Fuel 115 (2014) 46–53

Contents lists available at SciVerse ScienceDirect

Fuel

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

# In-line monitoring of the transesterification reactions for biodiesel production using NIR spectroscopy

Suzana Moreira de Lima [a,b], Bárbara Furtado Alves Silva [a], Dácio Vieira Pontes [a], Claudete Fernandes Pereira [c], Luiz Stragevitch [a], Maria Fernanda Pimentel [a,*]

[a] *Departamento de Engenharia Química, Laboratório de Combustíveis (LAC), Universidade Federal de Pernambuco – UFPE, Avenida Prof. Arthur de Sá s/n, Cidade Universitária, 50740-521 Recife, PE, Brazil*

[b] *Departamento Acadêmico de Controle e Sistemas Industriais (DACI), Instituto Federal de Educação, Ciência e Tecnologia de Pernambuco – IFPE, Avenida Prof. Luiz Freire 500, Cidade Universitária, 50740-540 Recife, PE, Brazil*

[c] *Departamento de Química, Laboratório de Análise Química, Inorgânica e Sensores (LAQIS), Universidade Federal Rural de Pernambuco – UFRPE, Rua Dom Manuel de Medeiros s/n, Dois Irmãos, 52071-030 Recife, PE, Brazil*

## HIGHLIGHTS

- In-line NIR monitoring of the transesterification reactions for biodiesel production.
- Methyl ester, mono-, di- and triglycerides contents predicted by PLS and MLR models.
- Multivariate control chart for monitoring the progress of biodiesel production.

## ARTICLE INFO

*Article history:*
Received 8 April 2013
Received in revised form 27 June 2013
Accepted 28 June 2013
Available online 12 July 2013

*Keywords:*
In-line monitoring
NIR spectroscopy
Biodiesel
Transesterification reaction
Control charts

## ABSTRACT

In this work, methods for in-line monitoring of the transesterification reactions of soybean oil with methanol using NIR spectroscopy are developed. Gas chromatography was employed for determining the contents of methyl ester (ME), monoglycerides (MG), diglycerides (DG) and triglycerides (DG) during the transesterification reaction and used as reference to build partial least squares regression (PLS) and multiple linear regression (MLR) calibration models employing NIR spectra. Some pre-processing and variable selection strategies of the spectral data were evaluated. Satisfactory RMSEP values were obtained for the external prediction subset. In addition, a multivariate control chart based on latent variables to monitor the progress of the reactions is also presented.

## 1. Introduction

Biodiesel is a fuel consisting of long chain alkyl esters, generally produced by the transesterification of triglycerides with short chain alcohols (usually methanol or ethanol) in the presence of a catalyst (KOH, NaOH, $H_2SO_4$, etc.). This transesterification is a sequence of reversible reactions, which contain triglycerides, diglycerides and monoglycerides (as intermediates), and methyl or ethyl esters and glycerol (a by-product of reaction) as components.The reaction is usually carried out in batch reactors and various factors affect the process such as: variability of feedstocks, free fatty acid (FFA) content, type of alcohol, type and speed of stirring, concentration and type of catalysts and temperature. To ensure the quality of the biodiesel produced, fuel international standards (e.g. European Standard – EN 14214) establish limits for glycerol, mono-, di- and triglycerides, alcohol, moisture and free fatty acid. These residues may be present due to partial conversion or insufficient purification of the reaction mixture and can lead to engine deposits and fuel deterioration during storage [1,2].

The conventional approach to assess biodiesel quality is by analyzing samples, collected at the end of each batch or after the purification step, using off-line and lengthy analytical methods. Then, based on the obtained results, a decision about the acceptance, rejection or reprocessing the batch is taken. This strategy results in high production costs due to possible process-stops, time consuming analyses and batch reprocessing. In the literature, several analytical methods for off-line monitoring of biodiesel production and the quality of the final product have been reported, such as the use of chromatography [3–6], Raman [7], laser [8], and infrared [9] and near infrared (NIR) based methods [10–16].

---

* Corresponding author. Tel./fax: +55 81 2126 7235.
*E-mail addresses:* mfp@ufpe.br, mfernanda.pimentel@gmail.com (M.F. Pimentel).

0016-2361/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.fuel.2013.06.057

Analytical methods which allow continuous real time quality assurance, enable timely corrective actions as well as process understanding have become a necessity, as stated by the Process Analytical Technology (PAT) framework [17]. In general, spectroscopic techniques have been chosen for this goal, because they are fast, nondestructive and can be adapted for in- or on-line monitoring of the transesterification of vegetable oils. Nuclear Magnetic Resonance (NMR) methods have been reported, which are robust, rapid and allow the quantification of different chemical species (triacylglycerol, diacylglycerol, fatty acid methy esters) [18–20]. On the other hand, Infrared (IR) and Near Infrared (NIR) spectroscopies associated with multivariate calibration are being increasingly used as described below.

On-line monitoring of transesterification of soybean oil with ethanol using FTIR spectroscopy (3707–814 cm$^{-1}$) has been described by Trevisan et al., in 2008 [21]. Ethyl esters contents were obtained by $^1$H NMR spectroscopy. Models based on PLS were developed using several preprocessing strategies and evolving factor analysis (EFA) for the selection of the spectral variables. The use of EFA produced relative errors under 3%.

In 2011, Killner et al. [22] developed a method based on a Partial Least Squares (PLS) regression and NIR spectroscopy to monitor the progress of the transesterification reactions of soybean oil with methanol. A spectroscopic flow cell was used for on-line monitoring the transesterification reaction.$^1$H NMR was employed for estimation of the conversion of glycerides to methyl esters during the transesterification reaction, and used as reference to build a PLS calibration model employing data from NIR spectroscopy. Different NIR spectral ranges were tested to build calibration models to determine the conversion to methyl esters. The best calibration model was built employing the spectral range between 5928 and 5959 cm$^{-1}$ (RMSEC = 1.09% and RMSECV = 1.32% of conversion). This calibration model was used to predict the convertion of reactions carried out with temperatures at 20 °C and 55 °C and the RMSEP values obtained were 0.74% and 1.27% of conversion, respectively.

