Short communication

Screening analysis of biodiesel feedstock using UV–vis, NIR and synchronous fluorescence spectrometries and the successive projections algorithm

Matías Insausti [a], Adriano A. Gomes [b], Fernanda V. Cruz [b], Marcelo F. Pistonesi [a], Mario C.U. Araujo [b],*, Roberto K.H. Galvão [c], Claudete F. Pereira [d], Beatriz S.F. Band [a]

[a] FIA Lab., INQUISUR–CONICET, Sector Química Analítica, Universidad Nacional del Sur. Av. Alem 1253, B8000CPB Bahía Blanca, Argentina
[b] Universidade Federal da Paraíba, Departamento de Química, Laboratório de Automação e Instrumentação em Química Analítica/Quimiometria (LAQA), Caixa Postal 5093, CEP 58051-970–João Pessoa, PB, Brazil
[c] Instituto Tecnológico de Aeronáutica, Divisão de Engenharia Eletrônica, 12228-900, São José dos Campos, SP, Brazil
[d] Departamento de Química, Universidade Federal Rural de Pernambuco, CEP: 52171-900-Recife, PE, Brazil

ARTICLE INFO

ABSTRACT

This paper investigates the use of UV–vis, near infrared (NIR) and synchronous fluorescence (SF) spectrometries coupled with multivariate classification methods to discriminate biodiesel samples with respect to the base oil employed in their production. More specifically, the present work extends previous studies by investigating the discrimination of corn-based biodiesel from two other biodiesel types (sunflower and soybean). Two classification methods are compared, namely full-spectrum SIMCA (soft independent modelling of class analogies) and SPA-LDA (linear discriminant analysis with variables selected by the successive projections algorithm). Regardless of the spectrometric technique employed, full-spectrum SIMCA did not provide an appropriate discrimination of the three biodiesel types. In contrast, all samples were correctly classified on the basis of a reduced number of wavelengths selected by SPA-LDA. It can be concluded that UV–vis, NIR and SF spectrometries can be successfully employed to discriminate corn-based biodiesel from the two other biodiesel types, but wavelength selection by SPA-LDA is key to the proper separation of the classes.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Nowadays, biodiesel has become increasingly important in the energy matrix of many countries. In Brazil, Argentina, United States and several members of the European Union, transport fuel suppliers are now required to distribute petroleum diesel blended with biodiesel, which has resulted in a substantial increase in the production of this renewable fuel [1–3].

Biodiesel can be produced from a variety of raw materials, including animal fats (beef tallow, pork lard and chicken fat, for instance), residual oil and vegetable oils [4–7]. However, the quality of the resulting fuel may vary significantly, because the properties of the transesterification product depends on the fatty acid composition of the base oil or fat [8–10]. Indeed, typical crude oil contains up to 1.6% unsaponifiable matter, which is not affected by ester preparation and is likely to be present in similar amounts in the resulting biodiesel [11]. Moreover, unsaturations in the triglyceride chains may lead to polymerisation, whereas the

presence of saturated compounds increases the melting point of the fuel and hampers its use in regions with cold climate. In addition, such saturated compounds affect the cetane number, heat of combustion, oxidative stability and lubricity of the biodiesel [6]. The use of inadequate raw materials may also cause gum formation in the engine, as well as the liberation of harmful components and particulate matter in the exhaust gases [10].

The importance of feedstock type for the quality of biodiesel has motivated the development of analytical methods to certify the base oil employed in the production process. Within this scope, several investigations have been concerned with the use of near-infrared (NIR) or UV–vis spectroscopy [12–14]. Balabin and Safieva [12], Veras et al. [13], and Veras et al. [14] employed biodiesel samples produced from nine (sunflower, coconut, palm, rapeseed, cottonseed, castor, jatropha, linseed, used frying oil), four (cotton, sunflower, soybean, canola), and three (cotton, soybean, sunflower) types of vegetable oil, respectively. The samples were then discriminated by applying multivariate classification models to the NIR/UV–vis spectra.

