Emissions and By-products" covers topics related to biodiesel quality, performance of combustion engines that use biodiesel and the emissions they generate. New routes to determinate biodiesel properties are proposed and the process how the raw material source, impurities and production practices can affect the quality of the biodiesel is analyzed. In relation to the utilization of biofuel, the performance of combustion engines fuelled by biodiesel and biodiesels blends are evaluated. The applications of glycerol, a byproduct of the biodiesel production process as a feedstock for biotechnological processes, and a key compound of the biorefinery of the future is also emphasized.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

INTECH
open science | open minds