

Document downloaded from:

<http://hdl.handle.net/10251/102085>

This paper must be cited as:



The final publication is available at

<http://dx.doi.org/10.1016/j.compchemeng.2016.08.013>

Copyright Elsevier

Additional Information

Multivariate statistical process control charts for batch monitoring of transesterification reactions for biodiesel production based on near-infrared spectroscopy

Rafaella Figueiredo Sales^a, Raffaele Vitale^b, Suzana Moreira de Lima^c, Maria Fernanda Pimentel^{a*}, Luiz Stragevitch^a, Alberto Ferrer^b.

^aDepartamento de Engenharia Química, Universidade Federal de Pernambuco, Avenida Prof. Arthur de Sá s/n, Cidade Universitária, 50740-521 Recife, PE, Brazil.

^bGrupo de Ingeniería Estadística Multivariante, Departamento de Estadística e Investigación Operativa Aplicadas y Calidad, Universitat Politècnica de València, Camino de Vera s/n, 46022, Valencia, Spain.

^cInstituto Federal de Pernambuco, Av. Prof. Luís Freire, 500 - Cidade Universitária, Recife, PE, Brazil.

* Corresponding author. Tel.: +55(81) 21267235. E-mail address: mfp@ufpe.br or mfernanda.pimentel@gmail.com (Maria Fernanda Pimentel).

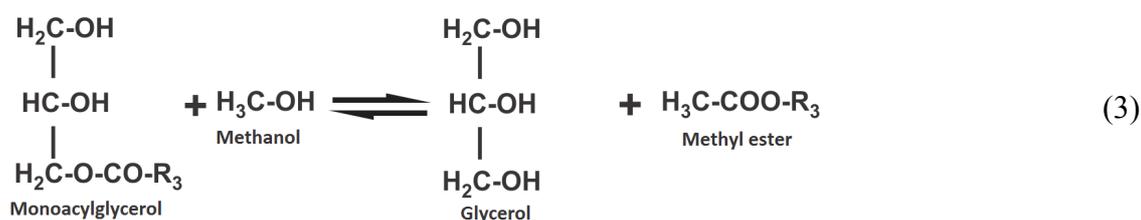
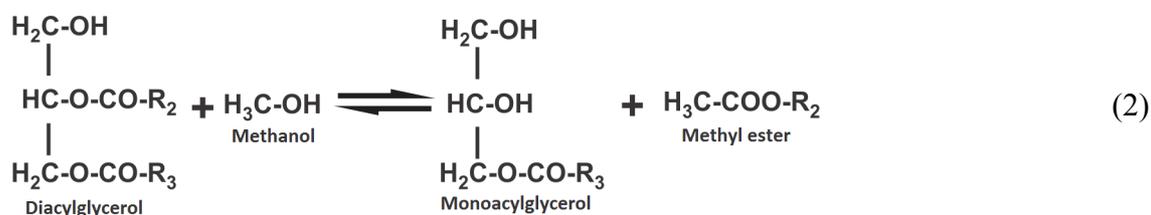
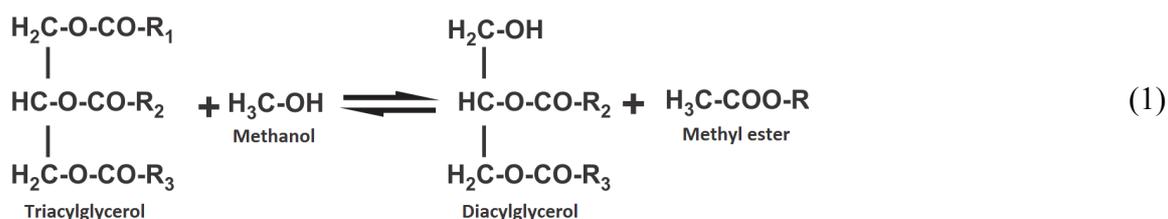
Abstract: This work describes an application of Multivariate Statistical Process Control to monitor soybean oil transesterification. For the development of multivariate control charts, near infrared spectra were acquired in-line during the evolution of ten batches under Normal Operating Conditions. They were then organized in a three-way array (batch×spectral variable×time). This structure was analysed by the two most commonly used approaches to develop batch monitoring schemes for handling such kind of data, referred to as Nomikos-MacGregor (NM) and Wold-Kettaneh-Friden-Holmberg (WKFH), respectively. To assess the performance of the approaches, eight test batches, during which specific interferences were induced, were manufactured. When applied for off-line monitoring, both NM and WKFH correctly pointed out such intentionally produced failures. On the other hand, concerning on-line monitoring, NM exhibited a better fault detection capability than WKFH. Contribution plots were found to highlight the spectral region mostly affected by the disturbances regardless of the modelling strategy resorted to.

Keywords: biodiesel; batch process monitoring; Near Infrared Spectroscopy (NIRS); Multivariate Statistical Process Control (MSPC); soybean oil methanolysis.

1. Introduction

Research into biofuels, which can be used as alternative energy sources to fossil fuels, has recently been encouraged by concerns about levels of pollution, environmental safety and oil supply. Among biofuels, biodiesel is a potential substitute for petroleum fuels, showing also high biodegradability, low toxicity and capacity to low emission of sulfates, aromatic compounds and other hazardous products. Moreover, it has replaced mineral diesel in numerous applications, e.g. in combustion engines (Arias et al., 2012; Atabani et al., 2012).

