Talanta 83 (2010) 565–568

Contents lists available at ScienceDirect

Talanta

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

Classification of biodiesel using NIR spectrometry and multivariate techniques

Germano Veras[a,*], Adriano de Araujo Gomes[a], Adenilton Camilo da Silva[a], Anna Luiza Bizerra de Brito[a], Pollyne Borborema Alves de Almeida[b], Everaldo Paulo de Medeiros[b]

[a] Universidade Estadual da Paraíba, Departamento de Química, 58.429-500, Campina Grande, PB, Brazil
[b] Empresa Brasileira de Pesquisa Agropecuária, Centro Nacional de Pesquisa de Algodão, 58.428-095, Campina Grande, PB, Brazil

ARTICLE INFO

Article history:
Received 23 August 2010
Received in revised form
29 September 2010
Accepted 30 September 2010
Available online 8 October 2010

Keywords:
Biodiesel
NIR
Pattern recognition

ABSTRACT

This article describes the classification of biodiesel samples using NIR spectroscopy and chemometric techniques. A total of 108 spectra of biodiesel samples were taken (being three samples each of four types of oil, cottonseed, sunflower, soybean and canola), from nine manufacturers. The measurements for each of the three samples were in the spectral region between 12,500 and 4000 $\mathrm{cm}^{-\mathbf{1}}$. The data were preprocessed by selecting a spectral range of 5000–4500 cm^{-1}, and then a Savitzky–Golay second-order polynomial was used with 21 data points to obtain second derivative spectra. Characterization of the biodiesel was done using chemometric models based on hierarchical cluster analysis (HCA), principal component analysis (PCA) and soft independent modeling of class analogy (SIMCA) elaborated for each group of biodiesel samples (cotton, sunflower, soybean and canola). For the HCA and PCA, the formation of clusters for each group of biodiesel was observed, and SIMCA models were built using 18 spectral measurements for each type of biodiesel (training set), and nine spectral measurements to construct a classification set (except for the canola oil which used eight spectra). The SIMCA classifications obtained 100% accurate identifications. Using this strategy, it was feasible to classify biodiesel quickly and nondestructively without the need for various analytical determinations.

1. Introduction

Interest in fuels that replace petroleum derivates has increased because of their potential for reduced emissions and renewability [1]. This interest has also grown dramatically since the 1970s when the Organization of Petroleum Exporting Countries (OPEC) decided to restrict the amount of oil to be made available, causing a substantial increase in the price of oil per barrel and two major world crises. Since then the search for renewable energy sources (with less environmental impact), has become part of the agenda for governments around the world [1]. A renewable fuel, biodiesel is a good alternative to petroleum-derived diesel, and is able to significantly reduce the emission of many greenhouse gases like carbon dioxide and sulfur. It can also be used alone by itself, or mixed with diesel in any proportion [2]. Biodiesel consists of esters of fatty acids, obtained by transesterification of vegetal oils and animal fats with short-chain alcohols (methanol or ethanol), in the presence of sodium hydroxide as a catalyst.

Along with growing interest and the increase in worldwide production, studies and quality assessments of biodiesel have multiplied for both stand alone [3–14] and blended biodiesel–diesel [15–21].

Several analytical techniques are used in biodiesel analysis, among which NIR spectrometry [22–27], a rapid and nondestructive analytical technique based on the absorption of electromagnetic energy between 400 and 14,000 cm^{-1} is useful to predict various properties. However, despite successful application of NIR techniques to the classification of beverages [28,29], foods [29–32], cigarettes [33] and fossil fuels [34–43], few articles have been published for biodiesel [24].

Considering the lack of procedures to identify the raw materials used in obtaining a certain biodiesel (i.e. government verification for tax exemptions), this study united NIR spectrometry, and pattern recognition for biodiesels in order to identify which vegetable oils are used in production. Principal component analysis (PCA), and hierarchical cluster analysis (HCA) were used for unsupervised pattern recognition while self modeling independent cluster analysis (SIMCA), was used for supervised pattern recognition.

2. Experimental

2.1. Reagents, samples and apparatus

Four vegetable oils: cottonseed, soybean, canola, and sunflower oils, twelve samples of each from different manufacturer

---

[*] Corresponding author at: Departamento de Química, Universidade Estadual da Paraíba, 58.429-500, Campina Grande, PB, Brazil. Tel.: +55 83 3315 3356.
E-mail address: germano.veras@pq.cnpq.br (G. Veras).

