Process modeling and control applied to real-time monitoring of distillation processes by near-infrared spectroscopy

Rodrigo R. de Oliveira[1,2], Ricardo H. P. Pedroza[2], Adriano O. Sousa[2], Kássio M. G. Lima[2,3], Anna de Juan[1*]

[1]Chemometrics Group, Department of Analytical Chemistry, Universitat de Barcelona, Diagonal 645, 08028, Barcelona, Spain

[2] LabPVT, Federal University of Rio Grande do Norte, Av. Senador Salgado Filho, 3000, Natal 59078-970, RN, Brazil

[3] Institute of Chemistry, Biological Chemistry and Chemometrics, Federal University of Rio Grande do Norte, Av. Senador Salgado Filho, 3000, Natal 59078-970, RN, Brazil

[*]Corresponding author. Tel.: +34 934 03 9778.

e-mail address: anna.dejuan@ub.edu (A. de Juan).

Abstract: A distillation device that acquires continuous and synchronized measurements of temperature, percentage of distilled fraction and NIR spectra has been designed for real-time monitoring of distillation processes. As a process model, synthetic commercial gasoline batches produced in Brazil, which contain mixtures of pure gasoline blended with ethanol have been analyzed. The information provided by this device, i.e., distillation curves and NIR spectra, has served as initial information for the proposal of new strategies of process modeling and multivariate statistical process control (MSPC). Process modeling based on PCA batch analysis provided global distillation trajectories, whereas multiset MCR-ALS analysis is proposed to obtain a component-wise characterization of the distillation evolution and distilled fractions. Distillation curves, NIR spectra or compressed NIR information under the form of PCA scores and MCR-ALS concentration profiles were tested as the seed information to build MSPC models. New on-line PCA-based MSPC approaches, some inspired on local rank exploratory methods for process analysis, are proposed and work as follows: a)MSPC based on individual process observation models, where multiple local PCA models are built considering the sole information in each observation point; b) Fixed Size Moving Window – MSPC, in which local PCA models are built considering a moving window of the current and few past observation points; and c) Evolving MSPC, where local PCA models are built with an increasing window of observations covering all points since the

beginning of the process until the current observation. Performance of different approaches has been assessed in terms of sensitivity to fault detection and number of false alarms. The outcome of this work will be of general use to define strategies for on-line process monitoring and control and, in a more specific way, to improve quality control of petroleum derived fuels and other substances submitted to automatic distillation processes monitored by NIRS.

Keywords: Near-infrared spectroscopy; on-line multivariate statistical process control - MSPC; process modeling; distillation process; petroleum.

1. Introduction

Distillation curves are frequently used for quality control of petroleum products. The evolution and shape of these curves is directly related to the composition and chemical characteristics of these products and, hence, a temperature deviation from normal distillation behavior may be an indicator of adulteration. ASTM D86 [1] is the standard test method required to obtain distillation curves and classical process control is made by comparing the temperature at specific distillation points with standard specification limits.

However, distillation curves, based only on boiling temperature monitoring, are not conclusive to identify adulterations in product composition. Adulterants can nowadays be chosen so that the modified petroleum products show normal distillation curve behavior. Near-infrared spectroscopy (NIRS) may help to overcome such scenario because of the rich physicochemical information associated with this spectroscopic technique and the existence of many NIR sensors designed for on-line process monitoring. Along this line, distillation devices that incorporate NIR sensors and collect synchronized distillation temperatures and related NIR absorption spectra measurements, as proposed by Pasquini and Scafi, are a suitable solution [2]. Thus, the fiber optic probes coupled to NIR spectrometers can be located directly in the distillation process stream, allowing continuous real-time in-process measurements [2–4]. Therefore, information representing both physical and chemical properties of the distilled sample can be derived from each distillation batch.

In Brazil, commercial gasoline is blended with ethanol. Thus, gasoline derived directly from refineries without ethanol addition is denominated “type A”. Gasoline “type C” is the commercial mixture of gasoline “type A” and (27 ± 1)% of ethanol (% v/v) [5]. As a process model for this work, a study of quality control of Brazilian gasolines regarding ethanol content specification is proposed. To do the experimental process monitoring, an improved version of the automatic distillation device monitored by NIRS proposed by Pasquini and Scafi [2], which allows continuous and synchronized data acquisition and storage of distillation temperatures, distilled mass and related NIR spectra, is proposed. Detailed description of the experimental setup is found in section 2 below.

The distillation curves and NIR spectra collected from distillation batch processes can be modelled with principal component analysis (PCA) [6] and multivariate curve resolution – alternating least squares (MCR-ALS) [7] for better process understanding and use of this information in further process control. PCA batch analysis provides global distillation

trajectories, whereas MCR-ALS offers the additional value of describing the temperature-dependent evolution and characterization of the different distilled fractions during the process.

MSPC has being used to control processes related to very diverse fields, such as pharmacy [8–11], petrochemistry [12–14] and biotechnology [15,16]. Batch MSPC using NIRS has been described in recent works [3,4,8–11,16,17]; however, no MSPC using NIRS to monitor batch distillation process has been reported in the literature.

In this study, different off-line process control models are studied using complete batch information collected during the distillation process monitored by NIRS. Distillation curves, original NIR spectra, as well as the compressed spectral information contained in PCA scores or MCR-ALS concentration profiles are used to build off-line PCA-based multivariate statistical process control (MSPC) models. To our knowledge, there is no report in the literature about using the concentration profiles from MCR-ALS analysis as starting information to build PCA-based MSPC models. In this framework, the description of the separate components of the process provided by MCR-ALS would allow for using all concentration profiles on the MSPC model or profiles of selected compounds that could be envisioned as more specific indicators of process evolution.

NIR measurements obtained from distillation processes are also used to build on-line batch MSPC models. On-line batch MSPC approaches commonly used are based on the methods proposed by Nomikos and MacGregor [18–20] and Wold et. al. [21]. Other approaches are proposed by Rännar et. al. for adaptive batch monitoring using hierarchical PCA [22], by Zhao et. al using multiple PCA models for local model building at each observation point [23] and using moving window [24]. In this work, chemometric tools typically used to perform local exploratory analysis of the evolution of processes, such as evolving factor analysis (EFA) or fixed size moving window - EFA (FSMW-EFA) [25,26], have been adopted to propose new on-line batch MSPC strategies. The performance of these on-line MSPC approaches has been studied in terms of sensitivity to fault detection and number of false alarms.

The outcomes of this study will be of general applicability, as guidelines for process modeling and control based on spectroscopic measurements, and suppose a significant improvement on the specific field of quality control based on distillation processes, both from the instrumental point of view and from the way to handle the derived information from coupled temperature-NIR distillation curves.

2. Experimental

2.1. Automatic distillation device setup

The automatic distillation device designed is shown in Figure 1. It is formed by a distillation glassware setup (125 ml), a transmittance flow cell connected through optical fibers to a FT-NIR spectrophotometer (Rocket, ARCoptix ANIR, Switzerland), an analytical balance (XS204, Mettler-Toledo, Switzerland), a thermocouple and heater controlled by a data acquisition device and a personal computer with a data acquisition software that connects and controls the distillation setup. Heating mantle power applied is automatically controlled based on a feedback controller to keep distillation rate constant rather than keeping constant power as in [2].

(Insert Figure 1)

2.1. Batch distillation process

For every distillation batch, 100 mL from the suitable sample, previously weighed, are introduced in the distillation flask. The heater is started and once the initial boiling point (IBP) is automatically detected, distillation process starts and synchronized measurements of temperature, distillation recovered percentage (wt%) of initial sample weight and NIR absorption spectra (900 – 2600 nm) are taken every five seconds until the end point (EP) is reached. Data are stored in MATLAB format in such a way that values every 1 wt% are saved. Temperature and NIR spectra are averages of all measurements recorded during every 1 wt% distillation interval.

Synthetic gasoline (type C) batches were distilled using the designed automatic distillation device. The gasoline batches were prepared by mixing ethanol AR (99% Sigma-Aldrich) and pure gasoline (type A, from Petrobras refinery) at different ratios. A set of 23 blends was performed: 11 samples containing 27 %(v/v) ethanol (on-specification gasolines) and 12 with 10-25 %(v/v) and 30-40 %(v/v) ethanol (off-specification gasolines). Table 1 describes the gasoline batches prepared with their related composition. These batch ID labels will be used to identify the batches throughout the manuscript.

(Insert Table 1)

3. Data treatment

3.1. Raw data and preprocessing

Temperature, distilled weight and NIR spectra were obtained synchronously every 5 s and averaged measures were stored every 1 wt% of distilled weight increment from IBP until EP. The final process range considered was from 5 to 90 wt% distilled weight, which corresponded to K = 86 observation points. Observations at the beginning (< 5 wt%) and end (> 90 wt%) of the distillation process were unstable and, therefore, not used for process control. NIR spectra working wavelength range was 1103 – 2228 nm due to high noise observed in measurements out of these wavelength boundaries. This range contained J = 573 spectral channels. For each distillation batch, a column-vector sized (K x 1) with the temperatures associated with the distillation curve and a matrix sized (K x J) with the related NIR infrared spectra were obtained.

