Talanta 87 (2011) 30–34

Contents lists available at SciVerse ScienceDirect

Talanta

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

Determination of biodiesel content in biodiesel/diesel blends using NIR and visible spectroscopy with variable selection

David Douglas Sousa Fernandes[a], Adriano A. Gomes[b], Gean Bezerra da Costa[c], Gildo William B. da Silva[a], Germano Véras[a,b,*]

[a] Programa de Pós-Graduação em Ciências Agrárias, Universidade Estadual da Paraíba, 58.429-500 Campina Grande – PB, Brazil
[b] Departamento de Química, CCEN, Universidade Federal da Paraíba, 58.051-970, João Pessoa – PB, Brazil
[c] Departamento de Química, Centro de Ciências e Tecnologia, Universidade Estadual da Paraíba, 58.429-500 Campina Grande – PB, Brazil

ARTICLE INFO

Article history:
Received 22 June 2011
Received in revised form
14 September 2011
Accepted 14 September 2011
Available online 19 September 2011

Keywords:
Biodiesel
Diesel
Near infrared spectroscopy
Visible spectroscopy
Variable selection

ABSTRACT

This work is concerned of evaluate the use of visible and near-infrared (NIR) range, separately and combined, to determine the biodiesel content in biodiesel/diesel blends using Multiple Linear Regression (MLR) and variable selection by Successive Projections Algorithm (SPA). Full spectrum models employing Partial Least Squares (PLS) and variables selection by Stepwise (SW) regression coupled with Multiple Linear Regression (MLR) and PLS models also with variable selection by Jack-Knife (Jk) were compared the proposed methodology. Several preprocessing were evaluated, being chosen derivative Savitzky-Golay with second-order polynomial and 17-point window for NIR and visible-NIR range, with offset correction. A total of 100 blends with biodiesel content between 5 and 50% (v/v) prepared starting from ten sample of biodiesel. In the NIR and visible region the best model was the SPA-MLR using only two and eight wavelengths with RMSEP of 0.6439% (v/v) and 0.5741 respectively, while in the visible-NIR region the best model was the SW-MLR using five wavelengths and RMSEP of 0.9533% (v/v). Results indicate that both spectral ranges evaluated showed potential for developing a rapid and nondestructive method to quantify biodiesel in blends with mineral diesel. Finally, one can still mention that the improvement in terms of prediction error obtained with the procedure for variables selection was significant.

1. Introduction

Biodiesel has proven to be an increasingly viable proposal as a substitute for petrodiesel due to features such as high flash point, excellent lubricity and high cetane number, in addition to economic and environmental advantages, by coming from a renewable matrix, being biodegradable and helping reduce the emission of gases causing the greenhouse effect [1–15].

Biodiesel insertion in the Brazilian energy matrix is occurring in a gradual and progressive way driven by the National Program for Production and Use of Biodiesel (Law No. 11097 of January 13, 2005). Currently, five percent biodiesel is added to diesel, blend called B5, which represents a national daily production of about 17,000 m^3/day according to data from the National Agency of Petroleum, Natural Gas and Biofuels (ANP) [16], Brazilian federal government agency responsible for regulating the fuel sector.

As important as the mastery of technologies for synthesis and storage of biodiesel is having rapid analytical methodologies, nondestructive and low cost for use in quality control of biofuels. In Brazil, the ANP is responsible for setting standards for certifying the biodiesel quality, aiming to establish permissible limits of contaminants, which cause no harm to combustion process performance in engines and content of toxic gases emitted [17]. These parameters are mostly based on European (EN) and North American Standards, developed by the American Society of Testing and Materials (ASTM), which make use of time-consuming techniques that require and high consumption of reagents, such as Inductively Couple Plasma Optical Emission Spectrometry (ICP-OES), High Temperature Gas Chromatography (HTGC) and High-Performance Liquid Chromatography (HPLC) [18].

In contrast, the molecular absorption spectroscopy in the ultraviolet–visible and near-infrared regions are quick, nondestructive and highly accurate techniques and require little or no pretreatment of samples [6–11,19]. Together with the analytical techniques mentioned above is necessary the use of multivariate analysis tools in view of the large volume of data generated by these methods. In the literature various methods of multivariate regression are reported, such as Principal Component Regression (PCR), Partial Least Squares (PLS) and Multiple Linear Regression (MLR) [20]. When applied to high-dimensionality data set the existence of multicolinearity among variables harm the MLR models performance, which does not occur with the PCR and PLS models

* Corresponding author. Tel.: +55 83 3315 3356.
E-mail address: germano.veras@pq.cnpq.br (G. Véras).