0039-9140/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.talanta.2010.09.060

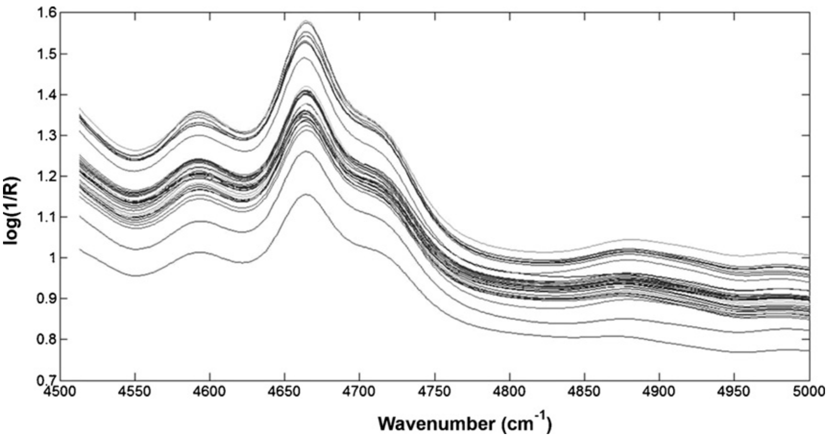

Fig. 1. Average gross spectra from biodiesel samples of cotton, canola, sunflower and soybean in the spectral region of 5000 and 4500 cm^{-1}.

batches were acquired in the city of Campina Grande, Paraíba, Brazil.

From these 12 samples of four vegetable oils, three biodiesel samples each were synthesized. Transesterification reaction with ethanol and homogeneous catalysis using KOH and a molar ratio of 6:1 of ethanol/oil ensured excess ethanol with 1% (w/w) catalyst. Afterwards, the biodiesel samples were washed with distilled water and hydrochloric acid solution 0.1 mol L^{-1}, for neutralization at a pH of 7.0.

2.2. NIR spectra measurements

Diffused reflectance spectroscopic measurements were performed by using an XDS Master Lab spectrometer (FOSS), equipped with quartz cells (4 mm optical path). The spectral resolution was 1.10 cm^{-1}. For each sample, three spectra were recorded in the 12,500–4000 cm^{-1} wavelength range, after which the three spectra were treated with chemometrics pattern recognition techniques and HCA classification, carried out on Statistics 9.0, and PCA and SIMCA, using Unscrambler 9.8.

2.3. Chemometrics study

The spectra were preprocessed by selecting a range of wavelengths, 5000–4500 cm^{-1} as a working region, because in this region overtones of C–H and C=O stretching combination and C–H and C=C asymmetric stretching combination occur. To remove noise the spectra were then treated using Savitzky–Golay first derivative procedure with a second-order polynomial and a 21-point window in the Unscrambler software. The resulting derivative spectra comprised 380 variables. PCA and HCA analysis were applied to the data matrix verifying the clustering of biodiesel samples from the cottonseed, soybean, canola, and sunflower vegetable oils.

The HCA dendrogram was conducted in complete linkage mode (amalgamation rule), using 1-Pearson *r* distances (distance measure). The PCA classification model was validated by (leave-one out) cross-validation procedure.

The sample set was constructed using 12 biodiesel samples from each vegetable oil.

The SIMCA models were built from 18 spectral measurements for each biodiesel. The classification set was composed of 36 spectral measurements (9 for each type of biodiesel). The optimal number of principal components (PC) for each model were one for cottonseed e soybean and two for canola and sunflower.

3. Results and discussion

3.1. Preprocessing of spectral data

The average spectra of biodiesel samples are shown in Fig. 1, in the working spectra range of (4500–5000 cm^{-1}). In Fig. 2, the signals of the preprocessed data, to which were applied the Savitzky–Golay first derivative procedure with a second-order polynomial and a 21-point window are shown.

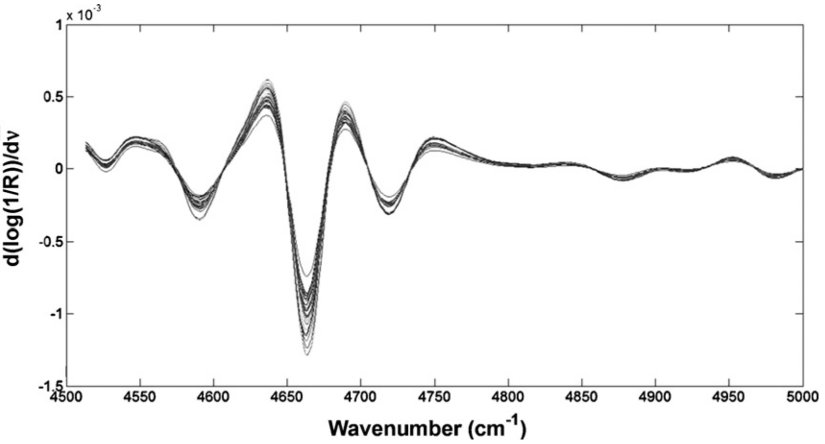

Fig. 2. Average preprocessed spectra of biodiesel samples of cotton, canola, sunflower and soybean in the spectral region of 5000 and 4500 cm^{-1}.