Richard et al. [23] in 2011 developed a fast analytical method to monitor on-line the transesterification reaction between triglycerides and ethanol using near infrared spectroscopy. Gas chromatography was employed as the reference method. The transesterification reactions were carried out in a one liter batch reactor, monitored with a NIR probe. PLS regression models using different data pre-processing methods were developed to determine the ethyl ester and monoglycerides contents using the spectral range from 4500 to 9000 cm$^{-1}$. The RMSEP obtained were 1.74% and 1.49% for ethyl ester and monoglycerides contents, respectively. In 2013, Richard et al. [24] also described the use of a continous microstrutured device to carry out the ethanolysis of high oleic sunflower oil. The device they developed allows on-line monitoring (using a NIR probe) of the transesterification reaction with ethanol. Partial Least Square regression models were built between NIR spectra data (spectral regions: 8649–8085 cm$^{-1}$, 7097–5508 cm$^{-1}$ and 5050–4476 cm$^{-1}$) and reference analytical data (gas chromatography with flame ionization detection, CG-FID). The ethyl ester contents were predicted with RMSEP values of 4.10% and 3.52% using reflection and transflectance probes, respectively.

In the present work, NIR methods for in-line monitoring of the transesterification reaction with methanol are described. For this purpose, PLS and MLR models using different strategies of preprocessing of spectra data were developed to determine methyl ester (ME), monoglycerides (MG), diglycerides (DG) and triglycerides (TG) content. In addition, multivariate control charts based on continuous NIR spectra recorded during the reactions were employed as a multivariate statistical process control tool.

## 2. Material and methods

### *2.1. Samples and monitoring of the transesterification reactions*

Commercial soybean oils (Liza, acidity 0.29 mg KOH/g, water content 691.6 ppm, density at 20 °C 0.91882 g/cm$^3$, kinematic viscosity 31.60 mm$^2$/s) were bought in local markets and used in all transesterification reactions performed in this work. Analytical grade methanol (Merck), sodium hydroxide (Merck), glacial acetic acid (Dinâmica) and anhydrous sodium sulfate (Dinâmica) were also employed. The transesterification reaction was carried out in a 500 mL batch reactor with temperature controlled by a thermostatized water bath. This reactor was equipped with a thermocouple and a mechanical stirrer. In addition, an NIR fiber-optic transflectance probe was immersed directly in the reaction mixture for in-line monitoring of the transesterification reaction.

A 6:1 methanol to oil molar ratio was used with different amounts of catalyst (NaOH 0.5; 0.75 and 1 w/w% compared to oil mass). Nineteen batches with temperatures of 20, 45 and 55 °C were produced. These different conditions, which were chosen based on Killner et al. [22], allowed variability in the process, resulting in samples with different conversions. Table 1 shows the experimental conditions employed; the number of batches produced under each experimental condition; the time range in which the samples were taken for further chomatographic analysis; and the number of samples collected.

For each batch, 400 g of soybean oil were transferred to the reactor and the mechanical stirring was turned on until the desired temperature was reached. A mixture of methanol and sodium hydroxide was added to the oil. At the end of each batch, the reaction medium obtained was allowed to stand in order to separate the organic phase from glycerol. To evaluate the methyl ester and glycerides contents, samples (20 mL) of the reaction mixture were collected at different reaction times and the reaction was immediately stopped by adding 2.0 mL of glacial acetic acid. Sequentially, these samples were washed with distilled water and centrifuged three times. At room temperature (20 ± 1 °C),

**Table 1**
Experimental conditions for each transesterification batch and the results obtained by the reference methods.

| T (°C) | Catalyst, NaOH (wt%) | Time range (min)[a] | Number of batches | Number of samples | Range of concentration values (wt%) | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | ME | MG[b] | DG[b] | TG[b] |
| 20 | 1.0 | 2–200 | 3 | 21 | 34.1–95.3 | 0.26–2.65 | 0.13–9.67 | 0.34–12.41 |
| 20 | 0.75 | 2–30 | 1 | 7 | 16.8–83.4 | 1.15–3.03 | 3.03–9.94 | 7.59–17.59 |
| 20 | 0.5 | 2–30 | 1 | 7 | 5.1–64.1 | 0.33–2.00 | 5.50–9.27 | 18.94 |
| 45 | 0.75 | 2–90 | 4 | 12 | 66.1–95.3 | 0.54–3.55 | 0.34–6.51 | 0.68–15.75 |
| 55 | 1.0 | 2–80 | 4 | 28 | 83.9–96.0 | 0.22–2.08 | 0.05–2.90 | 0.05–4.88 |
| 55 | 0.75 | 30–90 | 6 | 7 | 95.1–97.1 | – | – | – |

[a] Time range used for sample acquisition for further chromatographic analysis.

[b] Only the results are shown which were within the adapted reference method concentration range.

