Determination of Biodiesel Adulteration with Raw Vegetable Oil from ATR-FTIR Data using Chemometric Tools

Itânia P. Soares,*a Thaís F. Rezende,a Rita de Cássia. C. Pereira,b Cláudio G. dos Santos,c and Isabel C. P. Fortesb

*aEmbrapa Agroenergia, Parque Estação Biológica, Av. W3 Norte s/n, 70770-901 Brasília-DF, Brazil
bLaboratório de Ensaio de Combustíveis, Departamento de Química, Universidade Federal de Minas Gerais, Av. Antônio Carlos, 6627 - Campus Pampulha, 31270-901 Belo Horizonte-MG, Brazil
cDepartamento de Química, Universidade Federal de Ouro Preto, Rua Diogo de Vasconcelos 122, 35400-000 Ouro Preto-MG, Brazil

Três origens diferentes de biodiesel (algodão, mamona e palma) foram adulteradas com óleo de soja cru, em concentrações variando de 1 a 40% (m/m). Estas amostras foram analisadas por espectrometria de infravermelho médio (MIR) e os seus espectros foram estudados em três diferentes faixas espectrais: espectro inteiro (4000-665 cm⁻¹), e nas faixas de 1800-1700 cm⁻¹ e 1800-1000 cm⁻¹. Para determinar a origem do biodiesel utilizado no sistema adulterado, os dados espectrais foram analisados pela ferramenta análise de componentes principais (PCA) e a melhor segregação das origens foi obtida para o espectro inteiro (4000-665 cm⁻¹). A variação explicada foi de 99%, para os três primeiros componentes. Para quantificar o óleo de soja cru foi aplicada a ferramenta mínimos quadrados parciais (PLS). Os melhores resultados foram obtidos para a região espectral de 1800-1000 cm⁻¹, com valores de RMSEP (erro médio quadrático de previsão) variando de 1,10 a 1,47% (m/m).

Introduction

Biodiesel is usually produced by the transesterification of vegetable oil or animal fat with a short alcohol chain in the presence of a catalyst. Since the CO₂ released during combustion is captured by the oleaginous plant, biodiesel represents an important fuel alternative. This contributes to the reduction in the emission of CO₂, which is the main responsible for the greenhouse effect. Combustion of biodiesel also reduces particulate material and SOₓ emission when compared to conventional fossil fuel. Biodiesel is mainly produced from rapeseed oil in Europe and other countries in the world. In Brazil, there is a large number of oleaginous plants, which could be used for biodiesel production and, because of agroclimatic zoning, some of these oleaginous cultures concentrate in specific regions. Palm, for example, is more common in the northern area while castor is easier to find in the northeast area, and soybean culture develops better in the southern and south eastern areas. This shows the great potential of Brazil as a

*e-mail: itania.soares@embrapa.br
world producer and exporter of this commodity. A common use of biodiesel is in blends with conventional mineral diesel fuel. In Brazil, diesel has been commercialized with the addition of 5% volume of biodiesel since January 2010.

One of the biggest problems of the current fuel (gasoline, ethanol and biodiesel) scenario in Brazil is adulteration, apart from the tax-evasion involved with this practice. Adulteration also results in increasing environmental pollution, as well as, in consumer harm, since the product does not meet the regular specifications, with potential to cause several problems to car engines. In the particular case of biodiesel, government subsidy is different when compared to other fuels. This type of differentiated subsidy can lead to unreal declarations of the biodiesel source and, consequently, to tax evasion. Another problem that can occur in the process of biodiesel production is the addition of raw oil to B100 (pure biodiesel), since the process costs are still very significant. Thus, it is imperative to solve or try to minimize these problems, by developing methodologies which allow both the identification of biodiesels source and the determination of its adulteration.