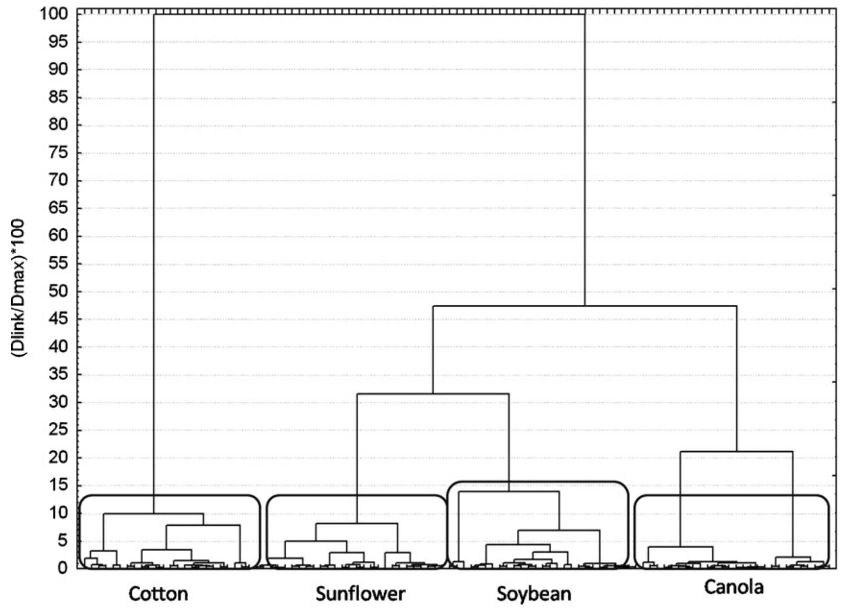

Fig. 3. Dendrogram analysis (HCA) for biodiesel samples of cotton, canola, sunflower and soybean.

3.2. Clustering using HCA and PCA

Fig. 3 shows the dendrogram for the study of biodiesel samples, based on HCA techniques. For this, the dendrogram was conducted in complete linkage mode (amalgamation rule) using 1-Pearson *r* distances (distance measure). The clusters were formed based on the relationship between binding distance, and the maximum distance binding: cotton cluster (binding distance of 21%), canola cluster (10%), sunflower cluster (14%) and soybean (8%). It was also observed that the groups of biodiesel samples differed by the following distances; soybean cluster to sunflower (binding distance of 32%), soybean/sunflower cluster to cotton (47%), soybean/sunflower/cotton cluster to canola (100%).

In Fig. 4 the scores graph PC1 (93%) × PC2 (6%) for the biodiesel samples is studied, clearly demonstrating four clusters of biodiesel samples from the oils. PC1 separates cotton biodiesel from soybean biodiesel. The difference between the two oils is the presence of linolenic, eicosanoic and eicosenoic acids in soybean oil, which obviously produce esters. PC2 can separate biodiesel samples from canola, sunflower and cotton/soybean. The difference in the fatty acids of the oils is in the proportion of oleic and linoleic acids in soybean and sunflower oils and linoleic acid in cotton/soybean, virtually the only differences between these oils.

3.3. SIMCA classification

The training set for each type of biodiesel was based on 21 spectral measurements. The model validation was based on full cross validation (leave-one-out). The classification set was constituted of six spectra for each type of biodiesel. It was verified that in all cases, the samples from the set were classified correctly, with no occurrence of type I and II errors. As an example of class separation, the Coomans plot at the 95% confidence level, which shows the orthogonal distances of samples from the classification set from the models of cotton biodiesel versus sunflower biodiesel is presented in Fig. 5. Similar results are obtained with the other Coomans plots.