It is worth noting that none of these papers [12–14] was concerned with the use of corn oil, which is an important feedstock for biodiesel production. Indeed, according to Szulczyk and

McCarl [15], corn oil is the primary feedstock for biodiesel in the United States, followed by soybean oil. The use of corn oil has been mainly motivated by the growth of the wet corn mill industry, driven by the increasing demand for ethanol production. This scenario is part of the focus currently placed on the use of biofuels made from crops, such as corn, sugar cane, and soybeans, for use as renewable energy sources, as discussed by Balat [16]. In this context, the present study extends previous works [12–14] by investigating the discrimination of corn-based biodiesel from other biodiesel types with the use of spectrometric techniques and multivariate classification methods. More specifically, the investigation includes samples produced from sunflower and soybean oils, which are commonly employed in the biodiesel literature [17].

In the context of spectral data classification, one of the most popular chemometrics methods is SIMCA (Soft Independent Modelling of Class Analogy). In SIMCA, principal component analysis (PCA) is applied to a set of training samples in order to obtain models for each class under consideration. These models can be subsequently employed to classify unknown samples [18]. The use of PCA facilitates the handling of high-dimensional datasets, which typically arise in spectrometry. However, in some analytical problems, the information of relevance for classification purposes may be concentrated in specific spectral regions. If this is the case, using the entire spectrum in the modelling process may actually mask the discriminatory information of interest. As an alternative, variable selection techniques may be used to choose a subset of wavelengths to be used in the classification model [19–21]. The present work investigates the potential advantages of using variable selection to improve the discrimination of the biodiesel samples under study. For this purpose, the Successive Projections Algorithm (SPA) is adopted.

SPA is a variable selection technique originally designed for multivariate calibration purposes [22], which was later adapted to handle classification problems [23]. SPA is aimed at selecting variables with minimal multicollinearity that convey appropriate information related to the analytical problem. Applications have included determination of metals in steel alloys by plasma emission spectrometry [24], determination of quality parameters of vegetable oils by near-infrared (NIR) spectrometry [25], quantification of phenolic compounds in sea water by molecular absorption spectrometry [26], classification of cigarettes employing NIR spectrometry [27], classification of vegetable oils using square-wave voltammetry [28], classification of soil samples using LIBS [29], classification of coffee samples on the basis of molecular absorption spectra [30], screening analysis of beer samples with respect to ageing state [31], detection of adulteration in diesel/biodiesel blends using NIR spectrometry [32], among several others.

SPA can be divided into two phases. In phase 1, chains of variables are generated by using projection operations involving the matrix of instrumental responses. Each variable in a chain is selected to display the least collinearity with the preceding ones. In phase 2, candidate subsets of variables are extracted from the chains generated in phase 1 and then evaluated according to a suitable metric. Within the scope of classification, the metric proposed in [23] was defined as an average risk of incorrect classification evaluated by using the Mahalanobis distance. At the end, the variables selected by SPA are employed to build a linear discriminant analysis (LDA) classifier.

Three instrumental techniques are employed in the present study, namely NIR, UV–vis and synchronous fluorescence (SF) spectrometry. All these techniques are fast, non-destructive and do not require the use of chemicals for sample treatment. NIR and UV–vis spectrometry have already been used to discriminate biodiesel samples with respect to feedstock, as described above.

SF is included in the study to investigate its potential as an alternative to NIR and UV–vis spectrometries. Indeed, the biodiesel production process generates fluorescent pigments that could act as intrinsic markers for a particular type of feedstock [33]. Moreover, SF spectrometers are relatively simple instruments commonly found in chemical analysis laboratories.

2. Experimental

2.1. Samples

The biodiesel samples were produced from corn, sunflower and soybean oils of different lots and manufacturers, which were acquired in Paraíba (Brazil). For the NIR/UV–vis study, ethyl esters were produced by transesterification of oils with ethanol in the presence of a catalyst (KOH). The glycerine byproduct was separated and the resulting biodiesel was washed with water and dried. A total of 78 samples (27, 24 and 27 from corn, sunflower and soybean, respectively) were produced. A similar biodiesel synthesis procedure was employed for the SF spectrometry study, the only difference being the use of methanol instead of ethanol. In this case, a total of 60 samples (20 from each oil type) were produced.

2.2. Spectrum acquisition

The UV–vis and NIR spectra were acquired by using a UV–vis/NIR FOSS XDS MasterLab spectrometer with a resolution of 0.5 nm. The working ranges were adopted as in previous works, namely 400–780 nm for UV–vis [14] and 2000–2222 nm for NIR [13].