One of the analytical techniques which have been mostly used to monitor quality of biodiesel and petrodiesel blends is infrared (IR) spectroscopy, due to its many advantages. It is non-destructive, very reliable and allows direct and fast determination of several properties without sample pretreatment. In recent years a number of reports has appeared on the use of multivariate analysis applied to near infrared spectroscopy (NIR) and Fourier-transform infrared spectroscopy (FTIR). By using this approach, Pereira et al. have determined gasoline adulteration; Che Man and Setiowat have determined fatty acid in palmitolein using calibration for partial least squares (PLS), which was also applied by Knothe to monitor the completion of the transesterification reaction of biodiesel. Calibration methods based on FTIR, MIR and NIR spectroscopy have also been developed for the determination of the methyl ester content in biodiesel blends and the content of biodiesel in diesel fuel blends, taking the presence of raw vegetable oil into account. The application of multivariate models to the analysis of biodiesel is valuable because the IR spectra of vegetable oils and their corresponding esters are very similar, resulting in an overlapping band. Nevertheless, this methodology has not been able to segregate biodiesel from different sources or to quantify the biodiesel adulteration with raw vegetable oil. Oliveira et al. used FTIR and NIR spectroscopy to design calibration models for the determination of the methyl ester content in biodiesel blends (methyl ester + diesel).

Other analytical techniques have also been utilized for the characterization of biodiesel profile. Monteiro and co-workers obtained good results by H NMR technique to determine the biodiesel/diesel proportion using samples of soy and castor derived biodiesel mixed with diesel form three different batches. Catharino et al. fingerprinted several origins of biodiesel using electrospray ionization mass spectrometry (ESI-MS). Also using ESI-MS, Eide and Zahlsen fingerprinted biodiesel origins and mixtures of diesel, classifying them with multivariate analysis.

In a previous work, the adulteration of biodiesel with vegetable oil was determined using FTIR with the accessory of attenuated total reflectance (ATR) and PLS calibration, with variable selection. However, when this method is applied there is a need to optimize the model according to each biodiesel origin. In the present work, an even simpler alternative is presented, which uses a spectral range where there is a high correlation between the IR absorbance and the grade of adulterant that can be applied to biodiesel produced from any origin. Principal components analysis (PCA) was used to classify the biodiesel origin also using different spectral ranges.

**Experimental**

**Samples**

Raw soybean oil was purchased from a local market. Biodiesel used in the experiments were from companies and/or Universities, which have already produced them for the market or on a bench scale, kindly donated. Castor oil ester was supplied by Santa Cruz State University (Bahia State, Brazil), palm oil ester by Agropalma Company (Pará State, Brazil), and cotton oil ester was obtained from Soyminas (Minas Gerais State, Brazil). The samples were produced by classic basic transesterification reaction using methylic route. Samples were characterized according to current parameters established by the National Agency for Petrol, Natural Gas and Biofuels (ANP), Resolution 07/2008. The assays were done at the Laboratory for Fuel Assays (LEC) of Federal University of Minas Gerais (Minas Gerais State, Brazil).

A total of 120 samples were prepared by mixing biodiesel from different sources with raw soybean oil in percentages varying from 1 to 40% m/m with 1% m/m increments. These samples were used as classifying for PCA and calibration set. The external validation set comprised other 15 samples of each source which were prepared in the same way as the calibration set, but the percentage of raw soybean oil added was randomly chosen, resulting in 45 samples.

**ATR-FTIR analysis**

ATR-FTIR spectra were obtained from an ABB Bomen IR spectrometer model MB 102 equipped with an
ATR sampling accessory with a deuterated triglycerine sulfate detector. All spectra were collected at 16 ± 1 °C using an average of 16 scans, with spectral resolution of 2 cm⁻¹. The background spectra were obtained using a clean ATR accessory with an average of 100 scans. After recording each spectrum, the cell was cleaned by successive treatments with heptane. The average spectra in the range 4000-665 cm⁻¹ from triplicate analysis were treated chemometrically using MINITAB software®, version 14.

Modeling and data analysis

PCA is a well-known tool in multivariate data analysis for visualizing information from large data sets. PCA relies on the linear transformation of the original set of measurements into a substantially smaller set of uncorrelated variables while retaining as much information present in the original data set as possible.20,21 The original data set is substituted by two matrices that contain information about the weight of the original variable in the PC space (loading matrix) and the scattering of the samples in this space (score matrix). Thus, graphical presentation of the pair-wise components allows the natural grouping of the samples to be observed indicating the similarity between samples and allowing different groups of samples to be identified. In this work, PCA was employed to verify the possibility of classifying samples biodiesel from cotton, castor and palm oils, adulterated by different levels of raw soybean oil. Since all variables considered in this study were within the same scale, PCs were obtained from the covariance matrix.