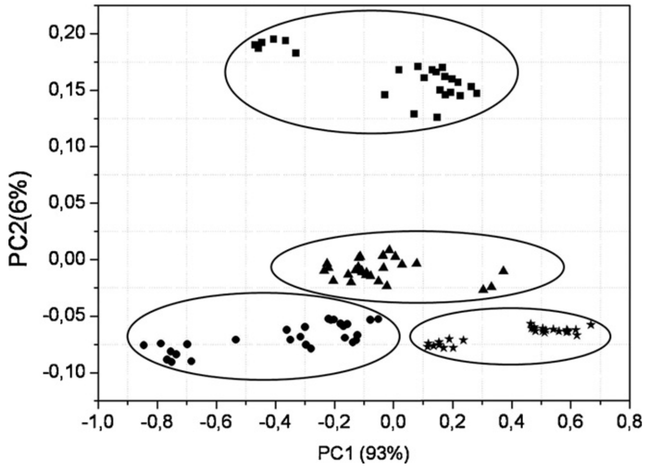

Fig. 4. PC1 × PC2 score plot for the overall set of 108 biodiesel samples (● cotton; ⋆ soybean; ▲ sunflower; ■ canola). The explained variance by each principal component is indicated in parenthesis.

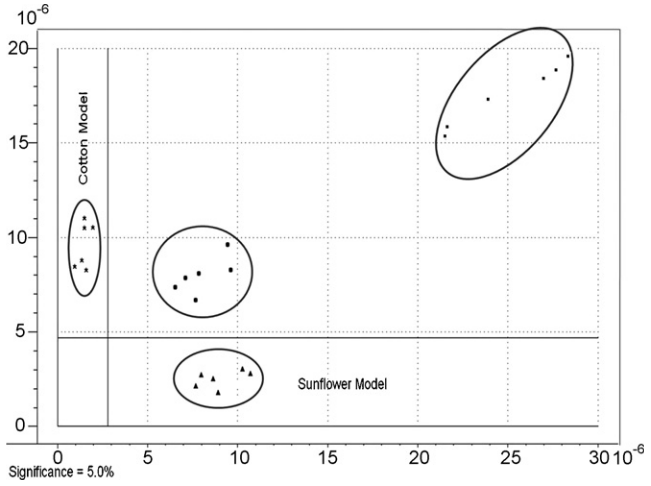

Fig. 5. SIMCA Coomans plot of soybean and cotton biodiesel models for classification set.

4. Conclusions

This paper proposes a classification method based on chemometric modeling and NIR spectrometry, for identifying the oil(s) originated the biodiesel. The results from PCA and HCA unsupervised classification techniques show the formation of distinct clusters for each type of biodiesel while demonstrating that the canola group differed notably from the others. It was also shown that the SIMCA models allowed classification of biodiesel samples based on the seed oil used without type I or type II errors.

The classification models using SIMCA techniques were reliable showing that the samples have structures relating each biodiesel to its seed oil. The results also suggest that the proposed methods hold promise for the assessment of biodiesel samples using spectroscopic measurements and chemometric techniques alone.

Acknowledgment

This work was supported by CNPq (Process no. 576416/2008-8).

