

Full Length Article



Monitoring biodiesel and its intermediates in transesterification reactions with multivariate curve resolution alternating least squares calibration models

Werônica de Lima Furtado^a, Camila Nunes Costa Corgozinho^a, Romá Tauler^b, Marcelo Martins Sena^{a,c,*}

^a Departamento de Química, Instituto de Ciências Exatas (ICEx), Universidade Federal de Minas Gerais (UFMG), 31270-901 Belo Horizonte, MG, Brazil

^b Department of Environmental Chemistry, IDAEA-CSIC, C/Jordi Girona, 18 26, 08034, Barcelona, Spain

^c Instituto Nacional de Ciência e Tecnologia em Bioanálítica, 13083 970 Campinas, SP, Brazil

ARTICLE INFO

Keywords:

MCR-ALS
Correlation constraint
Chemometrics
Transesterification reaction
Biofuel
FTIR

ABSTRACT

The monitoring of biodiesel production processes is currently performed by methods requiring high reagent consumption and long response times. Aiming to overcome these drawbacks, this article developed chemometric models combining principal component analysis (PCA), multivariate curve resolution alternating least squares (MCR-ALS) and mid infrared spectroscopy to determine simultaneously the concentrations of the main components, methyl esters and triglycerides, along transesterification reactions. This methodology also tried to model the intermediates (fatty acids plus monoglycerides, and diglycerides), thus providing semi-quantitative predictions for these components. These reactions were carried out in different experimental conditions according to a factorial design with two factors, the batch of soybean oil (raw material) and the alcohol:oil molar ratio. The concentration values estimated by high performance liquid chromatography (HPLC), as the reference method, were the inputs for the correlation constraint used as a key step in the multivariate calibration MCR-ALS models. The spectral profiles recovered for the methyl esters and for the triglycerides have shown high correlations (greater than 0.985) in comparison to the reference spectra. In addition, the spectra calculated for fatty acids, monoglycerides and diglycerides showed absorption bands characteristic of functional groups present in these molecules, mainly the hydroxyl and carbonyl stretching bands. The root mean square errors of calibration and prediction estimated for the triglycerides and for the methyl esters were within the range from 1.9% to 6.3%.

1. Introduction

The growth of the world population and the rapid industrialization have increased fossil fuels demand about 43% in the last 41 years. Besides, CO₂ emissions in 2030 are estimated to be 80% above 2007 levels [1]. In the last years, several countries have created political strategies to promote the use of biofuels aiming to reduce environmental impacts and their dependence on fossil fuels [2]. In Brazil, fuels from renewable sources have been used for many years. Some examples are sugarcane ethanol, used in automotive engines, and Eucalyptus sp. charcoal, used by steel-making industries. As of 2005, biodiesel was incorporated into the Brazilian energy matrix. Since then, the minimum percentage of biodiesel added to diesel oil has been compulsorily increased and is

currently 11% [3]. Biodiesel is a mixture of straight-chain alkyl esters obtained by a transesterification reaction, in which triglycerides of vegetable oil and animal fat react with short-chain alcohols in a catalytic process [4].

The transesterification reaction is usually carried out in alkaline medium, under heating and stirring. Fatty acid esters are the product of interest and glycerol is a co-product. However, triglycerides, diglycerides, and monoglycerides might be present if the conversion is not complete [4]. This situation is undesirable once only the esters have similar properties to diesel oil and can be blended to it to be used in automotive engines. The National Agency of Petroleum, Natural Gas and Biofuels (ANP) is the Brazilian regulatory agency that defines the fuel specifications and the standard methods to its quality control. According

* Corresponding author at: Departamento de Química, Instituto de Ciências Exatas (ICEx), Universidade Federal de Minas Gerais (UFMG), 31270-901 Belo Horizonte, MG, Brazil.

E-mail address: marcsen@ufmg.br (M.M. Sena).

<https://doi.org/10.1016/j.fuel.2020.119275>

Received 29 July 2020; Received in revised form 12 September 2020; Accepted 15 September 2020

Available online 8 October 2020

0016-2361/© 2020 Elsevier Ltd. All rights reserved.

Table 1
Gradient elution program used in the reference HPLC method.

Time (min)	Methanol (%)	Isopropanol:Hexane 5:4 v/v (%)
0.01	100	0
15.0	50	50
20.0	0	100
23.0	0	100
26.0	100	0
36.0	100	0

to ANP, the content of esters in biodiesel must be determined by gas chromatography (GC) and the result must be greater than 96.5% w/w [5,6]. However, if triglycerides, monoglycerides, and diglycerides are present, these low volatile substances can damage the capillary columns and other parts of the chromatographic system [7–9]. Then, the quantification of these products requires a previous derivatization step to convert them into volatile substances [5,8,10–13]. The maximum amount of mono-, di- and triglycerides in a biodiesel sample must be 0.7% w/w, 0.2% w/w and 0.2% w/w, respectively [5,6].

All the analytical techniques mentioned above are expensive and time-consuming. Then, several alternative methodologies have been developed to analyze the products of transesterification reactions. Among the developed methods, the most widespread have been those based on high-performance liquid chromatography (HPLC) and hydrogen nuclear magnetic resonance (^1H NMR) spectroscopy [6,14,15]. Although these techniques provide relevant information about the biodiesel production process, they are still costly and make difficult to follow the reaction quickly. In this context, studies using vibrational spectroscopy in the near infrared (NIR) region have become popular. They aim to develop methodologies that allow monitoring transesterification reaction reducing analysis time as well as associated cost, not demanding reagents or solvents, nor generating chemical waste [15–17].

Pinzi et al. [16] have applied chemometric methods to determine the yield of methyl esters and the content of monoglycerides, diglycerides, triglycerides, free and total glycerol in biodiesel samples using NIR spectra. The samples were clustered in a principal component analysis (PCA) model according to the feedstock oil used for transesterification (maize, sunflower or olive–pomace). Using partial least squares (PLS) and reference concentrations obtained by GC the reaction components were determined with errors close to those provided by conventional methodologies. Lima et al. [17] have developed a method for in-line monitoring of the transesterification reactions of soybean oil with methanol. Multivariate calibration models employing NIR spectra were built for determining the contents of methyl esters and other components. This method presented some drawbacks related to the limitations of the reference technique (GC). Other authors have also developed in-line monitoring of transesterification reactions. An article has developed control charts based on NIR data, enabling the identification of spectral regions mostly affected by faults in the process [18]. In another article, the variations in the Raman band of methanol were identified by a PCA model as the main factor to monitor the reaction development [19].

Vibrational spectroscopic techniques, such as those previously mentioned, provide data with overlapped signals and demand the use of chemometric methods to extract relevant information. Specifically for quantitative modeling, the most widely used multivariate calibration method has been PLS. However, PLS (and other methods, such as artificial neural networks) presents some drawbacks, requiring a minimum of some tens of samples to build the model, and not allowing to estimate/recover pure spectra of the components. By contrast, multivariate curve resolution alternating least squares (MCR-ALS) can be used to build models with a relative smaller number of samples, and allows the recovery of pure spectra of the system constituents [20].

MCR-ALS is a curve resolution method suitable for monitoring

mixtures. It aims to decompose mathematically a set of instrumental responses for a mixture in pure contributions of each component present in the system [21–23]. For the application of MCR-ALS to a dataset, it is necessary to provide an initial estimate for the concentration or the spectral profile. From this point, the ALS algorithm acts in the optimization step until a pre-established convergence criterion is reached. If no mathematical constraints are adopted, an infinite number of feasible solutions will be obtained for the MCR decomposition, leading to the so called rotational ambiguity [24,25]. In addition to it, intensity and permutation ambiguities may also be present in the MCR-ALS solutions. The main constraints that can be applied to reduce/circumvent ambiguities are non-negativity, unimodality, closure, known pure spectra/concentration profiles, selectivity, hard modeling, and correlation. The correlation constraint is one recently proposed tool as a key step for quantitatively processing first-order data [26,27], which was more recently also extended to second order data [28]. This constraint has expanded the applications of MCR-ALS, previously mostly restricted to qualitative models, to multivariate calibration.