PLS regression is a popular multivariate calibration method aiming to assess the degree of relationship between a set of x-predictor variables and a set of y-outcome variables.1 It has been widely applied to multicomponent spectral analysis, especially in IR, NIR and Raman spectroscopy. PLS is a full spectral calibration method and has built-in capacity to deal with specific problems of full spectrum calibration.22 However, the selection of wavelength or wavenumber region is still very important.23,24 An important goal is to search for informative spectral regions for multicomponent spectral analysis. Informative regions mean that they contain useful information for building a PLS model and are helpful to improve the performance of the model.20

PLS is a powerful approach for the analysis of mixtures and was employed to determine the concentration of soybean oil in biodiesel, using leave-one-out cross validation method. Predictive residual error sum of squares (PRESS) is a commonly used criterion for LVs number selection.22 For every set of data from each biodiesel source, a PLS model with a selected LV number is built, and root mean square error of calibration for cross validation (RMSECV). Once the external validation is made, the root mean square error of prediction (RMSEP) can be calculated.

Results and Discussion

Physical-chemical assays

In order to simulate an adulterated system it is fundamental to begin with samples that are considered within specification, according to Brazilian legislation (Resolution 07/2008). Thusly, the physicochemical parameters were previously obtained for all the biodiesel samples used in this study. The assays were performed according to the national and international standards (ABNT/ASTM /EN), as presented in Table 1. The specified values by the regulatory agency and the results obtained of the biodiesel samples from three distinct sources are showed in Table 1. The results show that the samples used meet the requirements of current legislation. It is noteworthy that all samples presented grades of ether higher than 96.5%, which is considered threshold for it to be marketed.

ATR-FTIR analysis

MIR spectra of castor, palm and cotton biodiesel are very similar to that of non-esterified soybean oil (Figure 1), showing absorption bands in the regions 3700 to 3000 cm⁻¹, 1900 to 1500 cm⁻¹ and 1800 to 800 cm⁻¹. Another important feature in these spectra comes from the distinctive band at 3333 cm⁻¹ in the castor oil spectrum, which can be assigned to axial stretching vibrations of hydroxyl O–H bond in ricinoleic acid.24 Bands around 1200 cm⁻¹ may be assigned to the antisymmetric axial stretching vibrations of CC(=O)–O, bonds of the ester group, while those around 1183 cm⁻¹ may be assigned to asymmetric axial stretching of O–C–O bonds. Carbonyl absorption of saturated aliphatic esters usually appears from 1750 to 1735 cm⁻¹, while that for α, β-unsaturated esters from 1730 to 1715 cm⁻¹. In monomers and dimers of carboxylic acids carbonyl absorptions appear on 1760 cm⁻¹ and from 1720 to 1706 cm⁻¹, respectively.24

Carboxylic acids show in-plane bending of C–O–H bond in 1408 cm⁻¹ and axial deformation for dimer C–O bond in 1280 cm⁻¹. Carboxylic acid dimer shows an intense and α

β-unsaturated esters from 1750 to 1735 cm⁻¹, while that for α, β-unsaturated esters from 1730 to 1715 cm⁻¹. In monomers and dimers of carboxylic acids carbonyl absorptions appear on 1760 cm⁻¹ and from 1720 to 1706 cm⁻¹, respectively.24

Carboxylic acids show in-plane bending of C–O–H bond in 1408 cm⁻¹ and axial deformation for dimer C–O bond in 1280 cm⁻¹. Carboxylic acid dimer shows an intense and widespread absorption in the region 3300 to 2500 cm⁻¹, usually centered in 3000 cm⁻¹. This absorption may be due to the hydroxyl of ricinoleic acid from castor oil, fatty acids, glycerin and mono-and diglycerides. In Figure 1, the
Determination of Biodiesel Adulteration with Raw Vegetable Oil

hydroxyl absorption is observed only in castor oil biodiesel (spectrum b).

