Journal of Chromatography A, 1225 (2012) 150–157

Contents lists available at SciVerse ScienceDirect

Journal of Chromatography A

journal homepage: www.elsevier.com/locate/chroma

Determination of vegetable oils and fats adulterants in diesel oil by high performance liquid chromatography and multivariate methods

Luiz Filipe Paiva Brandão[a], Jez Willian Batista Braga[b,*], Paulo Anselmo Ziani Suarez[b]

[a] Agência Nacional do Petróleo, Gás Natural e Biocombustíveis, SGAN 603, Módulo H, 70830-030 Brasília, DF, Brazil
[b] Instituto de Química, Universidade de Brasília, Campus Universitário Darcy Ribeiro, 70904-970 Brasília, DF, Brazil

ARTICLE INFO

Article history:
Received 13 October 2011
Received in revised form
21 December 2011
Accepted 22 December 2011
Available online 30 December 2011

Keywords:
Biodiesel
HPLC
Oils and fats
PCA
KNN
PLS

ABSTRACT

The current legislation requires the mandatory addition of biodiesel to all Brazilian road diesel oil A (pure diesel) marketed in the country and bans the addition of vegetable oils for this type of diesel. However, cases of irregular addition of vegetable oils directly to the diesel oil may occur, mainly due to the lower cost of these raw materials compared to the final product, biodiesel. In Brazil, the situation is even more critical once the country is one of the largest producers of oleaginous products in the world, especially soybean, and also it has an extensive road network dependent on diesel. Therefore, alternatives to control the quality of diesel have become increasingly necessary. This study proposes an analytical methodology for quality control of diesel with intention to identify and determine adulterations of oils and even fats of vegetable origin. This methodology is based on detection, identification and quantification of triacylglycerols on diesel (main constituents of vegetable oils and fats) by high performance liquid chromatography in reversed phase with UV detection at 205 nm associated with multivariate methods. Six different types of oils and fats were studied (soybean, frying oil, corn, cotton, palm oil and babassu) and two methods were developed for data analysis. The first one, based on principal component analysis (PCA), nearest neighbor classification (KNN) and univariate regression, was used for samples adulterated with a single type of oil or fat. In the second method, partial least square regression (PLS) was used for the cases where the adulterants were mixtures of up to three types of oils or fats. In the first method, the techniques of PCA and KNN were correctly classified as 17 out of 18 validation samples on the type of oil or fat present. The concentrations estimated for adulterants showed good agreement with the reference values, with mean errors of prediction (RMSEP) ranging between 0.10 and 0.22% (v/v). The PLS method was efficient in the quantification of mixtures of up to three types of oils and fats, with RMSEP being obtained between 0.08 and 0.27% (v/v), mean precision between 0.07 and 0.32% (v/v) and minimum detectable concentration between 0.23 and 0.81% (v/v) depending on the type of oil or fat in the mixture determined.

1. Introduction

Biodiesel is considered an alternative to conventional fossil fuels and has been increasingly mixed with commercial diesel because of the advantage it offers in terms of environmental sustainability. As a strategy to encourage the use and production of biodiesel, the legislation of several countries requires that biodiesel (B100) should be added to all diesel fuel (in Brazil is called *diesel A*) sold in the national territory on a pre-established percentage [1,2]. Despite the numerous economic, social and environmental benefits, this situation also brings a new concern to the fuel market, related to ensuring the quality of diesel oil sold in the country. In Brazil, the biodiesel to be used blended with diesel fuel must have up to 96.5% of fatty acid methyl or ethyl esters and match the specification established by specific regulation by the National Agency for petroleum, natural gas and biofuels (ANP) [1]. Therefore, the identification of the presence of oil in diesel in any amount features a type of misrepresentation and a nonconforming product, subjecting violators to penalties provided by law.

Irregular practices related to the addition of vegetable oils to diesel A instead of biodiesel may occur and are, in most of the times, motivated by the high price of biodiesel compared to some low cost vegetable oils available in the market. Moreover, the physico-chemical similarity between vegetable oils and biodiesel compositions contribute also to this type of practice. Losses from illegal practices involving diesel oil have widely affected the domestic market, because of the extensive road system in the country used in the transport of domestic production. Much of the current analytical methodologies proposed for the identification and quantification of vegetable oils in diesel apply spectroscopy

* Corresponding author. Tel.: +55 61 31073840.
E-mail address: jez@unb.br (J.W.B. Braga).

0021-9673/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2011.12.076

techniques [3–7], often chosen because of the cost of instrumentation relatively low, speed and precision of analysis. However, these techniques may be susceptible to interferences (chemical or spectral), especially when the analyte is present in complex matrices, such as diesel oil. Other papers also propose NMR [8] and mass spectrometry [9] techniques for quality control of diesel/biodiesel blends. It is important to note that most of these papers apply chemometric methods to enable the proper data analysis and determine the analytes.

Separation techniques show great versatility and are recommended to minimize the problems arising from matrix effects, facilitating interpretation and processing the data. The current regulation in Brazil uses GC techniques in order to determine biodiesel contamination, especially for pure biodiesel (B100) specification. However, it is needed two analyses: one to determine the amount of biodiesel and the second to analyze the presence of acylglycerides. It is also worth to mention that each analysis is done in different conditions and takes at least half an hour. Using HPLC we could determine in only one analysis both glycerides and biodiesel. The high performance liquid chromatography (HPLC) has great applicability in several branches of industry and science, with features that has become a preference among many analysts in the world, such as high resolving power, rapid separations, continuous monitoring of the eluent, accurate quantitative measurements, repetitive and reproducible analysis with the same column and the possibility of automation of the analytical procedure and data processing. Different chromatographic techniques have been described for identification and quantification of acylglycerides, such as gas chromatography [10–12] and HPLC with mass spectrometry detection [8–13], light scattering [14,15], flame ionization detection [16], refraction index [17] and ultraviolet detection [18]. Over the past 40 years, liquid chromatography has considerably contributed to obtain detailed knowledge about the composition of vegetable oils, mainly using the separation mechanisms in reversed-phase (with non-polar C18 column) [18].