A computer-controlled spectrofluorimeter SLM Aminco Bowman series 2, equipped with a xenon discharge light source (150 W), was used to obtain the SF spectra. Wavelength accuracy and wavelength repeatability were ± 0.5 and ± 0.25 nm, respectively. Excitation and emission slits of 8 nm were used. A standard quartz cuvette (600 μl working volume) was employed. The scan rate was 5 nm s^{-1}. For each sample, ten synchronous spectra were obtained by scanning both monochromators simultaneously at constant wavelength differences ($\Delta\lambda = \lambda_{\text{emission}} - \lambda_{\text{excitation}}$) of 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 nm. The excitation range 280–600 nm was the same for all spectra, whereas the emission range varied from 285–605 nm to 330–650 nm according to the wavelength difference ($\Delta\lambda$) employed.

2.3. Software

Spectral pre-processing (Savitzky-Golay smoothing), Principal Component Analysis (PCA) and SIMCA were carried out in The Unscrambler® 9.7. Sample set partitioning (Kennard-Stone algorithm) and SPA-LDA modelling were implemented in MatLab 2010a®.

The Kennard-Stone algorithm [34] was applied to the spectra in order to divide the samples into training, validation and test sets for SIMCA and SPA-LDA modelling. In SIMCA, the training samples were employed to build a model for each class, with a number of factors optimised according to the validation set. In SPA-LDA, the training samples were used to obtain an LDA model which discriminates all classes simultaneously. In this case, the validation samples were used to guide the variable selection process in order to minimise the risk of incorrect classification. Finally, the test samples were used as an external set to assess the classification performance of the resulting SIMCA and SPA-LDA models.

3. Results and discussion

3.1. UV–vis/NIR spectra

Figs. 1a and b present the UV–vis and NIR spectra of the biodiesel samples, respectively. As can be seen, the spectra display systematic baseline variations. Such problems were corrected by using a first-derivative Savitzky-Golay filter with a second-order polynomial and a 21-point window. Figs. 1c and d present the resulting derivative spectra, which were employed throughout the study. The Kennard-Stone algorithm was employed to divide the data into training, validation and test sets, as indicated in Table 1.

3.2. SF spectra

Fig. 2 presents the synchronous fluorescence spectra of the biodiesel samples for all $\Delta\lambda$ values.

The spectra for each $\Delta\lambda$ were initially smoothed using a Savitzky-Golay filter with a second-order polynomial and an 11-point window. The Kennard-Stone algorithm was then employed to divide the data into training, validation and test sets, as indicated in Table 1. The spectra for each $\Delta\lambda$ were evaluated according to the number of training and validation errors obtained by using SPA-LDA. As can be seen in Fig. 3a, the best choices of $\Delta\lambda$ would be 10 nm or 15 nm, for which all training and validation samples were correctly classified. Due to parsimony considerations, $\Delta\lambda = 10$ nm was selected because the corresponding model has a smaller number of spectral variables (eight) as compared to $\Delta\lambda = 15$ nm (nine), as shown in Fig. 3b.

The spectra for $\Delta\lambda = 10$ nm, which were employed in all subsequent calculations, are indicated by an arrow in Fig. 2. These spectra extend from 295 to 615 nm. According to the literature, this range comprises intense fluorescence bands of polyphenols [35] and pherols [36] between 300 and 330 nm, as well as low intensity emission bands of vitamins between 350 and 600 nm [37].

3.3. SIMCA classification

Table 2 presents the number of classification errors for SIMCA, which can be either of type 1 (sample not included in its own

class) or type 2 (sample included in an incorrect class). As can be seen, most errors were of type 2. For the SF data, a large number of type-2 errors were obtained for the three biodiesel classes, which means that many samples were assigned to more than one class. Such a result is not adequate because the final classification for these samples would be undecided. In the UV–vis case, good results were obtained for the sunflower class, but the other two classes still exhibited a significant number of type-2 errors. Finally, the use of NIR data provided a perfect classification of the corn-based samples and a good result for soybean. However, the number of type-2 errors for the sunflower class was the largest in comparison with SF and UV–vis.