The overlapped bands in the fingerprint region (1300 to 900 cm⁻¹) indicate that univariate calibration models may cause significant prediction error in the quantification of biodiesel samples with different concentration when raw oil is present. Those models are also inadequate for identifying the presence of raw oil in a spoiled blend either because of incomplete conversion during esterification reaction or the illegal addition of raw oil. Zagonel et al.¹⁰ also observed overlapped bands in the MIR spectra of soybean oil and its corresponding ester. These authors used multivariate calibration of the bands in the region 1800 to 1700 cm⁻¹, corresponding to axial stretching vibrations of carbonyl groups to distinguish soybean from its ester.

Classification of biodiesel groups by PCA

PCA was used in an attempt to evaluate if biodiesel from different sources (cotton, castor or palm oil) exhibited distinguishing features that could make the identification of these sources easier, even though they were spoiled by raw soybean oil.

In an attempt to obtain models with a more efficient segregation of groups the whole spectrum as well as some of its regions was considered. According to literature, the best region was assigned to the axial deformation of carbonyl group (1800-1700 cm⁻¹). Zagonel et al.¹⁰ have shown that there is a displacement of carbonyl band of biodiesel and the raw oil when a first derivative spectrum is obtained in this region. Thus, from literature information and visual analysis PCA models were built up considering the full spectrum (4000-665 cm⁻¹) and two spectral ranges (1800-1700 and 1800-1000 cm⁻¹), encompassing both the carbonyl as CC(=O)-O vibrations. These spectral ranges were labeled as follows: model 1, 2 and 3, respectively.
The variance explained by the first ten PCs for each model is shown in Figure 2. For model 2 and 3, the first three principal components captured around 98% of the total data while for model 1 the response is 99%. The results suggest four independent variation sources, indicating that each biodiesel source contributes differently to IR data.

Figure 2. Variance captured for the first principal components.

Figure 3 shows a three-dimensional plot of the first three principal components for each set of data studied. It is possible to distinguish, in each model, three different groups of samples (cotton, castor and palm biodiesel). This demonstrates unequivocally that segregation between samples was very efficient and confirms that the IR data really contains enough information to aggregate the samples according to its biodiesel source. However, a general trend to dispersion among samples belonging to the same group can be noticed. Figure 3a shows the smallest dispersion for a given group, and the largest distance between each group. The best results for classification and identification of the biodiesel source were obtained from the full spectrum (4000-665 cm⁻¹). The region 1800-1700 cm⁻¹, corresponding to carbonyl vibrations, presented a greater dispersion than the full spectrum, which also proved to be a good region to work with multivariate analysis. According to Figure 3c, the data from spectral range 1800-1000 cm⁻¹ shows to be inefficient for source separation when the three spectral ranges were compared. It demonstrated a close proximity between samples from castor and cotton biodiesel.

**PLS models**

In order to quantify the amount of raw soybean oil added to the different biodiesel sources, multivariate calibration models were built, using the same MIR spectra employed in PCA analysis by PLS. Thus, one model was built: for each spectral range studied: model 1 (full spectrum, 4000-665 cm⁻¹), model 2 (1800-1700 cm⁻¹) and model 3 (1800-1000 cm⁻¹).

Table 1 lists the root mean square error of calibration for cross validation (RMSECV) of the model, correlation coefficient (R), latent variable (LV) and root mean square error of prediction (RMSEP) of the PLS model, considering all set of samples (cotton, castor and palm biodiesel) and different spectral ranges studied. Because of an algorithm limitation in the software related to the numbers of variables that can be used, the RMSEP values were only calculated for model 2 and 3. The maximum numbers of variable is 1000 for a set of calibration and validation, falling short from the 1488 variables needed for the calibration and prediction in model 1, precluding the external validation and calculation of RMSEP for this spectral range.

According to Table 2, considering both the spectral range as the biodiesel source, a slight variation from 0.972 to 0.999 was observed for the R values. A similar behavior was observed for the LV values, which fluctuated between 5 and 7. While the R and LV values did not varied noticeably between the models, the spectral range, 1800-1000 cm⁻¹, gave the smallest values of RMSECV for all of the three sources of biodiesel.