Nowadays there is a great variety of raw materials (vegetable oils, fats and animal fats) for biodiesel production and, consequently, different kinds of adulterations can occur in diesel/biodiesel blends commercialized. Therefore, it is necessary to develop analytical methods able to provide a safe and an accurate identification and quantification of the possible adulteration in these commercial blends. This paper proposes and describes a simple analytical method to detect, identify and quantify oils and fats (not transesterified) in diesel oil, applying HPLC with ultraviolet detection associated with multivariate methods. Two independent methods are described. The first one able to identify individual types of adulterations present in the sample by principal component analysis (PCA) and *K*th Nearest Neighbor (KNN) classification, where the presence of commercial oils (soy, corn, cotton and frying oil) and palm-tree fats (palm kernel, *Elaeis* sp., and Babassu, *Orbignia* sp.) in the blends was studied and a posterior determination by univariate regression. The second method was applied to determine up to three types of adulterations present in a sample applying multivariate calibration based on Partial Least Square Regression (PLS).

2. Materials and methods

2.1. Solvents and chemicals

Two different types of diesel oil A obtained from a licensed producer were used for the sample preparations, S1800 and S500, which presented 1800 and 500 mg L^{-1} sulfur content, respectively, according to the specification provided by different fuel suppliers in Brazil. Commercial oils (soybean, corn, cotton and frying oil) and fats (palm kernel and babassu) were purchased at the commercial market and the biodiesel was obtained from a licensed producer. Methanol and 2-propanol (HPLC grade) obtained from J. T. Baker and *n*-hexane (UV/HPLC spectroscopic grade) 97% (w/w) from VETEC were also used for dilution of the samples and preparation of the mobile phase. Glyceryl trioleate (triolein), a triacylglycerol standard (99%, w/w purity) from Sigma–Aldrich was used as reference for the method development.

2.2. Instrumentation and chromatographic conditions

The analysis was performed by injecting 10.0 μL of sample in a HPLC Shimadzu Prominence, equipped with a UV–vis detector (SPD-20 A) operating at 205 nm, a quaternary pump (LC-20 AT) and a Shim-Pack VP-ODS column (C18 nonpolar, 250 mm, 4.6 mm internal diameter) kept at 40 °C. Chromatography conditions of composition and programming of the mobile phase used were proposed by Holcapek et al. [19], which applied a binary solvent gradient at a constant flow of 1.0 mL min^{-1} and programmed as follows: 100% (v/v) methanol and 0% (v/v) 2-propanol:hexane 5:4 (v/v) solution at 0 min, progressing linearly to 50% (v/v) methanol and 50% (v/v) of 2-propanol:hexane 5:4 (v/v) for 10 min. Subsequently, the gradient was kept isocratic until $t = 18.5$ min, returning to the initial gradient at 20 min and maintained at this composition for 3 min, in a total time of 23 min of analysis.

2.3. Procedure of analysis

The chromatographic system was daily submitted to cleaning by one or more injections of the solvent used to dilute the sample (2-propanol:hexane 5:4, v/v solution) until no interfering peak was observed at baseline of the solvent chromatogram. This step was repeated between successive analyses of samples in order to ensure that no carry over is observed (presence of peaks of compounds derived from previous runs in subsequent chromatograms).

Before injection in the chromatographic system, blanks, standards and samples were previously diluted with a mixture of 2-propanol and *n*-hexane 5:4 (v/v) at a rate of 25.0 μL to 2.0 mL, respectively. It was necessary to avoid saturation of the UV detector and all runs were performed in triplicate. The volumes transferred were measured with automatic micropipettes calibrated in order to minimize errors in sample preparation.

All the calibration samples were prepared with a S1800 diesel oil A as a matrix. However, before the step of calibration, the matrix effects were evaluated for the other three different types of diesel matrices: S500 diesel oil A and two mixtures B3 (diesel oil B), one blended with S1800 and the other with S500 diesel oil A. Mixtures B3 were prepared from the addition of 3% (v/v) of a soy biodiesel (B100) to S1800 and S500 of diesel oil A. In order to improve the evaluation of the influence of the matrix of biodiesel (B100) used, the content of residual triacylglycerols was calculated, according to the methodology proposed in the standard ASTM D6584 [20].

2.4. Peak integration

All peaks in the chromatogram corresponding to triacylglycerols were manually integrated, applying the integration system available in the equipment software. The integration parameters were adjusted to a slope of 10,000 mV min^{-1}, rejecting peak areas less than 1000 units of measurement and a window of ±5% in retention time to identify the peaks of the chromatogram. After this step, the data of retention times and peak areas of triacylglycerols of each run were exported from the equipment software to a spreadsheet or Matlab®, where the calculations for univariate or multivariate calibration were performed.

2.5. Development of the identification and quantification methods

Two methods were developed, one for the analysis of diesel oil samples adulterated by a single type of oil or fat and the other for diesel oil samples adulterated with mixtures of up to three types of oils or fats.

In the first method, applied to samples of diesel with adulteration of only one type of oil or fat, PCA was firstly applied to explore the variation in the data. After that, KNN method was applied using the four nearest neighbors of a sample to classify the type of oil or fat adulteration. In the KNN algorithm samples, the four nearest neighbors belonging to different classes were associated with the class of the first nearest neighbors. The data matrix used for model development showed thirty samples with known origin, arranged in six different classes, each one containing five samples at different concentrations of oil or fat. Each row of the data matrix represented a specific sample and is formed by the areas of 9 peaks of triacylglycerols obtained in the chromatogram. The calculations were performed in the Matlab® version 7.8 using the PLS Toolbox package version 5.3.

After the classification, the concentration of the oils and fats were quantified in a reference calibration curve, developed with a high purity triacylglycerol, triolein (C18:1) 99%. This reference curve was used for quantification of all six types of materials studied in this method (soybean, corn, cotton, palm kernel, babassu and frying oil). In this case, the application of correction factors (CF) to compensate the different response factors of these oils and fats in relation to triolein in the UV detector were determined by specific curves developed for each kind of oil and fat. The correction response factor for each oil and fat was calculated by the ratio of the response factor (slope of the analytical curve) of triolein and the response factor obtained in other analytical curves developed for each raw material.