Table 1
Sample set partitioning.

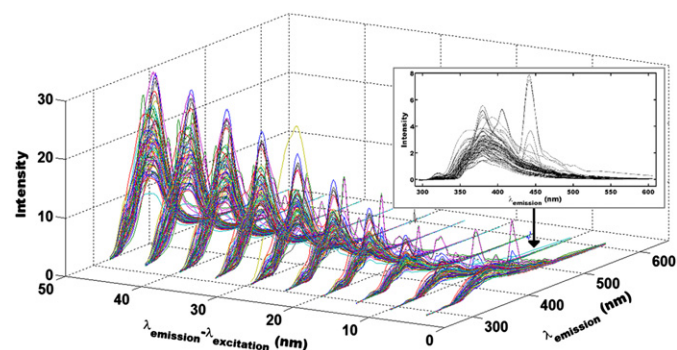| Class	UV-Vis/NIR			SF		
	Training	Validation	Test	Training	Validation	Test
Corn	14	6	7	10	5	5
Sunflower	12	6	6	10	5	5
Soybean	14	6	7	10	5	5
Total	40	18	20	30	15	15

Fig. 2. Synchronous fluorescence spectra obtained by scanning both monochromators simultaneously at constant wavelength differences ($\Delta\lambda = \lambda$emission − λexcitation) of 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 nm. The arrow indicates the spectra for $\Delta\lambda = 10$ nm, which were employed in all calculations.
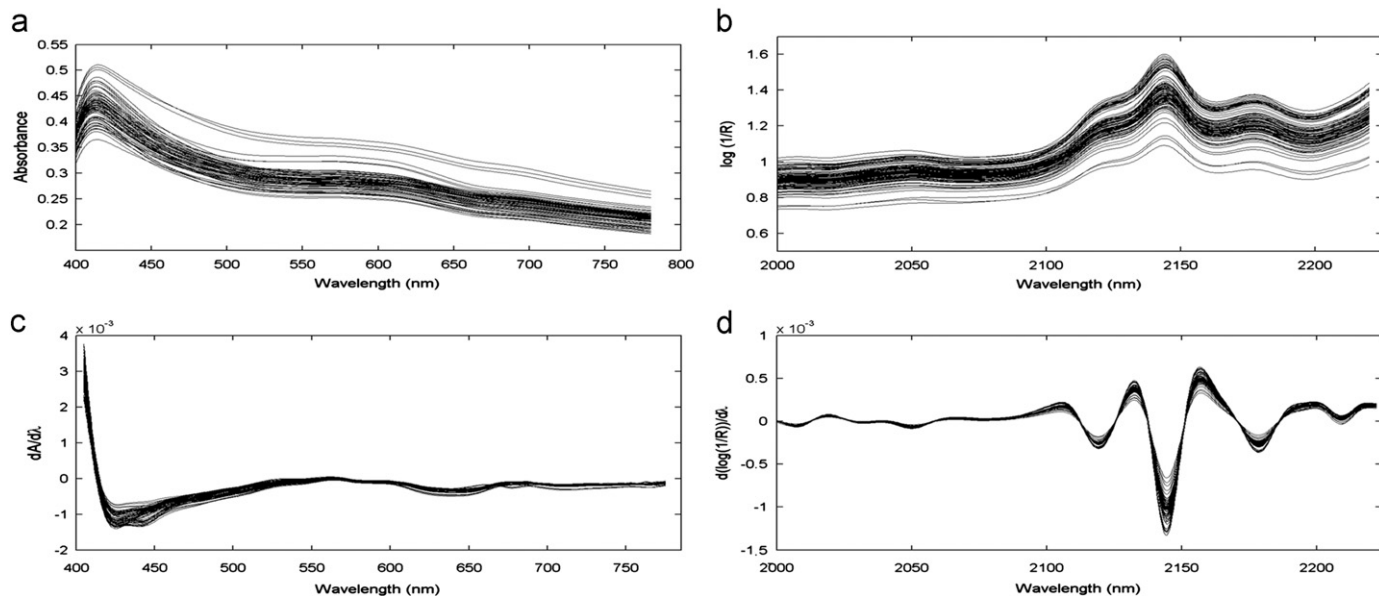