0039-9140/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.talanta.2011.09.025

that promote an orthogonal transformation of data to reduce dimensionality, being the new variables mutually orthogonal [21].

However, when MLR is used in conjunction with variable selection algorithms such as Genetic Algorithm [22,23], Simulated Annealing [24], Stepwise [25], for instance, the model is considerably improved, in many cases with similar or better results to those obtained with PCR and PLS models [26,27]. PLS models based in variable selection also are described in the literature, for example. GA-PLS, interval PLS, PLS-Jk, among others [28–30].

MLR models based on variables selection have the advantage of be simple to interpret, by acting in the original domain data, although losing to the full-spectrum methods in sensitivity, by using a lower number of analytical channels. However, they are superior in selectivity by selecting variables with higher correlation with parameters to be determined and may also guide the construction of photometer dedicated based on Light Emitting Diodes (LED) [31].

In the context of variables selection, Araújo et al. [32] proposed the Successive Projections Algorithm (SPA) to circumvent the problem of multicolinearity in Multiple Linear Regression. Given the matrix of instrumental responses $X(_{\mathrm{mxn}})$, with m calibration of objects measured in n sensors. The SPA starts with the variable x_i (i = 1,2,3,. . .,k), projecting in the other subspace (z^i) orthogonal to x_i. Assuming that the selected variable has index i equal to j, the next step is to project the remaining variables in the subspace z^j, orthogonal to x_j. The process continues until a maximum number ($N_{\max}$) of variables has been included in the chain, for the case of MLR regression the maximum amount of variables is $m - 1$.

In the second phase, the SPA tests the correlation of the variables chains, generated in phase 1, with the dependent variable, constructing an MLR model for each chain (i = 1–$N_{\max}$). The parameter of interest is estimated to external validation set. The validation error (RMSEV) is used as a metric for choosing the best variables chain.

Several modifications were carried out in other works, such as the version with cross-validation [33], used in this work, where a third step to improve the algorithm parsimony was proposed, whose variables which does not contribute significantly to lower RMSEV are excluded [34]. A more recently modification to the SPA was proposed so that it could operate even in the presence of uncalibrated interferences [35]. In the literature there are many reports of successful SPA applications, whose sulfur determination in diesel [36], determination of phenols in sea water [37], simultaneous determination of metals in multivitamin [38] and quality of insulating oils [39] are some examples.

Since errors and possible adulteration may occur during the mixing process of biodiesel to diesel in the proportions to be added in distributors, it is important to have methodologies able to quantify the biodiesel content in diesel. Thus, MLR models were constructed to quantify the biodiesel content in biodiesel/diesel blends, using SPA as a tool for selecting variables and exploring the regions of visible and NIR, separately and together. For comparison, SW-MLR, PLS-Jk based in variable selection and PLS full-spectrum models were build.

2. Material and methods

2.1. Biodiesel samples

Biodiesel samples were obtained by transesterification reaction via methanol route using potassium hydroxide as catalyst. 8:1 alcohol:oil ratio and 1.5% catalyst percentage in relation to the oil mass were used.

The reaction mixture was subjected to magnetic stirring and heating to 45 °C for 15 min. Subsequently, there was a rest time followed by separation, drying and purification of biodiesel samples. Physical and chemical characterization of samples was conducted to assess their quality. Petrodiesel was provided by Petrobras Distribuidora, located in Cabedelo, state of Paraiba.

The 100 biodiesel/diesel blends were prepared using ten biodiesel samples by adding biodiesel to diesel at 5, 10.15, 20, 25, 30, 35, 40, 45 and 50 (%v/v).

2.2. Spectra acquisition

The spectra of biodiesel/diesel blends were obtained in triplicate at range from 441 to 1551 nm with 1 nm resolution, however average spectrum was used in modeling the data. The data matrix resulting is composed of 100 spectra recorded in 1110 wavelength. The measures was carried out using and Perkin Elmer spectrophotometer 750 Lambda model, equipped with quartz cell with 1 cm optical path, tungsten source and R928 photomultiplier tube and Peltier-cooled PbS detection systems.