In general, biodiesel, composed of fatty acid alkyl esters, is produced by transesterification of triacylglycerols with an alcohol, most commonly methanol. Renewable raw materials such as animal fats and vegetable oils are frequently used as feedstock. The methanol-mediated transesterification reaction (methanolysis) proceeds via three reversible and consecutive steps: diacylglycerols and methyl esters are obtained from triacylglycerols (Eq. 1); monoacylglycerols and methyl esters are obtained from diacylglycerols (Eq. 2); monoacylglycerols are converted to methyl esters and glycerol (Eq. 3). Their reversibility motivates the use of methanol in excess with respect to the initial amount of triacylglycerols. In addition, a catalyst is always resorted to for ensuring the rapid achievement of the maximum conversion rate (Tubino et al., 2014).



At the beginning of the methanolysis process, the reaction system consists of two highly immiscible phases: the oil phase, rich in triacylglycerols, and the methanol phase. With the progress of the reaction, methyl ester formation increases the solubility of triacylglycerols in methanol as well as the overall mass transfer rate, intensifying the reactant mixing. After few minutes, the mixture appears as an emulsion in which small droplets of triacylglycerols are in the methanol phase. In this step, an acceleration in the methyl ester production can be observed (Csernica and Hsu, 2012). Afterwards, the reaction reaches the equilibrium. In general, transesterification reactions present a fast conversion and, under certain conditions, e.g. high temperatures, the initial stage is shortened and the acceleration in the methyl ester production is achieved sooner (Tubino et al., 2014).

Transesterification reactions for biodiesel production are usually carried out in batch reactors. The process variables mostly affecting these reactions are temperature, catalyst concentration and type, stirring speed, reactant purity, feedstock variability and free fatty acid content (Arias et al., 2012; Atabani et al., 2012; Csernica and Hsu, 2012), whose setting and monitoring is of utmost importance to ensure biodiesel quality uniformity and fulfil its commercialization requirements (Killner et al., 2011).

A number of analytical techniques like chromatography are commonly applied for monitoring transesterification reactions. However, these methods are not suitable for real time process monitoring, which could be extremely useful in such circumstances owing to the usually observed fast conversion rate (Torrissi and Sabino, 2013; Tubino et al., 2014). Due to the recent development of new analytical technologies, a high number of tools can be exploited for process monitoring purposes. Among them, Near Infrared Spectroscopy (NIRS) has a wide applicability. It is commonly used for in-line, at-line, on-line or off-line process parameter control and product attribute determination. Moreover, NIRS is a rapid and non-destructive technique and does not require labour-intensive sample treatments, saving time and effort. Consequently, it can play an important role, especially for real-time applications (Huang and Qu, 2011; Kona et al., 2013).

NIRS is a well-established spectroscopic technique that can provide qualitative and quantitative information for a wide range of organic compounds. NIR spectra show bands (mainly overtones and combinations bands) related to transitions between the levels of vibrational energy of a molecule. The information present in the NIR spectra can be used to determine the composition of either a sample or a bulk or its physical

properties (such as density, viscosity, refractive index, particle size, temperature, homogeneity *etc.*) when these have some effect on the intensity and/or on the shape of the recorded signal (Pasquini, 2003; Shao et al., 2010; Workman and Weyer, 2007). Such features have recently motivated the application of this technique for process monitoring purposes in different fields of application (Li et al., 2016; Lima et al., 2014; Richard et al., 2013; Killner et al., 2011; Kona et al., 2013).

Several studies describing the use of NIRS in the biodiesel field have recently been published. Research works have demonstrated its effectiveness for the development of multivariate calibration models and the estimation of biodiesel content in diesel fuel blends (Alves and Poppi, 2013; Pimentel et al., 2006) or to determine specific chemical products throughout the biodiesel production (Lima et al., 2014; Richard et al., 2011; Richard et al., 2013). Applications of NIRS to predict biodiesel and diesel-biodiesel blend properties, such as oxidative stability and density, and quality have also been described (Lira et al., 2010; Oliveira et al., 2009).

Concerning the application of NIRS for on- and in-line monitoring of biodiesel production, this technique is well established when associated with Partial Least Squares (PLS) regression models to determine the composition of the reaction mixture (Killner et al., 2011; Lima et al., 2014; Richard et al., 2011; Richard et al., 2013). In general, such studies satisfactorily described the use of PLS calibration models to estimate the content of alkyl ester and glycerides as well as the conversion ratio of glycerides during transesterification reactions.

Apart from biodiesel property determination, NIRS can also be applied in combination with Multivariate Statistical Process Control (MSPC) methods based on projection techniques for monitoring both continuous and batch processes. However, as far as the authors are concerned, no application of batch MSPC approaches to the biodiesel field is detailed in the scientific literature. For this reason, in the present work, methanol-mediated soybean oil transesterification was chosen as an example to illustrate the effectiveness of coupling NIRS and batch MSPC to monitor biodiesel production. NIRS-based multivariate control charts were developed applying the two most commonly used approaches for dealing with such kind of data, referred to as Nomikos-MacGregor (NM) and Wold-Kettaneh-Friden-Holmberg (WKFH), respectively, to a set of data collected during ten batches evolving under Normal Operating Conditions (NOC). Eight *out-of-control* runs, affected by disturbances in the temperature and stirring conditions as well as in the catalyst content, were manufactured

to assess the off-line and on-line fault detection capability of the resulting monitoring schemes. Classical contribution plots (Kourti and MacGregor, 1996) were additionally resorted to for enabling fault diagnosis, i.e. the identification of the nature of each failure.

Therefore, this work aims to explore the potential of NIRS associated with batch MSPC to establish monitoring schemes for biodiesel production. As mentioned above, an advantage of this method is its fast measurements that allows real time process monitoring, which is meaningfully important to the industrial production. In addition, it provides chemical and physical information about the system under study. In this way, the present work presents a comprehensive comparison for off-line and on-line process monitoring of NM and WKFH approaches. To the best of the author's knowledge, this kind of comparison was only presented by Aguado et al. (2007). Nevertheless, the authors described the development of control charts based on process variables and applied them for a completely different application (wastewater treatment).