Fig. 1. Raw (a, b) and derivative (c, d) spectra in the UV–vis (a, c) and NIR (b, d) ranges.
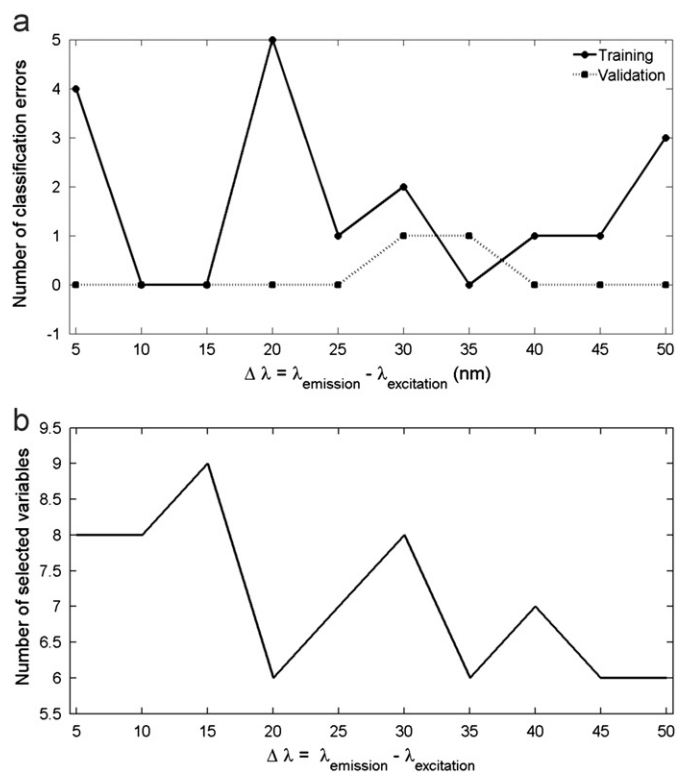
a

b

Fig. 3. Choice of $\Delta\lambda$ for SPA-LDA modelling. (a) Number of classification errors in the training and validation sets. (b) Number of variables selected by SPA-LDA.

Table 2
Number of SIMCA classification errors (type 1 [a]/type 2 [b]) for the Training (Train), Validation (Val) and Test sets using the default significance level (5%).

Class	UV-Vis			NIR			SF		
	Train	Val	Test	Train	Val	Test	Train	Val	Test
Corn	0/7	1/4	0/2	0/0	0/0	0/0	0/7	0/4	0/10
Sunflower	0/2	0/2	0/0	0/21	0/11	0/12	0/14	0/6	0/9
Soybean	0/9	0/5	0/5	0/3	0/1	0/2	0/5	0/6	0/8

[a] Sample not included in its own class.
[b] Sample included in an incorrect class.

These classification results are corroborated by the PC1 × PC2 score plots for the UV–vis, NIR and SF data, which are presented in Figs. 4a, b and c, respectively. In the UV–vis case, Fig. 4a reveals that sunflower is reasonably well separated from corn and soybean, but these two classes overlap each other. In the NIR case, Fig. 4b indicates a good separation between corn and soybean. However, since the corn class exhibits a large dispersion, the SIMCA corn model encompasses the sunflower samples, causing the type-2 sunflower errors observed in Table 2. At last, no separation among the classes is apparent for the SF data in Fig. 4c. As can be seen, regardless of the spectrometric technique employed, an appropriate discrimination of the three biodiesel types was not obtained. It may be argued that the use of full spectrum models is masking the discriminatory information of interest. In order to investigate this hypothesis, SPA-LDA was employed to choose a suitable subset of wavelengths for classification purposes, as described below.