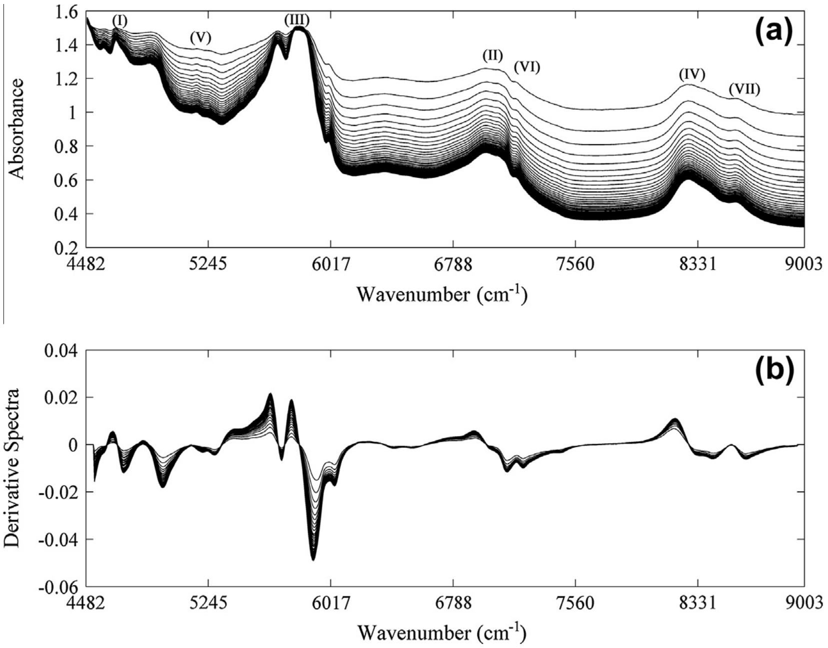

**Fig. 1.** In-line monitoring of transesterification reaction of biodiesel: (a) raw (b) derivative spectra.

anhydrous sodium sulfate was added to the biodiesel and the mixture was then filtered with qualitative filter paper (80 g Unifil, diameter 12.5 cm, ash content max 0.1%). These samples were analyzed by off-line gas chromatography using a Shimadzu 17-A instrument with automatic injection. The methyl ester content was quantified according to method EN14103. An adapted ASTM D6584-10a method was used to quantify the MG, DG and TG contents. The concentration range was extended because some samples had presented high amounts of these compounds at the beginning of the reactions. The last four columns of Table 1 show the concentration range of ME, MG, DG and TG (using the adapted reference methods) obtained for the samples collected under each experimental condition. It is important to notice that only the results which were within the concentration range allowed by the adapted methods are shown. For example, in the batch produced at 20 °C and 0.5 wt% of catalyst, all the samples (except the one, collected at 30 min) showed TG concentration values outside the linear dynamic range of the adapted reference method.

Eight batches “under control” conditions (500 rpm, 55 °C, 0.75 w/w% of catalyst) were employed to build a multivariate control chart based on latent variables. One additional “under control batch” and one batch “out of control” (the temperature control failed for a few minutes) were used to test the generated multivariate control chart.

### 2.2. NIR spectra acquisition

In-line NIR spectra (13,988–3799 $cm^{-1}$) of the transesterification reaction mixture were acquired using a fiber-optic transflectance probe (Solvias) with a 5.0 mm optical path length (resolution of 16 $cm^{-1}$), connected to the FTLA 2000-160 FTIR spectrometer (ABB Bomem). One spectrum every 20 s (as an average of 64 scans) was obtained during the reaction using the equipment kinetic mode. The background spectra were obtained using an empty transflectance probe. For multivariate regression models of the ME, MG, DG and TG contents, spectral data were registered at the moment of the samples collection from the reaction mixture for chromatograph analysis.

### 2.3. Chemometric procedure and software

All spectral data were mean-centered before modeling procedures. Some pre-processing strategies of the spectral data were evaluated for the determination of ME content, such as: baseline correction plus smoothing, first derivative with Savitzky–Golay smoothing (second order polynomial using different window points), Multiplicative Scatter Correction (MSC and EMSC) and Standard Normal Variate (SNV). The best strategies obtained for the determination of methyl ester content were also tested for MG, DG and TG determination.

The sample set was divided into calibration and prediction subsets using the SPXY (sample set partitioning based on joint *x*–*y* distances) algorithm [25]. The SPXY is a Kennard-Stone algorithm modification. In the SPXY algorithm the sample distances are based on the independent variable ($\boldsymbol{X}$) and on the dependent variable ($\boldsymbol{y}$) space for the parameter under consideration. These distances are divided by their maximum values in the data set, in order to assign equal importance to the distribution of the samples in the $\boldsymbol{X}$ and $\boldsymbol{y}$-spaces. Detection and elimination of outliers were carried out using score, residual ($\boldsymbol{X}$ and $\boldsymbol{y}$) and leverage plots. All spectral data were mean-centered before modeling procedures.

Four different strategies of multivariate calibration were evaluated: (1) PLS models using all spectral data (PLS full spectrum); (2) PLS models performed by using the variables which correspond to the significant regression coefficients selected by the Jack-Knife algorithm (PLS/JK); (3) *i*PLS algorithm applied by dividing the derivative spectra into 10 non-overlapping interval of equal size, allowing the construction of PLS models for each interval and the best range (or combination of ranges) selected on the basis of the lowest value of the root-mean square error of cross validation (RMSECV); (4) multiple linear regression using the successive projections algorithm for variable selection (MLR/SPA) [26].

Full cross-validation (CV) was carried out to select the number of PLS factors or variables for MLR models. The predictive ability of the models was evaluated according to the root-mean-square error of prediction (RMSEP) obtained for the external validation set. An *F*-test at a confidence level of 95% was used to assess the statistical