2. Multivariate statistical monitoring of batch processes

Batch processes show strong dynamic nature and high variability among different runs, thus their monitoring is essential to always guarantee on-specification products are manufactured. To address batch process monitoring by spectroscopic techniques like NIRS, projection methods, namely Principal Component Analysis (PCA) or Partial Least Squares regression (PLS), are required. These multivariate statistical tools project the information carried by the extremely correlated spectral channels onto low-dimensional subspaces defined by few orthogonal Latent Variables (LVs), easing the analysis of very large, ill-conditioned datasets. PCA works by defining a small set of uncorrelated factors (principal components), which are linear combination of the original variables and explain most of their variability. On the other hand, PLS addresses this dimensionality reduction by maximising the covariance between the original data matrix and an array containing the values of particular response variables (Bersimis et al., 2007; Ferrer, 2007; Kerkhof et al., 2012; Kourti, 2003; Li et al, 2016).

Constructing a batch process monitoring scheme can be specifically looked at as a three-step procedure: i) collecting and modelling data associated to NOC; ii) detecting possible faults (by checking whether new observations are consistent with the NOC

data); iii) diagnosing detected faults that is identifying which ones of the measured variables are responsible for them (by tools like contribution plots). In this respect, it is important to highlight that monitoring schemes are quite different from classical control schemes, in which the process is regulated to automatically recover NOC (Nomikos and MacGregor, 1995; Wold et al., 1998).

In general, batch process data can be organized in a three-dimensional array with dimensions ($I \times J \times K$), where I equals the number of analysed batches, during which J variables are measured at K time points. In order to develop PCA- or PLS-based multivariate control charts, first, such three-way array needs to be unfolded into a matrix. Among the various techniques to develop batch monitoring schemes, Nomikos-MacGregor (NM) (Nomikos and MacGregor, 1995) and Wold-Kettaneh-Friden-Holmberg (WKFH) (Wold et al., 1998) approaches are extensively applied and available in commercial software packages. The former unfolds the data batch-wise, so that each row of the resulting matrix contains all the information related to one batch (Fig. 1a). The final $I \times JK$ two-way array is then centred and subjected to PCA. NM permits to capture both the linear dynamics of the analysed process and possible changing relationships among process variables. Furthermore, data centring removes the strongly non-linear trend of their evolution, which allows their variation about their average trajectory to be modelled, reducing the process analysis to a stationary problem.

On the other hand, WKFH unfolds the data variable-wise, by rearranging the single sub-matrices associated to the evolution of each batch preserving the variable direction. WKFH leads to a new array with dimensions $IK \times J$ (Fig. 1b). Such array is then centred and used to build a PLS regression model with a single response variable: the *local batch time*, a measure of the batch maturity index. At this step, WKFH does not take into account the dynamics of the process under study and has been found to be a valid strategy only when its correlation structure is more or less constant along the batch run. Data centring does not remove the average variable trajectories, thus differences among batches are initially overlooked. In a second modelling stage, the *batch level*, the resulting PLS-scores are unfolded batch-wise, re-centred and analysed by PCA. A good critical discussion on these methods can be found in Aguado et al. (2007), Camacho et al. (2008) and Wold et al. (2009).

INSERT FIGURE 1

NM and WKFH can be used for both off-line and on-line monitoring of batch processes. Concerning the former, once NOC models have been constructed, new complete batches are classified as normal or faulty. For their methodological principles, NM can directly address this task, while WKFH requires the aforementioned batch level, which may be particularly useful in case huge numbers of variables have to be preliminarily compressed for being subsequently analysed (e.g. in spectroscopic or chromatographic case-studies). Regarding on-line monitoring, the evolution of new batches is assessed in real time, allowing abnormal behaviours to be possibly corrected before the batch completion, thus preventing the manufacturing of out of specification products. Here, NM needs at each consecutive time instant the imputation of future unknown observations to always deal with complete data matrices. This imputation can be carried out in many distinct ways. For the purposes of this article, the *Single Component Projection* (SCP) approach proposed in Nelson et al. (1996) was appealed to. Conversely, WKFH does not necessitate any imputation (Wold et al., 2009). However, this implies assuming all the batches under study feature the same evolution pace, which is rather uncommon in practical cases. In such circumstances, WKFH may exhibit the critical limitations already comprehensively detailed in González-Martínez et al. (2014).

3. Materials and methods

3.1 Transesterification reactions

Transesterifications were carried out in a Mettler Toledo OptiMax synthesis workstation, provided with a 1000 mL glass reactor. The OptiMax system enabled an accurate control of the agitation speed and reaction temperature, which also allowed specific disturbances to be properly induced. Fig. 2 shows a schematic representation of the described equipment.

For the development of multivariate control charts, ten NOC batches (ID number: 1-10) were produced. The normal operating conditions (0.75 w/w% of catalyst with respect to the amount of oil, temperature of 55°C and stirring speed of 500 rpm) were set based on Lima et al. (2014). For each batch, 400 g of commercial soybean oil (Liza) were transferred to the reactor and stirred until the temperature of 55°C was reached. A mixture of methanol (Dinâmica) and catalyst (Sodium Hydroxide,

Dinâmica) was then added so that the molar ratio between methanol and soybean oil was equal to 6:1. At the same time, the NIR spectra acquisition was started. All the batches were monitored for sixty minutes.

INSERT FIGURE 2

Eight new batches (ID number: 11-18) were then manufactured following the same experimental procedure, but varying the reaction conditions (catalyst concentration, temperature and agitation speed). Specifically, except for batch 18, they were subjected to interferences intentionally generated approximately 90s after their starting point (see Table 1 for details about the nature of the various disturbances).