MCR-ALS has been used in the context of process analytical technology (PAT) to monitor different chemical systems, such as complexation reactions, esterification, alcoholic fermentation, and polymerization, including processes on an industrial scale [28–30]. Other articles have reported the advantages of MCR-ALS associated with different analytical techniques, such as GC [31], Raman [32], near and mid infrared spectroscopies [33] for the identification and quantification of adulterants in biodiesel–diesel blends, and for the determination of antioxidants in biodiesel mixtures from different vegetable sources using UV–Vis–NIR spectroscopy [27]. More specifically, two articles have already been published reporting the application of MCR-ALS combined with NIR spectroscopy to monitor transesterification reactions. In the study of Rouchi et al., the initial estimate of the number of factors and their spectral profiles was performed by evolving factor analysis (EFA). Despite estimating the concentration and spectral profiles of biodiesel and raw material during the transesterification reaction, this article did not determine the true concentrations of the components, since no reference method was used for quantification. In addition, the developed MCR-ALS model was limited for quantitative applications, because only the non-negativity constraint was adopted [34]. Sales et al. have built a MCR-ALS model applying the correlation constraint only to the biodiesel component/factor, and its reference values were obtained by GC. Due to the limitations of the reference technique, the range of biodiesel concentration predicted by this MCR-ALS model was restricted to above 64% [35]. None of these articles have determined the concentrations of the intermediate components of the reaction.

The main originality of the present work is the monitoring of the transesterification reaction by modeling simultaneously the components of the system. The main components, triglycerides and methyl esters, were directly quantified, without the need of physical separation, while semi-quantitative predictions were provided for the intermediates, diglycerides and the mixture of monoglycerides and fatty acids. For this aim, attenuated total reflectance Fourier transform infrared spectroscopy (ATR-FTIR) associated with MCR-ALS and the correlation constraint were used. The reference values for the components were obtained by HPLC. A factorial design was used to scan the reaction conditions, and a PCA model was built as a preliminary tool for characterizing this monitoring.

2. Material and methods

2.1. Sampling and characterization of soybean oil

Three samples of commercial soybean oils were purchased in duplicate on the local market and identified as 010, 020 and 100. They were characterized according to the following parameters: density at 20 °C [36], water content [37], kinematic viscosity [38], oxidative

Table 2
Characterization of the soybean oils from three different batches (N = 3).

Parameters	Soybean oil batches		
	020	010	100
Saponification index(mg KOH / g sample)	188 ± 3	188 ± 1	188 ± 1
Average molar weight(g/mol)	895 ± 14	894 ± 3	897 ± 6
Water content(mg/kg)	666 ± 20	695 ± 29	689 ± 34
Density at 20 °C(kg/m ³)	0.9189 ± 0.0002	0.9199 ± 0.0001	0.9199 ± 0.0002
Kinematic viscosity(mm ² /s)	32.00 ± 0.01	32.18 ± 0.01	31.82 ± 0.01
Oxidative stability(h)	8.8 ± 0,1	6.6 ± 0.1	6.9 ± 0.2
Acid number(mg KOH/g sample)	0.13 ± 0.01	0.36 ± 0.01	0.28 ± 0.01

stability [39], acid number [40] and saponification index [41]. All determinations were performed in triplicate.

2.2. Transesterification reactions

Five transesterification reactions were performed, varying the molar ratio of the reagents and the batch of soybean oil according to an experimental design 2² with a central level for the variable molar ratio and a different oil with intermediate value of oxidative stability (Table 1S, Supplemental materials). The two factors were the batch of soybean oil (raw material) and the alcohol:oil molar ratio (methanol:oil between 5:1 and 7:1, molar ratio). The temperature was maintained at 20 °C and the catalyst was sodium methoxide 1% w/w. The purity of the reagents (methanol and NaOH from Sigma-Aldrich) and their amount used in each reaction are presented in Table 2S. The initial stirring speed was 1000 rpm until the complete addition of sodium methoxide to the oil. Then, the speed was changed to 1200 rpm and the reaction monitoring started from this point.

The sampling was divided into two steps. In the first step, 500 µL of the reaction mixture was removed every 1 min until to complete 10 min of reaction. From then on, sampling was performed at intervals of 5 min until the final period of 120 min. Thus, 33 samples were obtained for each transesterification reaction, totaling 165 samples.

Acetic acid p.a. (50 µL, Sigma-Aldrich) was used to inactivate the catalyst, stopping the reaction. The collected products were centrifuged for 30 min at 7000 rpm using a specific micro tube rotor for Eppendorf tubes (Centrifuge 5804, rotor FA-45-30-11, Eppendorf, Hamburg, Germany). After centrifugation, the collected samples were separated into two phases. Then, the upper phase was analyzed by ATR-FTIR and HPLC. After 120 min of reaction, the biodiesel (target product) was separated, washed with warm water, and dried with anhydrous sodium sulfate. The purified biodiesel was also analyzed by ATR-FTIR and HPLC.

2.3. Reference concentrations by HPLC

The relative concentrations of fatty acids and monoglycerides (as a mixture, since their peaks are not resolved), diglycerides, triglycerides and methyl esters in the samples were determined using a Shimadzu liquid chromatograph, model LC-20AT, equipped with an automatic sampler SIL-20HT, a UV-Vis detector SPD-M20A and a Shimadzu CLC-ODS® column (M) (250 mm × 4.6 mm × 5 µm). The method used was adapted from the article by Holcapek et al. [42]. The analysis conditions were as follows. Detection wavelength, 205 nm; oven temperature, 40 °C; injection volume, 6 µL; mobile phase flow, 1 mL/min; gradient elution using methanol and isopropanol:hexane 5:4 v/v, as shown in Table 1; and run time: 36 min. The sample preparation for chromatographic analysis consisted of its dilution in the 1:20 ratio, in isopropanol:hexane 5:4 v/v, followed by filtering using a hydrophobic PTFE filter with a diameter of 13 mm and pore diameter of 0.22 µm.

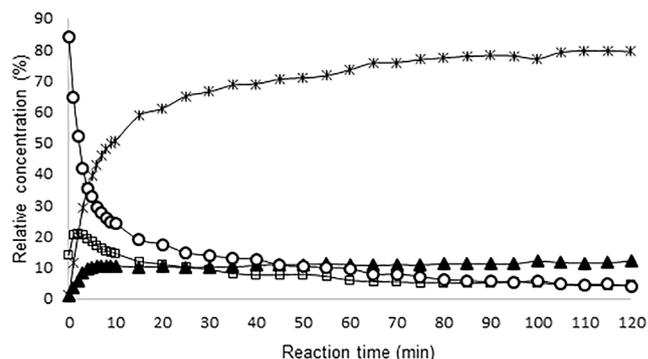


Fig. 1. Concentration profiles for components of the reaction 010-7 obtained by HPLC-DAD. Fatty acids and monoglycerides (full up triangles), methyl esters (asterisks), diglycerides (empty squares) and triglycerides (empty circles).

Table 3

Relative concentrations determined by HPLC-DAD for purified biodiesel samples obtained in the final of the transesterification reactions.

Reaction components*	Range of retention time (minutes)	Reaction codes				
		010-5	010-7	020-5	020-7	100
		Relative concentrations (%)				
FA + MG	3.0 – 4.5	7.10	7.51	5.07	2.54	9.05
ME	4.6 – 7.0	80.43	82.10	80.99	90.00	79.94
DG	7.1 – 15.9	4.71	6.00	6.89	4.96	4.55
TG	16.0 – 24.0	7.76	4.40	7.05	2.49	6.46

*FA + MG: fatty acids and monoglycerides; ME: methyl esters; DG: diglycerides; TG: triglycerides.

2.4. FTIR spectra acquisition

Molecular absorption spectra in the mid infrared region were recorded for the 165 samples in a Perkin Elmer Frontier spectrometer equipped with an ATR accessory with a diamond crystal. The analysis was performed with 20 µL of sample in the range of 650 to 4000 cm⁻¹, with 16 scans and a resolution of 4 cm⁻¹.