2.3. Software and data analysis

The spectral ranges visible and NIR were evaluated separately and combined to check the synergistic effects. In visible range base line correction was evaluated by offset and linear. In NIR and visible-NIR ranges approaches to base line correction and noise removal were evaluated: base line correction by offset coupled with smooth Savitzky–Golay and derivate Savitzky–Golay. In all case second order polynomial was applying and windows of 11, 15, 17 and 21 point were used.

Preprocessing calculations, PLS and PLS-Jk models were performed in Unscrambler® v9.8 and calculations involving MLR models with variables selection by SPA and SW were carried out in MatLab environment, version R2011.a.

3. Results and discussion

Fig. 1a shows the spectra of 100 samples of biodiesel/diesel blends. The absorption band in the visible with maximum at about 530 nm is highlighted, corresponding to the methyl esters containing conjugated double bonds and non-transesterified triglycerides traces and the bands in the NIR region with peaks in 1400 and 1200 nm, corresponding to the first and second C–H, respectively [40].

In Fig. 1a can be seen that the spectra show deviation from the baseline and the NIR region is noisier. The preprocessing technique chose for visible-NIR (Fig. 1b) and only NIR (Fig. 1c) ranges was derivative Savitzky–Golay with second-order polynomial and 17-point window, while for visible (Fig. 1d) region offset correction showed best results.

3.1. Outlier detection

An important step in building a multivariate calibration model is the identification of possible anomalous samples since such samples can affect the final quality of models and should be removed beforehand. Fig. 2 shows the graph of leverage versus residual studentized concentration. The leverage and residual concentration were obtained using PLS regression with full cross-validation, leave-one-out, for the entire data set.

Horizontal lines represent the studentized residue limits to 95% confidence, while the vertical line represents the influence critical value (critical leverage: h_c), defined as 3*k/n, where k is the number of factors and n the number of samples [41,42]. Based on the graphic above no sample was considered anomalous in all spectral region evaluated, thereby remaining the data set with 100 samples.

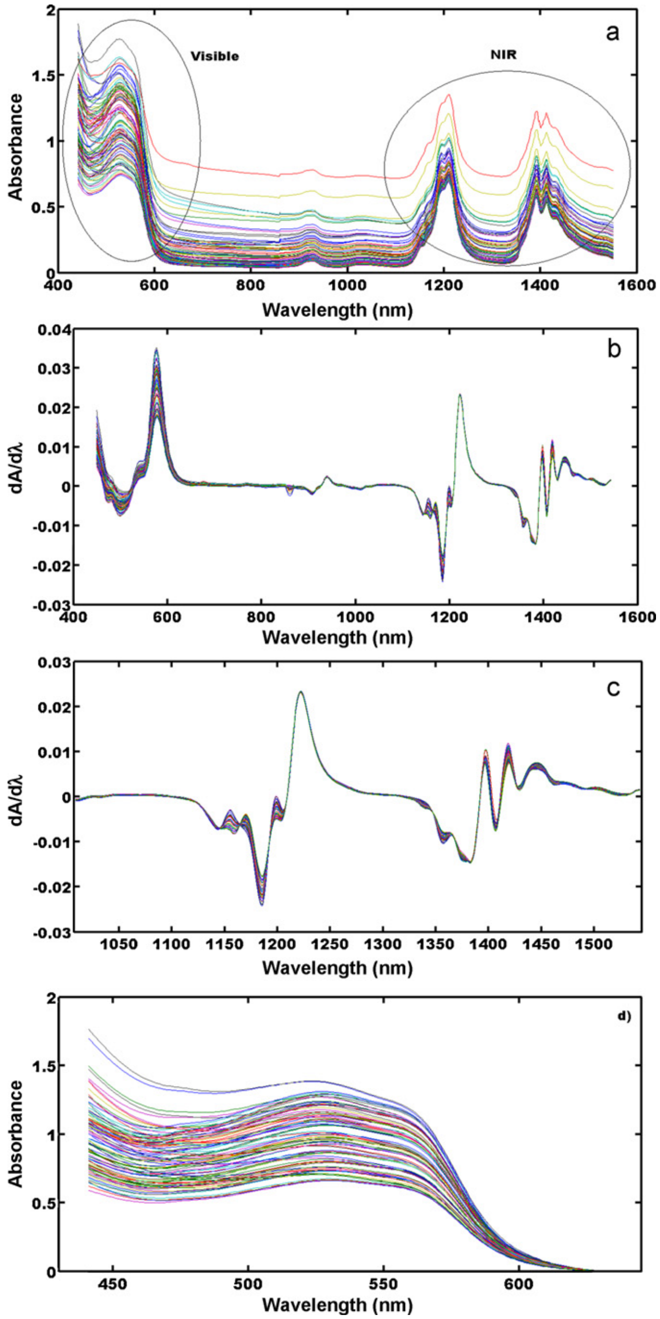

Fig. 1. Spectra of 100 blends biodiesel/diesel, (a) raw (b) visible-NIR range preprocessing (c) NIR preprocessing (d) visible preprocessing.