INSERT TABLE 1

3.2 NIR data acquisition

A fiber-optic transmittance probe (Solvias) connected to a FTLA 2000-160 FTIR spectrometer (ABB Bomem) was used for NIR data collection. The profiles, recorded as average of 64 scans at 20-second time intervals in the wavenumber range within 14000 and 3800 cm^{-1} with an optical path length of 5 mm and a spectral resolution of 16 cm^{-1} , were acquired by the software GRAMS/AI (Thermo Scientific). Background signal was measured by placing the probe into the empty reactor.

3.3 Multivariate control charts development

Multivariate control charts were constructed from *in-control* NIR spectra collected in-line during soybean oil methanolysis. For each reaction, 587 variables (absorbance values at specific wavenumbers) were acquired at 166 time points. As ten NOC batches were produced, a training three-way array of dimensions (10×587×166) was obtained.

Before the modelling phase, different preprocessing algorithms for spectral data were tested: Standard Normal Variate (SNV) (Barnes et al., 1989), Multiplicative Scatter Correction (MSC) (Martens et al., 1983; Geladi et al., 1985), and first and second derivative with Savitzky–Golay smoothing (Savitzky and Golay, 1964) (adopting a

second order polynomial function and an 11-point window). The three-dimensional structures were then analysed by using either NM or WKFH for comparison purposes. Unfolding and preprocessing were performed by Matlab® R2010a 1997.10.0.499 (Mathworks), while multivariate control charts were developed using the software SIMCA 13.0.3 (Umetrics) and ProSensus MultiVariate 15.08 (ProSensus).

Once built both the off-line and on-line monitoring schemes, the data related to the eight faulty process runs (ID number: 11-18), organized in a test set of dimensions ($8 \times 587 \times 166$), were used in order to assess their performance. The quality of these runs was evaluated resorting to the Squared Prediction Error (*SPE*)/Distance to the Model (*DModX*) and the Hotelling's T^2 statistics, considered as process wellness indices.

The *SPE* and *DModX* values associated to the n -th row vector, \mathbf{x}_n^T , of a generic ($N \times M$) data matrix, \mathbf{X} , can be obtained by Eq. (4) and (5), respectively:

$$SPE_n = (\mathbf{x}_n - \mathbf{x}_n^*)^T (\mathbf{x}_n - \mathbf{x}_n^*) \quad (4)$$

$$DModX_n = c \sqrt{\frac{SPE_n}{M-A}} \quad (5)$$

where \mathbf{x}_n^* is the reconstruction \mathbf{x}_n returned by the model, A represents the number of extracted latent variables and c is a correction factor, which is a function of N and A (Ferrer, 2007; Eriksson, 2013).

Clearly, *DModX* (provided by SIMCA 13.0.3) and *SPE* (provided by ProSensus MultiVariate) can be interpreted in a similar fashion. In fact, *out-of-control* signals in their corresponding control charts point out a breakage in the correlation structure of the original measured variables, usually highlighting the occurrence of an actual fault.

The Hotelling's T^2 statistic is instead derived for \mathbf{x}_n^T as:

$$T_n^2 = \mathbf{t}_n^T \mathbf{S}^{-1} \mathbf{t}_n \quad (6)$$

Where \mathbf{t}_n^T corresponds to the row vector containing the projection coordinates of \mathbf{x}_n^T on the A latent variable of the constructed model, while \mathbf{S} denotes the covariance matrix of the respective PCA- or PLS-scores. Unusual Hotelling's T^2 values are generally due to abnormal deviations within the model space (e.g. changes in the standard operational conditions), not affecting the aforementioned variable correlation structure.

4. Results and discussion

4.1 Spectroscopic data

Fig. 3a shows the raw spectra collected during a typical batch evolved under NOC. Since wavenumber regions below 9003 and above 4482 cm^{-1} were characterized by saturated signals and missing information, they were discarded before the data modelling. Furthermore, also the first three spectra of each batch, recorded during the reactant mixing step, were not taken into account when building the monitoring schemes.

The main differences observed in the registered profiles are evidently related to baseline shifts. This effect occurs because of the complex nature of the two-phase reaction mixture, whose properties such as viscosity and refractive index change as the transesterification progresses (Tubino et al., 2014; Torrisi and Sabino, 2013). In such circumstances, the application of an appropriate preprocessing is fundamental to assess the chemical information present in the NIR data. To this end, the aforementioned pretreatment algorithms were exploited. The most powerful monitoring schemes (in terms of fault detection capability) were obtained by applying a MSC filter (Fig. 3b) for WKFH and a first derivative Savitzky-Golay filter (Fig. 3c) for NM.

INSERT FIGURE 3

Only very slight time variations were detected for each spectral variable among runs and the paces of all the different batches were found to be pretty similar (results not shown). For this reason, no previous synchronization was required.

Finally, the pre-processed data were centred for *in-control* model calibration.

4.2 Reference models

Table 2 displays an overview of the parameters of the reference models constructed by NM and WKFH, for both off-line and on-line process monitoring. R^2 (cum) represents the portion of so-called *explained* data variance. It gives an idea of how well each model fits the experimental data. Regarding the WKFH-based off-line

monitoring scheme, mind that R^2 (cum) relates to the PLS-scores rearranged batch-wise in the batch level (Aguado et al., 2007).

INSERT TABLE 2

Clearly, both the approaches are able to properly describe the batch data, with R^2 (cum) values pretty close to 1.

4.3 Off-line monitoring

In the off-line monitoring, data from completed batches are available. Therefore, the new batches (ID number: 11-18) can be classified as normal or faulty using the models based on NOC behavior. For WKFH approach, the off-line analysis is evaluated by rearranging the PLS-scores in batch-wise and performing a PCA in the resulting matrix (batch level).