2.5. Chemometric modeling

All the data processing was performed in MATLAB environment (The MathWorks, Natick, MA, USA). Spectra obtained for the five reactions (Table 1S) were firstly organized in a matrix (165x3351), which was preprocessed by multiplicative scatter correction (MSC), aiming to eliminate non-linear baseline deviations caused by light scattering [43], and then mean centered. A PCA model was built using PLS_Toolbox (Eigenvector Research, Manson, WA, USA). PCA is a non-supervised classification method widely used in multivariate exploratory analysis [44]. PCA has been combined with vibrational spectroscopy for the monitoring of processes, such as transesterification reactions [16,19] and biodiesel degradation by thermo-oxidation [45].

The MCR-ALS GUI 2.0 [46] was used for building MCR-ALS models. This toolbox for Matlab is freely available for download at www.mcrals.info. MCR calibration models were built for each of the five reactions, splitting 2/3 of the samples for the calibration set and 1/3 for the validation set. The spectral data were arranged in a matrix with the first lines corresponding to the calibration samples and the last lines to the validation samples. The arrangement of the reference concentration matrix, obtained by HPLC, was performed similarly, with each row referring to a sample and each column corresponding to reaction components. For the samples in the validation set, the reference concentration values were replaced by NaN (not a number), which indicates the absence of a defined numerical value. After the model estimating the

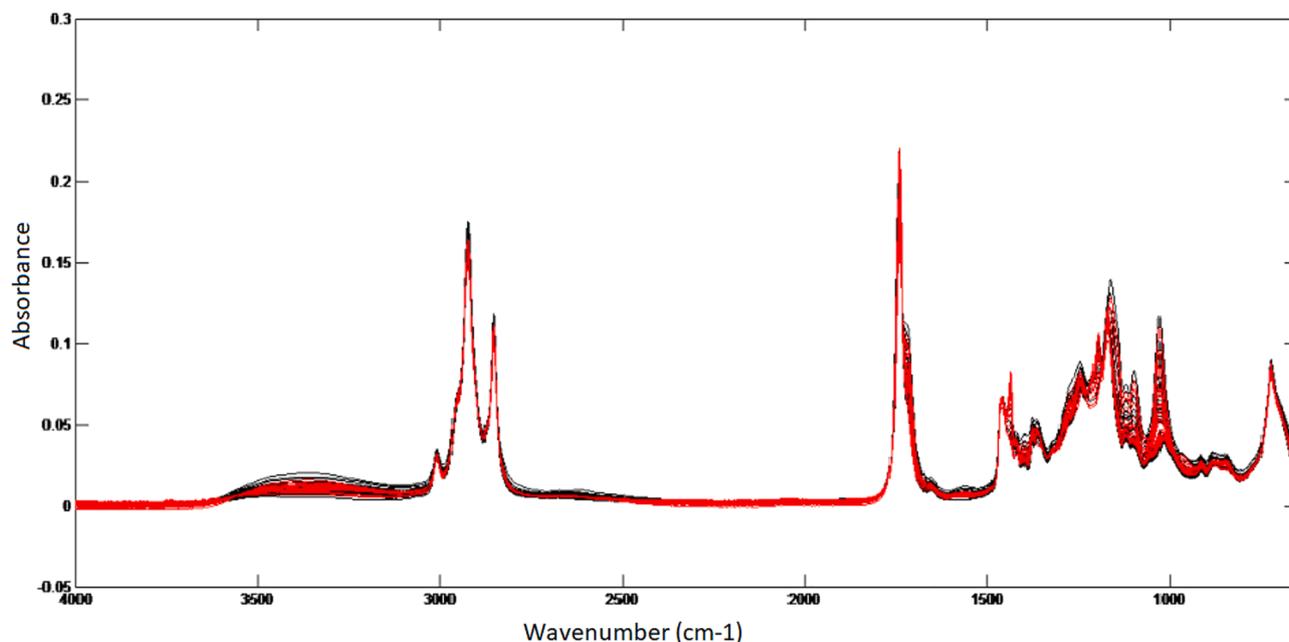


Fig. 2. ATR-FTIR spectra of 33 samples analyzed for transesterification reaction 010–7. Raw spectra in black, and spectra preprocessed by MSC in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

concentration values of each of the components of the validation samples, they were compared with the reference values (HPLC) through the figures of merit described in section 3.4.1.

Additionally, another MCR model was built using a column-wise augmented matrix, in which the spectral data from reactions 020–7, 010–7, 020–5 and 010–5 (the first number of this code means for the number of the batch of the raw material, and the second number means for the ratio of oil to methanol) comprised the calibration set, and the spectra from reaction 100 formed the validation set. In all these models, datasets were preprocessed by MSC. Four components were used for all the MCR models, according to the previous knowledge of the system, accounting for between 99.90% and 99.95% of the total data variance. The initial spectral profiles were estimated by the selection of the purest spectra using a method similar to SIMPLISMA (SIMPLe-to-use Interactive Self-modeling Mixture Analysis) [47]. This initialization was used for all the components and is based on a search for pure variables for each component. These purest spectral profiles (alternatively, concentration profiles could be used) are used as the starting point in iterative resolution methods, and are either the most dissimilar in the dataset or those spanning the data space most efficiently according to the criterion of the selection method [21]. Other important aspect is that the amount of filtering noise for the initial estimates has been recommended to be above the noise level of the data. When the selected noise level is higher than the experimental one, selected profiles are not noisy and have the expected spectral features [46]. In this study, we adopted 10% of filtering noise. Non-negativity (spectral and concentrations profiles) and correlation (in the concentrations profile) constraints were applied. Concentrations determined by HPLC were used as reference values in the calibration set. The convergence criterion adopted was a difference between the relative change of the standard deviation of the residuals between two consecutive iterations lower than 0.1%. For evaluating the performance of the estimated models, the percentages of explained variance and the lack of fit (LOF) were calculated.

2.6. MCR-ALS with correlation constraint

MCR-ALS is a chemometric method that aims at recovering the relative concentration values, contained in the C matrix, and the pure spectra, in the S^T matrix, of the components present in a mixture through

the decomposition of its instrumental response matrix (D), as described in Eq. (1). The matrix E contains non-modeled/residual information [21], and the symbol T indicates a transposed matrix.

$$D = C \cdot S^T + E \quad (1)$$

The first steps in MCR-ALS are the choice of the number of components present in the system and the initial estimates for the spectra or concentration profiles. Then, appropriate constraints must be applied, aiming to reduce the ambiguities associated with the resolution of D matrix. Correlation is the constraint highlighted in this study. By using this constraint, reference values (here obtained by HPLC) for samples of the calibration set are given as input. Thus, at each optimization iteration using the ALS algorithm, the reference values for the analyte in the calibration set (contained in the vector c_{ref}^{cal}) are correlated with the relative concentration values of the analyte calculated by the MCR model (contained in the vector c_{ALS}^{cal}) through a linear regression step (Eq. (2)) [27]. In this sequence, the parameters of the linear regression, slope, b , and offset, b_0 , estimated by Eq. (2) are used to calculate the absolute concentrations vector for the analyte in the samples of the validation set (\hat{c}_{val}), also from the relative concentration values provided by the MCR model (c_{ALS}^{val}) (Eq. (3)). Thus, the reference values for the calibration samples (c_{ref}^{cal}) and the calculated values for the validation samples (\hat{c}_{val}) are used to construct the input vector in the next ALS iteration. This iterative process continues until the convergence criterion is reached [26].

$$c_{ALS}^{cal} = b \cdot c_{ref}^{cal} + b_0 \quad (2)$$

$$\hat{c}_{val} = (c_{ALS}^{val} - b_0) / b \quad (3)$$

3. Results and discussion

3.1. Soybean oil characterization

Soybean oil samples from three different batches were characterized by physico-chemical parameters, and the average values jointly with the respective standard derivations are shown in Table 2. The saponification index values from soybean oil samples were used to calculate their average molar mass, using an analytical method of the Adolfo Lutz