3.4. SPA-LDA classification

By applying SPA-LDA, the number of selected variables for the UV–vis, NIR and SF data sets was 31, 3 and 8, respectively. These
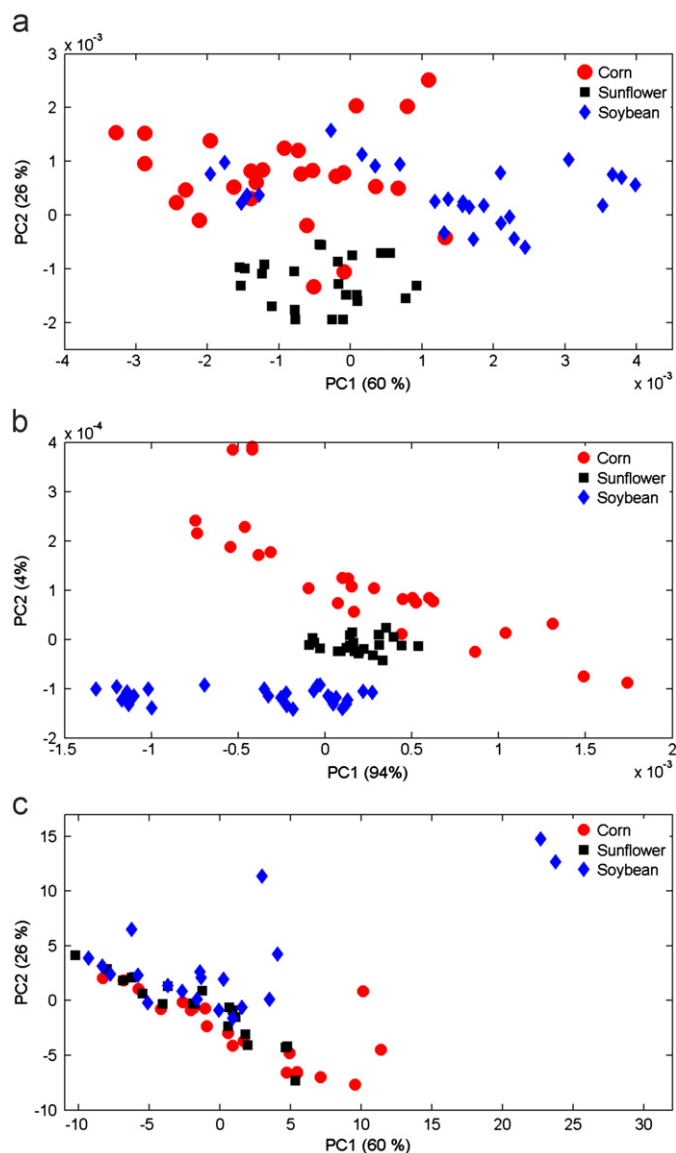
a

b

c

Fig. 4. PC1 × PC2 score plot for the UV–vis (a), NIR (b) and SF (c) spectral data. The percentage of explained variance is indicated in parenthesis at each axis.

variables were employed to build the Fisher discriminant functions employed in the LDA classification. Since the problem involves three classes (corn, sunflower, soybean), two discriminant functions [38] are defined for each data set (UV–vis, NIR, SF). The resulting SPA-LDA models correctly classified the training, validation and test samples in all cases. The proper separation of the three biodiesel types is illustrated in Fig. 5a (UV–vis), 5b (NIR) and 5c (SF), which present a score plot of the overall data set (training, validation and test) using the two Fisher discriminant functions. As can be seen, the linear discriminant boundaries perfectly separate the samples with respect to the base oil.