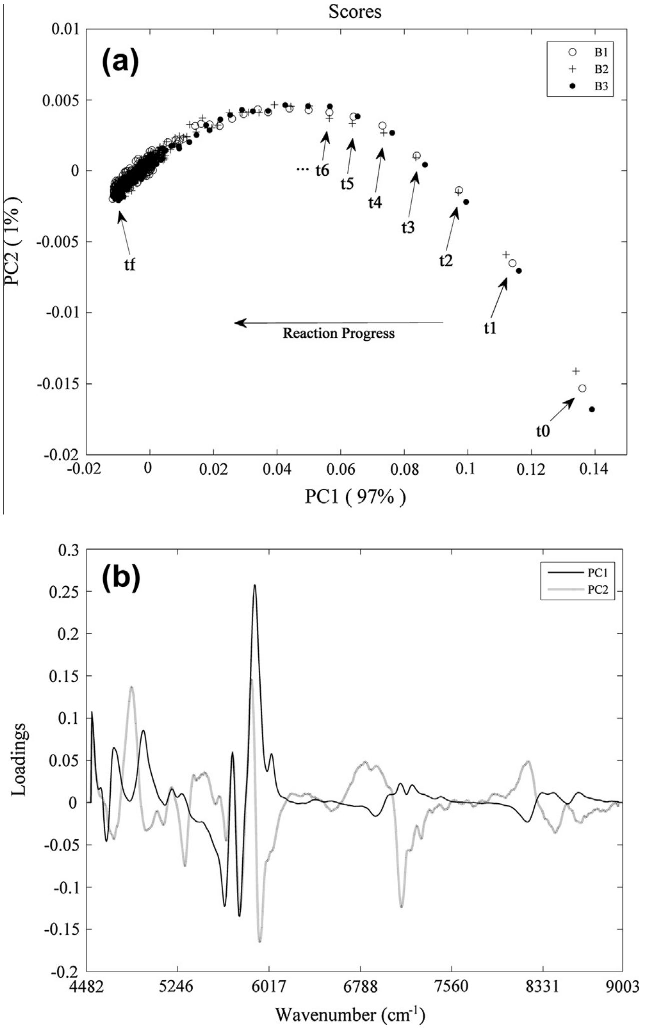

**Fig. 2.** (a) Scores on PC1 versus scores on PC2 for PCA analysis of the spectral data collected in-line during the soybean transesterification reaction (55 °C. oil/methol ration = 1/6, 0.75 wt% catalyst).$t_0$ = initial time, $t_f$ = final time and (b) loadings of PC1 and PC2.

significance of differences between RMSEP values. The *F*-values were calculated as the ratio of the squares of the largest and smallest RMSEP values. This ratio was compared with the critical *F*(*n*, *n*, 0.95) value. In our case *n* equaled the numbers of samples in the prediction set, which were 23, 14, 19 and 16 for ME, MG, DG and TG, respectively. Therefore, *F* values were 2.05, 2.58, 2.22 and 2.40 for ME, MG, DG and TG, respectively.

A multivariate control chart based on latent variables of (Principal Component Analysis, PCA) was built using the spectra collected from the eight batches produced "under control" conditions.

All computations were performed using the Unscrambler® X.1 (Camo), Matlab®R2010a 1997.10.0.499 (Mathworks) and SIMCA-P + 12.0.1 software platforms.

## 3. Results and discussion

During the transesterification of vegetable oils using methanol, complex mixtures with two immiscible phases are obtained. The conversion of vegetable oil to its fat acid methyl ester results in changes in the characteristics of the mixture, such as viscosity reduction. Because of this complex two phase mixture, a large scattering of radiation with baseline shifts are observed in the spectra (Fig. 1a). This scattering is reduced along the reaction as a consequence of various effects, for example, changes in viscosity, solubility, and refrative index. Thus, this physical effect could be used to monitor the reaction evolution. In this work, however, baseline shifts were corrected using pre-processing procedures. Fig. 1b shows derivative spectra of one batch taken as an example. The spectral regions below 4480 $cm^{-1}$ and above 9003 $cm^{-1}$ showed saturated and poor absorption signals (results not shown), respectively, and so they have been excluded in the present work.

Spectra presented in Fig. 1a show absorption bands assigned to combination (I: 4540–5060 $cm^{-1}$), first overtone combination (II: 6700–7500 $cm^{-1}$), first overtone (III: 5300–6150 $cm^{-1}$) and second overtone (IV: 8000–8900 $cm^{-1}$) regions of C–H stretching. In addition, bands assigned to second, third and fourth overtones of the C = O appear at about 5170 $cm^{-1}$ (V), 7000 $cm^{-1}$ (VI) and 8621 $cm^{-1}$ (VII), respectively [27].

Principal component analysis (PCA) of derivative NIR spectra for transesterification reactions, carried out at 55 °C with 0.75 w/w% of catalyst, is shown in Fig. 2a. Spectra were acquired from 80 s ($t_0$) to 5433s ($t_f$). As can be seen, PC1 and PC2 are sufficient to predict 98% of the total variance and PC1 scores allow monitoring the progress of transesterification reactions with the time indicated. In addition, adequate reproducibility of the transesterification reactions for three different batches, taken as examples, can be observed. In Fig. 2b, the loadings of PC1 and PC2 are shown. For PC1 the main loadings are assigned to first overtone of C–H stretching and second overtone of the C = O bonds. PC2 also show important loadings assigned to the first, third and fourth overtones of C = O.

As described before, detection and elimination of outliers were carried out using score, residual (***X*** and ***y***) and leverage plots. In general, most of the outliers were samples collected at the beginning of the reactions. This fact can be attributed to: (a) the system might not have been sufficiently homogenized owing to the complexity of the reaction mixture, with two immiscible phases, (b) as the reaction is very fast, the sample analyzed by the

**Table 2**
Best results for the calibration and validation steps for the PLS models developed to predict content of methyl ester (ME).