A better picture of the calibration results can be seen in Figure 4, resulting from the models developed for the three sources of biodiesel as a function of the predicted values for model 3 (1800-1000 cm⁻¹). This spectral range was chosen because of the smallest values of RMSECV and RMSEP presented. The plot shows only a very small dispersion which is comparable to all three sources. Cross validated results were quite close to the calculated value, if not coincident. These findings strongly support region 1800-1000 cm⁻¹ as the most reliable to detect adulteration of biodiesel by non esterified oil.

One way to observe qualitatively the linearity of a model is through the chart of residuals versus concentration of the samples. Residuals should be randomly distributed along
Table 2. PLS calibration results for biodiesel samples mixed with raw soybean oil

<table>
<thead>
<tr>
<th>Model</th>
<th>Range</th>
<th>Biodiesel</th>
<th>R</th>
<th>RMSECV (% m/m)</th>
<th>RMSEP (% m/m)</th>
<th>Variables for each spectral range</th>
<th>LV</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4000-665 cm(^{-1})</td>
<td>cotton</td>
<td>0.994</td>
<td>4.34</td>
<td>1.41</td>
<td>744</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>castor</td>
<td>0.995</td>
<td>2.85</td>
<td></td>
<td></td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>palm</td>
<td>0.990</td>
<td>2.85</td>
<td></td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1800-1700 cm(^{-1})</td>
<td>cotton</td>
<td>0.982</td>
<td>2.51</td>
<td>2.92</td>
<td>52</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>castor</td>
<td>0.972</td>
<td>2.23</td>
<td>2.36</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>palm</td>
<td>0.973</td>
<td>2.74</td>
<td>3.14</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1800-1000 cm(^{-1})</td>
<td>cotton</td>
<td>0.996</td>
<td>1.22</td>
<td>1.47</td>
<td>414</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>castor</td>
<td>0.999</td>
<td>1.02</td>
<td>1.10</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>palm</td>
<td>0.997</td>
<td>1.23</td>
<td>1.39</td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

*Number of columns exceeded the software processing capacity.

The advantage of this methodology in comparison to the previous one\(^ \text{15}\) is the speed and simplicity for building the models. The spectral range, 1800-1000 cm\(^{-1}\), generated prediction errors comparable to those obtained in the previous work\(^ \text{4}\) using the selection of variables with best results approach, with RMSEP varying from 1.10 to 1.47% (m/m) and from 0.65 to 1.40% (m/m) respectively.

Conclusion

PCA has shown that MIR spectra data contain information to differentiate samples of biodiesel, according to their source, even if there is some amount of raw soybean oil present. This chemometric tool showed to be suitable to classify blends of biodiesel/raw soybean distinguishing the groups from different sources very
well. The best result was obtained when the full MIR infrared spectrum was used.

PLS model based on MIR spectra developed in this work proved to be suitable as a practical analytical method to predict raw soybean content in biodiesel blends from 1 to 40% m/m. The spectral range 1800-1000 cm^{-1}, showed to the best region to develop a PLS calibration model for quantification of raw oil in biodiesel samples giving the lowest values for RMSECV and RMSEP. The advantage of using this spectral range to build models is that these models can be applied to biodiesel of different sources. In contrast, the variable selection method must be optimized for each source of biodiesel.

The advantage of this methodology is that it is very fast in determining the origin of the biodiesel, and whether or not it is adulterated, if so the level of the adulterant.

Acknowledgements

The authors would like to thank Agência Nacional de Petróleo, Gás Natural e Biocombustíveis (ANP) and Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) for the financial support and to LEC (Laboratório de Ensaios de Combustíveis, UFMG) and its staff for the technical support.

References

1. Pimentel, M. F.; Ribeiro, G. M. G. S.; Cruz, R. S.; Stragerwich, L.; Pacheco Filho, J. G. A.; Teixeira, L. S. G.; Microchem. J. 2006, 82, 201.

Submitted: September 10, 2010
Published online: February 24, 2011