Data obtained from on-specification batch B07 are used to illustrate the typical data obtained at the end of a batch distillation run. Figure 2(a) shows the distillation curve with the recorded boiling temperatures, Figure 2(b) the related raw NIR spectra and Figure 2(c) the raw NIR spectra at the four observation points indicated in Figure 2(a).

(Insert Figure 2)

NIR spectra were preprocessed for baseline correction by Savitzky-Golay derivative [27] (1^{st} order derivative, 2^{nd} order polynomial function and 9 points window) followed by signal intensity fluctuation corrected by spectral normalization, see Figure 3(b).

(Insert Figure 3)

3.2. Data analysis

Process modeling. Principal Component Analysis (PCA) and Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS).

The matrices with the NIR data from each on-specification batch were arranged one on top of each other into a column-wise augmented multiset structure, **D**, and modeled using principal component analysis (PCA) and multivariate curve resolution-alternating least squares (MCR-ALS). PCA provided a global model of trajectories explaining the overall process evolution, whereas MCR-ALS provided a model describing the evolution and

chemical identity of each component (distinct distilled fraction) in the distillation batches analyzed.

PCA was used to reduce the dimensionality of the spectral data from the distillation processes by compressing the high-dimensional mean-centered original NIR data matrix into a low-dimensional subspace of principal components. These components explain most of the data variability and are orthogonal linear combinations of the original spectroscopic variables [6]. The PCA model of column-wise augmented matrix **D** is expressed as: $\mathbf{D} = \mathbf{TP}^{\mathbf{T}}$, where **T** are the scores, related to the observations of the distillation process and $\mathbf{P}^{\mathbf{T}}$ are the loadings, related to the importance of the NIR wavelengths in the description of the principal components. The scatter plot of scores provides the global trajectories of the processes analyzed.

The same multiset structure was modeled using multivariate curve resolution - alternating least squares (MCR-ALS). MCR-ALS assumes a bilinear model, $\mathbf{D} = \mathbf{CS}^{\mathbf{T}}$, which is the multiwavelength extension of the Lambert-Beer's law [7,28–30]. $\mathbf{S}^{\mathbf{T}}$ contains the pure spectra of the components needed to describe the distillation process and **C** the concentration (distillation profiles). In contrast to PCA, MCR-ALS gives real meaningful concentration and spectral profiles of pure components of the system. MCR-ALS works by alternatingly optimizing **C** and $\mathbf{S}^{\mathbf{T}}$ under constraints. Initial estimates of $\mathbf{S}^{\mathbf{T}}$ were performed by using a pure variable selection method based on SIMPLISMA [31]. Constraints applied in this work were non-negativity and unimodality, i.e., presence of a single maximum per profile, for the concentration (**C**) profiles. Local rank constraints, i.e., setting the absence of certain compounds in observations of the concentration profiles, were used to improve the quality of the resolved spectral signatures [32]. This was done by appending pure ethanol NIR spectra to the column-wise multibatch structure (in this case, only the ethanol was set to be present in the concentration elements linked to the appended pure ethanol spectra).

MCR-ALS provides a much more detailed description of the process than PCA in terms of characterization of process profiles and spectral signatures, related to distillation fractions in this case. However, the single process trajectory provided by the scatter score plot of PCA is a global description of process evolution and a quick visual way to observe when a batch process evolves as NOC batches or does differently. Being complementary views about the evolution of a process, we found relevant to include both in this study. Both PCA scores

and MCR **C** profiles are afterwards used as starting information for off-line batch MSPC models described in the next section.

Process control

From the batches analyzed, nine on-specifications or NOC (Normal Operation Conditions) batches (batches B01-09), were selected to build PCA-based MSPC models (see Table 1). These models were afterwards used to detect whether a new batch (or observations within it) is in or out of control [33]. Two on-specification batches (B10-11) and twelve off-specification batches (B12-23) were used to test the MSPC models.

The PCA-based MSPC model is built using the preprocessed and mean-centered data matrix of NOC batches, $\mathbf{X_{NOC}}$, sized (nr. of NOC batches × observed measurements per batch) according to the equation below,

$$\mathbf{X_{NOC}} = \mathbf{T_{NOC}}\mathbf{P_{NOC}^{T}} + \mathbf{E_{NOC}} \quad \textbf{(1)}$$

where $\mathbf{T_{NOC}}$ is the scores matrix of all NOC batches and $\mathbf{P_{NOC}^{T}}$ is the loadings matrix. The number of components used in an MSPC model is a critical parameter and has been established by cross-validation [34].

The scores for new batches are obtained multiplying the measured preprocessed batch information, $\mathbf{X_{NEW}}$, with the loadings matrix $\mathbf{P_{NOC}^{T}}$ from the model built with the NOC batches, using the following equation:

$$\mathbf{T_{NEW}} = \mathbf{X_{NEW}}\mathbf{P_{NOC}} \quad \textbf{(2)}$$

Then, the residuals are obtained using the new batch scores, as:

$$\mathbf{E_{NEW}} = \mathbf{X_{NEW}} - \mathbf{T_{NEW}}\mathbf{P_{NOC}^{T}} \quad \textbf{(3)}$$

From the PCA model built with NOC batches, two MSPC control charts can be built, in which observations of new batches are represented: a) Hotelling's T^2 chart, usually referred as *D*-statistic ($D_{stat.}$), represents the estimated Mahalanobis distance from the center of the latent subspace, representing the average in control conditions of a batch, to the projection of a new batch (or observation) onto this subspace and the b) *Q*-statistic chart ($Q_{stat.}$)accounts for the residual part of the process variation not explained by the PCA model.

The Hotelling statistic, $D_{stat.}$, was calculated using the following equation:

$$D_{stat.} = \mathbf{t}^{\mathrm{T}}\Theta^{-1}\mathbf{t} \quad \textbf{(4)}$$

Where $\mathbf{t}$ is the vector containing the scores of a new given batch with the A retained principal components (PC's), and Θ is the scores covariance matrix with $(A \times A)$ size. The control limit for this chart is calculated according to the equation proposed by Jackson [35].

$$D_{CL} = \frac{A(I-1)}{I-A} F(A, I-A, \alpha) \quad \textbf{(5)}$$

where I is the number of in control batches used to build the model with A PC's and $F(A, I-A, \alpha)$ is the $100(1-\alpha)$ percentile of the corresponding F distribution.

The $Q_{stat.}$ for the *i*th new batch $\boldsymbol{x}_i$ is given by

$$Q_{stat.} = \boldsymbol{e}_i^T \boldsymbol{e}_i \quad \textbf{(6)}$$

where $\boldsymbol{e}_i$ is the residual vector of the *i*th new batch from the PCA model. Regarding the control limit for the $Q_{stat.}$ chart, Jackson and Mudholkar[36] showed that an approximate $Q_{stat.}$ critical value at significance level α is given by

$$Q_{CL} = \theta_1 \left[\frac{z_\alpha \sqrt{2\theta_2 h_0^2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right]^{1/h_0} \quad \textbf{(7)}$$

where, $\theta_k = \sum_{j=A+1}^{rank(X)} \lambda_j^k$ and $h_0 = 1 - (2\theta_1\theta_3 / 3\theta_2^2)$, λ_j are the eigenvalues of the PCA residual covariance matrix and z_α is the $100(1-\alpha)\%$ standardized normal percentile.

Two MSPC approaches were applied in this work, devoted to off-line and on-line control, respectively. Both approaches and related control charts are explained below.

Off-line batch MSPC

Off-line batch MSPC charts were built using data provided from completed distillation processes. Different models were built according to the starting information used, either temperatures from distillation curves or information derived from NIR spectra, Figure 4.