All analytical curves of the samples contaminated with only one type of oil or fat were obtained using five standard samples prepared in the concentration range from 0 to 5% in S1800 diesel oil A matrix. It was considered that this matrix in its purest form (blank) was originally triacylglycerols free. Therefore, it was assumed to be zero for the interception in the corresponding regression equation, resulting in this following equation $y = Ax$, where the y values are the sum of the areas of all peaks of the triacylglycerols in each standard and x values the reference concentrations (%, v/v) of the triacylglycerols (TAG) in diesel.

The standards were prepared by adding the respective oil or fat to diesel oil in a desired concentration (%, v/v) and further homogenization. For the triolein reference analytical calibration, a stock solution with 10% (v/v) of triolein in diesel was prepared. The standards at concentrations of approximately 1.0, 2.0, 3.0, 4.0 and 5.0% (v/v) were obtained by dilution of the stock solution of triolein in diesel.

For the contaminated or adulterated samples with up to three types of oils or fats, a second method was developed based on partial least squares regression with more than one dependent variable (PLS2) used for identification and quantification of each species in the mixture. For the model development, the cross-validation technique was used to determine the number of latent variables, where the lowest root mean square error of cross validation (RMSECV) indicates the most appropriate number of latent variables to model each dependent variable [21–23]. The calibration samples were composed of nine samples used in the first method, which contain only one type of adulterant (soybean, cotton and palm kernel oils at 1, 3 and 5%, v/v); three binary mixtures containing 2% (v/v), each one of the two adulterations present; and three ternary mixtures prepared from soybean oil, cotton and palm kernel oil ranging proportions of these three types of oils from 3 and 1% (v/v). These three oils were chosen for the study by their different physical properties, composition, relevance and applicability for the production of biodiesel in Brazil.

2.6. Methods validation

The first method was validated for accuracy with a test set composed by eighteen samples containing the six kinds of adulterants in three different concentrations each (1, 3 and 5%, v/v), where the concentrations 1 and 5% (v/v) were prepared with S1800 diesel oil A and the sample at 3% (v/v) with S500 diesel oil. The precision test was performed in three consecutive days. 10 successive injections of the same samples, at the concentrations 1.0, 3.0 and 5.0% (v/v), respectively were performed each day, interspersed by no injection of solvent cleaning.

The validation of the PLS2 model was accomplished with a test set composed by three ternary mixtures and fifteen samples adulterated with only one type of oil or fat (soybean, cotton and oil palm kernel 1–5%).

The average accuracy, precision, minimum detectable concentration and goodness of fit were determined as figures of merit for validation of both methods.

The root mean square errors (RMSE) express the average agreement between the reference and estimated values. For multivariate models, this parameter can be determined in three distinct forms: (RMSEC) by the prediction of the same samples used in the model development or calibration, (RMSECV) when the cross-validation is employed for the calibration samples and (RMSEP) when an independent validation/test set is used for its estimation. In general, the RMSE can be estimated as [21–23]:

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

where ν is the number of degrees of freedom, y_i is the reference concentration value of the ith sample and $\hat{y}_{i,j}$ the estimative for the ith sample. For RMSEP, RMSECV and RMSEC, the number of degrees of freedom (ν) is equal to n, n and $n - A - 1$, respectively; where n is the number of samples of the data set and A the number of latent variables in the PLS model.

Following the ICH definition, precision represents the degree of scatter between a series of measurements for the same sample under prescribed conditions. It is usually expressed as a standard deviation of a series of measurements [24]. When different concentration levels are used for its estimation, the mean precision can be calculated as:

$$\text{Mean precision} = \sqrt{\frac{\sum_{i=1}^{L}\sum_{i=1}^{m}(\bar{y}_{i,j} - \bar{y})^2}{Lm}} \quad (2)$$

where L is the number of samples and m the number of replicates.

The goodness of fit was evaluated by a linear fitting of the estimated values of the % (v/v) of TAG versus the reference concentrations and the calculation of the intercept and the slope of the regression line. The slope and intercept for this linear fit should be ideally equal to 0 and 1, respectively. Therefore, the confidence intervals of the slope and intercept were estimated according to a required confidence level and the accuracy and goodness of fit attested if the estimated values for slope and intercept are statistically equivalent to the ideal values. By the way, if the slope or intercept was not equivalent to the ideal values, proportional or constant systematic error are observed in the results [25].

The minimum detectable concentration (MDC) can be determined by application of the ISO 11843-2 norm [26]. For multivariate calibration methods, such as PLS, an extended application of the

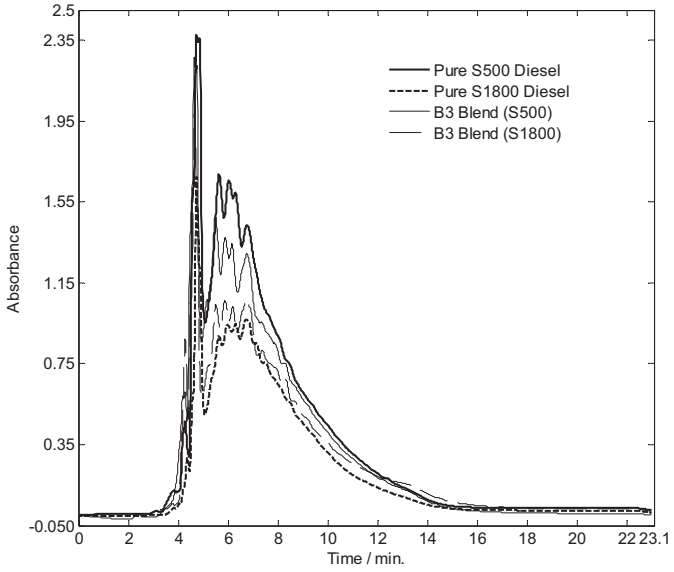

Fig. 1. Chromatographic profiles of the pure diesel matrix evaluated (S500 and S1800) and B3 blends.