| Modelling strategy | Calibration step (*N* = 50) | | | Validation step (*N* = 23) | |
|---|---|---|---|---|---|
| | RMSECV | $R_{Cal}$ | LVs | RMSEP | $R_{pred}$ |
| PLS full spectrum baseline/smothing | 2.98 | 0.92 | 3 | 1.99 | 0.97 |
| PLS/JK baseline/smothing | 2.88 | 0.93 | 3 | 1.99 | 0.97 |
| PLS full spectrum (first deriv.11 window points) | 2.39 | 0.95 | 6 | 1.51 | 0.98 |
| PLS/JK (first deriv.11 window points) | 2.22 | 0.96 | 4 | 1.60 | 0.98 |
| PLS full spectrum SNV/smoothing | 2.44 | 0.95 | 9 | 1.91 | 0.98 |
| PLS/JK SNV/smoothing | 2.13 | 0.96 | 8 | 2.81 | 0.94 |
| PLS full spectrum MSC | 2.94 | 0.93 | 8 | 2.25 | 0.97 |
| PLS/JK MSC | 2.51 | 0.97 | 6 | 2.68 | 0.99 |
| PLS full spectrum EMSC | 2.12 | 0.96 | 9 | 2.31 | 0.98 |
| PLS/JK EMSC | 1.78 | 0.97 | 5 | 2.60 | 0.99 |

**Table 3**
Best results for the calibration and validation steps for the PLS models developed to predict monoglycerides (MG), diglycerides (DG) and triglycerides (TG) contents.

| Parameter | Modeling strategy | Calibration step | | | | Prediction step | | |
|---|---|---|---|---|---|---|---|---|
| | | $N_{\text{Cal}}$ | RMSECV | $R_{\text{Cal}}$ | LVs | $N_{\text{pred}}$ | RMSEP | $R_{\text{pred}}$ |
| MG (wt%)[a] | PLS/JK (Deriv. 9 window points) | 50 | 0.23 | 0.95 | 4 | 14 | 0.24 | 0.96 |
| DG (wt%)[b] | PLS/JK (Deriv. 9 window points) | 50 | 0.58 | 0.98 | 3 | 19 | 0.35 | 0.99 |
| TG (wt%)[c] | PLS/JK (Deriv. 9 window points) | 45 | 1.18 | 0.97 | 2 | 16 | 0.68 | 0.98 |

Concentration range.

[a] 0.22–3.03 wt%.

[b] 0.05–9.94 wt%.

[c] 0.05–18.94 wt%.

reference method might not have been representative of the spectrum obtained. As already mentioned, samples with concentration values outside the adapted reference method concentration range have also been excluded.

Table 2 summarizes the best results obtained for ME content using PLS regression with different pre-processing strategies. RMSEP values varied between 1.51 and 2.81 %w/w. As can be seen, the better predictive ability was obtained using first derivative with Savitzky–Golay filter (second-order polynomial and 11 window points). It is worth noting that when this pre-processing strategy is used with Jack-Knife algorithm, a statistically similar result is obtained using fewer latent variables, according to the F-test at 95% confidence level. Models using baseline correction also present results similar to those obtained with derivative spectra using fewer latent variables. However, analyzing the spectra after this pre-processing, it could be seen that this strategy did not effectively correct the effect of radiation scattering (results not shown). Therefore, this was not considered for the present purpose. All other strategies, SNV, MSC and EMSC show worse results, which are statistically different from those obtained with derivative and baseline strategies according to the F-test at 95% confidence level.

Comparing the results obtained using PLS/JK derivative spectra with those described by Richard et al. [23,24] for on-line NIR monitoring of ethanolic transesterification using batch reactors or microreators, the present results are better. It is important to mention that the reaction with methanol gives a more complex two phase mixture than that with ethanol. On the other hand, the present results are slightly worse than those described by Killner et al. [22] for on-line monitoring of the transesterification reactions of soybean oil with methanol, whose RMSEP values were 0.74% and 1.27% of conversion for two different batches. The results of the present work were obtained, however, using samples collected from different reaction times and batches, which introduced more variability to the external validation data set. The present results for ME content is better than that described by Pinzi et al. [14] for off-line NIR monitoring of transesterification using multiple feedstocks (sunflower, maize and olive–pomace oils).

The best results obtained for monoglycerides (MG), diglycerides (DG) and triglycerides (TG) are presented in Table 3. In general, all the evaluated modeling strategies show similar prediction abilities among themselves at 95% confidence level (results not shown). However, it was observed that the PLS/JK models using Savitzky–Golay derivative spectra required a fewer number of latent variables, which show similar results independent of the number of window points. Therefore, these models are more parsimonious. For this reason, only the results obtained by PLS/JK models using derivative spectra with 9 window points are presented.

The RMSEP (0.24 wt%) obtained for MG content is better than that described in the literature, when the reaction was monitored on-line using ethanol [23] or off-line using methanol [14]. Up to now, no works have been found in literature, which use NIR in-line monitoring of the DG and TG contents. The present results for both, DG and TG contents are better than those obtained off-line by Pinzi et al. [14] who worked with three different feedstocks.