(Insert Figure 4)

a) Off-line batch MSPC models using distillation curves

The distillation curves from the 9 NOC distillation batches were arranged in a matrix ($I \times K$), with I = 9 rows and K = 86 observation points of the distillation curve. This matrix was mean-centered and decomposed by PCA to obtain the model loadings and MSPC limits,

see Figure 4(a). New batch data were projected into the model to obtain the related statistical parameters ($D_{stat.}$ and $Q_{stat.}$).

b) Off-line batch MSPC models using NIRS data

Different off-line MSPC models were built with the NIRS-derived information. All models were built on data sets with I = 9 rows and a variable number of columns depending on the kind of NIRS-derived information, see Figure 4(b). This gave rise to three different MSPC models:

i. **Models based on the original preprocessed NIR data matrix.** This model is done using a matrix containing the NIR readings from each individual NOC batch row-wise unfolded into a vector, i.e. the matrix of a batch with dimensions (K×J), where K = 86 are batch observation points and J = 573 wavelengths, is arranged in a row vector with dimension (1 ×KJ), with K = 86 and J = 573. Then, the information of I = 9 NOC batches was arranged in a matrix sized (I×KJ), on which the MSPC model was built.

ii. **Models based on the batch scores from PCA decomposition of the NOC multiset structure.** The information of a NOC batch are the scores obtained in the PCA model of the related NIR spectra, row-wise unfolded into a vector sized (1 ×KA) with A being the number of retained principal components. Then, the information of I = 9 NOC batches was arranged in a matrix sized (I×KA), on which the MSPC model was built.

iii. **Models based on the resolved concentration profiles from MCR-ALS decomposition of the NOC multiset structure.** The information of a NOC batch are the concentration profiles obtained in the MCR-ALS model of the related NIR spectra, row-wise unfolded into a vector sized (1 ×KN) with N being now the number of MCR contributions needed to describe the process. Then, the information of I = 9 NOC batches was arranged in a matrix sized (I×KN), on which the MSPC model was built. Please note that, generally speaking, the use of only some of the concentration profiles modeled in a batch could be an option to build the MSPC model, provided that the selected profiles were proven to be very specific indicators of the

process evolution or that the discarded profiles belonged to spurious process contributions, e.g., modeled background contributions if existing. Please note that even if the use of C-profiles implies a noise-filtered compression of the original information, the size of the unfolded profiles, sized (1 ×KN) per each NOC batch, requires a PCA-based MSPC model for easier interpretability.

The MSPC PCA models built with the different kinds of starting information were used to extract the related $D_{stat.}$ and $Q_{stat.}$ charts limits. Suitable data from new batches, not used to build the model, were projected onto the MSPC PCA model to test the performance of the models built.

On-line batch MSPC

Different on-line MSPC monitoring charts were developed using the data provided from NIRS measurements. As in the off-line approach, the same unfolded NOC matrix with the original NIR variables was used in the on-line approach. However, three on-line MSPC approaches were proposed using multiple PCA models based on different intervals of observation points, as described below:

a) On-line MSPC based on individual process observation models

This approach is the most straightforward method. An individual model is built per each observation point using historical data from on-specification completed batches as illustrated in Figure 5(a). Thus, during a new batch, the new on-line data obtained (NIR spectrum of current observation) is projected into the respective observation point model and the statistical parameters compared with the control chart limits.

(Insert Figure 5)

b) On-line MSPC based on Evolving MSPC models

MSPC models with increasing number of observation points are built adding the new current distillation point in every new model until all distillation process is covered. As illustrated in Figure 5(b), the first MSPC model is built using only the NIRS data matrix of the NOC historical data batches at the first recovered point (5 wt%), the second model using two observation points (5 and 6 wt%) and so on. For new batch monitoring, the data up to the current observation point are projected into the model for the related observations points and statistically tested.

c) On-line MSPC based on Fixed Size Moving Window, FSMW-MSPC, models

Several MSPC models built with a fixed size window (FSMW) including the current observation and several consecutive past observation points are built using the NOC historical data. The window slides one observation ahead in each new model until all observation points are covered. For instance, in Figure 5(c) the window moves from k to k = k + 1 and so on until k = K. For new batch monitoring, the data from the observation points covered by the moving window are projected into the model for the respective observations points and statistically tested.

The three approaches aim at on-line process control, but there are important differences due to the use of the different information in the models. Thus, the modality looking at individual process points does not take into account the neighbouring past observations and, hence, the evolution of the process. In the modalities FSMW-MSPC and evolving MSPC, the process evolution is taken into account and not only the new process observation of interest. In the case of the FSMW-MSPC model, only the recent past observation points (those within the window) are taken into account and the window size established is related to the number of relevant neighbouring process observations. Instead, the evolving-MSPC takes into account all process evolution until the present observation, giving potentially the same importance to all the past observations analyzed.

4. Results and discussion

4.1. Visual interpretation of distillation curve and NIR process data

Prior to chemometric analysis, the distillation curve and raw NIR spectra obtained during the distillation process were visually interpreted. Figure 2(a) illustrates the distillation curve of an on-specification batch (B07), the related process raw NIR spectra, Figure 2(b), and NIR spectra selected at four specific distillation points, Figure 2(c).

A sudden change in temperature can be observed through a simple visual inspection of the distillation curve between 60 and 70 wt%. This behavior is observed in gasoline-ethanol blends due to the formation of azeotropes of ethanol and hydrocarbons [37–40]. Distillation curve for gasoline-ethanol blend show three distinct regions: a plateau or azeotropic region (ethanol–hydrocarbon azeotropes are boiled) in the beginning of the distillation process, a transition region (sudden change in temperature) and a dilution-only region at the end of the distillation (after all added ethanol is boiled-off), as observed by French and Malone[38].

Four observation points (10, 50, 70 and 90 wt%) at the start and end of each distillation region were chosen to visualize the changes in NIR spectra with the evolution of distillation process. Figure 2(c) shows the complexity of the many superimposed absorption bands of the NIR spectra acquired during the distillation process. The bands around 1180 nm correspond to the second overtone, around 1400 nm to the 1st overtone combination and around 1700 nm to the first overtone region of carbon-hydrogen (C-H) bonds present in all points observed. The band around 2080 nm observed in the fractions at 10 and 50 wt% is related to the absorption of a combination of oxygen-hydrogen (O–H) stretching and bending from ethanol added to the gasoline. An absorbance increment in the band around 2080 nm was observed as the distillation was evolving from 10 to 50 wt%, mainly related to the increase of the ethanol relative concentration in the distilled fractions. A new band around 2170 nm appears in the spectra of the fraction at 70 and 90 wt%. This new band is related to absorption of aromatic compounds in the heavy fractions of the gasoline [41–43].

4.2. Process modeling of NIR data

Global process description (PCA model)

The NIR data were mean-centered and decomposed by PCA. Venetian blinds cross-validation method was used to find the number of principal components. 3 principal components explained 98.98 % (PC1 84.64%, PC2 13.11% and PC3 1.23%) representing a good summary of batch variability.

Figure 6(a) shows the principal components score plot distribution for PC1 and PC2 extracted from NIRS data of batches B01-09 used to build the PCA model (blue dots). The distribution of scores illustrates the process trajectory of on-specification batches and its variability. Because of the unstable distillation rate at the start of the distillation process, more variation was observed in these observation points as compared to the rest of the process. In addition, the NIRS data collected from the distillations of on-specification gasoline batch B11, not used in the PCA process modeling, and off-specification batch B13, which had only 15 %(v/v) of ethanol added, were projected in the PCA model. The scores obtained from PCA projection allowed the observation of the process trajectory of the new projected batches. Batch B11 (in magenta circles) was observed to follow the same on-specification process trajectory, while batch B13 (in red triangles) deviated from NOC trajectory, as illustrated in Figure 6(a). On-specification batch B10 when projected to the PCA model showed the same behavior as B11. Off-specification batches B12, B14-23 also deviate from NOC trajectory as

batch B13, (data not shown for clarity). The deviation becomes larger when the ethanol content is further from the ethanol specification level of NOC batches.

MCR-ALS

The dataset decomposition through MCR-ALS provides a model of process components easy to interpret and complementary to the global process description provided by PCA. The multibatch structure with the preprocessed (not mean-centered) data obtained from the distillation batches was decomposed by MCR-ALS. Four components were found through singular value decomposition, which agrees with the three contributions found in PCA of mean-centered data, since the rank decreases in one when mean centering is performed.

The four components concentration (distillation) and spectral profiles obtained after MCR-ALS decomposition of the multiset structure are shown in Figure 6(b). The components resolved from the distillation process are related to the main distilled fractions of gasolines "type C": First, light hydrocarbons; second, ethanol; third and fourth, mid to high molecular weight (MW) hydrocarbons and aromatic compounds, as reported elsewhere [44]. The identity of these compounds is confirmed when looking at the spectral features found in the related pure spectra and at the temperature distillation range.

The low MW hydrocarbons fraction is mainly distilled together with ethanol as azeotropes at the beginning of the distillation, i.e., at lower temperatures, as observed in the concentration profiles of components (1) and (2), see Figure 6(b). After 70 wt% of the distillation process, almost all ethanol, component (2), was boiled-off remaining most of the mid to high MW fractions of gasoline, rich in aromatic compounds, components (3) and (4). This region was observed in the distillation curves and is characterized by an increase in the slope of the distillation curve, as observed in Figure 2(a).

For comparison, Figure 6(b) shows the distillation profiles of B13 (with only 15 %(v/v) ethanol). Although the component spectra are the same, all distillation profiles are shifted to lower wt% of distillate, as expected for a batch with lower ethanol content.

4.3. Process control

Off-line batch process control

Off-line batch MSPC charts were built working with data coming from completed distillation batches. Specificity and sensitivity were adopted as quality parameters to assess the performance of MSPC charts for off-line batch process control. Specificity stands for the ratio of NOC batches (on-specification) correctly identified over the total NOC batches used to test the MSPC charts. Sensitivity is derived as the ratio of out of NOC (off-specification) batches correctly identified as out of NOC over the total out of NOC tested.