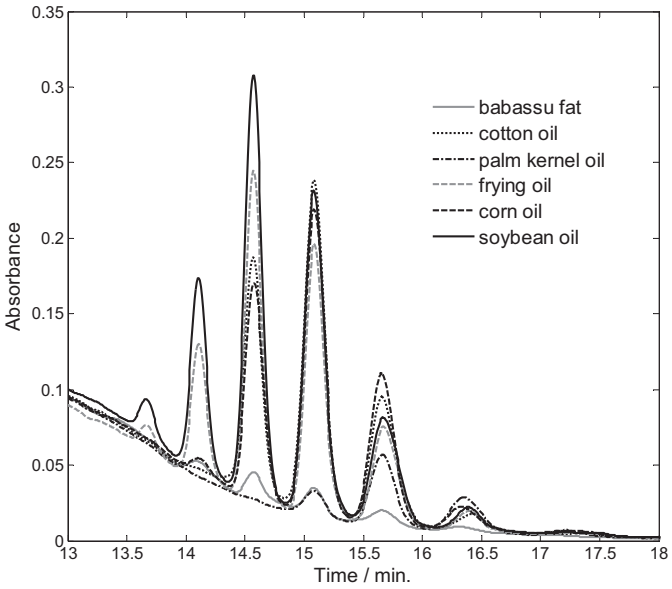

Fig. 2. Chromatographic profiles of five samples of S500 diesel oil A adulterated with 5% (v/v) of soybean, corn, cotton, palm kernel, babassu and frying oil between the region of 13 and 17 min.

same norm can be used, as suggested by Ortiz et al. [27]. Accordingly to this approach, the MDC values were evaluated by Eq. (3):

$$\mathrm{MDC} = \delta_{\alpha,\beta,\nu}\frac{s}{b}\sqrt{\frac{1}{m}+\frac{1}{n}+\frac{\bar{y}^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \quad (3)$$

where s is the standard deviation of the residuals (s) of the linear regression estimated for the goodness of fit estimation, at which the estimated concentration of the univariate or PLS method follows a linear dependence with the reference concentrations for the calibration samples, b is the slope, n is the number of calibration samples, m is the number of replicate measurements performed for each sample, $\bar{y}$ is the mean concentration, and $\delta_{\alpha,\beta,\nu}$ is the noncentrality parameter from the noncentral t-distribution with probabilities α and β for false positive and negative errors, respectively, and $\nu = n-2$ degrees of freedom. In this work, both probabilities of α and β were considered to be 0.05 (95% of confidence level), so $\delta_{\alpha,\beta,\nu}$ was equal to 4.320.

3. Results and discussion

3.1. Evaluation of diesel matrices

Figs. 1 and 2 present the chromatograms obtained for the two pure diesel matrixes evaluated (S1800 and S500), B3 Biodiesel/Diesel blends and samples adulterated with 5% (v/v) of soybean, corn, cotton, palm kernel, babassu and frying oil. It can be observed in Fig. 1 that the diesel oil matrix presents a high absorption from 2.5 to 13 min and the region of elution of the triacylglycerols is located in the decreasing ramp signal caused by diesel oil fuel (between 13 and 20 min). Therefore, the signal from the baseline was disregarded and the peaks were assumed to be fully resolved in the integration. Comparing the chromatograms obtained for S500 and S1800 diesel oil A, the least absorption was observed for S500 between 2.5 and 13 min. However, almost the same behavior was observed for the region of elution of triacylglycerols. It was observed some variation in the baseline between the chromatograms presented in Fig. 1, but no chromatographic peak was observed in the mixtures B3 (3% biodiesel), indication that these samples are free from oils or fats, showed the same chromatographic profile of pure diesel.

The interest region of the chromatogram was considered to be only between 13 and 20 min (Fig. 2), corresponding to the region of elution of triacylglycerols. In this region, no interference was observed for the matrices of S1800 and S500 diesel oil or eventual impurities of the biodiesel. Fig. 2 also shows that the six types of oils or fats present different concentrations of the triacylglycerols, a characteristic that makes possible the identification and classification of the type of adulteration by the chemometric models.

3.2. Determination of diesel adulteration by a single type of oil or fat

3.2.1. Identification of the oil or fat

Fig. 3 presents the PCA plot of the first three principal components obtained from the analysis of the thirty samples (five

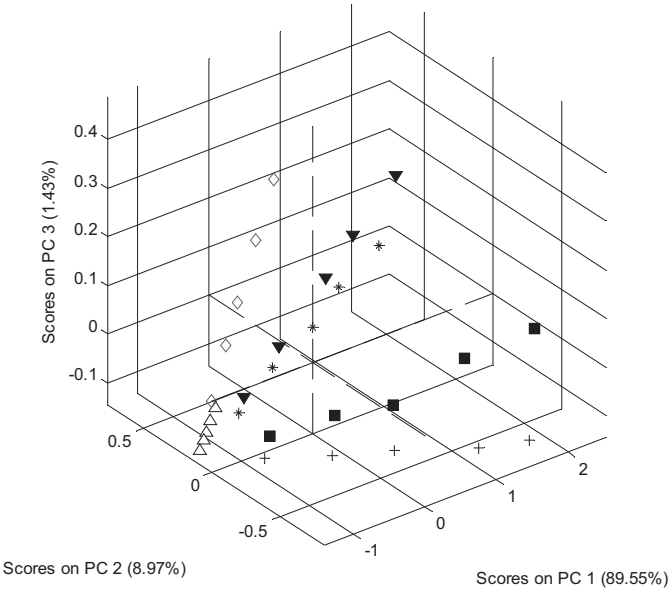

Fig. 3. PCA graphic to representation of calibration samples in three dimensional spaces. (▼) Soybean, (■) corn, (+) cotton, (Δ) babassu, (◊) palm kernel and (*) frying oil.

Table 1
Results of the KNN model for the classification of the adulterations in diesel oil.