References

[1] J. Janaun, N. Ellis, Renew. Sust. Energy Rev. 14 (2010) 1312.
[2] Y.H. Huang, J.H. Wu, Renew. Sust. Energy Rev. 12 (2008) 1176.
[3] G. Knothe, J. Am. Oil Chem. Soc. 76 (1999) 795.
[4] F. Saloua, C. Saber, Z. Hedi, Bioresour. Technol. 101 (2010) 3091.
[5] R.A. Ferrari, V.S. Oliveira, A. Scabio, Quim. Nova 28 (2005) 19.
[6] J.R.O. Lima, R.B. Silva, C.C.M. Silva, L.S.S. Santos, J.R. Santos Jr., E.M. Moura, C.V.R. Moura, Quim. Nova 30 (2007) 600.
[7] R.A. Ferrari, W.L. Souza, Quim. Nova 32 (2009) 106.
[8] L. Fjerbaek, K.V. Christensen, B. Norddahl, Biotechnol. Bioeng. 102 (2009) 1298.
[9] R. Falate, K. Nike, P.R. Costa Neto, E. Cação Jr., M. Muller, H.J. Kalinowski, J.L. Fabris, Quim. Nova 30 (2007) 1677.
[10] G. Knothe, Eur. J. Lipid Sci. Technol. 108 (2006) 493.
[11] B.R. Moser, In Vitro Cell. Dev. Biol. Plant 45 (2009) 229.
[12] J.P. Davis, D. Geller, W.H. Faircloth, T.H. Sanders, J. Am. Oil Chem. Soc. 86 (2009) 353.
[13] P.C. Smith, Y. Ngothai, Q.D. Nguyen, B.K. O'Neill, Renew. Energy 35 (2010) 1145.
[14] R.R. Catharino, H.M.S. Milagre, S.A. Saraiva, C.M. Garcia, U. Schuchardt, M.N. Eberlin, R. Augusti, R.C.L. Pereira, M.J.R. Guimarães, G.F. Sá, J.M.R. Caixeiro, V. Souza, Energy Fuels 21 (2007) 3698.
[15] C.M. Reddy, J.A. Demello, C.A. Carmichael, E.E. Peacock, L. Xu, J.S. Arey, Environ. Sci. Technol. 42 (2008) 2476.
[16] G. Knothe, J. Am. Oil Chem. Soc. 78 (2001) 1025.
[17] M.R. Monteiro, A.R.P. Ambrozin, M.S. Santos, E.F. Boffo, E.R. Pereira-Filho, L.M. Lião, A.G. Ferreira, Talanta 78 (2009) 660.
[18] N.N. Mahamuni, Y.G. Adewuyi, Energy Fuels 23 (2009) 3773.
[19] L.L.N. Guarieiro, A.C. Pinto, P.F. Aguiar, N.M. Ribeiro, Quim. Nova 31 (2008) 421.
[20] J.A. Cramer, R.E. Morris, B. Giordano, S.L. Rose-Pehrsson, Energy Fuels 23 (2009) 894.
[21] D.H. Qi, H. Chen, L.M. Geng, Y.Z.H. Bian, X.C.H. Ren, Appl. Energy 87 (2010) 1679.
[22] P. Baptista, P. Felizardo, J.C. Menezes, M.J.N. Correia, Anal. Chim. Acta 607 (2008) 153.
[23] A.F.C. Pereira, M.J.C. Pontes, F.F. Gambarra Neto, S.R.B. Santos, R.K.H. Galvão, M.C.U. Araújo, Food Res. Int. 41 (2008) 341.
[24] F. Gautius, J. Lloveras, J. Ferran, J. Puy, J. Agric. Sci. 142 (2004) 517.
[25] L.F.B. Lira, F.V.C. Vasconcelos, C.F. Pereira, A.P.S. Paim, L. Stragevitch, M.F. Pimentel, Fuel 89 (2010) 405.
[26] P. Felizardo, P. Baptista, J.C. Menezes, M.J.N. Correia, Anal. Chim. Acta 595 (2007) 107.
[27] I.K. Oliveira, W.F.C. Rocha, R.J. Poppi, Anal. Chim. Acta 642 (2009) 217.
[28] M.J.C. Pontes, S.R.B. Santos, M.C.U. Araújo, L.F. Almeida, R.A.C. Lima, E.N. Gaião, U.T.C.P. Souto, Food Res. Int. 39 (2006) 182.
[29] T. Woodcock, G. Gowney, C.P.O. Donnell, J. Near Infrared Spectrosc. 16 (2008) 1.
[30] D. Toher, G. Downey, T.B. Murphy, Chemom. Intell. Lab. Syst. 89 (2007) 102.
[31] L.A. Berrueta, R.M. Alonso-Salces, K. Héberger, J. Chromatogr. A 1158 (2007) 196.
[32] A.M.C. Davies, J.G. Franklin, A. Grant, N.M. Griffiths, R. Shepherd, G.R. Fenwick, Vib. Spectrosc. 2 (1991) 161.
[33] E.D.T. Moreira, M.J.C. Pontes, R.K.H. Galvão, M.C.U. Araujo, Talanta 79 (2009) 5.
[34] R.M. Balabin, R.Z. Safieva, Fuel 87 (2008) 1096.
[35] 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, Chemom. Intell. Lab. Syst. 78 (2005) 11.
[36] M. Kim, Y.H. Lee, C. Han, Comput. Chem. Eng. 24 (2000) 513.
[37] R.M. Balabin, R.Z. Safieva, E.I. Lomakina, Anal. Chim. Acta 671 (2010) 27.
[38] R.M. Balabin, R.Z. Safieva, E.I. Lomakina, Neural Comput. Applic. 18 (2009) 557.
[39] R.M. Balabin, R.Z. Safieva, E.I. Lomakina, Chemom. Intell. Lab. Syst. 93 (2008) 58.
[40] R.M. Balabin, R.Z. Safieva, E.I. Lomakina, Chemom. Intell. Lab. Syst. 88 (2007) 183.
[41] R.M. Balabin, R.Z. Safieva, Fuel 87 (2008) 2745.
[42] J.J. Kelly, C.H. Barlow, T.M. Jinguji, J.B. Callis, Anal. Chem. 61 (1989) 313.
[43] R.M. Balabin, R.Z. Safieva, J. Near Infrared Spectrosc. 15 (2007) 343.