Fig. 4 shows the NM- and WKFH-based $SPE/DModX$ and Hotelling's T^2 control charts for NM and WKFH approaches for the eight new test batches process runs. Regarding the SPE control charts for NM (Fig. 4a) and $DModX$ control chart for WKFH (Fig. 4c), batches 12-18 were correctly identified as faulty by both the approaches. Therefore, the $SPE/DModX$ statistic permitted to point out particular deviations from NOC related to stirring speed, temperature and catalyst content changes.

Concerning the Hotelling's T^2 control chart for NM (Fig. 4b), the eight test batches did not exceed the corresponding control limit. For WKFH approach (Fig. 4d), only batches 15 and 18 were found to be beyond the corresponding 95% confidence limit. Therefore, Hotelling's T^2 was barely able to identify dramatic failures due to modifications in the agitation speed (interruption of agitation for 10 min) and to the change in the catalyst content.

INSERT FIGURE 4

It can be then concluded that both the off-line monitoring schemes adequately detected most of the induced faults. In general, they generated a breakage in the covariance structure of the measured variables, leading to *out-of-control* signals in the $SPE/DModX$ control chart. Furthermore, an increasing trend of the values of this

statistic depending on the magnitude of the various failures (see Table 1 and Fig. 4a and 4c) was clearly observed. Conversely, the disturbances in the agitation speed produced during batch 11 were probably not strong enough to be properly highlighted by either the *SPE/DModX* or the Hotelling's T^2 control chart.

For off-line monitoring, both the approaches presented similar results. This is a consequence of the fact that WKFH (encompassing the *batch level*) is here able to capture the process dynamics as NM.

4.4 On-line monitoring

Fig. 5 shows the on-line *DModX* and Hotelling's T^2 control charts resulting from WKFH for the 8 test batches. The process deviations intentionally produced in batches 11, 12 and 13 were not detected by either of them, probably because these runs were subjected to only slight interferences in the agitation speed for a short period of time (Fig. 5a and 5b). Batch 11 just displayed some *out-of-control* signals that are presumably not related with the disturbance, since it was induced at the beginning of the reaction. The spikes in the *DModX* control chart (Fig. 5a) are probably due to the presence of air bubbles in the reaction system, which generated particular NIR signal distortions.

In batch 14, the agitation speed was reduced to 100 rpm for 15 min. As result of this more pronounced fault, both control charts displayed an *out-of-control* signal. *DModX* (Fig. 5c) and Hotelling's T^2 (Fig. 5d) started to exceed their corresponding control limit approximately 8 and 10 minutes after the disturbance was produced, respectively. Thus, a lag between the induction of the disturbance and the time point at which the process deviation was detected was observed. This time lag occurs due to the time taken to break the stable emulsion formed between oil and methanol when the agitation speed is sharply reduced. In contrast, when the agitation speed returned to a normal value (500 rpm), *DModX* and Hotelling's T^2 assumed again *in-control* values. This could be probably due to the fact that the disturbance induced during batch 14 was not sufficiently strong to modify significantly the methanolysis reaction rate. NIRS was then able to temporarily detect such fault because of the local changes in the physical properties of the reaction mixture.

In batch 15, then agitation was stopped for about 10 min. Also in this case, the fault was clearly and instantaneously detected by both control charts (Fig. 5c and 5d).

Even after the stirring speed was set again to 500 rpm, the *DModX* statistic continued fluctuating around its control limit, while the Hotelling's T^2 values were all found to be back *in-control*. Here, the induced failure may have appreciably affected the reaction rate and, therefore, the *DModX* values for this batch did not return below the corresponding control limit. Furthermore, due to the relatively high magnitude of the disturbance, a lag between its induction and the time point at which the process deviation was detected was not observed for this batch.

During batches 16 and 17, the reaction temperature was temporarily reduced to 50 and 45°C, respectively. In both cases, the *DModX* and Hotelling's T^2 statistic values started to slowly increase almost at the same time at which the disturbance was induced. Nevertheless, the first *out-of-control* signals were observed only slightly later. For the former, only the *DModX* statistic slightly exceeded its corresponding control limit (Fig. 5e); for the latter, the fault was clearly detected by both control charts. Specifically, *DModX* was always found to be beyond its control limit, while Hotelling's T^2 (Fig. 5f) assumed again *in-control* values when the temperature returned to a normal value (55°C).

Batches 16 and 17 were successfully detected as outliers mainly because NIRS is particularly sensitive to temperature changes. In fact, NIR spectra show features related to the molecular structure of the samples under study, as well as absorbance bands associated to inter- or intra-molecular interactions (e.g. hydrogen bonds), whose shape and intensity are affected by the temperature conditions (Shao et al., 2010; Workman and Weyer, 2007). Probably, for batch 17, the change in the reaction temperature also generated a reduction in the conversion rate, which affected the composition of the reaction mixture and led to *out-of-control* *DModX* statistic values albeit the normal temperature conditions were restored (Noriega et al., 2014).

Lastly, batch 18, manufactured to check how variations in the catalyst concentration could influence the process evolution, was clearly detected as outlier by the *DModX* control chart (Fig. 5e). As detailed in the scientific literature, a decrease in the catalyst concentration reduces the methanolysis reaction rate (Noriega et al., 2014). This probably influenced the composition of the reaction mixture, which allowed NIRS to capture this kinetic phenomenon.