Sample	Class of the four nearest neighbors[a]				Predicted class	True class	Estimated % TAG	Reference % TAG
	1°	2°	3°	4°				
1	1	2	3	4	1	1	1.1	1.0
2	1	2	1	2	1	1	3.2	3.0
3	1	1	2	1	1	1	4.9	5.0
4	3	4	2	6	3	3	1.0	1.0
5	3	4	3	4	3	3	2.9	3.0
6	3	4	4	3	3	3	4.8	5.0
7	6	6	6	5	6	6	0.9	1.0
8	6	6	6	6	6	6	2.9	3.0
9	6	6	6	6	6	6	5.1	5.0
10	2	1	4	3	2	2	1.2	1.0
11	2	2	1	1	2	2	3.3	3.0
12	2	1	1	1	1	2	4.0 (5.1)	5.0
13	4	3	2	1	4	4	1.0	1.0
14	4	3	3	4	4	4	3.1	3.0
15	4	3	4	3	4	4	4.8	5.0
16	5	5	6	6	5	5	1.1	1.0
17	5	5	5	5	5	5	2.8	3.0
18	5	5	5	5	5	5	5.1	5.0

[a] 1: soybean; 2: frying; 3: corn; 4: cotton; 5: palm kernel; 6: babassu.

concentrations for each one of the six types of oils and fat studied). In this representation, a clear distinction among the six classes was observed, where cotton, corn and palm kernel oil were the most distinct classes. On the other hand, differentiation between soybean and frying oil was more difficult, once these two classes were very close to each other. This proximity is justified by the fact that the frying oil is in fact soybean oil used (degraded). Therefore, these two classes present a similar composition.

It was also observed in Fig. 3 that for the lowest concentrations of the samples, the classes seemed to converge to the same region of space defined by principal components, which makes it difficult to differentiate them. The class babassu is located in this region of convergence. At the opposite extreme, for samples that presented the highest concentrations, the classes had a better distinction, even in the case of soybean and frying oil, where the separation between classes showed a better distinction.

After this first pattern recognition of the data, a classification model was developed based on KNN, trained with the samples presented in Fig. 3 and the four nearest neighbors of a sample. Table 1 presents the results obtained by KNN classification for the eighteen samples used for the validation of the model. The analysis of the table reveals that the results are directly related to the graph of PCA in Fig. 3. It can be observed in the Table 1 that for most of the validation samples presenting concentration 1% (v/v) (1, 4, 10 and 13) the 4 nearest neighbors belong to 4 different classes. For this reason, it was necessary to use the tie-breakers for the classification of these samples, since each class has 1 vote. The difficult of the KNN model to classify samples presenting low concentrations can be explained by similarity of the peak areas of the triacylglycerols at these concentration levels, which can be observed in the PCA graph. The validation of sample 16 (1%, v/v) also required the tiebreaker for its classification. However, in this case only two classes were tied with two neighbors each, where the first two nearest neighbors belong to the true class. At the lowest concentration level (1%, v/v), the only sample that was classified by a majority of the nearest neighbors was sample 7 (three votes for class 6). In the end, despite the relative difficulty encountered by the KNN, all samples of 1% were correctly classified.

Samples presenting 5% (v/v) of adulteration (3, 9, 12 and 18) were classified by a majority votes, except for samples 6 and 15, which required the tiebreaker because they had two votes for each class, in both cases. The only sample incorrectly classified by KNN model was the sample number 12 (concentration 5%, v/v), classified by majority vote as a soybean oil instead of frying. It is important to note that the frying oil is, in fact, used (degraded) soybean oil, which explains the great similarity between these classes (observed in Fig. 1) and the misclassification error at the higher concentration level studied.

Table 2
Correction factors calculated for each type of oil or fat.

Oil/fat	Ratio of response factors[a]	Correction factors
Soybean	0.240/1.560	0.154
Frying	0.240/1.235	0.194
Corn	0.240/1.155	0.208
Cotton	0.240/1.116	0.215
Palm kernel	0.240/0.263	0.913
Babassu	0.240/0.125	1.920

[a] In relation to the response of triolein 99%.

It is important to emphasize that part of the difficulty of classification by the KNN can be also attributed to the small number of training samples available. The use of a greater number of training samples may allow a better delineation of the classes and minimize the use of tiebreaker for classification of the samples. However, the low rate of errors, even with few training samples, indicates the potential of this method for the classification of types of adulteration with the oil or fat tested.

3.2.2. Obtainment of calibration curves, response factors and correction factors

Table 2 lists the correction factors calculated from the ratio of response factors (slope) obtained in the analytical curves constructed to each type of oil and fat. Two distinct groups were distinguishable: the group of oil, represented by soybean, frying, cotton and corn oil presented the highest sensitivity and the fat represented by palm and babassu, the lowest sensitivity. From the oilseeds, cotton was the oil that presented the highest content of saturated fatty acids (1.5% (w/w) C14:0 + 22% (w/w) C16:0 + 5% (w/w) C18:0) and therefore, it was less sensitive to UV detector, followed by corn oil (7% (w/w) C16:0 and 3% (w/w) C18:0). Moreover, soy was one that showed the highest percentage of unsaturated fatty acids (C18:1, C18:2 and C18:3) and consequently, a greater sensitivity. The frying oil, which in this case soybean oil was used, probably had its sensitivity reduced due to the hydrolytic oxidation (nonenzymatic) occurred in the process of frying, where double bonds were cleaved to form acids, peroxides, and other compounds and free fatty acids were formed by hydrolysis of ester groups.

For adulterations with fats, it was found that the response factor obtained for the palm kernel oil was practically the same as the one obtained for triolein. This similarity can be demonstrated by the correction factor used for the palm kernel of 0.913, where 1.000 represents the total identity at the UV detector response. The highest content of palmitic acid (C16:0) present in palm kernel oil contributes considerably to its low sensitivity. For babassu, the composition of fatty acids is mostly saturated (about 60–70%, w/w). Therefore, a lower response in the UV detector was obtained when compared to the others adulterations. This was the only one that presented a correction factor greater than 1 (1.920), which means that the peak of triolein in the UV detector was almost two times higher than the sum of the areas of all peaks of triacylglycerols in the chromatogram of babassu oil in a given concentration.