After evaluating the possible existence of anomalous samples, the set of 100 samples was partitioned into calibration (60 samples) and external validation (40 samples) using the Kernnard–Stone (KS) algorithm adapted by Galvão et al. [43], which unlike classic KS, takes into account the statistics of *X* and *Y* matrices.

3.2. Variables selection

Algorithms of variables selection used were SPA and SW Jk. Results are shown in Fig. 3, where wavelengths selected by each algorithm for each region studied were indexed to the average spectrum over all set of samples.

In visible-NIR region SPA and Jk select wavelengths in the visible and NIR. SW selects only the NIR region. The best results in terms of RMSEP for visible-NIR spectra were obtained to the SW (Table 1). Possibly the synergistic effect prejudice calibration models to this case.

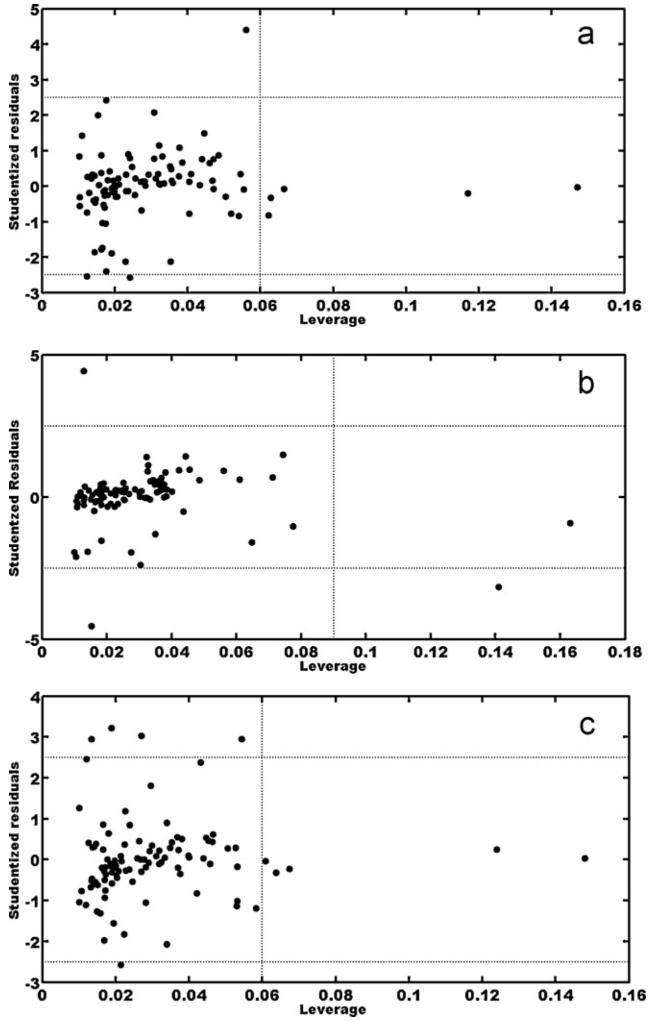

Fig. 2. Leverage versus studentized residuals: (a) visible-NIR range (b) NIR range (c) visible range.

For NIR region, SPA only selected two wavelengths corresponding to 1184 and 1507 nm, while the SW selected four wavelengths, all in the second overtone region, which is less intense than the first. We can also highlight SPA parsimony front the SW, selecting only two wavelengths. For the visible band stepwise was more parsimonious than the SPA, selecting half the number of variables in the SPA. For both the NIR and visible the Jack-Knife criterion selects a large number of variables, getting the prediction results similar to those obtained with full spectrum PLS.

3.3. Determination of biodiesel content

The determination of biodiesel content was performed by four regression models for each spectral range evaluated. All models were built employing full cross-validation, leave-one-out.

Variable selection by SW-MLR use mixed method with one step forward followed by a backward step starting from the most correlated variable with the parameter to be determined, the alpha value for inclusion or exclusion of a variable was kept constant and equal to 0.05.