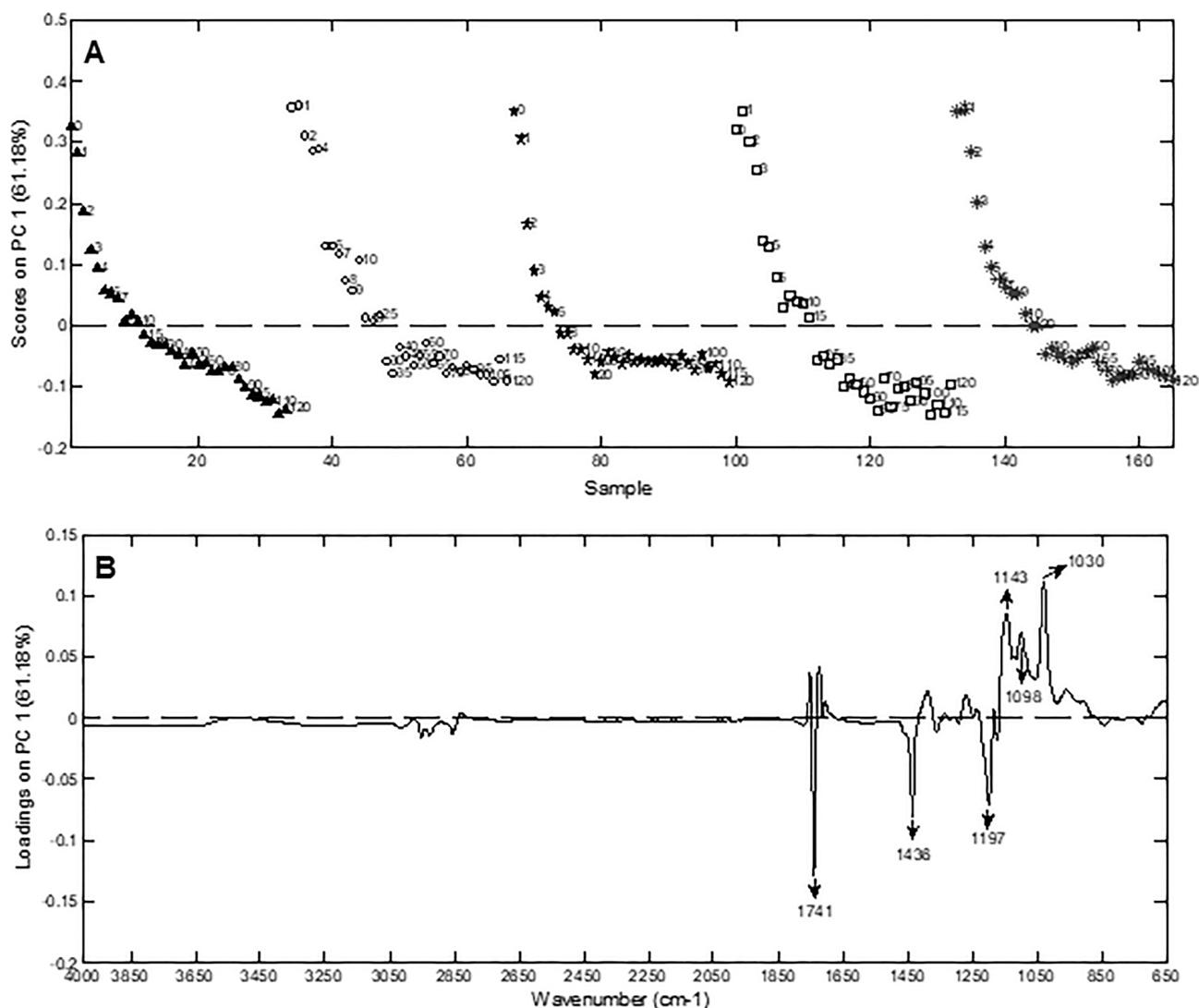


Fig. 3. PC1 (61.18%) (A) scores and (B) loadings obtained from the PCA model for the monitoring of the five reactions. Reactions 010–5 (full up triangles), 010–7 (empty circles), 020–5 (full stars), 020–7 (empty squares) and 100 (asterisks).

Institute [41]. The average molar mass of the oils is a relevant parameter to calculate the stoichiometric amounts of sodium hydroxide and methanol used in the transesterification reaction. The results for these parameters were close to those observed in the literature, 188.6 mg KOH / g and 896.88 g/mol for saponification index value and average molar mass, respectively [48]. The results among all soybean oil samples were similar for the parameters kinematic viscosity and density. These parameters are related to the chemical composition of oils. Since soybean oil has carbon chains derived from palmitic (C16: 0), oleic (C18: 1) and linoleic (18: 2) acids, it has a density close to 0.91 g/cm³ [49].

The soybean oil from batch number 020 presented the highest value for oxidative stability. This result suggests that this oil will spend more time than the others to start its oxidation process. Besides, this sample had lower values for the acid number and water content, what characterizes this oil as having the best quality for biodiesel production among those analyzed. The presence of water leads to the hydrolysis of the triglycerides, producing free fatty acids and consequently increasing the acid number of the vegetable oil. These conditions contribute to the undesirable formation of soap during the transesterification reaction. Finally, it was observed that, in general, soybean oil from batch 100 presented intermediate parameters between batches 010 and 020. For this reason, this batch was chosen for the central coordinate level of the adopted experimental factorial design.

3.2. Transesterification reactions

Transesterification reactions were performed as described in section 2.2. The choice of the factorial design presented in Table 1S aimed at adding variability to the monitored system and providing robust chemometric models. The use of different batches of soybean oil, the most used oil for biodiesel production, incorporated variations related to its physico-chemical parameters, due to climatic conditions, storage and transportation. The variation in the alcohol:oil molar ratio included some fluctuations associated with the biodiesel production process on an industrial scale, such as losses of alcohol by evaporation or small errors in the measurement of the volume of the reagents. Thus, soybean oils with different quality parameters were used as raw materials, and the molar ratio of methanol:oil was varied around 6:1 (from 5:1 to 7:1), which is the most industrially used ratio [50].

The conversion of triglycerides into intermediates and methyl esters occurs very quickly in the first ten minutes of the reaction. To monitor this initial variation in the system composition, two strategies were adopted: choosing a low reaction temperature (20 °C) and a higher sampling frequency at the beginning of the reaction. The lower temperature decreased the kinetics of the reaction and favored its monitoring. Thus, it was possible to incorporate into the chemometric models the variation in the composition occurring in the first moments of the