The results found for the determination of the % (v/v) of adulterations in the eighteen unknown validation samples are also shown in Table 1. The agreement of the estimated and the reference values confirms the validity of the proposed method for the quantification of adulterated diesel samples by one type of oil or fat. Except for the validation samples 10 (1%, v/v of frying oil) and 12 that were incorrectly classified, all relative errors were lower or equal to 10.0% in comparison to the reference concentration. For the validation of sample 12, it was presented two predicted values of TAG in diesel in Table 1: 4.0% (v/v), considering the class found erroneously indicated by KNN (soybean oil), and the value of 5.1% (v/v), that would be the one found for the concentration if KNN had been correctly classified this sample as a frying oil.

Table 3 shows the results for RMSEP, minimum detectable concentration (MDC) and the goodness of fit (intercept and slope) calculated for each type of oil or fat. The determination of babassu presented the lowest average prediction error, RMSEP (0.10%, v/v), whereas frying oil presented the highest one (0.22%, v/v). Unfortunately, in the determination of soybean, babassu, frying and cotton oil the intercept values did not contain the ideal value of zero. Also, the values of slope of babassu and frying oil were not statistically equivalent to the ideal value of 1, which may indicate the presence of constant and proportional systematic errors, respectively, in relation to the reference values. The presence of these systematic errors may be explained by the fact that all validation samples have been analyzed more than two weeks after the calibration samples, which can introduce some systematical variation of the chromatograph in the analysis. However, as this systematic error can be corrected or eliminated in routine analysis by the development of the analytical curves in shorter periods of time, the observed systematic error in the proposed method is not considered a disadvantage of the method, but a result of the experimental conditions in this particular work.

The lowest MDC was obtained for corn oil (0.12%, v/v), whereas soybean oil presented the highest one (0.48%, v/v). It should be noted that the calculations of the MDC take into account not only the sensitivity, but also all the errors of the regression model, which explain the fact of the soybean present a higher sensitivity but lower MDC. The results showed that method is able to determine low adulterations levels in diesel.

3.3. Determination of diesel adulteration by up to three types of oil or fat

The variation of RMSECV as a function of number of latent variables for the model PLS2 developed with samples adulterated with soybean, cotton and palm kernel oil, respectively are shown in Fig. 4. It was observed that if these calculations were performed with three independent PLS1 models, the same results could be obtained. As it can be seen in Fig. 4, three latent variables are sufficient to build the model, which is consistent with the fact that three different/independent types of adulterations are present in the calibration samples. The models built by PLS1 and PLS2 presented almost the same values obtained for concentration of TAG. For this reason, the model PLS2 was chosen for the method, once the three types of adulteration could be simultaneously determined using a single calibration model.

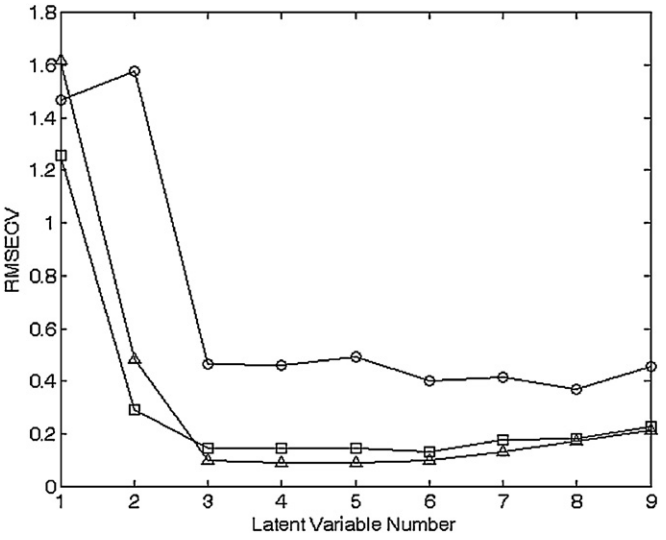

Fig. 4. Root mean square error of cross-validation (RMSECV) as a function of the number of latent variables. (○) Soybean, (□) cotton and (Δ) palm kernel oil.

Figs. 5–7 present the graphs of the reference concentrations versus the estimated ones for the calibration and validation samples obtained by PLS2 for soybeans, cotton and palm kernel oil, respectively. The dispersions show that for soybeans and cotton a good agreement was obtained between the reference and the estimated concentrations for both calibration and validation samples. Also, a larger dispersion was observed for palm kernel. The largest errors obtained for this model are probably associated with low area values obtained for the triacylglycerols comprising palm kernel oil, which makes the quantitative determination more susceptible to instrumental variations and harder in the case of multiple adulterations.

Table 4 presents the figures of merit for the PLS2 model, such as: the root mean square error (RMSEC, RMSECV and RMSEP), average precision, minimum detectable concentration (MDC) and the

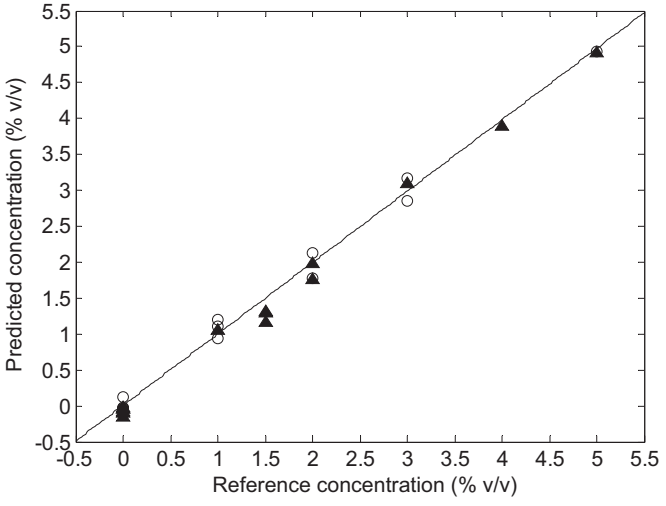

Fig. 5. Graph correlation between reference and estimated concentrations for the calibration and validation of samples adulterated with soybean oil. (○) Calibration samples, (Δ) validation samples, (–) linear fit of the calibration samples.

Table 3
Mean error of prediction (RMSEP), minimum detectable concentration (MDC) and values of intercept and slope obtained for each oil or fat.