Table 1 shows results for calibration and external validation sets. For all models evaluated were obtained RMSEP values of low and high explained variance and for visible-NIR range SW-MLR showed

Table 1
Parameters of the external validation set for models.

Range	Parameter RMSEC[a](% v/v)	RMSECV[a] (% v/v)	*R*-square[a]	RMSEP[b] (%v/v)	bias[b]	SVD[b]	*t*[b]
Vis-NIR							
PLS (1)[c]	1.3672	1.3979	0.9905	1.0195	0.6390	1.4836	2.6898
$\text{PLS}_{\text{Jack-Knife}}$ (1)[c]	1.4517	1.4898	0.9893	1.1034	0.5464	1.5071	2.2642
SPA-MLR (5)[d]	1.0403	1.3992	0.9945	0.9594	0.0948	0.9857	0.6006
WS-MLR (5)[d]	0.8786	0.9979	0.9961	0.9533	0.0218	0.9663	0.1409
NIR							
PLS (2)[c]	1.2287	1.2758	0.9922	0.6615	0.2720	0.8224	2.0920
$\text{PLS}_{\text{Jack-Knife}}$ (2)[c]	1.2496	1.2950	0.9919	0.6841	0.2760	0.8452	2.0394
SPA-MLR (2)[d]	1.1509	1.1915	0.9932	0.6439	0.0848	0.7400	0.6830
WS-MLR (4)[d]	0.9963	1.1374	0.9949	0.7225	0.0586	0.7400	0.4920
Visible							
PLS (2)[c]	1.7376	1.8088	0.9846	0.6939	0.0655	0.7120	0.5745
$\text{PLS}_{\text{Jack-Knife}}$ (2)[c]	1.7376	1.8088	0.9846	0.6938	0.0677	0.7127	0.5932
SPA-MLR (8)[d]	1.0516	1.2127	0.9944	0.5741	0.1605	0.6460	1.5516
WS-MLR (5)[d]	1.0021	1.1028	0.9949	0.6554	0.1373	0.7043	1.2175

$t_{\text{crit.}}$ = 2.021 for 95% of statistical significance.
[a] Calibration set.
[b] External validation set.
[c] Latent variables used in the model.
[d] Wavelength used in the model.

best results, RMSEP of 0.9533 (% v/v), employing four wavelength, all belonging to the NIR region. This result suggests that using the two spectral regions together did not improve the models when compared to results obtained for the two spectral bands separately.

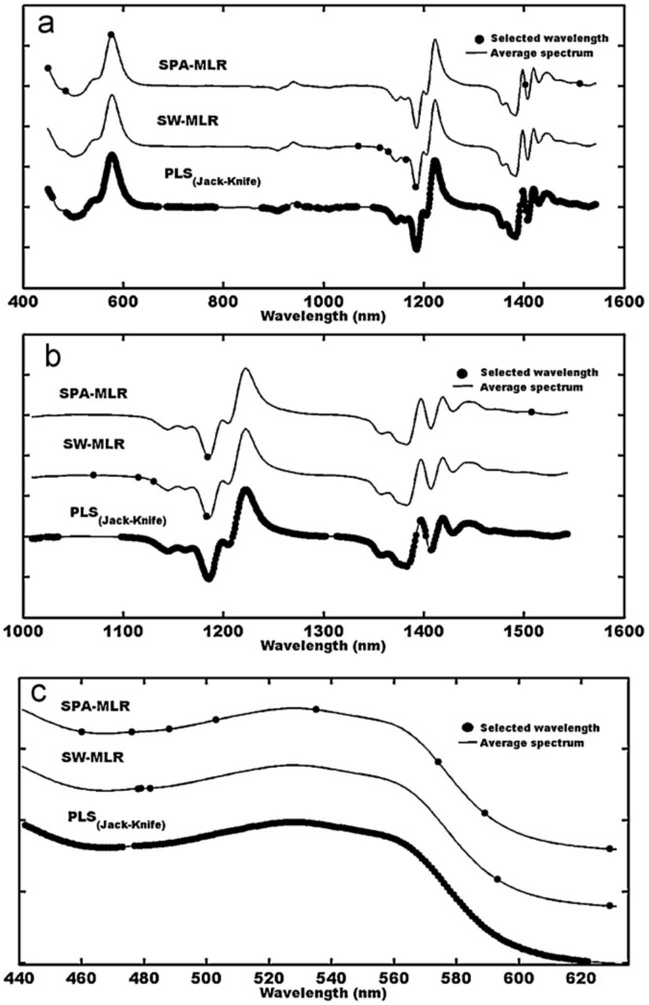

Fig. 3. Selected wavelength: (a) visible-NIR range (b) NIR range (c) visible range.