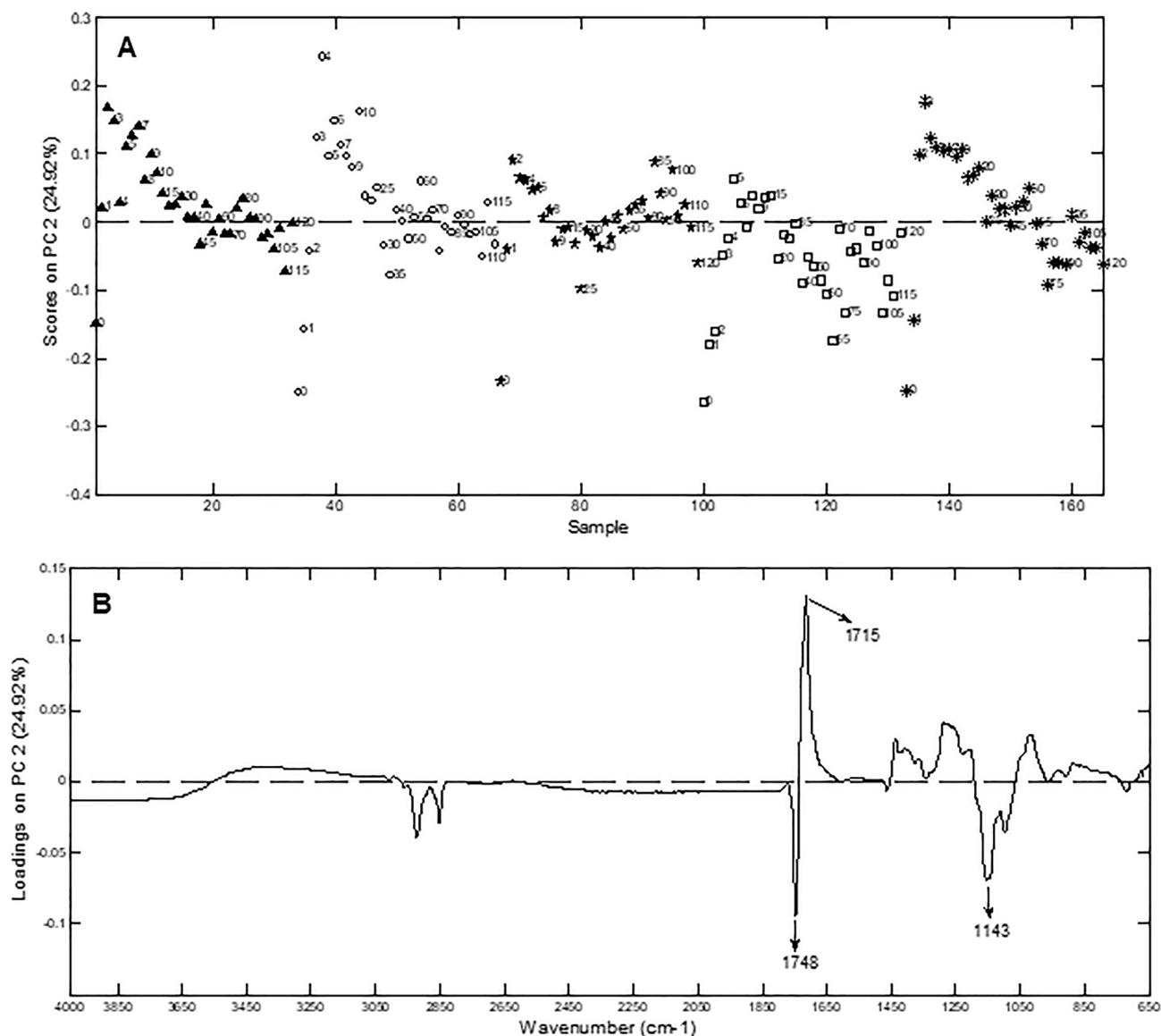


Fig. 4. PC2 (24.92%) (A) scores and (B) loadings obtained from the PCA model for the monitoring of the five reactions. Reactions 010–5 (full up triangles), 010–7 (empty circles), 020–5 (full stars), 020–7 (empty squares) and 100 (asterisks).

reaction. All samples collected from the reactions were analyzed using both the techniques, HPLC-DAD and ATR-FTIR. Fig. 1 shows the concentration profiles for the samples from reaction 010–7. The results for the five reactions were quite similar (Table 3S).

The concentration values for all the components were arranged in data matrices and used as inputs for the correlation constraint in MCR-ALS models. In these matrices, each row corresponds to a sample and each column to one of the four monitored components.

The biodiesel obtained at the end of each reaction was separated, washed with warm water, dried with anhydrous sodium sulfate, and analyzed. Table 3 shows the concentrations of fatty acids and monoglycerides (FA + MG), methyl esters (ME), diglycerides (DG), and triglycerides (TG) for the purified biodiesel samples obtained in the five reactions. The HPLC reference method is not able to separate FA and MG since they eluted in the same interval (3.0–4.5 min) (Fig. 1S, Supplemental materials). Therefore, they will be chemometrically modeled as only one component, not affecting the goals of this study. The highest yield of methyl esters was achieved for the biodiesel produced in the reaction 020–7. The soybean oil used in this reaction has the best quality among all the oils, as already previously discussed (section 3.1). Another factor that may have improved the conversion in this reaction was the

molar ratio methanol:oil of 7:1. The greater excess of methanol can shift the reaction equilibrium to the production of methyl esters.

3.3. Exploratory PCA model

Fig. 2 shows ATR-FTIR spectra of 33 samples analyzed for one of the reactions (010–7). A PCA model was built with ATR-FTIR spectra obtained for 165 samples, including all the five reactions. A model built with five PC accounted for 97.19% of the total data variance. PC1 accounted for 61.18% of variance and was interpreted as a component describing the progress of the transesterification reaction. Fig. 3A shows the scores of PC1 as a function of the samples arranged in the sequence of the five reactions. As can be observed in this plot, the scores for each reaction presented similar profiles, suggesting that the products of the transesterification were the main source of variability in the system. For each reaction, scores values decreased with the progress of the reaction varying from positive to negative values. It should be stressed that this variation is arbitrary and only makes sense for spectral interpretation by the comparison of these scores with the respective PC loadings, whose positive or negative values will be related to wavenumbers' intensities decreasing or increasing during the progress of the reaction,

Table 4
Parameters estimated for MCR-ALS models.

Parameters	Reaction code				Augmented matrix	
	020-5	020-7	010-7	100		
% Explained variance	99.8984	99.9281		99.9473	99.7842	
% LOF	3.1882	2.6811		2.2959	4.6451	
Standard deviation of residuals	0.001065	0.000916	99.9125	0.000790	0.001577	
FA + MG	RMSEC (%)	1.7	0.8	2.9577	2.1	–
	R	0.8054	0.8686	0.001000	0.8445	–
ME	RMSEP (%)	2.15	1.03	0.68	2.08	–
	RMSEC (%)	5.4	4.7	4.1	3.0	5.5
DG	R	0.9605	0.9826	0.9835	0.9886	0.9620
	RMSEP (%)	6.3	4.4	4.9	4.1	4.1
TG	RMSEC (%)	11.2	6.1	2.1	1.6	–
	R	0.4221	0.6203	0.9336	0.9622	–
R _{rc} ME	RMSEP (%)	11.4	4.9	2.2	1.3	–
	R	0.9633	0.9966	0.9862	0.9939	0.9780
R _{rc} TG	RMSEP (%)	5.2	3.7	3.6	1.9	2.8
	R	0.9910	0.9947	0.9857	0.9854	0.9905
		0.9859	0.9893	0.9973	0.9877	0.9862

FA + MG: fatty acids and monoglycerides; ME: methyl esters; DG: diglycerides; TG: triglycerides; R_{rc}: correlation coefficient (R) between the reference and the estimated spectral profiles.

respectively.

The observation of the PC1 loadings plot allows realizing that the spectral variation associated with the progress of the transesterification is fully contained in the region between 650 and 1850 cm^{-1} (Fig. 3B). The most negative loadings peaks are observed at 1741 cm^{-1} , 1436 cm^{-1} , and 1197 cm^{-1} . The absorption bands between 1425 and 1447 cm^{-1} and 1188–1200 cm^{-1} presented small differences when the spectra of soybean oil and its respective biodiesel are compared [51]. Nevertheless, the intensities of these spectral bands increased with the reaction progress as can be interpreted from PC1. The strong absorption band observed between 1750 and 1730 cm^{-1} showed more intensely this trend, and was assigned to the stretching vibration of ester carbonyl bonds of methyl esters [52]. The most positive loadings are present at 1030 cm^{-1} , 1098 cm^{-1} , and 1143 cm^{-1} . This is in accordance with the literature, in which the disappearance of spectral bands centered at 1026 cm^{-1} and 1095 cm^{-1} during the transesterification has been reported, with the consequent appearance of new bands at 1435 cm^{-1} and 1195 cm^{-1} , indicating the conversion of soybean oil into biodiesel [52].

PC2 accounted for 24.92% of the data variance. Samples corresponding to the initial reaction time ($t = 0$) presented more negative scores on PC2, although most of the other scores are distributed along the origin of this axis (Fig. 4A). This distinct behavior may be associated with the nature of the reagents. Soybean oil and methanol are partially miscible, requiring a certain stirring time for diffusion between the phases in order to effectively start the reaction. Fig. 4B shows that the most negative loadings are present as bands centered at 1748 cm^{-1} and 1143 cm^{-1} . In transesterification reactions, the presence of a spectral band at 1748 cm^{-1} has been associated with the presence of oil (raw material). As the oil converts to biodiesel, the most intense band shifts from 1748 cm^{-1} to 1741 cm^{-1} [52]. PC3, PC4, and PC5 accounted for small variances (5.88%, 3.74%, and 1.47%, respectively), and are related to samples showing non-significant particular behaviors.

3.4. Multivariate calibration MCR-ALS models

3.4.1. Models for individual reactions

After performing MSC preprocessing of spectral datasets, specific MCR-ALS multivariate calibration models were built for each reaction. Each model was built with four components from a total of 33 samples that were split in 22 for the calibration set and 11 for the validation set. Validation samples were selected in equally spaced intervals through the whole time reaction range.