Type of oil/fat	RMSEP[a,b]	MDC[b]	Goodness of fit	
			Intercept	Slope
Soybean	0.14	0.48	0.239 ± 0.085	0.950 ± 0.055
Corn	0.13	0.12	0.039 ± 0.078	0.950 ± 0.051
Babassu	0.10	0.24	−0.244 ± 0.075	1.067 ± 0.049
Frying	0.22	0.24	0.261 ± 0.125	0.983 ± 0.082
Cotton	0.13	0.13	0.153 ± 0.097	0.942 ± 0.064
Palm kernel	0.14	0.32	−0.050 ± 0.085	1.017 ± 0.055

[a] RMSEP of frying oil calculated considering the value of 5.1% (v/v) to sample 12, at Table 1.
[b] Expressed as a percentage (v/v) (95% confidence), $\nu = 13$.

Table 4
Figures of merit estimated in PLS2 model for the determination of soybean, cotton and palm oil in the mixture.

Parameter		Soybean	Cotton	Palm kernel
RMSEC[a]		0.12	0.07	0.31
RMSECV[a]		0.15	0.10	0.46
RMSEP[a]		0.08	0.14	0.27
Mean precision[a,c]		0.13	0.07	0.32
Minimum detectable concentration (MDC)[a,b]		0.23	0.43	0.81
Goodness of fit	Intercept	−0.06 ± 0.05[b]	0.04 ± 0.10[b]	0.05 ± 0.20[b]
	Slope	1.00 ± 0.03[b]	±0.05[b]	1.03 ± 0.10[b]

[a] Expressed in % (v/v).
[b] 95% confidence level with 13 degrees of freedom.
[c] 4 degrees of freedom.

goodness of fit of the model for the three types adulterations evaluated (soybean, cotton and palm kernel oil). The determination of cotton seed oil in the calibration mixtures presented the lowest RMSECV and RMSEC values, being 0.10 and 0.07% (v/v), respectively. On the other hand, the largest RMSECV and RMSEC results were observed for the determination of palm kernel oil (0.46 and 0.31%, v/v, respectively). Regarding the RMSEP, the determination of soybean and cotton oils in mixtures showed the lowest errors. For palm kernel oil, a larger error (0.27%, v/v) was obtained. However, it can also be considered acceptable for identification and determination of adulterations on mixtures.

The MDC values obtained for the determination of soybean oil and cotton in mixtures were 0.23 and 0.43% (v/v), respectively; which can be considered to be statistically equivalent to the values estimated when only one type of adulteration is present in the sample. Furthermore, these values were acceptable considering that all errors of the calibration model were taken into account. For palm kernel oil, it was obtained 0.81%, which begins to become significant for determinations of low levels of adulterations of this fat. However, the high cost of this vegetal fat makes its application for adulteration impracticable, and consequently, the introduction of palm kernel oil in this work was held to challenge the PLS2 method with its fat characteristic.

According to the values of slope and intercept of the regression between the reference versus estimated values by PLS2 model, presented in Table 4, we can see that no systematic errors, constant or proportional, were obtained for cotton and palm kernel oils, since the values of intercept and slope contain the expected values of 0 and 1, respectively. For soybean, it was observed that the validation samples showed a small negative bias, but a significant 95% confidence, as the confidence interval of the intercept did not contain zero.

The multivariate model described in this method was able to determine the adulteration in mixtures, even using a small number

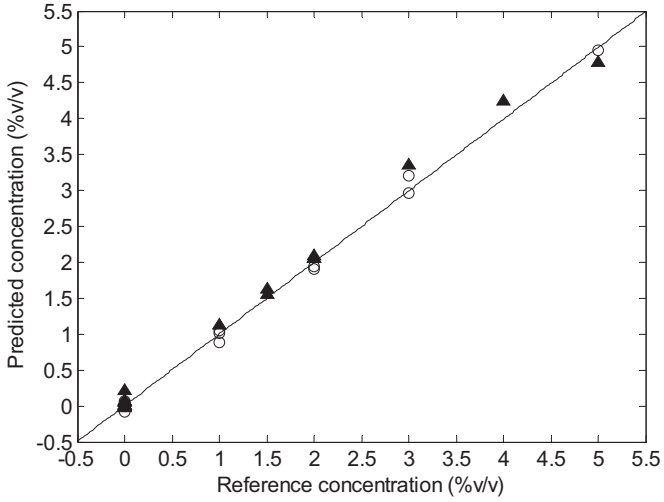

Fig. 6. Graph correlation between reference and estimated concentrations for the calibration and validation of samples adulterated with cotton oil. (○) Calibration samples, (Δ) validation samples, (–) linear fit of the calibration samples.

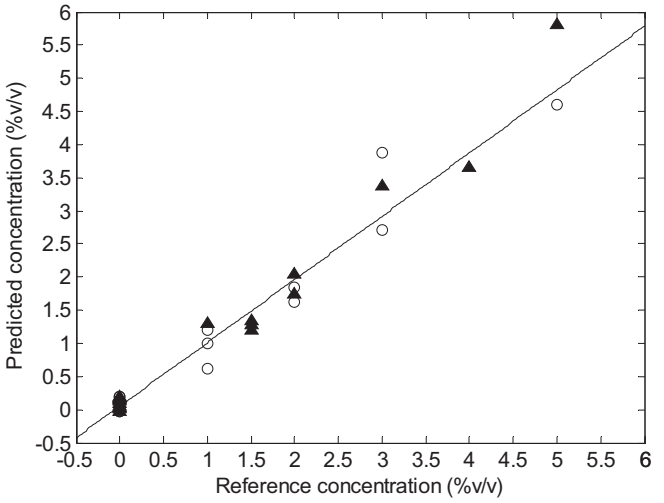

Fig. 7. Graph correlation between reference and estimated concentrations for the calibration and validation of samples adulterated with palm kernel oil. (○) Calibration samples, (Δ) validation samples, (–) linear fit of the calibration samples.

of calibration samples (fifteen), and of these fifteen, only six were really mixtures.

Despite the reduced number of calibration samples (fifteen) used in the developed of the PLS2 model, good results could be obtained for the eighteen validation samples. These results prove that the multivariate model may be successfully used for determination of up to three types of simultaneously adulterations in diesel. It is also important to note that the development of the PLS2 model required the same number of standards solutions (calibration samples) of the individual univariate regression models for the three types of adulterations evaluated in this case, and so the time required to build the models was almost the same.