For the visible and NIR regions separately the best results were obtained using the SPA-MLR. RMSEP of 0.6439 (% v/v) was obtained for the NIR and to the visible 0.5741 (% v/v), using 2 and 8 wavelengths, respectively.

The systematic error (bias) and standard deviation validation (SVD) were calculated according with Equations [44] below:

$$\text{bias} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)}{n} \tag{1}$$

$$\text{SDV} = \sqrt{\frac{\sum_{i=1}^{n}[(y_i - \hat{y}_i) - \text{bias}]^2}{n-1}} \tag{2}$$

Subsequently, a *t* test at 95% confidence and $n - 1$ degrees of freedom was applied to evaluate the existence of significant systematic error for each model; the *t* value shown in Table 1 was calculated based on Eq. (3):

$$t = \frac{|\text{bias}|\sqrt{n}}{\text{SDV}} \tag{3}$$

The test of systematic errors at 95% statistical confidence indicated the existence of significant bias only for the PLS and PLS-Jk models in the NIR and visible-NIR bands.

4. Conclusions

In this work was present a rapid and non destructive method to determine the content of biodiesel in diesel employing visible-NIR, NIR and visible spectra and SPA-MLR. With the results obtained can be shown that it is possible to determine the biodiesel content in blends with petrodiesel employing both the region. It is also possible to observe that the procedure of variables selection improves the predictive power when compared to PLS models, which allows the construction of dedicated photometers, given the low number of selected wavelengths.

Acknowledgments

This work was supported by CNPq (Process n. 576416/2008-8).

References

[1] M.F. Ferrão, M.S. Viera, R.E.P. Pazos, D. Fachini, A.E. Gerbase, L. Marder, Fuel 90 (2011) 701–706.
[2] R.M. Balabin, R.Z. Safieva, Anal. Chim. Acta 689 (2011) 190–197.