INSERT FIGURE 5

On the other hand, Fig. 6 shows the on-line *SPE* and Hotelling's T^2 control charts resulting from NM for the 8 test batches. As they can be interpreted similarly to the WKFH-based control charts, the discussion will be simply focused on the differences between the outcomes obtained by the two approaches under study. In contrast to WKFH, the process deviations intentionally produced in batches 12 and 13 were clearly detected by the *SPE* control chart (Fig. 5a). Batch 11 presented some *out-of-control* signals in the *SPE* control chart that are probably not associated with the deviation intentionally induced along the process. These *out-of-control* signals, also detected for WKFH, are probably related to the position of the probe in relation to the stirrer for this batch. Nevertheless, no *out-of-control* signal was displayed by the Hotelling's T^2 control chart.

For batches 14 and 15, the *SPE* statistic values were found to be above the respective control limit all along the run evolution. For batch 15, Hotelling's T^2 exceeded its confidence limit for a certain time period but returned below it slightly after the agitation speed was set again to 500 rpm.

Batches 16, 17 and 18 were detected as faulty by the *SPE* control chart. The Hotelling's T^2 control chart instead pointed out an *out-of-control* situation only at the beginning of batches 17 and 18. There is a clear difference between faults 11-17 and fault 18. The first batches are characterised by a distinct evolution over time from the beginning of the runs (mainly *SPE* continuously increases till exceeding its control limit), while for fault 18 all the observations have high and approximately constant *SPE* values, which may be due to the fact that the interference was induced before starting the batch. It could be also realized for WKFH approach.

INSERT FIGURE 6

In this specific on-line monitoring case-study, NM showed a better performance than WKFH, as it enabled a correct detection of all the failures intentionally produced during the test batch runs. This is probably associated to the fact that, unlike WKFH (not including the aforementioned *batch level*), NM is able to capture the dynamics of the transesterification process.

In contrast, the NM-based control charts tended to spot deviations from NOC even after the induced disturbances were interrupted. This effect is typical of this kind

of monitoring schemes and is due to the strong effect of the auto-correlation on the *SPE* and/or Hotelling's T^2 statistic (Camacho and Picó, 2006; Camacho et al., 2009).

4.5 Contribution plots

Once a particular fault has been detected, its root causes can be highlighted by tools like contribution plots (Kourti and MacGregor, 1996). A contribution plot shows the contribution values of each original variable to the statistic used for monitoring purposes. Variables with a high contribution are those mainly related to the *out-of-control* signal detected by the corresponding control chart (Kerkhof et al., 2012; Nomikos and MacGregor, 1995; Li Vigni et al., 2009).

Although contribution plots for spectral data are not always easily interpretable and cannot precisely point out the root causes of the detected failures (there exist no univocal relation between the changes in the spectral profiles and the variation of the physical parameters leading to the fault), they can anyway be resorted to for highlighting which spectral regions are those most affected by the *out-of-control* situation, i.e. at which spectral regions the signal changes the most when the process is deviating from NOC.

INSERT FIGURE 7

In this section, as an example, the contribution plots resulting from the WKFH-based on-line monitoring scheme and associated to the time instant at which the Hotelling's T^2 and *DModX* statistic values were observed to be out of control for batches 14 and 17 (approximately 11 min after their starting point) are shown. Because of the huge number of spectral variables to be taken into account for each process run, contribution plots derived from the NM approach are not shown, but they enabled an equally correct fault diagnosis.

Batches 14 and 17 were subjected to different types of disturbance. During the former, the stirring speed was reduced to 100 rpm for 15 min, while the temperature was temporarily decreased to 45°C 90 seconds after the beginning of the latter. As shown in Fig. 7 and Fig. 8, the spectral regions associated to the highest contribution values are: (I) 7324 – 7116 (aliphatic C-H combinations); (II) 6222 – 6099 cm^{-1} (first overtone of olefinic C-H stretching); (III) 5900 – 5580 cm^{-1} (first overtone of aliphatic

C-H stretching); (IV) 5244 – 5106 cm^{-1} (second overtone of C=O stretching in esters); (V) 4874 – 4744 (methanol O-H and C-O stretching and bending combinations). The band attribution was carried out according to Workman and Weyer (2007).

In particular, for batch 14 (Fig. 7), the absorption bands related to methanol O-H and C-O stretching and bending combinations (region V) are characterized by negative contributions to both monitoring statistics. The bands related to the first overtone of olefinic C-H stretching (region II), which are proper of soybean oil composition, exhibit negative contributions to Hotelling's T^2 . On the other hand, the aliphatic C-H combination bands (region I) show a slightly positive contribution to $DModX$.

INSERT FIGURE 8

For batch 17 (Fig. 8), the absorption bands related to the first overtone of olefinic C-H stretching (region II) are characterized by positive contributions to Hotelling's T^2 . The most negative contributions to $DModX$ are possibly related to methanol O-H and C-O stretching and bending combination bands (region V). Also the aliphatic C-H combination spectral region (I) show a negative contribution to the $DModX$ statistic, while the second overtone of C=O stretching in esters (region IV) contributed positively to it.

One can clearly see that the disturbances affecting the reaction temperature and the stirring speed produced specific deviations in very similar spectral ranges. Methanol O-H and C-O combinations showed the most negative contributions to Hotelling's T^2 and $DModX$ for batch 14 and to $DModX$ for batch 17. Aliphatic C-H bond-related absorption bands were found to be characterized by positive contributions to both statistics for batch 14 and negative contributions to $DModX$ for batch 17. Finally, the spectral region related to the first overtone of olefinic C-H stretching contributed negatively for batch 14 and positively for batch 17 to Hotelling's T^2 .