4. Conclusion

The present paper extended previous works by investigating the discrimination of corn-based biodiesel from two other biodiesel types (sunflower and soybean) on the basis of UV–vis, NIR and SF spectrometries. The results indicate that wavelength selection is key to the proper discrimination of the samples. In fact, all samples employed in the study were correctly classified
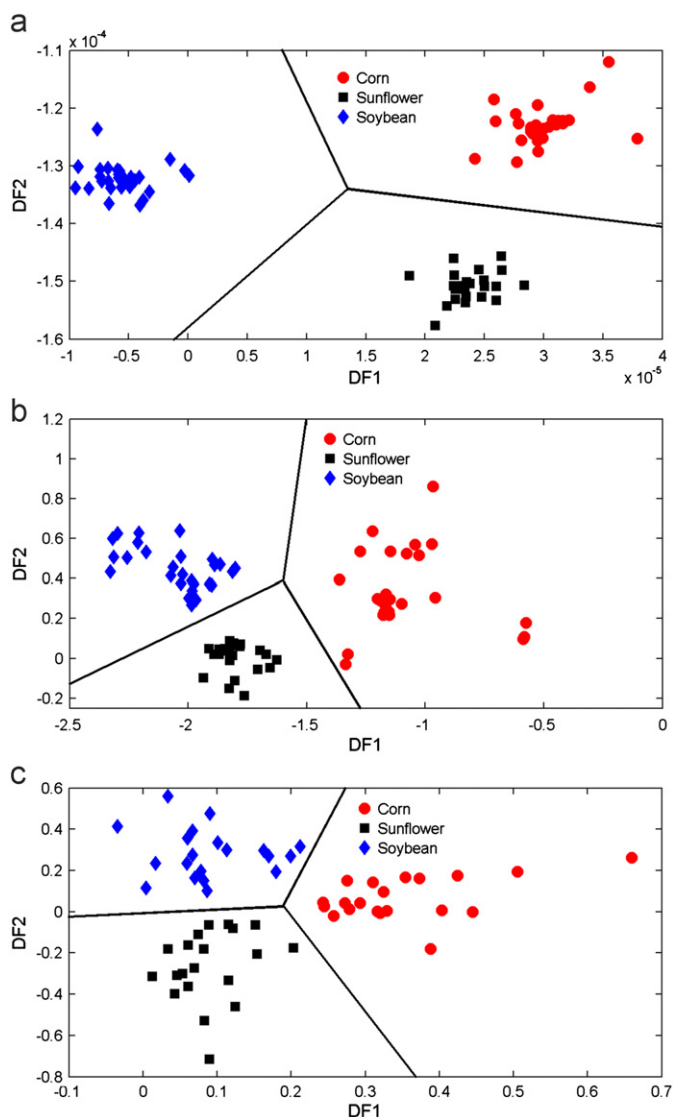
Fig. 5. Score plot of the two Fisher discriminant functions (DF1, DF2) for the UV–vis (a), NIR (b) and SF (c) spectral data. The three straight lines correspond to the boundaries resulting from linear discriminant analysis.

on the basis of a reduced number of wavelengths selected by SPA-LDA. On the other hand, full-spectrum SIMCA did not provide an appropriate discrimination of the three biodiesel types, regardless of the spectrometric technique employed.

Future works may be concerned with the application of the proposed method for the analysis of biodiesel produced from blends of feedstock.