Process control starting information

The starting information used to build off-line batch MSPC models came either from distillation curves or NIR process data. The different starting information is depicted in section 3.2. Full distillation curves or observations within a selected temperature range were used to build off-line PCA-based MSPC models. Derivative form of the distillation curves was also used to improve the models. As for NIR information, full original preprocessed NIR spectra or selected spectral ranges were used to build the models. MSPC models were also built with the PCA scores, extracted from the process modeling by PCA, with all the distillation concentration profiles or only with the component related to ethanol, extracted from MCR-ALS decomposition, as described in section 3.2.

Off-line batch MSPC results

Table 2 shows the summary of the results using the different starting information to build and test off-line batch MSPC models.

(Insert Table 2)

An MSPC PCA model with mean-centered full distillation curve data (5-90 wt%) from NOC batches was built with 2 PC's and explained 90.91 % of data variance. MSPC chart based on $Q_{stat.}$. parameter correctly identified NOC and off-spec batches used to test the control charts as observed in Table 2 (row #1 has 100% specificity and sensitivity of $Q_{stat.}$). However, despite $D_{stat.}$. chart correctly identifies NOC batches, some off-spec batches are below the $D_{stat.}$ limit, see Figure 7(b), the sensitivity observed was 73.33%, Table 2 row #1. This may have happened because distillation curves of off-specification batches with ethanol concentration near to the on-specification level, 27 %(v/v), have extensive distillation ranges

with similar behavior (except for the points in the steepest zone of the curve) and, when considered the full curve, stay within the accepted variability of the NOC batches.

Another PCA model was built using the same data used previously, but this time preprocessed by Savitzky-Golay derivative and mean-centered. Results showed an improvement on the sensitivity, but still some batches were misidentified in the $D_{stat.}$chart, as reported in Table 2 row #2. All off-specification batches could be correctly identified using the derivative curve data only in the distillation range between 25 and 75 wt%, (Table 2 row #3). This range showed most of the variation in the distillation curves due to different ethanol content and avoided the unstability and, hence, undesired and non-composition related variability in the beginning of the distillation.

As observed in the MSPC charts built with the distillation curve data, the specificity and sensitivity of the $Q_{stat.}$. charts for all models built with NIRS data were 100%. However, different strategies were necessary to improve the sensitivity of $D_{stat.}$. MSPC charts.

The off-line PCA-based batch MSPC charts built using information from NIR spectra are explained below. Table 2, row #4, shows the results from a model built using the full preprocessed spectra (1103 – 2228 nm) and distillation (5-90 wt%) range. Despite of the 100% specificity in $Q_{stat.}$. chart, none of the off-specification batches was detected as faulty by the $D_{stat.}$. chart. The $D_{stat.}$ MSPC chart sensitivity was significantly improved to 75% when the NIRS data were reduced taking only the NIR observations within the distillation range from 60 wt% to 70 wt%, see Table 2, row #5. NIRS data were also reduced by selecting the most expressive spectral bands related mainly to hydrocarbons (1600-1800 nm) and ethanol absorption regions (2000-2200 nm). The MSPC chart built with this reduced spectral and distillation range improved the $D_{stat.}$ sensitivity to 83% (row #6), but still some samples with composition similar to the on-specification batches were missed by the control chart.

Off-line MSPC models were built with the NIR information compressed by PCA and MCR-ALS. Similar results were observed. The sensitivity for $D_{stat.}$. MSPC charts built with concatenated PCA scores or MCR-ALS concentration profiles (row #7 and #10) improved when compared with full spectral and distillation range data without data compression (row #4), see Table 2. MSPC models built with the compressed information extracted from the NIR observations within the 60-70 wt% distillation range showed an expressive improvement of the $D_{stat.}$ sensitivity to 91.67% (row #8 and #11). $D_{stat.}$. charts (row #9 and #12) showed 100% sensitivity when MSPC models were built using the Savitzky-Golay derivative of the

PCA scores or the MCR-ALS concentration profiles within the same distillation range (60-70%). MSPC models were built also using only the ethanol distillation profile. Results are show in Table 2, rows #13 and #14. The $D_{stat.}$ sensitivity was higher than in models built with all four components for models built with the full distillation range. Moreover, when the derivative ethanol profile in the 60-70 wt% distillation range, 100 % specificity in $D_{stat.}$ chart was achieved. The improvement of results when using only the ethanol concentration profile might be related to the better definition of this compound in the MCR-ALS results.

On-line batch MSPC on the NIR data

On-line batch MSPC control charts were built following the strategies described in section 3.2. For the distillation batches studied, PCA models were calculated for each observation point (86 models) following each one of the strategies described using the mean-centered data collected from NOC batches. Individual observation models (see Figure 5(a)) and evolving models (see Figure 5(b)) were calculated as described. For FSMW evolving models (see Figure 5(c)), the window selected enclosed 15 neighbouring observations. Thus, for observations nr. 1 to 14, PCA models were calculated as in the evolving strategy (see Figure 5(b)), whereas from observation nr. 15 and on, the full sliding window of 15 points was applied, as seen in Figure 5(c). PCA models for individual observation and FSMW evolving strategies were built with one principal component for all observation points, while in evolving models, one PC was used in evolving models from 5 to 20 wt% and three PC's in the remaining observation points. A confidence interval of 99% was considered to calculate the MSPC charts limits, $D_{CL99\%}$ and $Q_{CL99\%}$, for each model, as described earlier in section 3.2.

The NIR measurements for a new batch observation were mean-centered according to the mean of NOC batches and each observation (or set of observations) projected into the PCA model built for each strategy to extract the MSPC statistics, $D_{stat.}$ and $Q_{stat.}$.

At this point, it is important to stress the difference between off-line and on-line MSPC control charts.

Off-line MSPC control charts are based on a single PCA model built on the completed NOC batches. The final $D_{stat.}$ and $Q_{stat.}$ charts represent the values of these statistics *vs.* the batch index of each analyzed new batch. Every new batch is represented by a point.

On-line MSPC control charts display simultaneously the information of many PCA models, as many as observations in each batch, see Figure 5. Therefore, each new observation (NIR spectrum) acquired in a new batch is tested to see whether it is in- or out of control on a different PCA model. The process control is done at an observation level and not at a full batch level, as in the off-line approach. As a consequence, every new batch has a full $D_{stat.}$ and a full $Q_{stat.}$ plot, where the x-axis refers now to the different observations studied along the process evolution.

Control limits in $D_{stat.}$ and $Q_{stat.}$ would change per each new observation analyzed, since a different PCA model is used for projection every time. To facilitate visualization, the y-axis represents scaled values of $D_{stat.}$ and $Q_{stat.}$, defined as $D_{stat.}/D_{CL}$ and $Q_{stat.}/Q_{CL}$. In this way, a flat line at value 1 represents the control limits for all models used in $D_{stat.}$ and $Q_{stat.}$ for all observations. On-line $D_{stat.}$ and $Q_{stat.}$ charts, which represent the evolution of the related scaled statistics as a function of the observation (% distillate) analyzed, allow not only identifying on- and off-specification batches, but to know when the anomaly in an abnormal batch starts.

The results after monitoring new batches through the three different on-line MSPC strategies (individual observation model, FSMW MSPC evolving models and evolving MSPC models) are summarized in Table 3. Table 3 shows whether a new batch was diagnosed as on-specification or not and which MSPC chart ($D_{stat.}$, $Q_{stat.}$ or both) detected the fault. (Please note that the behavior of the full distillation batch is analyzed in this section for a better comparison of the three approaches. In a real on-line control context, the distillation would be stopped as soon as found to be out of specification).

Observing the information summarized in Table 3, $Q_{stat.}$ on-line MSPC charts detected correctly a fault in all off-specification batches by using any of the three on-line strategies. $D_{stat.}$ charts worked generally well, except when using evolving models, which were not able to detect fault in off-specification batches with ethanol content very approximate to the accepted specification and above, i.e. batches B17 to B23, see Table 1 and 3. The on-specification batch B11 was wrongly detected as faulty by the $Q_{stat.}$ on-line MSPC chart using the individual observation model MSPC strategy.

(Insert Table 3)

Figure 8 shows the D and Q statistics on-line control charts from distillation batches B11 (on-specification, with 27% ethanol added) and B18 (off-specification, with 25% ethanol added) for the three different on-line batch MSPC strategies.