[3] R.M. Balabin, E.I. Lomakina, R.Z. Safieva, Fuel 90 (2011) 2007–2015.
[4] M. Meira, C.M. Quintella, A.S. Tanajura, H.R.G. Silva, J.D.S. Fernando, P.R. Costa Neto, I.M. Pepe, M.A. Santos, L.L. Nascimento, Talanta 85 (2011) 430–434.
[5] F.D. Andrade, A.M. Netto, L.A. Colnago, Talanta 84 (2011) 84–88.
[6] R.M. Balabin, E.I. Lomakina, Analyst 136 (2011) 1703–1712.
[7] K.M. Pierce, S.P. Schale, Talanta 83 (2011) 1254–1259.
[8] R.M. Balabin, R.Z. Safieva, E.I. Lomakina, Anal. Chim. Acta 671 (2010) 27–35.
[9] R.M. Balabin, R.Z. Safieva, Fuel 87 (2008) 1096–1101.
[10] M.A. Cantarelli, I.G. Funes, E.J. Marchevsky, J.M. Camina, Talanta 80 (2009) 489–492.
[11] P. Baptista, P. Felizardo, J.C. Menezes, M.J.N. Correia, Talanta 77 (2008) 144–151.
[12] R.M. Balabin, R.Z. Safieva, E.I. Lomakina, Microchem. J. 98 (2011) 121–128.
[13] R.Z. Syunyaev, R.M. Balabin, I.S. Akhatov, J.O. Safieva, Energy Fuels 23 (2009) 1230–1236.
[14] R.M. Balabin, R.Z. Safieva, E.I. Lomakina, Chemom. Intell. Lab. Syst. 93 (2008) 58–62.
[15] R.M. Balabin, R.Z. Safieva, J. Near Infrared Spectrosc. 15 (2007) 343–349.
[16] Agência Nacional do Petróleo, Gás Natural e Biocombustíveis – ANP, BOLETIM MENSAL DE BIODIESEL Março de 2011.
[17] I.P. Lôbo, S.L.C. Ferreira, Quim. Nova 32 (2009) 1596–1608.
[18] M.R. Monteiro, A.R.P. Ambrozin, L.M. Lião, A.G. Ferreira, Talanta 77 (2008) 593–605.
[19] G. Veras, A.A. Gomes, A.C. Silva, A.L.B. Brito, P.B.A. Almeida, E.P. Medeiros, Talanta 83 (2010) 565–568.
[20] K.R. Beebe, R.J. Pell, B. Seasholtz, Chemometrics – A Pratical Guide, New York: Wiley, 1998.
[21] M. Andersson, J. Chemom. 23 (2009) 518–529.
[22] H.C. Goicoechea, A.C. Olivieri, J. Chemom. 17 (2003) 338–345.
[23] H.C. Goicoechea, A.C. Olivieri, J. Chem. Inf. Comput. Sci. 42 (2002) 1146–1153.
[24] U. Hörchner, J.H. Kalivas, J. Chemom. 9 (1995) 283–308.
[25] R.M. Balabin, S.V. Smirnovb, Anal. Chim. Acta 692 (2011) 63–72.
[26] A.F.C. Pereira, M.J.C. Pontes, F.F.G. Neto, S.R.B. Santos, R.K.H. Galvão, M.C.U. Araujo, Food Res. Int. 41 (2008) 341–348.
[27] L.F.B. Lira, M.S. Albuquerque, J.G.A. Pacheco, T.M. Fonseca, E.H.S. Cavalcanti, L. Stragevich, M.F. Pimentel, Microchem. J. 96 (2010) 126–131.
[28] R. Leardi, A.L. Gonzáles, Chemom. Intell. Lab. Syst. 41 (1998) 195–207.
[29] A. Hoskuldsson, Chemom. Intell. Lab. Syst. 55 (2001) 23–38.
[30] H. Martens, M. Martens, Food Qual. Pref. 11 (2000) 5–16.
[31] G. Veras, E.C. Silva, W.S. Lyra, S.F.C. Soares, T.B. Guerreiro, S.R.B. Santos, Talanta 77 (2009) 1155–1159.
[32] M.C.U. Araújo, T.C.B. Saldanha, R.K.H. Galvão, T. Yoneyama, H.C. Chame, V. Visani, Chemom. Intell. Lab. Syst. 57 (2001) 65–73.
[33] R.K.H. Galvão, M.C.U. Araújo, E.C. Silva, G.E. José, S.F.C. Soares, H.M. Paiva, J. Braz. Chem. Soc. 18 (2007) 1580–1584.
[34] R.K.H. Galvão, M.C.U. Araújo, W.D. Fragoso, E.C. Silva, G.E. José, S.F.C. Soares, H.M. Paiva, Chemom. Intell. Lab. Syst. 92 (2008) 83–91.
[35] S.F.C. Soares, R.K.H. Galvão, M.C.U. Araújo, E.C. Silva, C.F. Pereira, S.I.E. Andrade, F.C. Leite, Anal. Chim. Acta 689 (2011) 22–28.
[36] M.C. Breitkreitz, I.M. Raimundo Jr., J.J.R. Rohwedder, C. Pasquini, H.A.D. Filho, G.E. José, M.C.U. Araújo, Analyst 128 (2003) 1204–1207.
[37] M.S.D. Nezio, M.F. Pistonesi, W.D. Fragoso, M.J.C. Pontes, H.C. Goicoechea, M.C.U. Araújo, B.S.F. Band, Microchem. J. 85 (2007) 194–200.
[38] H.A.D. Filho, Ê.S.O.N. Souza, V. Visani, S.R.R.C. Barrosa, T.C.B. Saldanha, M.C.U. Araújo, R.K.H. Galvão, J. Braz. Chem. Soc. 16 (2005) 58–61.
[39] M.J.C. Pontes, A.M.J. Rocha, M.F. Pimentel, C.F. Pereira, Microchem. J. 98 (2011) 254–259.
[40] Z. Xiaobo, Z. Jiewen, M.J.W. Povey, M. Holmes, M. Hanpin, Anal. Chim. Acta 667 (2010) 14–32.
[41] A.M.K. Pedro, M.M.C. Ferreira, Anal. Chem. 77 (2005) 2505–2511.
[42] M.M.C. Ferreira, A.M. Antunes, M.S. Melgo, P.L.O. Volpe, Quim. Nova 22 (1999) 724–731.
[43] 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) 736–740.
[44] Annual Book of ASTM Standards, Standards practices for infrared, multivariate, quantitative analysis, E1655, vol. 03.06, ASTM International, West Conshohocken, Pennsylvania, USA, 2000.