4. Conclusions

PCA and KNN showed great effectiveness to identify the types of vegetable oils and fats present in adulterated diesel oil samples. A simple linear regression was able to determine individual adulterations of all types of oils and fats with a single standard analytical curve of triolein, regardless of chemical composition. In this case, average prediction errors (RMSEP) between 0.10 (babassu) and 0.22% (frying) and MDC between 0.12 (corn) and 0.48% (v/v) (soybean) were obtained.

For diesel oil samples adulterated with up to three types of vegetable oils or fats, a regression model based on PLS2 was efficient in the determinations, even using a small number of calibration samples. The results for the determination of palm kernel in mixtures showed the highest mean errors (RMSEC, RMECV and RMSEP of 0.31, 0.46 and 0.27%, v/v, respectively) and the highest MDC (0.81%, v/v) compared to soybean oil and cotton. Besides, for the other two vegetable oils errors at the same magnitude to the univariate method were obtained, which is restricted the determination of adulteration of diesel by only one type of oil or fat.

Once different kinds of oils were used in this study, it was concluded that no significant interference was observed in the region of interest triacylglycerols, which means that the proposed methodology can be applied with equal effectiveness for S500 and S1800 diesel oil A pure or blended with biodiesel.

The proposed methodology based on HPLC combined with multivariate techniques showed great potential and reliability for the determination of fats and oils in samples of diesel, therefore, it can be safety used as a routine method for quality control of this fuel.

Acknowledgements

We would like to thank CNPq, FAPDF and Fundação Banco do Brasil for partial financial support. PAZ Suarez is in debt with CNPq for his research fellowship.

References

[1] National Agency of Petroleum, Natural Gas and Biofuel, RANP 42/2009. http://nxt.anp.gov.br/nxt/gateway.dll/leg/resolucoes_anp/2009/dezembro/ranp%2042%20-%202009.xml (accessed at 13.09.10).
[2] Directive 2003/30/EC, Off. J. Eur. Union L123 (2003) (17 May) 42/46.
[3] M.F. Pimentel, G.M.G.S. Ribeiro, R.S. da Cruz, L. Stragevitch, J.G.A.P. Filho, L.S.G. Teixeira, Microchem. J. 82 (2006) 201.
[4] C.N.C. Corgozinho, V.M.D. Pasa, P.J.S. Barbeira, Talanta 76 (2008) 479.
[5] F.C.C. Oliveira, C.R.R. Brandão, H.F. Ramalho, L.A.F. Costa, P.A.Z. Suarez, J.C. Rubim, Anal. Chim. Acta 587 (2007) 194.
[6] I.P. Soares, T.F. Rezende, I.C.P. Fortes, Energy Fuels 23 (2009) 4143.
[7] I.P. Soares, T.F. Rezende, R.C. Silva, E.V.R. Castro, I.C.P. Fortes, Energy Fuels 22 (2008) 2079.
[8] N.N. Mahamuni, Y.G. Adewuyi, Energy Fuels 23 (2009) 3773.
[9] M.R. Monteiroa, A.R.P. Ambrozina, M.S. Santos, E.F. Boffo, E.R. Pereira-Filho, L.M. Lião, A.G. Ferreira, Talanta 78 (2009) 660;
P.V. Abdelnur, L.S. Eberlin, G.F. de Sá, V. de Souza, M.N. Eberlin, Anal. Chem. 80 (2008) 7882.
[10] E.M. Goh, R.E. Timms, J. Am. Oil Chem. Soc. 62 (1985) 730.
[11] C. Plank, E. Lorbeer, J. Chromatogr. A 697 (1995) 461.
[12] M. Lechner, C. Bauer-Plank, E. Lorbeer, J. High Resolut. Chromatogr. 20 (1997) 581.
[13] H.R. Mottram, S.E. Woodbury, R.P. Evershed, Rapid Commun. Mass Spectrom. 11 (1997) 1240.
[14] T.A. Foglia, K.C. Jones, J.G. Philips, Chromatographia 62 (2005) 115.
[15] A.J. Palmer, F.J. Palmer, J. Chromatogr. 465 (1989) 369.
[16] K.V.V. Nurmela, L.T.J. Satama, Chromatographia 435 (1988) 139.
[17] R.V. Flor, L.T. Hecking, B.D. Martin, J. Am. Oil Chem. Soc. 70 (1993) 199.
[18] E.J.C. van der Klift, G. Vivó-Truyols, F.W. Claassen, F.L. van Holthoon, T.A. van Beek, J. Chromatogr. A 1178 (2008) 43.
[19] M. Holcapek, P. Jandera, J. Fischer, B. Prokes, J. Chromatogr. A 858 (1999) 13.
[20] ASTM D6584, Standard Test Method for Determination of Total Monoglyceride, Total Diglyceride, Total Triglyceride, and Free and Total Glycerin in B-100 Biodiesel Methyl Esters by Gas Chromatography, ASTM Annual Book of Standards, West Conshohocken, PA, 2000.
[21] R.G. Brereton, Analyst 125 (2000) 2125.
[22] H. Martens, T. Naes, Multivariate Calibration, John Wiley & Sons, New York, 1989.
[23] P. Valderrama, J.W.B. Braga, R.J. Poppi, J. Agric. Food Chem. 55 (2007) 8331.
[24] International Conference on Harmonization (ICH), Validation of Analytical Procedures: Text and Methodology, Q2(R1), 2005.
[25] J.N. Miller, J.C. Miller, Statistics and Chemometrics for Analytical Chemistry, 5th ed., Pearson Prentice Hall, London, 2005.
[26] ISO, 11843-2; Capability of Detection, International Standards Organization, Geneva, Switzerland, 2000.
[27] M.C. Ortiz, L.A. Sarabia, A. Herrero, M.S. Sánchez, M.B. Sanz, M.E. Rueda, D. Giménez, M.E. Meléndez, Chemom. Intell. Lab. Syst. 69 (2003) 21.