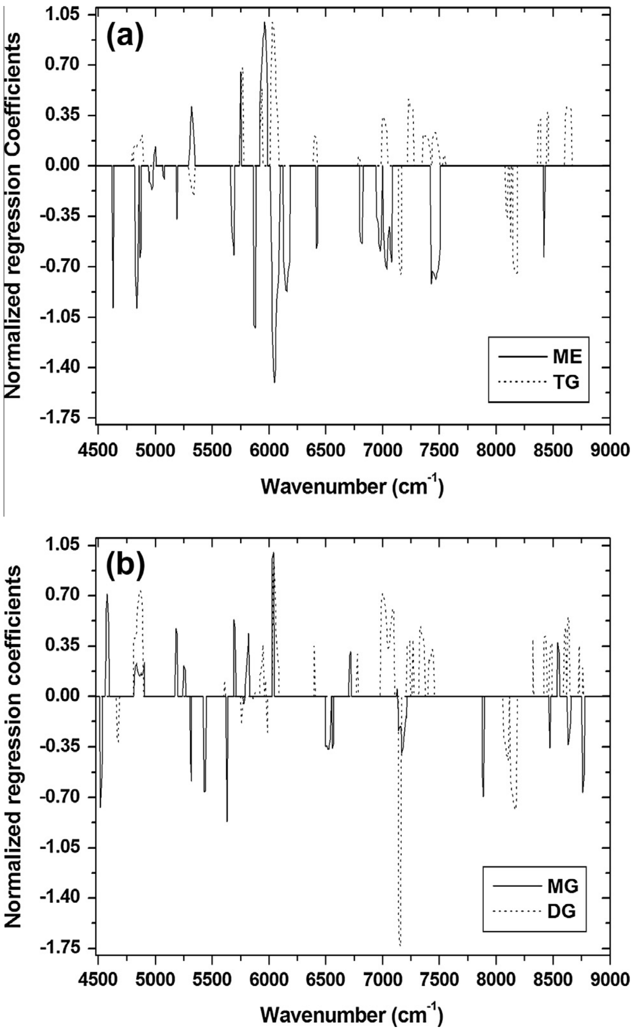

**Fig. 3.** Normalized significant PLS regression coefficients selected by the JK algorithm for ME, MG, DG and TG determination.

The significant PLS regression coefficients selected by the JK algorithm are shown in Fig. 3 for PLS ME, MG, DG and TG models. As can be seen, different significant regression coefficients and consequently different spectral information are used for ME, MG, DG and TG models. In general, the most significant regression coefficients for the ME content correspond to variables in the first overtone region (5920–6190 cm$^{-1}$) and first overtone combination (6790–7097 cm$^{-1}$) of C–H stretching (Fig. 3a). It is worth noting

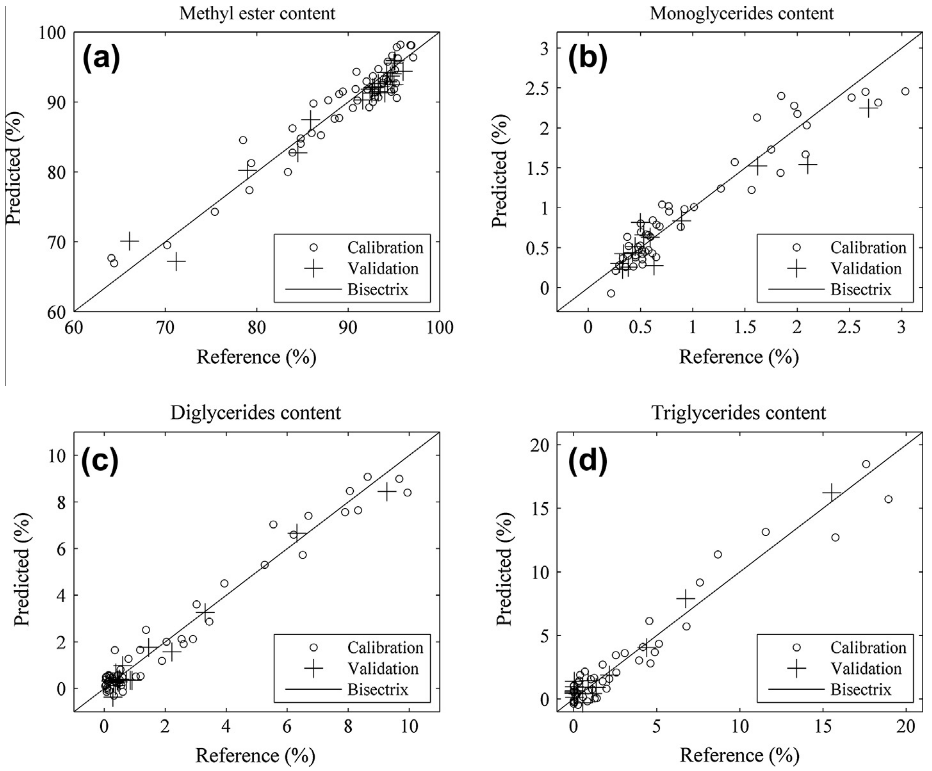

**Fig. 4.** Predicted versus reference plots obtained with the best modeling strategy for: (a) methyl ester (ME), (b) monoglycerides (MG), (c) diglycerides (DG) and (d) triglycerides (TG).

**Table 4**
Best results for the calibration and validation steps for the PLS models (using *i*PLS) and MLR models (using SPA algorithm) to predict methyl ester (ME), monoglycerides (MG), diglycerides (DG) and triglycerides (TG) contents.

| Parameter | *i*PLS | | | | | | SPA/MLR | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Spectral range (cm$^{-1}$) | RMSECV | $R_{Cal}$ | LVs | RMSEP | $R_{pred}$ | $N$[a] | RMSECV | $R_{Cal}$ | RMSEP | $R_{pred}$ |
| ME (wt%) | 5847–6302 | 2.66 | 0.94 | 4 | 3.43 | 0.91 | 9 | 2.11 | 0.96 | 1.98 | 0.97 |
| MG (wt%) | 4937–5392 | 0.28 | 0.93 | 3 | 0.26 | 0.97 | 14 | 0.24 | 0.95 | 0.29 | 0.95 |
| DG (wt%) | 5847–9003 | 0.69 | 0.97 | 2 | 0.41 | 0.99 | 7 | 0.51 | 0.99 | 0.40 | 0.99 |
| TG (wt%) | 5847–6302 | 1.13 | 0.97 | 3 | 0.72 | 0.99 | 3 | 0.85 | 0.98 | 0.54 | 0.99 |

[a] Number of selected variables by SPA.