Some comments need to be done for each on-line strategy according to the observed results.

a) Individual process observation models

$Q_{stat.}$ MSPC charts were observed to be very sensitive to fault detection. Besides, they show clearly the point where the batch starts to be anomalous. However, $Q_{stat.}$ charts were more prone to show false alarms, Figure 8(a). This happens because each model was built using a single observation point and a slight variation in an individual observation for a new batch process leads to a fault detection.

b) Evolving MSPC models

The evolving MSPC strategy considered the evolution of the process since the start, building models with increasing number of observations. This caused less sensitive $D_{stat.}$ charts, since past NOC observations may have a lot of weight in the models and batches can be detected easily as faulty only when the fault observations occurred at the beginning of the distillation process (batches B11-16).Batches with ethanol concentration near to the specification value and above were not detected by $D_{stat.}$ charts since the fault occurred too late and was not large enough to compensate the weight of the large number of initial NOC observations. Despite of this fact, the evolution of $D_{stat.}$ values for undetected off-specification batches show a different trend (a clear increase when the abnormal behavior starts) as compared to on-specification batches (presenting a flat constant tendency), as observed in $D_{stat.}$ charts Figure 8(b). This may suggest that the $D_{stat.}$ chart could still be used in these instances if the control limits were set empirically. $Q_{stat.}$ charts performed more satisfactorily in fault detection. However, faults were detected later than in individual observation models due to the excessive weight of past NOC observations as well.

c) Fixed size moving window, FSMW-MSPC models

The FSMW strategy considered only a few past observation points, set according to the window size (15 observations were used in this study). This feature produced more sensitive $D_{stat.}$ charts because past NOC observations had less weight in models, Figure 8(c). FSMW strategy was observed to be less prone to false alarms on $Q_{stat.}$ charts than individual

observation charts, Figure 8(a), since individual point fluctuations have less impact in the window-based PCA models. This strategy has been found to be the most flexible of the three, showing efficient and easy detection of faults and avoiding false alarms. Obviously the performance of this approach may depend on the width of the window: if too small, false alarms may show up in analogy to what happens in individual observation models; if too big, sensitivity in $D_{stat.}$ chart may decrease because of the weight of too many past NOC observations in the chart. This inconvenient is clearly surmountable if the window width is set by using representative off-spec batches that may allow setting the correct window width to avoid the malfunctions described in the other two on-line MSPC approaches.

5. Conclusions

The present work provides an improvement of PAT technologies for distillation-based quality control procedures through the design of an automatic distillation device that allowed synchronized measurements of the distillated mass percentage, distillation temperature and NIR spectra during the distillation process.

Process modeling on NIR spectra by PCA and MCR-ALS allowed understanding the process evolution from a global (scores plot) and component-wise (distillation profiles) point of view, respectively. In this sense, MCR-ALS provides a good thermal and physicochemical characterization of distilled fractions, even if coming from a simple distillation process.

MSPC strategies based on the different kinds of data obtained from the designed device are proposed. Off-line models using distillation curves were able to detect off-specification batches when suitable preprocessing and distillation curve range were used. Successful off-line MSPC models were built with the NIR spectra information compressed into PCA scores or MCR-ALS concentration profiles. The possibility to perform a sensible selection of some of the MCR-ALS concentration profiles, linked to particularly relevant process contributions has proven to improve MSPC results.

On-line batch MSPC strategies were proposed for fault detection during the distillation process using the collected NIRS data. Individual process observations MSPC models showed $D_{stat.}$charts very sensitive to fault detection; however, false alarms were observed in the $Q_{stat.}$charts. Evolving MSPC models were able to solve the false alarms observed with the individual observation strategy, but failed to detect some off-specification batches with similar composition to NOC batches when using $D_{stat.}$charts. The FSMW-MSPC approached used a flexible combination of the other two strategies and succeeded to

detect all off-specification batches and correctly identify on-specification batches during the test with both $D_{stat.}$ and $Q_{stat.}$ control charts, avoiding false alarms.

6. Acknowledgements

R.R. de Oliveira acknowledges the EMQAL Grant. EMQAL is a Joint European Master Programme selected under Erasmus Mundus coordinated by University of Barcelona. Funding support from the European Community's Framework programme for Research and Innovation Horizon 2020 SPIRE - Integrated Process Control (ProPAT), grant Agreement no. 637232 and Spanish government through project CTQ2015-66254-C2-2-P are also acknowledged. K.M.G. Lima acknowledges the CNPq (Grant 305962/2014–4) for financial support. A.O. Sousa acknowledges the PETROBRAS (Grant 2012/00217-7) for financial support.

7. References

[1] ASTM D86-15, Standard Test Method for Distillation of Petroleum Products and Liquid Fuels at Atmospheric Pressure, in: ASTM B. Stand., 2015.

[2] C. Pasquini, S.H.F. Scafi, Real-Time Monitoring of Distillations by Near-Infrared Spectroscopy, Anal. Chem. 75 (2003) 2270–2275.

[3] V.A. Corro-Herrera, J. Gómez-Rodríguez, P.M. Hayward-Jones, D.M. Barradas-Dermitz, M.G. Aguilar-Uscanga, A.C. Gschaedler-Mathis, *In-situ* monitoring of *Saccharomyces cerevisiae* ITV01 bioethanol process using near-infrared spectroscopy NIRS and chemometrics, Biotechnol. Prog. 32 (2016) 510–517.

[4] H. Xiong, X. Gong, H. Qu, Monitoring batch-to-batch reproducibility of liquid-liquid extraction process using in-line near-infrared spectroscopy combined with multivariate analysis, J. Pharm. Biomed. Anal. 70 (2012) 178–187.

[5] RESOLUÇÃO ANP N^{o} 40, DE 25.10.2013 DOU 28.10.2013, http://nxt.anp.gov.br/NXT/gateway.dll/leg/resolucoes_anp/2013/outubro/ranp 40 - 2013.xml?f=templates$fn=document-frame.htm$3.0$q=$x=$nc=2230. Accessed August 1, 2016.

[6] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, Chemom. Intell. Lab. Syst. 2 (1987) 37–52.

[7] R. Tauler, M. Maeder, A. de Juan, Multiset Data Analysis: Extended Multivariate Curve Resolution, in: Compr. Chemom. Chem. Biochem. Data Anal. Four-Volume Set. Vol. 2, Chapter 2.24, S.D. Brown, R. Tauler, B. Walcz., Elsevier, 2009: pp. 473–505.

[8] Y. Jin, Z. Wu, X. Liu, Y. Wu, Near infrared spectroscopy in combination with chemometrics as a process analytical technology (PAT) tool for on-line quantitative monitoring of alcohol precipitation, J. Pharm. Biomed. Anal. 77 (2013) 32–39.

[9] M.C. Sarraguça, P.R.S. Ribeiro, A.O. dos Santos, J.A. Lopes, Batch Statistical Process Monitoring Approach to a Cocrystallization Process, J. Pharm. Sci. 104 (2015) 4099–4108.

[10] H. Howland, S.W. Hoag, Analysis of curing of a sustained release coating formulation by application of NIR spectroscopy to monitor changes physical-mechanical properties, Int. J. Pharm. 452 (2013) 82–91.

[11] R. Kona, H. Qu, R. Mattes, B. Jancsik, R.M. Fahmy, S.W. Hoag, Application of in-line near infrared spectroscopy and multivariate batch modeling for process monitoring in fluid bed granulation, Int. J. Pharm. 452 (2013) 63–72.

[12] T. Kourti, J.F.J.F. MacGregor, Process analysis, monitoring and diagnosis, using multivariate projection methods, Chemom. Intell. Lab. Syst. 28 (1995) 3–21.

[13] A. AlGhazzawi, B. Lennox, Monitoring a complex refining process using multivariate statistics, Control Eng. Pract. 16 (2008) 294–307.

[14] A. AlGhazzawi, B. Lennox, Model predictive control monitoring using multivariate statistics, J. Process Control 19 (2009) 314–327.

[15] T.C. Ávila, R.J. Poppi, I. Lunardi, P.A.G. Tizei, G.A.G. Pereira, Raman spectroscopy and chemometrics for on-line control of glucose fermentation by Saccharomyces cerevisiae, Biotechnol. Prog. 28 (2012) 1598–1604.

[16] J. Alves-Rausch, R. Bienert, C. Grimm, D. Bergmaier, Real time in-line monitoring of large scale Bacillus fermentations with near-infrared spectroscopy, J. Biotechnol. 189 (2014) 120–128.

[17] H. Huang, H. Qu, In-line monitoring of alcohol precipitation by near-infrared spectroscopy in conjunction with multivariate batch modeling, Anal. Chim. Acta 707 (2011) 47–56.

[18] P. Nomikos, J.F. MacGregor, Multivariate Statistical Process Control Charts for Monitoring Batch Processes, Technometrics 37 (1995) 41–59.

[19] P. Nomikos, J.F. MacGregor, Monitoring batch processes using multiway principal component analysis, AIChE J. 40 (1994) 1361–1375.

[20] P. Nomikos, J.F. MacGregor, Multi-way partial least squares in monitoring batch processes, Chemom. Intell. Lab. Syst. 30 (1995) 97–108.