For comparison purpose, contribution plots resulting from the NM-based on-line monitoring scheme for batches 14 and 17 were also analysed at the same time at which contribution plots from WKFH. Spectral regions similar to the aforementioned were mainly associated to the *out-of-control* signals in the *SPE* and Hotelling's T^2 control charts. In addition, a spectral region (6900 – 6800 cm^{-1}), that is probably related to the first overtone of the hydroxyl group, showed the most positive contributions to Hotelling's T^2 and *SPE* for batch 17.

The reliability of the fault diagnosis was in any case confirmed by the original spectral data. In fact, the most evident changes in the NIR profiles (from *in-control* to faulty conditions) were actually observed in the wavelength regions identified by the contribution plots as the most affected by the induced disturbances (results not shown).

5. Conclusions

NIR-based MSPC was applied for batch monitoring of soybean oil methanolysis. Various reactions carried out in a batch system generated a three-way data array, which was analysed using two different approaches: NM and WKFH. The capability of the developed monitoring schemes of detecting specific disturbances induced during the evolution of 8 test batches was assessed and compared.

The control charts resulting from both the modelling strategies showed a very similar performance in terms of off-line fault detection power. Concerning on-line monitoring, NM, unlike WKFH, allowed faults due to modifications in the stirring speed of the reaction mixture to be correctly pointed out. Contribution plots enabled a clear identification of the spectral region mostly affected by the faults when both the approaches were resorted to.

In general, most of the failures associated to the reaction temperature, catalyst content and stirring speed were properly highlighted. However, the monitoring schemes proved to be much more sensitive to changes in the catalyst concentration and in the temperature conditions, probably owing to the fact that they were sufficiently strong to cause a significant modification in the reaction rate and, thus, in the composition of the reaction mixture.

Such satisfactory results show the application of these MSPC tools can clearly enhance transesterification process understanding as well as the detection and diagnosis of faults occurring during soybean oil methanolysis and might possibly be extended to other branches of biodiesel production, in which different raw materials, catalysts and reaction conditions are involved.

Acknowledgments

The authors would like to thank FACEPE/NUQAAPE, CNPq/INCTAA science funding programs for partial financial support. Research fellowships granted by the Brazilian

agencies ANP/Petrobras and CNPq are also gratefully acknowledged. This research work was partially supported by the Spanish Ministry of Economy and Competitiveness under the project DPI2014-55276-C5-1R and Shell Global Solutions International B.V. (Amsterdam, The Netherlands).

References

Aguado D, Ferrer A, Ferrer J, Seco A. Multivariate SPC of a sequencing batch reactor for wastewater treatment. *Chemom. Intell. Lab. Syst.* 2007; 85: 82–93.

Alves JCL, Poppi RJ. Biodiesel content determination in diesel fuel blends using near infrared (NIR) spectroscopy and support vector machines (SVM). *Talanta* 2013; 104: 155-161.

Arias ELM, Martins PF, Munhoz AJ, Gutierrez-Rivera L, Filho RM. Continuous Synthesis and in Situ Monitoring of Biodiesel Production in Different Microfluidic Devices. *Ind. Eng. Chem. Res.* 2012; 51: 10755–10767.

Atabani AE, Silitonga AS, Badruddin IA, Mahlia TMI, Masjuki HH, Mekhilef S. A comprehensive review on biodiesel as an alternative energy resource and its characteristics. *Renew. Sust. Energ. Rev.* 2012; 16: 2070–2093.

Barnes R, Dhanoa M, Lister S. Standard Normal Variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Appl. Spectrosc.* 1989; 43: 772-777.

Bersimis S, Psarakis S, Panaretos J. Multivariate Statistical Process Control Charts: An Overview. *Qual. Reliab. Eng. Int.* 2007; 23: 517–543.

Camacho J, Picó J, Ferrer A. Bilinear modelling of batch processes. Part I: theoretical discussion. *J. Chemom* 2008; 22: 299-308.

Camacho J, Picó J, Ferrer A. On-line monitoring of batch processes based on PCA: does the modelling structure matter? *Anal. Chim. Acta* 2009; 642: 59-68.

Camacho J, Picó J. Online monitoring of batch processes using multi-phase principal component analysis. *J. Process Contr.* 2006; 16: 1021-1035.

Csernica AN, Hsu JT. The Phase Behavior effect on the Kinetics of Transesterification Reactions for Biodiesel Production. *Ind. Eng. Chem. Res.* 2012; 51: 6340-6349.

Eriksson L, Byrne T, Johansson E, Trygg J, Vikström C. Multi- and Megavariate Data Analysis. MKS Umetrics AB, Malmö, Sweden, 2013.

Ferrer A. Multivariate Statistical Process Control Based on Principal Component Analysis (MSPC-PCA): Some Reflections and a Case Study in an Autobody Assembly Process. *Qual. Eng.* 2007; 19: 311-325.

Geladi P, Macdougall D, Martens H. Linearization and scatter-correction for near-infrared reflectance spectra of meat. *Appl. Spectrosc.* 1985; 39: 491-500.

González-Martínez JM, Vitale R, De Noord OE, Ferrer A. Effect of synchronization on bilinear batch process modeling. *Ind. Eng. Chem. Res.* 2014; 53: 4339-4351.

Huang H, Qu H. In-line monitoring of alcohol precipitation by near-infrared spectroscopy in conjunction with multivariate batch modeling. *Anal. Chim. Acta* 2011; 207: 47-56.

Kerkhof PV, Gins G, Vanlaer J, VanImpe JFM. Dynamic model-based fault diagnosis for (bio)chemical batch processes. *Comput. Chem. Eng.* 2012; 40: 12– 21.

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: 3268-3273.

Kona R, Qu H, Mattes R, Jancsik B, Fahmy RM, Hoag SW. Application of in-line near infrared spectroscopy and multivariate batch modeling for process monitoring in fluid bed granulation. *Int. J. Pharm.* 2013; 452: 63– 72.