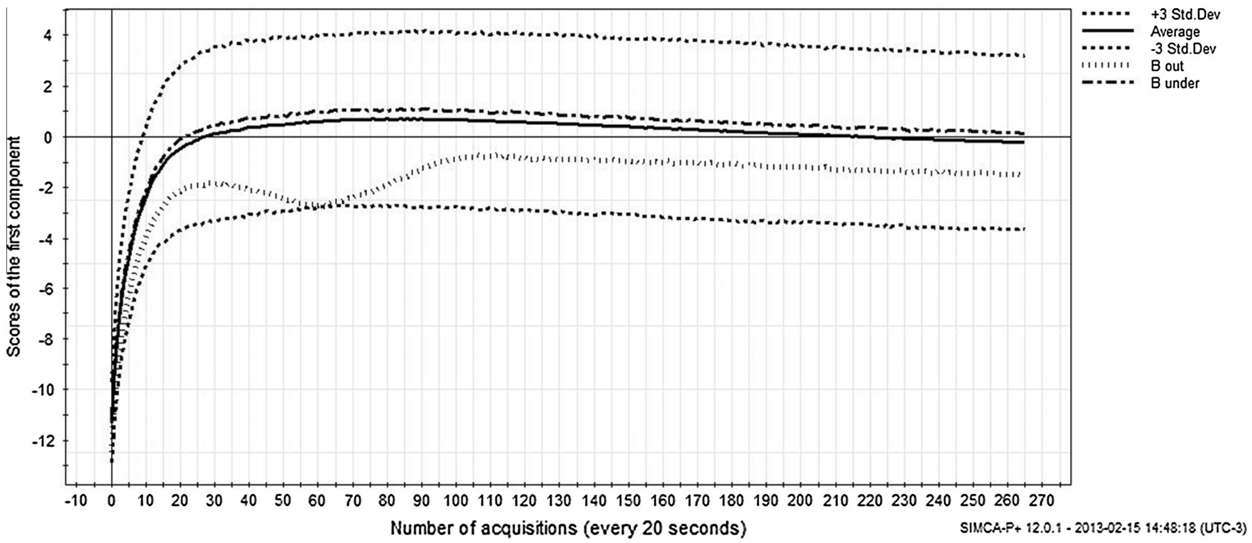

**Fig. 5.** Multivariate control chart. B out = batch “out of control”, B under = batch “under control”.

that the spectral range between 5920 and 6190 $cm^{-1}$ includes that described in the literature (5928.7–5959.6 $cm^{-1}$) for ME determination [13,22]. For MG, DG and TG modeling, the most significant regression coefficients correspond to variables also included in the second overtone (8000–8800 $cm^{-1}$) regions of C–H stretching. It is worth noting that the regression coefficients to MG, DG and TG are quite different among themselves.

Fig. 4 shows the predicted versus reference plots obtained from the best modeling strategies for predicting ME, MG, DG and TG contents. As can be seen, the points are randomly distributed around the bisectrix line along the entire range of *y*-values and no systematic error is present in the predictions.

Table 4 shows the best results obtained for ME, MG, DG and TG contents using *i*PLS and SPA/MLR strategies. As can be seen, for ME content the result obtained with *i*PLS strategy is worse and statistically different from the results obtained with PLS/JK (Table 2). On the other hand, SPA/MLR shows similar prediction ability (according to the *F*-test at 95% confidence level) to PLS/JK using only nine spectral variables, which are mainly assigned to first overtone (5300–6150 $cm^{-1}$) region of C–H stretching. This spectral region has been described for the determination of ME [22–23].

In general, for MG, DG and TG content (Table 4), both iPLS and SPA/MLR strategies show prediction abilities according to the *F*-test at 95% confidence level, similar to those obtained with the PLS/JK model (Table 3). It is worth noting that the SPA/MLR model for TG content shows the optimum RMSEP, using only three spectral variables, which are assigned to first overtone (5894 $cm^{-1}$) and first overtone combination (7152 and 7476 $cm^{-1}$) regions of C–H stretching. In addition, the fourth overtones of the C = O appear at about 7000 $cm^{-1}$.

Finally, for the multivariate control chart [28], statistical limits were estimated (at 3 standard deviations) using the scores of PC1 and the eight batches produced “under control” conditions (Fig. 5). The additional batch produced “under control” conditions shows scores within the control limits. It is also possible to identify the batch “out of control”, which changed its profile between 13 and 33 min due to a temperature perturbation produced during the reaction.

## 4. Conclusion

The NIR methods developed were successfully applied to in-line monitoring of the transesterification reaction of soybean oil, using methanol and NaOH. The composition (methyl ester, monoglycerides, diglycerides and triglycerides) of the reaction mixture during transesterification were predicted using PLS and MLR regression models, with satisfatory RMSEP, which were statistically similiar among themselves. SPA/MLR can be considered as a simpler modeling strategy because it uses a smaller subset of wavenumbers (e.g. three to fourteen spectral variables). In addition, control charts built using NIR spectra during the transesterification reaction proved to be a useful multivariate statistical process control tool, allowing real time identification of small changes in the process behavior.

These NIR multivariate models allow a more effective control of the transesterification reactions, which can reduce the time and cost of reactions and improve the quality of the final products.

## Acknowledgments

The authors thank the FINEP, FACEPE, CAPES and CNPq/INCTAA science funding programs for partial financial support. Research fellowships granted by the Brazilian agenciesANP/Petrobras, CNPq are also gratefully acknowledged.

## References

[1] Huang D, Zhou H, Lin L. Biodiesel: an alternative to conventional fuel. Energy Procedia 2012;16:1874–85.

[2] Atabani AE, Silitonga AS, Badruddin IA, Mahlia TMI, Masjukia HH, Mekhilef S. A comprehensive review on biodiesel as an alternative energy resource and its characteristics. Renew Sust Energ Rev 2012;16:2070–93.

[3] EN 14103. Fat and oil derivatives – fatty acid methyl esters (FAME) – determination of ester and linolenic acid methyl ester contents. Brussels, Belgium: European Standard; 2003.