[21] S. Wold, N. Kettaneh, H. Friden, A. Holmberg, Modelling and Diagnostics of Batch Processes and Analogous Kinetic Experiments, Chemom. Intell. Lab. Syst. 44 (1998) 331–340.

[22] S. Rännar, J.F. MacGregor, S. Wold, Adaptive batch monitoring using hierarchical PCA, Chemom. Intell. Lab. Syst. 41 (1998) 73–81.

[23] L. Zhao, T.-Y. Chai, G. Wang, A Nonlinear Modeling and Online Monitoring Method for the Batch Process Using Multiple Local PCA, (2003) 2–5.

[24] L. Zhao, T. Chai, Adaptive moving window MPCA for online batch monitoring, in: 2004 5th Asian Control Conf., 2004: p. 1290–1295 Vol.2.

[25] M. Maeder, Evolving Factor Analysis for the Resolution of Overlapping Chromatographic Peaks, Anal. Chem. 59 (1987) 527–530.

[26] M. Maeder, A. de Juan, Two-Way Data Analysis: Evolving Factor Analysis, Compr. Chemom. 2 (2010) 261–274.

[27] A. Savitzky, M.J.E. Golay, Smoothing and Differentiation of Data by Simplified Least Squares Procedures., Anal. Chem. 36 (1964) 1627–1639.

[28] A. de Juan, J. Jaumot, R. Tauler, Multivariate Curve Resolution (MCR). Solving the mixture analysis problem, Anal. Methods 6 (2014) 4964.

[29] A. de Juan, R. Tauler, Chemometrics applied to unravel multicomponent processes and mixtures, Anal. Chim. Acta 500 (2003) 195–210.

[30] R. Tauler, B.R. Kowalski, S. Fleming, Multivariate Curve Resolution Applied to Spectral Data from Multiple Runs of an Industrial Process, Anal. Chem. 65 (1993) 2040–2047.

[31] W. Windig, J. Guilment, Interactive Self-Modeling Mixture Analysis, Anal. Chem. 63 (1991) 1425–1432.

[32] J. Jaumot, A. de Juan, R. Tauler, MCR-ALS GUI 2.0: New features and applications, Chemom. Intell. Lab. Syst. 140 (2014) 1–12.

[33] T. Kourti, 4.02 - Multivariate Statistical Process Control and Process Control, Using Latent Variables, in: Compr. Chemom., 2009: pp. 21–54.

[34] S. Wold, Cross -Validatory Estimation of the Number of Components in Factor and Principal Components Models, Technometrics 20 (1978) 397–405.

671 [35] J.E. Jackson, A User's Guide to Principal Components, Wiley, New York, 1991.

672 [36] J.E. Jackson, G.S. Mudholkar, Control Procedures for Residuals Associated with
673 Principal Component Analysis, Technometrics 21 (1979) 341–349.

674 [37] E. V. Takeshita, R.V.P. Rezende, S.M.A.G.U. de Souza, A.A.U. de Souza, Influence of
675 solvent addition on the physicochemical properties of Brazilian gasoline, Fuel 87
676 (2008) 2168–2177.

677 [38] R. French, P. Malone, Phase equilibria of ethanol fuel blends, 229 (2005) 27–40.

678 [39] R.M. Balabin, R.Z. Syunyaev, S.A. Karpov, Quantitative measurement of ethanol
679 distribution over fractions of ethanol-gasoline fuel, Energy and Fuels 21 (2007) 2460–
680 2465.

681 [40] V.F. Andersen, J.E. Anderson, T.J. Wallington, S.A. Mueller, O.J. Nielsen, Distillation
682 curves for alcohol-gasoline blends, Energy and Fuels 24 (2010) 2683–2691.

683 [41] Z. Xiaobo, Z. Jiewen, M.J.W. Povey, M. Holmes, M. Hanpin, Variables selection
684 methods in near-infrared spectroscopy, Anal. Chim. Acta 667 (2010) 14–32.

685 [42] C. Pasquini, Near Infrared Spectroscopy: fundamentals, practical aspects and analytical
686 applications, J. Braz. Chem. Soc. 14 (2003) 198–219.

687 [43] D.A. Burns, E.W. Ciurczak, Handbook of near-infrared analysis, 3rd ed., Anal.
688 Bioanal. Chem. 393 (2009) 1387–1389.

689 [44] J.L. Burger, N. Schneider, T.J. Bruno, Application of the advanced distillation curve
690 method to fuels for advanced combustion engine gasolines, Energy and Fuels 29
691 (2015) 4227–4235.

692

8. List of Tables

Table 1 Description of Batch ID and their related composition, as used in this work.

Batch ID	%(v/v) Gasoline	%(v/v) Ethanol	Class
B01-B11	73	27	On-specification
B12	90	10	
B13	85	15	
B14-B16	80	20	
B17-B18	75	25	Off-specification
B19-B20	70	30	
B21-B22	65	35	
B23	60	40	

Table 2 Off-line batch MSPC results.

#	Description	NC	%EV	%Specificity		%Sensitivity	
				$D_{stat.}$	$Q_{stat.}$	D_{stat}	$Q_{stat.}$
	Distillation Curves						
1	DistRange (5-90 wt%)	2	90.91	100	100	66.67	100
2	DistRange(5-90 wt%)_1st diff.SG	2	94.83	100	100	91.67	100
3	DistRange(25-75 wt%)_1st diff.SG	2	98.36	100	100	100	100
	NIR information						
	Unfolded NIRS data						
4	FullSpec_DistRange(5-90%)	3	68.91	100	100	0.00	100
5	FullSpec_DistRange(60-70%)	2	92.56	100	100	75.00	100
6	SelSpec_DistRange(60-70%)	2	94.54	100	100	83.33	100
	scores from PCA modelling (3 PC's)						
7	DistRange(5-90%)	2	87.53	100	100	66.67	100
8	DistRange(60-70%)	2	96.52	100	100	91.67	100
9	DistRange(60-70%)_1st diff.SG	2	98.67	100	100	100	100
	C profiles from MCR-ALS modelling (4comp)						
10	DistRange(5-90%)	2	81.07	100	100	33.3	100
11	DistRange(60-70%)	2	94.33	100	100	91.67	100
12	DistRange(60-70%)_1st diff.SG	2	96.99	100	100	100	100
13	DistRange(5-90%)_EtOHcomp	2	90.60	100	100	58.33	100
14	DistRange(60-70%)_1st diff.SG_EtOHcomp	2	99.65	100	100	100	100

Model number, **NC** number of PCA principal components, **%EV** cumulative explained variance by NC principal components, **diff.SG** Savitzky-Golay derivative, **DistRange** Distillation Range used to build the model, **FullSpec** Complete NIRS measurement range, **SelSpec** Small more selective to ethanol signal, **EtOHcomp** component 2 related to ethanol.

Table 3 On-line batch MSPC results on test batches B10-23.

Test Batch	Method		
	Ind. Obs.[a]	FSMW[b]	Evolving[c]
on-spec			
B10	on-spec.[d]	on-spec.	on-spec.
B11	$Q_{stat.}$	on-spec.	on-spec.
off-spec			
B12	$Q_{stat.}$, $D_{stat.}$	$Q_{stat.}$, $D_{stat.}$	$Q_{stat.}$, $D_{stat.}$
B13	$Q_{stat.}$, $D_{stat.}$	$Q_{stat.}$, $D_{stat.}$	$Q_{stat.}$, $D_{stat.}$
B14	$Q_{stat.}$, $D_{stat.}$	$Q_{stat.}$, $D_{stat.}$	$Q_{stat.}$, $D_{stat.}$
B15	$Q_{stat.}$, $D_{stat.}$	$Q_{stat.}$, $D_{stat.}$	$Q_{stat.}$, $D_{stat.}$
B16	$Q_{stat.}$, $D_{stat.}$	$Q_{stat.}$, $D_{stat.}$	$Q_{stat.}$, $D_{stat.}$
B17	$Q_{stat.}$, $D_{stat.}$	$Q_{stat.}$, $D_{stat.}$	$Q_{stat.}$
B18	$Q_{stat.}$, $D_{stat.}$	$Q_{stat.}$, $D_{stat.}$	$Q_{stat.}$
B19	$Q_{stat.}$, $D_{stat.}$	$Q_{stat.}$, $D_{stat.}$	$Q_{stat.}$
B20	$Q_{stat.}$, $D_{stat.}$	$Q_{stat.}$, $D_{stat.}$	$Q_{stat.}$
B21	$Q_{stat.}$, $D_{stat.}$	$Q_{stat.}$, $D_{stat.}$	$Q_{stat.}$
B22	$Q_{stat.}$, $D_{stat.}$	$Q_{stat.}$, $D_{stat.}$	$Q_{stat.}$
B23	$Q_{stat.}$, $D_{stat.}$	$Q_{stat.}$, $D_{stat.}$	$Q_{stat.}$

[a] Ind.Obs., Individual observation MSPC models.
[b] FSMW, fixed size moving window MSPC models.
[c] Evolving MSPC models.
[d] On-spec means the batch is on-specification according to both D_{stat} and $Q_{stat.}$. D_{stat} means the batch is off-spec according to D_{stat} chart. $Q_{stat.}$ means the batch is off-spec. according to $Q_{stat.}$ chart.