Acknowledgements

References

[1] N. Law 11097, ANP- Agência Nacional de Petróleo Gás Natural e Bicombustíveis, Brasil, 2005.
[2] Law 26.093, Régimen de Regulación y Promocion para la Produccion y uso Sustentables de Biocombustibles, Argentina, 2006.
[3] G. Knothe, J.V. Gerpen, J. Kral, L.P. Ramos, Manual de Biodiesel, Edgard Blucher, São Paulo, 2006.
[4] J. Janaun, N. Ellis, Renew. Sustain. Energy Rev. 14 (2010) 1312.
[5] B.R. Moser, S.F. Vaughn, Biomass and Bioenergy 37 (2012) 31.
[6] B.R. Moser, In Vitro Cell. Dev. Biol. Plant 45 (2009) 229.
[7] A. Karmakar, S. Karmakar, S. Mukherjee, Bioresour. Technol. 101 (2010) 7201.
[8] S.K. Hoekmana, A. Broch, C. Robbins, E. Ceniceros, M. Natarajan, Renew. Sustain. Energy Rev. 16 (2012) 143.
[9] A. Sagiroglu, Ş. Selenisbilir, H. Ozcan, H. Paluzar, N.M. Toprakkiran, Chem. Ind. & Chem. Eng. Q 17 (2011) 53.
[10] K.R. Szulczyk, B.A. McCarl, Renew. Sustain. Energy Rev. 14 (2010) 2426–2433.
[11] G. Karavalakis, E. Bakeas, G. Fontaras, S. Stournas, Energy 36 (2011) 5328.
[12] R.M. Balabin, R.Z. Safieva, Anal Chim. Acta 689 (2011) 190.
[13] G. Veras, A.A. Gomes, A.C. Silva, A.L.B. Brito, P.B.A. Almeida, E.P. Medeiros, Talanta 83 (2010) 565.
[14] G. Veras, A.L.B. Brito, A.C. Silva, P. Silva, G.B. Costa, L.C.N. Félix, Quim. Nova 35 (2012) 315.
[15] K.R. Szulczyk, B.A. McCarl, Ren. Sustain. Energy Rev. 14 (2010) 2426.
[16] M. Balat, Energy Convers. and Manag. 52 (2011) 1479.
[17] A.C. Pinto, L.L.N. Guarieiro, M.J.C. Rezende, N.M. Ribeiro, E.A. Torres, W.A. Lopes, P.A.P. Pereira, J.B. Andrade, J. Braz. Chem. Soc. 16 (2005) 1313.
[18] K.R. Beebe, R.J. Pell, B. Seasholtz, Chemometrics – a pratical guide, Wiley, New York, 1998.
[19] H.C. Goicoechea, A.C. Olivieri, J. Chemom 17 (2003) 338.
[20] U. Horchner, J.H. Kalivas, J. Chemom 283 (1995) 308.
[21] R.K.H. Galvão, M.C.U. Araújo, in: B. Walczak, R. Tauler, S. Brown (Eds.), Comprehensive Chemometrics, 3, Elsevier, Oxford, 2009., pp. 233.
[22] M.C.U. Araújo, T.C.B. Saldanha, R.K.H. Galvão, T. Yoneyama, H.C. Chame, V. Visani, Chemom. Intell. Lab. Syst. 65 (2001) 57.
[23] M.J.C. Pontes, R.K.H. Galvão, M.C.U. Araújo, P.N.T. Moreira, O.D. Pessoa Neto, G.E. José, T.C.B. Saldanha, Chem. Intell. Lab. Syst. 78 (2005) 11.
[24] R.K.H. Galvão, M.F. Pimentel, M.C.U. Araújo, T. Yoneyama, V. Visani, Anal. Chim. Acta 443 (2001) 107.
[25] A.F.C. Pereira, M.J.C Pontes, F.F.G. Neto, S.R.B. Santos, R.K.H. Galvão, M.C.U. Araújo, Food Res. Int. 41 (2008) 341.
[26] M.S. Di Nezio, M.F. Pistonesi, W.D. Fragoso, M.J.C. Pontes, H.C. Goicoechea, M.C.U. Araujo, B.S.F. Band, Microchem. J. 85 (2007) 194.
[27] E.D.T. Moreira, M.J.C. Pontes, R.K.H. Galvão, M.C.U. Araujo, Talanta 79 (2009) 1260.
[28] F.F.G. Neto, G. Marino, M.C.U. Araújo, R.K.H. Galvão, M.J.C. Pontes, E.P. Medeiros, R.S. Lima, Talanta 77 (2009) 1660.
[29] M.J.C. Pontes, J. Cortez, R.K.H. Galvão, C. Pasquini, M.C.U. Araújo, R.M. Coelho, M.K. Chibad, M.F. Abreud, B.E Madarie, Anal. Chim. Acta 642 (2009) 12.
[30] U.T.C.P. Souto, M.J.C. Pontes, E.C. Silva, R.K.H. Galvão, M.C.U. Araújo, F.A.C. Sanches, F.A.S. Cunha, M.R.S. Oliveira, Food Chem. 119 (2010) 368.
[31] M. Ghasemi-Varnamkhasti, S.S. Mohtasebi, M.L. Rodriguez-Mendez, A.A. Gomes, M Cesar, U. Araújo, R.K.H. Galvão, Talanta 89 (2012) 286.
[32] M.J.C. Pontes, C.F. Pereira, M.F. Pimentel, F.V.C. Vasconcelos, A.G.B. Silva, Talanta 85 (2011) 2159.
[33] K. Kumar, A.K. Mishra, J. Fluoresc. 22 (2012) 339.
[34] R.K.H. Galvão, M.C.U. Araújo, G.E. José, M.J.C. Pontes, E.C. Silva, T.C.B. Saldanha, Talanta 67 (2005) 736.
[35] S. Schober, M. Mittelbach, Eur. J. of Lipid Sci. and Technol. 106 (2004) 382.
[36] A. Fröhlich, S. Schober, J. Am. Oil Chem. Soc. 84 (2007) 579.
[37] M. Zandomeneghi, L. Carbonaro, C. Caffarata, J. Agric. Food Chem. 53 (2005) 759.
[38] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, John Wiley, New York, 2001.