[4] ASTM D6584-10a. Standard test method for determination of total monoglycerides, total diglycerides, total triglycerides, and free and total glycerin in B-100 biodiesel methyl esters by gas chromatography. West Conshohocken, PA: ASTM International; 2011.

[5] Arzamendi G, Arguiñarena E, Campo I, Gandía LM. Monitoring of biodiesel production: simultaneous analysis of the transesterification products using size-exclusion chromatography. Chem Eng J 2006;122:31–40.

[6] Richard R, Tiebaud-Roux S, Prat L. Modelling the kinectis of transesterification reaction of sunflower oil with ethanol in microreators. Chem Eng Sci 2013;87:258–69.

[7] Ghesti GF, Macedo JL, Braga VS, Souza ATCP, Parente VCI, Figuerêdo ES, et al. Application of Raman spectroscopy to monitor and quantify ethyl esters in soybean oil transesterification. J Am Oil Chem Soc 2006;83:597–601.

[8] Boni LAB, Silva INL. Monitoring the transesterification reaction with laser spectroscopy. Fuel Process Technol 2011;92:1001–6.

[9] Mahamuni NN, Adewuyi YG. Fourier transform infrared spectroscopy (FTIR) method to monitor soy Biodiesel and soybean oil in transesterification reactions, petrodiesel–biodiesel blends, and blend adulteration with soy oil. Energ Fuel 2009;23:3773–82.

[10] Knothe G. Rapid monitoring of transesterification and assesing biodiesel fuel quality by near-infrared spectroscopy using a fiber-optic probe. J Am Oil Chem Soc 1999;76:795–800.

[11] Felizardo P, Baptista P, Menezes JC, Correia MJN. Multivariate near infrared spectroscopy models for predicting methanol and water content in biodiesel. Anal Chim Acta 2007;595:107–13.

[12] Baptista P, Felizardo P, Menezes JC, Correia MJN. Multivariate near infrared spctroscpy models for predicting the iodine value, CFPP, kinematic viscosity at 40 C and density at 15 C of biodiesel. Talanta 2008;77:144–51.

[13] Baptista P, Felizardo P, Menezes JC, Correia MJN. Multivariate near infrared spectroscopy models for predicting the methyl esters content in biodiesel. Anal Chim Acta 2008;607:153–9.

[14] Pinzi S, Alonso F, Olmo JG, Dorado MP. Near infrared reflectance spectroscopy and multivariate analysis to monitor reaction products during biodiesel production. Fuel 2012;92:354–9.

[15] Knothe G. Monitoring a progressing transesterification reaction by fiber-optic near infrared spectroscopy with correlation to $^{1}$H nuclear magnetic resonance spectroscopy. J Am Oil Chem Soc 2000;77:489–93.

[16] Lira LFB, Albuquerque MS, Pacheco JGA, Fonseca TM, Cavalcanti EHS, Stragevitch L, et al. Infrared spectroscopy and multivariate calibration to monitor stability quality parameters of biodiesel. Microchem J 2010;96:126–31.

[17] Menezes JC, Ferreira AP, Rodrigues LO, Brás LP, Alves TP. Chemometrics Role within the PAT context: examples from primary pharmaceutical manufacturing. In: Brown SD, Tauler R, Walczak B, editors. Comprehensive chemometrics: chemical and biochemical data analysis. Oxford: Elsevier Science; 2009. p. 313–55.

[18] Cabeça LF, Marconcini LV, Mambrini GP, Azeredo RV, Colnago LA. Monitoring the transesterification reaction used in biodiesel production, with a low cost unilateral nuclear magnetic resonance sensor. Energ Fuel 2011;25:2696–701.

[19] Anderson LA, Franz AQ. Real-time monitoring of transesterification by $^{1}$H NMR spectroscopy: catalyst comparison and improved calculation for biodiesel conversion. Energ Fuel 2012;26:6404–10.

[20] Kouame SB, Perez J, Eser S, Benesi A. $^{1}$H-NMR monitoring of the transesterification process of Jatropha oil. Fuel Process Technol 2012;97:60–4.

[21] Trevisan MG, Garcia CM, Shuchardt U, Poppi RJ. Evolving factor analysis-based method for correcting monitoring delay in different batch runs for use with PLS: On-line monitoring of a transesterification reaction by ATR-FTIR. Talanta 2008;74:971–6.

[22] Killner MHM, Rohwedder JJR, Pasquini C. A PLS regression model using NIR spectroscopy for on-line monitoring of the biodiesel production reaction. Fuel 2011;90:3268–73.

[23] Richard R, Li Y, Dubreuil B, Thiebaud-Roux S, Prat L. On-line monitoring of the transesterification reaction between triglycerides and ethanol using near infrared spectroscopy combined with gas chromatography. Bioresource Technol 2011;102:6702–9.

[24] Richard R, Dubreuil B, Thiebaud-Roux S, Prat L. On-line monitoring of the transesterification reaction carried out in microreators using near infrared spectroscopy. Fuel 2013;104:318–25.

[25] Galvão RKH, Araújo MCU, Jose GE, Pontes MJC, Silva EC, Saldanha TCB. A method for calibration and validation subset partitioning. Talanta 2005;67:736–40.
[26] Paiva HM, Soares SFC, Galvão RKH, Araújo MCU. A graphical user interface for variable selection employing the successive projections algoritm. Chemometr Intell Lab 2012;118:260–6.
[27] Workman J, Weyer L. Pratical guide to interpretative near-infrared spectroscopy. 1st ed. New York: CRC Press; 2008.
[28] Ferrer-Riquelme AJ. Statistical control of measures and process. In: Brown SD, Tauler R, Walczak B, editors. Comprehensive chemometrics: chemical and biochemical data analysis. Oxford: Elsevier Science; 2009. p. 97–126.