9. List of Figures

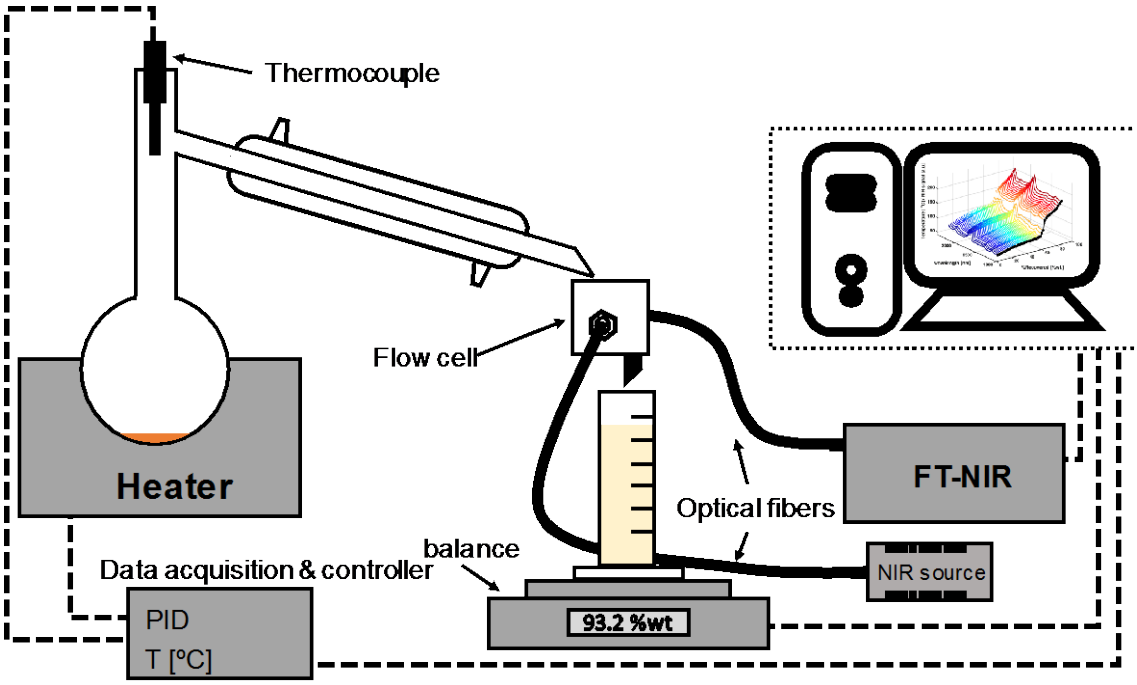

Figure 1 Experimental setup of the automatic distillation device with on-line NIRS monitoring.

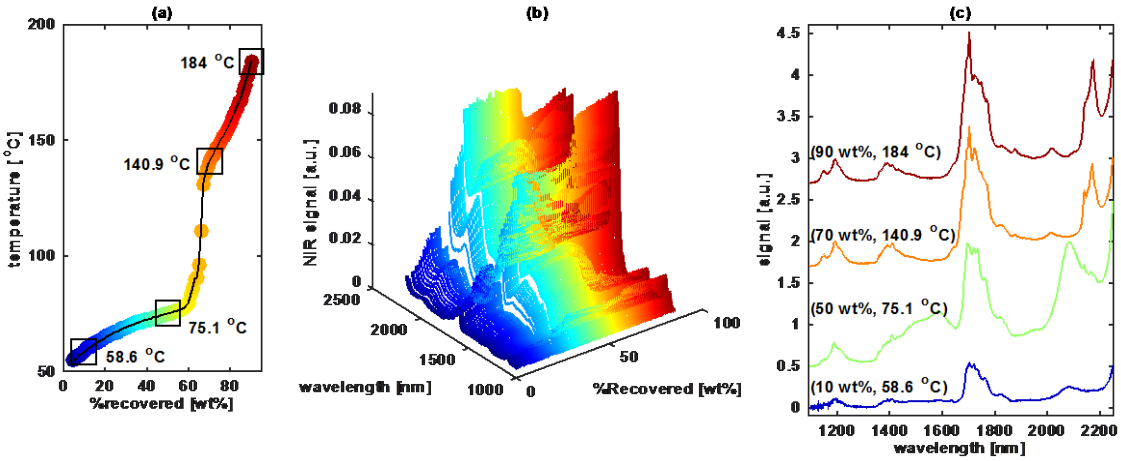

Figure 2 Process data from distillation batch B07. (a) Distillation curve, (b) On-line raw NIR spectra *vs.* percentage of the distilled fraction [wt%] and (c) raw NIR spectra at distilled fraction at 10, 50, 70 and 90 wt% indicated in (a), the spectra were vertically offsetted for clear comparison.

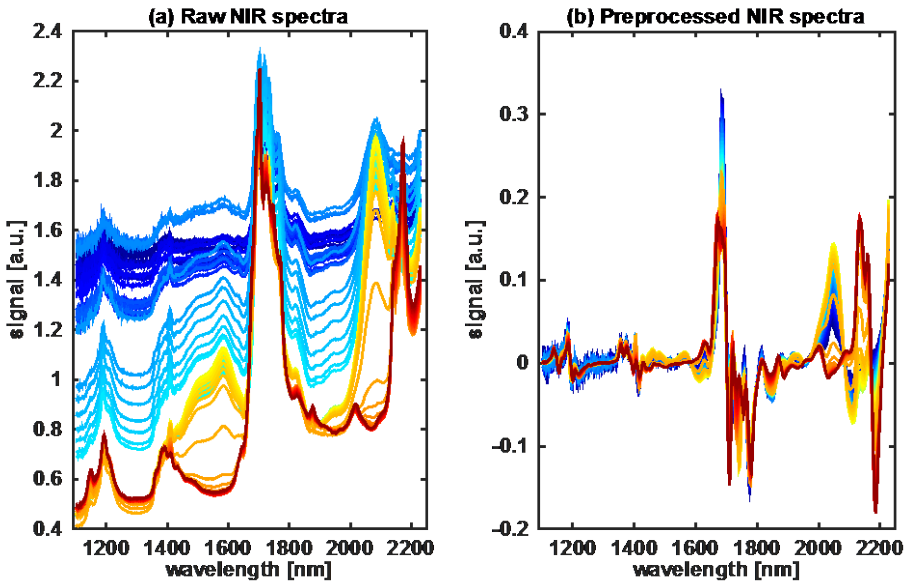

Figure 3 Plot of the (a) raw and (b) preprocessed NIR spectra obtained from the distillation batch B07 between 5 – 90 wt% with 1 wt% interval, 5 wt% (blue) → 90 wt% (red).

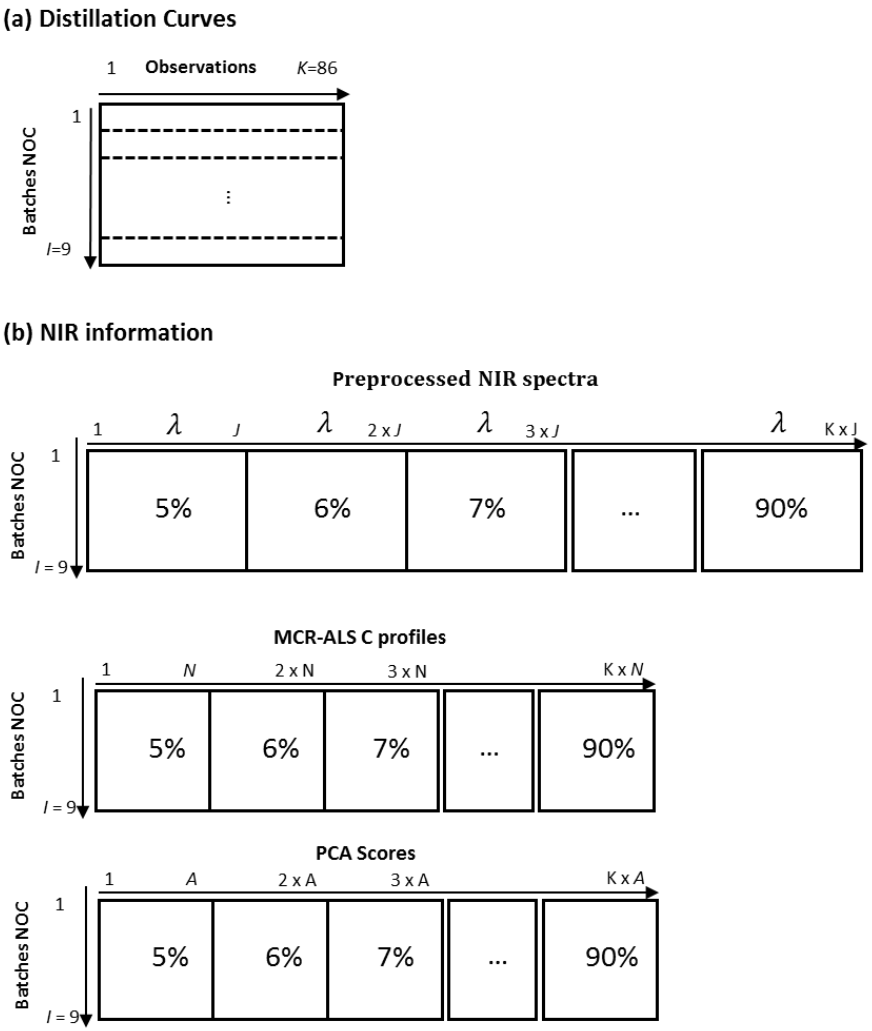

Figure 4 **Different starting information used to build off-line PCA-based batch MSPC models (a) Distillation curves, (b) NIR information, from top to bottom: Original preprocessed NIR variables, concentration profiles from MCR-ALS and scores from PCA extracted from the multibatch structure for process modeling.**