Non-negativity and correlation constraints were applied to all four

components. The convergence criterion was achieved, except for reaction 010-5. For this reaction, the divergence was assumed after 20 consecutive iterations without decreasing the percentage of LOF. Thus, spectral data of reaction 010-5 were used only in the construction of the augmented matrix model. In general, all models showed a high percentage of explained variance and a low percentage of LOF (Table 4). These results corroborate that the experimental data were well resolved by the MCR-ALS method.

Recovered concentration profiles for all components were very similar to the HPLC profiles shown in Fig. 1. Values of figures of merit for model accuracy, RMSEC (root mean square error of calibration) and RMSEP (root mean square error of prediction), for the quantification of ME and TG were considered satisfactory, indicating that the models can be applied to monitor the production of biodiesel. The reference values and the calculated concentration profiles for the calibration set were compared using the correlation coefficient (R). Correlation coefficients for ME and TG were always higher than 0.96. The concentrations of FA + MG, and DG presented small variations during the reactions (Table 3S). Consequently, MCR-ALS calibration models presented less accurate predictions for these two components, resulting in RMSEC and RMSEP higher than those for ME and TG.

Spectral profiles recovered by MCR-ALS for each component were very similar among all the reactions/models. Fig. 5 shows estimated and reference spectra for TG (Fig. 5A) and ME (Fig. 5B) for reaction 010-7. Spectra of soybean oil and purified biodiesel were used as reference spectra for these two components. Correlation coefficients between the reference spectra and those estimated by MCR-ALS were higher than 0.985, corroborating model accuracy. MCR-ALS estimated spectra for FA + MG, and DG are shown in Fig. 5C. Reference spectra for these intermediates were not available since these components are produced and consumed during the transesterification reaction.

It can be observed in all spectra of Fig. 5 the presence of a strong absorption band characteristic of the carbonyl group ($\nu_{\text{C}} = \text{O}$) of esters around 1750–1730 cm^{-1} . Stretching vibrations of CH_3 , CH_2 , and CH are present at 2980–2950 cm^{-1} , 2950–2850 cm^{-1} and 3050–3000 cm^{-1} , while bending vibrations of these same groups are present at 1475–1350 cm^{-1} , 1350–1150 cm^{-1} and 722 cm^{-1} , respectively [52]. Moreover, a large band around 3500 cm^{-1} , present only in the spectra of Fig. 5C, can be assigned to the stretching of O–H bonds present in FA, MG, and DG. This band is not observed in ATR-FTIR spectra of TG and ME, since these molecules do not have hydroxyl groups in their structure.

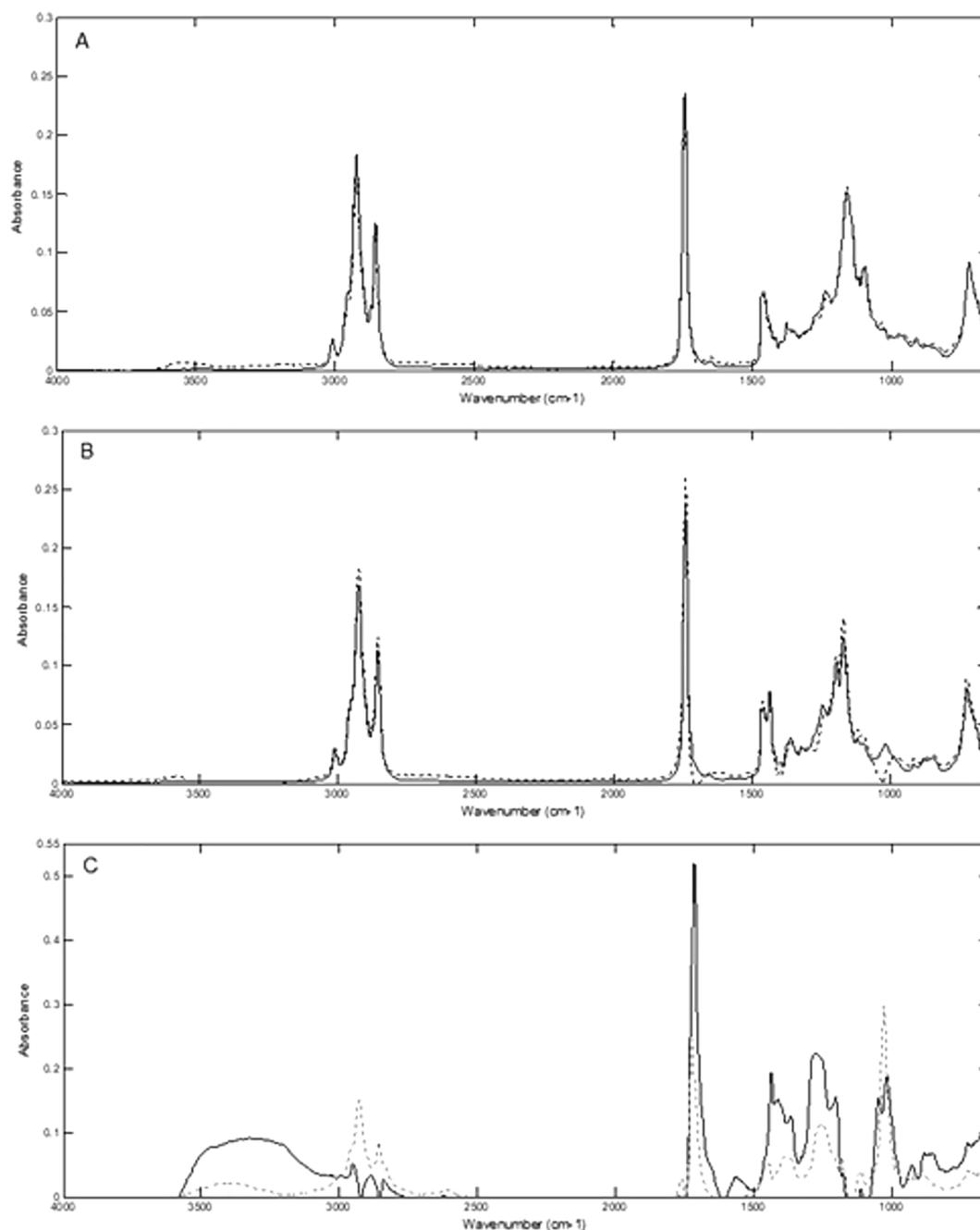


Fig. 5. Reference (solid lines) and estimated/recovered by MCR-ALS (dashed lines) ATR-FTIR spectra for the components of reaction 010–7. (A) Triglycerides. (B) Methyl esters. (C) ATR-FTIR spectra estimated/recovered by MCR-ALS for fatty acids and monoglycerides (solid lines) and for diglycerides (dashed lines).

3.4.2. Augmented matrix model

Spectral data of the five reactions were aligned in the direction of the columns forming an augmented matrix (162x3351). An initial attempt to build a model applying the correlation constraint to all the four components did not reach the convergence criterion. Thus, a new model was built applying the correlation constraint only to ME and TG. Spectra of four reactions (010–5, 010–7, 020–5 and 020–7) were used in the calibration set of this MCR-ALS model, which was built with four components and accounted for 99.78% of the data variance and presented a LOF of 4.65% (Table 4). Samples of the reaction 100 were used as the validation set. Fig. 6 shows a plot of the HPLC reference values ($C_{\text{reference}}$) versus concentration values predicted by the model (C_{ALS}). The ability of this model to monitor the conversion of triglycerides in methyl esters was verified by estimating proper figures of merit. High

correlation coefficients between reference and estimated values (0.962 for ME and 0.978 for TG) indicated good linearity, while RMSEC (5.5% for ME and 3.6% for TG) and RMSEP (4.1% for ME and 2.8% for TG) values indicated good accuracy. Individual relative errors for each sample were lower than 10%. Finally, reference and MCR-ALS calculated spectra also showed high correlation.

The developed model constructed with the augmented matrix included the contribution of the intermediates of the transesterification reaction as components of the system, although they were not quantified in this case. Once this method is validated, its main advantage is the possibility of determining the concentrations of the raw material and the product of the biodiesel synthesis using only ATR-FTIR spectra. It can be applied to any other batch of transesterification reaction carried out in this range of experimental conditions.