Kourti, T, MacGregor, J. Multivariate SPC methods for process and product monitoring. *J. Qual. Technol.* 1996; 28: 409-428.

Kourti T. Multivariate dynamic data modeling for analysis and statistical process control of batch process, start-ups and grade transitions. *J. Chemom.* 2003; 17: 93-109.

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.

Lira, LFB, De Vasconcelos FVC, Pereira CF, Paim APS, Stragevitch L, Pimentel MF. Prediction of properties of diesel/biodiesel blends by infrared spectroscopy and multivariate calibration. *Fuel* 2010; 89: 405-409.

Li Vigni M, Durante C, Foca G, Marchetti A, Ulrici A, Cocchi M. Near Infrared Spectroscopy and multivariate analysis methods for monitoring flour performance in an industrial bread-making process. *Anal. Chim. Acta* 2009; 642: 69-76.

Li W, Han H, Cheng Z, Zhang Y, Liu S, Qu H. A feasibility research on the monitoring of traditional Chinese medicine production process using NIR-based multivariate process trajectories. *SensorActuat. B-Chem.* 2016; 231:313-323.

Martens H, Jensen, S, Geladi, P. Multivariate linearity transformation for near-infrared reflectance spectrometry. In: *Proceedings of the Nordic Symposium on Applied Statistics*, StokkandForlag Publisher, Stavanger, Norway, p. 208-234, 1983.

Nelson PRC, Taylor PA, MacGregor JF. Missing data methods in PCA and PLS: score calculations with incomplete observations. *Chemom. Intell. Lab. Syst.* 1996; 35:45-65.

Nomikos P, Macgregor JF. Multivariate SPC charts for monitoring batch processes. *Technometrics* 1995; 37: 41-59.

Noriega MA, Narváez PC, Heinz C. Kinetics of Jatropha oil methanolysis. *Fuel* 2014; 134: 244-249.

Oliveira IK, Rocha WFC, Poppi RJ. Application of near infrared spectroscopy and multivariate control charts for monitoring biodiesel blends. *Anal. Chim. Acta* 2009; 642: 217–221.

Pasquini, C. Near infraed spectroscopy: Fundamentals, practical aspects and analytical applications. *J. Braz. Chem. Soc* 2003; 14:198-219.

Pimentel MF, Ribeiro GMGS, Da Cruz RS, Stragevitch L, Pacheco Filho JGA, Teixeira LSG. Determination of biodiesel content when blended with mineral diesel fuel using infrared spectroscopy and multivariate calibration. *Microchem. J.* 2006; 82: 201–206.

Richard R, Dubreuil B, Thiebaud-Roux S, Prat L. On-line monitoring of the transesterification reaction carried out in microreactors using near infrared spectroscopy. *Fuel* 2013; 4: 318–325.

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. *Bioresour. Technol.* 2011; 102:6702–6709.

Savitzky A, Golay M. Smoothing and differentiation of data by simplified least squares procedure. *Anal. Chem.* 1964; 36: 1627-1639.

Shao X, Kang J, Cai W. Quantitative determination by temperature dependent near-infrared spectra. *Talanta* 2010; 82:1017-1021.

Torrison NM, Sabino M. Biodiesel production control using PNN and Coriolis flowmeter. *Neural Comput. & Applic.* 2013; 23:1275-1282.

Tubino M, Junior JGR, Bauerfeldt GF. Biodiesel synthesis with alkaline catalysts: A new refractometric monitoring and kinetic study. *Fuel* 2014; 125: 164–172.

Wold S, Kettaneh N, Fridén N, Holmberg A. Modelling and diagnostics of batch processes and analogous kinetic experiments. *Chemom. Intell. Lab. Syst.* 1998; 44: 331–340.

Wold S, Kettaneh N, Macgregor JF, Dunn KG. Batch process modeling and MSPC. In: Brown SD, Tauler R, Walczak B, editors. *Comprehensive Chemometrics*, Oxford: Elsevier B.V.; 2009, p. 163-197.

Workman J, Weyer L. *Practical guide to interpretative near-infrared spectroscopy*. 1st ed. New York: CRC Press; 2007.

FIGURE CAPTIONS

Fig. 1: Schematic representation of (a) batch-wise and (b) variable-wise unfolding.

Fig. 2: Schematic representation of the Optimax synthesis workstation (Mettler Toledo).

Fig. 3: Raw spectra (a), MSC-preprocessed spectra (b) and first derivative spectra (c) collected during a typical batch evolved under NOC.

Fig. 4: Off-line monitoring: *SPE* (a) and Hotelling's T^2 (b) control charts resulting from NM for the eight faulty test batches. *DModX* (c) and Hotelling's T^2 (d) control charts resulting from WKFH for the eight faulty test batches. Black dashed lines correspond to their respective 95% confidence limits.

Fig. 5: WKFH-based on-line monitoring: *DModX* and Hotelling's T^2 control charts for batches 11-13 (a and b), 14-15 (c and d) and 16-18 (e and f). Red dashed lines correspond to their respective 95% confidence limits.

Fig. 6: NM-based on-line monitoring: *SPE* and Hotelling's T^2 control charts for batches 11-13 (a and b), 14-15 (c and d) and 16-18 (e and f). Red dashed lines correspond to their respective 95% confidence limits.

Fig. 7: WKFH-based on-line monitoring: Hotelling's T^2 (a) and *DModX* (b) instantaneous contribution plots for batch 14 at 11 minutes after the beginning of the run, respectively.

Fig. 8: WKFH-based on-line monitoring: Hotelling's T^2 (a) and *DModX* (b) instantaneous contribution plots for batch 17 at 11 minutes after the beginning of the run, respectively.