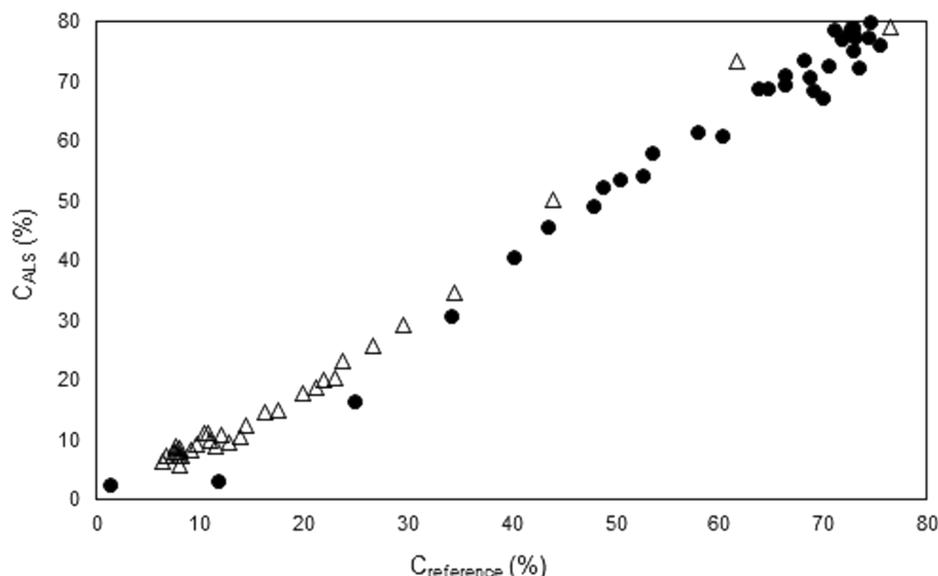


Fig. 6. Results obtained with the MCR-ALS model built with the augmented matrix for the predictions of the reaction 100. Estimated (C_{ALS}) versus reference ($C_{reference}$) concentration values for methyl esters (full circles) and triglycerides (empty up triangles).

4. Conclusions

MCR-ALS models have been mostly restricted to qualitative applications in the past, but the recently introduced correlation constraint has increased the scope of quantitative curve resolution methods. In this work, MCR-ALS method with the correlation constraint was applied to the ATR-FTIR quantitative monitoring and modelling of transesterification reactions for biodiesel production. Five reactions were carried out based on an experimental design aiming to build robust quantitative models. The raw material (triglycerides) and the product (methyl esters) of the reactions were qualitatively and quantitatively modeled in a wide analytical range utilizing a HPLC method for obtaining reference values. The intermediate components (diglycerides, and the mixture of fatty acids and monoglycerides) were also modeled based on HPLC reference values, providing semi-quantitative results. Individual MCR-ALS models were built for each reaction. Nevertheless, the most useful MCR-ALS model was obtained using an augmented matrix built with samples from four reactions. This model was successfully applied to monitor independent samples from another reaction, providing accurate and linear predictions. MCR-ALS modeling also allowed to recover spectral profiles that matched the reference spectra for individual components of the reaction, corroborating the selectivity of the method. A PCA model was also built as a preliminary tool for the spectral interpretation of the system.

The developed method was more rapid, simple, of lower cost and environmentally friendly in relation to the traditional GC or HPLC alternatives to monitor biodiesel production. In comparison to other works in the literature using NIR spectroscopy and multivariate calibration with PLS [16,17], our model presented better or similar linearity and worse accuracy (RMSEP), though those NIR methods have monitored narrower ranges for ME and TG. In addition, MCR-ALS has the already mentioned advantage of allowing to recover the pure spectra of all the components, providing a better qualitative monitoring of the reaction system. Finally, the proposed analytical strategy can lead to several advantages in the industrial context, such as the possibility of faster intervention of the analyst to readjust the parameters of the process.

CRedit authorship contribution statement

Werônica Lima Furtado: Conceptualization, Investigation,

Methodology, Formal analysis, Validation, Visualization, Data curation, Writing - original draft. **Camila Nunes Costa Corgozinho:** Conceptualization, Methodology, Resources, Supervision, Writing - original draft, Project administration. **Romá Tauler:** Software, Writing - review & editing. **Marcelo Martins Sena:** Conceptualization, Methodology, Resources, Data curation, Supervision, Project administration, Funding acquisition, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We gratefully acknowledge the Brazilian agency Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for the financial support and a Master's scholarship (WLF), and Laboratório de Ensaios de Combustíveis (LEC/UFMG) for the structural analytical support.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.fuel.2020.119275>.

References

- [1] Singh D, Sharma D, Soni SL, Sharma S, Kumar Sharma P, Jhalani A. A review on feedstocks, production processes, and yield for different generations of biodiesel. *Fuel* 2020;262:116553. <https://doi.org/10.1016/j.fuel.2019.116553>.
- [2] Moiola E, Salvati F, Chiesa M, Siecha RT, Manenti F, Laio F, Rulli MC. Analysis of the current world biofuel production under a water–food–energy nexus perspective. *Adv Water Resour* 2018;121:22–31.
- [3] Brazilian law 13263. Brasília: Presidência da República do Brasil; 2016. http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2016/lei/L13263.htm (accessed in July, 2020).
- [4] Murugesan A, Umarani C, Chinnusamy TR, Krishnan M, Subramanian R, Neduzchezain N. Production and analysis of bio-diesel from non-edible oils—A review. *Renew Sustain Energy Rev* 2009;13(4):825–34.
- [5] ANP. Regulation 45. Brasília: Agência Nacional de Petróleo, Gás Natural e Biocombustíveis; 2014. <http://legislacao.anp.gov.br/?path=legislacao-anp/resol-anp/2014/agosto&item=ranp-45-2014> (accessed in July, 2020).
- [6] Lôbo I, Ferreira S, Cruz R. Biodiesel: quality parameters and analytical methods. *Quim Nova* 2009;32:1596–608. doi:10.1590/S0100-40422009000600044.

- [7] ABNT. NBR 15764: Determination of total esters content by gas chromatography. Rio de Janeiro: Associação Brasileira de Normas Técnicas; 2015:13.
- [8] Leung DY, Wu X, Leung MKH. A review on biodiesel production using catalyzed transesterification. *Appl Energy* 2010;87(4):1083–95.
- [9] CEN. EN-14103: Determination of ester and linolenic acid methyl ester contents. Brussels: European Committee for Standardization; 2011:18.
- [10] CEN. EN 14105: Fat and oil derivatives. Fatty acid methyl esters (FAME). Determination of free and total glycerol and mono-, di-, triglyceride contents. Brussels: European Committee for Standardization; 2011:26.
- [11] ASTM. D6584: Determination of total monoglycerides, total diglycerides, total triglycerides, and free and total glycerin in B-100 biodiesel methyl esters by gas chromatography. West Conshohocken: American Society for Testing and Materials; 2017:9.
- [12] ABNT. NBR 15342: Determination of monoacylglycerols and diacylglycerols content of castor oil biodiesel by gas chromatography. Rio de Janeiro: Associação Brasileira de Normas Técnicas; 2016:7.
- [13] ABNT. NBR 15908: Determination of free glycerin, monoglycerides, diglycerides, triglycerides and total glycerin by gas chromatography. Rio de Janeiro: Associação Brasileira de Normas Técnicas; 2012:19.
- [14] Jin F, Kawasaki K, Kishida H, Tohji K, Moriya T, Enomoto H. NMR spectroscopic study on methanolysis reaction of vegetable oil. *Fuel* 2007;86(7-8):1201–7. <https://doi.org/10.1016/j.fuel.2006.10.013>.
- [15] Meher L, Vidyasagar D, Naik S. Technical aspects of biodiesel production by transesterification—a review. *Renew Sustain Energy Rev* 2006;10(3):248–68.
- [16] Pinzi S, Alonso F, García Olmo J, Pilar Dorado M. Near infrared reflectance spectroscopy and multivariate analysis to monitor reaction products during biodiesel production. *Fuel* 2012;92(1):354–9.
- [17] de Lima SM, Silva BFA, Pontes DV, Pereira CF, Stragevitch L, Pimentel MF. In-line monitoring of the transesterification reactions for biodiesel production using NIR spectroscopy. *Fuel* 2014;115:46–53.
- [18] Sales RF, Vitale R, de Lima SM, Pimentel MF, Stragevitch L, Ferrer A. Multivariate statistical process control charts for batch monitoring of transesterification reactions for biodiesel production based on near-infrared spectroscopy. *Comput Chem Eng* 2016;94:343–53.
- [19] Fontalvo-Gómez M, Colucci JA, Velez N, Romañach RJ. In-Line Near-Infrared (NIR) and Raman Spectroscopy Coupled with Principal Component Analysis (PCA) for in Situ Evaluation of the Transesterification Reaction. *Appl Spectrosc* 2013;67(10):1142–9.
- [20] Tauler R. Multivariate curve resolution applied to second order data. *Chemometrics and Intelligent Laboratory Systems* 1995;30(1):133–46.
- [21] de Juan A, Tauler R. Chemometrics applied to unravel multicomponent processes and mixtures. *Anal Chim Acta* 2003;500(1-2):195–210.
- [22] de Juan A, Tauler R. Multivariate Curve Resolution (MCR) from 2000: Progress in Concepts and Applications. *Crit Rev Anal Chem* 2006;36(3-4):163–76.
- [23] Ruckebusch C, Blanchet L. Multivariate curve resolution: A review of advanced and tailored applications and challenges. *Anal Chim Acta* 2013;765:28–36.
- [24] Tauler R. Calculation of maximum and minimum band boundaries of feasible solutions for species profiles obtained by multivariate curve resolution. *J. Chemometrics* 2001;15(8):627–46.
- [25] Golshan A, Abdollahi H, Beyramysoltan S, Maeder M, Neymeyr K, Rajkó R, Sawall M, Tauler R. A review of recent methods for the determination of ranges of feasible solutions resulting from soft modelling analyses of multivariate data. *Anal Chim Acta* 2016;911:1–13.
- [26] Azzouz T, Tauler R. Application of multivariate curve resolution alternating least squares (MCR-ALS) to the quantitative analysis of pharmaceutical and agricultural samples. *Talanta* 2008;74(5):1201–10.
- [27] de Oliveira RR, de Lima KMG, Tauler R, de Juan A. Application of correlation constrained multivariate curve resolution alternating least-squares methods for determination of compounds of interest in biodiesel blends using NIR and UV–visible spectroscopic data. *Talanta* 2014;125:233–41.
- [28] Neves AC, Tauler R, Lima KM. Area correlation constraint for the MCR–ALS quantification of cholesterol using EEM fluorescence data: A new approach. *Anal Chim Acta* 2016;937:21–8.
- [29] Blanco M, Castillo M, Peinado A, Beneyto R. Application of Multivariate Curve Resolution to Chemical Process Control of an Esterification Reaction Monitored by Near-Infrared Spectroscopy. *Appl Spectrosc* 2006;60(6):641–7.
- [30] Tauler R, Kowalski BR, Fleming S. Multivariate curve resolution applied to spectral data from multiple runs of an industrial process. *Anal. Chem.* 1993;65(15):2040–7.
- [31] Mogollon NGS, Ribeiro FAL, Lopez MM, Hantao LW, Poppi RJ, Augusto F. Quantitative analysis of biodiesel in blends of biodiesel and conventional diesel by comprehensive two-dimensional gas chromatography and multivariate curve resolution. *Anal Chim Acta* 2013;796:130–6.
- [32] Dantas WFC, Alves JCL, Poppi RJ. MCR-ALS with correlation constraint and Raman spectroscopy for identification and quantification of biofuels and adulterants in petroleum diesel. *Chemometrics and Intelligent Laboratory Systems* 2017;169:116–21.
- [33] Câmara ABF, Carvalho LS, Moraes CLM, Lima LAS, Araújo HOM, Oliveira FM, et al. MCR-ALS and PLS coupled to NIR/MIR spectroscopies for quantification and identification of adulterant in biodiesel–diesel blends. *Fuel* 2017;210:497–506.
- [34] Rouchi MB, Khorrami MK, Garmarudi AB, de la Guardia M. Application of infrared spectroscopy as Process Analytics Technology (PAT) approach in biodiesel production process utilizing Multivariate Curve Resolution Alternative Least Square (MCR-ALS). *Spectrochim Acta Part A Mol Biomol Spectrosc* 2019;213:347–53.
- [35] Sales RF, Lima SM, Stragevitch L, Pimentel MF, de Juan A. NIR Monitoring and modelling of soybean oil methanolysis with multivariate curve resolution-alternating least squares with correlation constraint. *J Braz Chem Soc* 2017;28:695–706. <https://doi.org/10.21577/0103-5053.20160218>.
- [36] ASTM. D4052–18a: Standard test method for density, relative density, and API gravity of liquids by digital density meter. West Conshohocken: American Society for Testing and Materials; 2018.
- [37] ASTM. D6304–16e1: Standard test method for determination of water in petroleum products, lubricating oils, and additives by coulometric Karl Fisher titration. West Conshohocken: American Society for Testing and Materials; 2016.
- [38] ASTM. D445–19a: Standard test method for kinematic viscosity of transparent and opaque liquids (and calculation of dynamic viscosity). West Conshohocken: American Society for Testing and Materials; 2019.
- [39] CEN. EN 14112: Fat and oil derivatives. Fatty acid methyl esters (FAME). Determination of oxidation stability (accelerated oxidation test). Brussels: European Committee for Standardization; 2016.
- [40] ASTM. D664–18e2: Standard test method for acid number of petroleum products by potentiometric titration. West Conshohocken: American Society for Testing and Materials; 2018.
- [41] Instituto Adolfo Lutz. Métodos Físico-Químicos para Análise de Alimentos. 4th ed. Zenebon O, Pascuet, NS, Tiglea, P (Coords.). São Paulo: IMESP; 2008.
- [42] Holcapek M, Jandera P, Fischer J, Prokeš B. Analytical monitoring of the production of biodiesel by high-performance liquid chromatography with various detection methods. *J Chromatogr A* 1999;858(1):13–31.
- [43] Rinnan A, van den Berg F, Engelsen SB. Review of the most common pre-processing techniques for near-infrared spectra. *TrAC, Trends Anal Chem* 2009;28(10):1201–22. <https://doi.org/10.1016/j.trac.2009.07.007>.
- [44] Bro R, Smilde AK. Principal component analysis. *Anal. Methods* 2014;6(9):2812–31.
- [45] Savi EL, Herculano LS, Lukasiewicz GVB, Torquato AS, Baesso ML, Astrath NGC, et al. Evaluation of thermo-oxidative stability of biodiesel. *Energy Fuels* 2017;31:7132–7. doi: 10.1021/acs.energyfuels.7b00696.
- [46] Jaumot J, de Juan A, Tauler R. MCR-ALS GUI 2.0: New features and applications. *Chemometrics and Intelligent Laboratory Systems* 2015;140:1–12.
- [47] Windig W. Spectral data files for self-modeling curve resolution with examples using the Simplisma approach. *Chemometrics and Intelligent Laboratory Systems* 1997;36(1):3–16.
- [48] Li Y, Qiu F, Yang D, Li X, Sun P. Preparation, characterization and application of heterogeneous solid base catalyst for biodiesel production from soybean oil. *Biomass Bioenergy* 2011;35(7):2787–95.
- [49] Lee AF, Bennett JA, Manayil JC, Wilson K. Heterogeneous catalysis for sustainable biodiesel production via esterification and transesterification. *Chem Soc Rev* 2014;43(22):7887–916.
- [50] Sun H, Duan J, Chen P, Lou H, Zheng X. Room temperature transesterification of soybean oil to biodiesel catalyzed by rod-like C₆₀SiO_x-2 solid base. *Catal Commun* 2011;12(11):1005–8.
- [51] Zhang WB. Review on analysis of biodiesel with infrared spectroscopy. *Renew Sustain Energy Rev* 2012;16(8):6048–58.
- [52] Tariq M, Ali S, Ahmad F, Ahmad M, Zafar M, Khalid N, et al. Identification, FT-IR, NMR (¹H and ¹³C) and GC/MS studies of fatty acid methyl esters in biodiesel from rocket seed oil. *Fuel Process Technol* 2011;92(3):336–41.