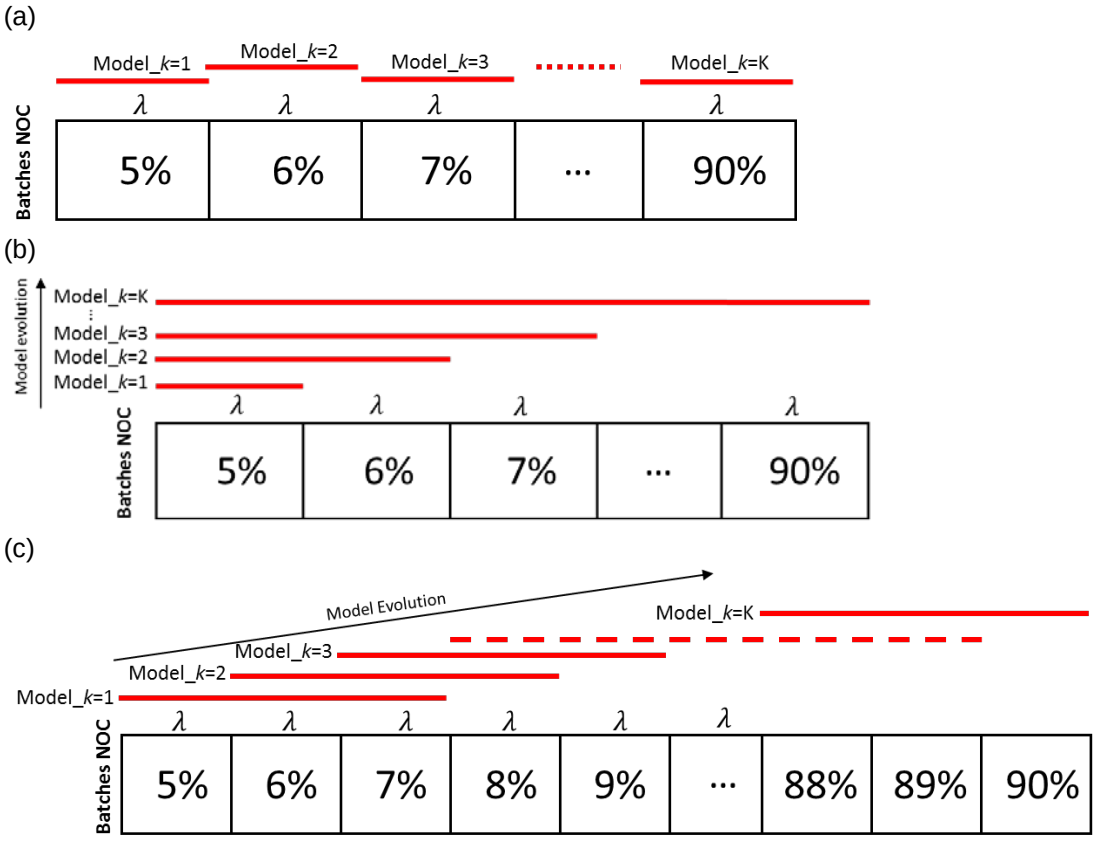

Figure 5 Different evolving on-line MSPC models approaches. a) Individual process observation models, b) Evolving MSPC models and c) FSMW-MSPC models.

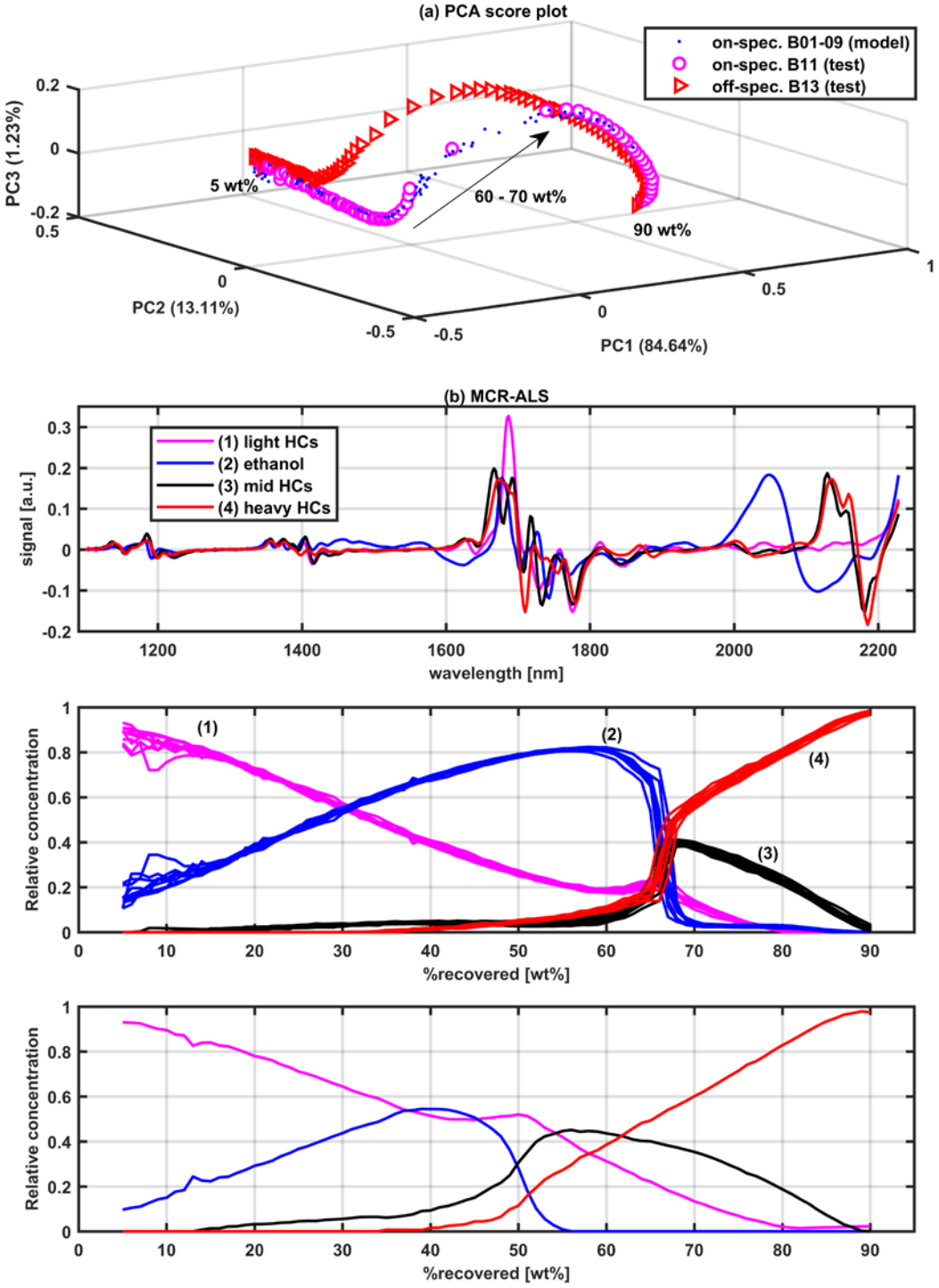

Figure 6 Process modeling. (a) PCA map of the 3PC's scores from on-specification batches B01-09 used to build PCA model, and after projection in the PCA model batches B11 (on-specification) and B13 (off-specification). Start (5 wt%), transition region (60-70 wt%) and end (90 wt%) of distillation process are indicated; (b) Multibatch MCR-ALS showing from top to bottom the pure spectral profiles, the superimposed concentration profiles for on-specification batches B01-09 and concentration profiles for off-specification batch B13 (components (1) to (4)).

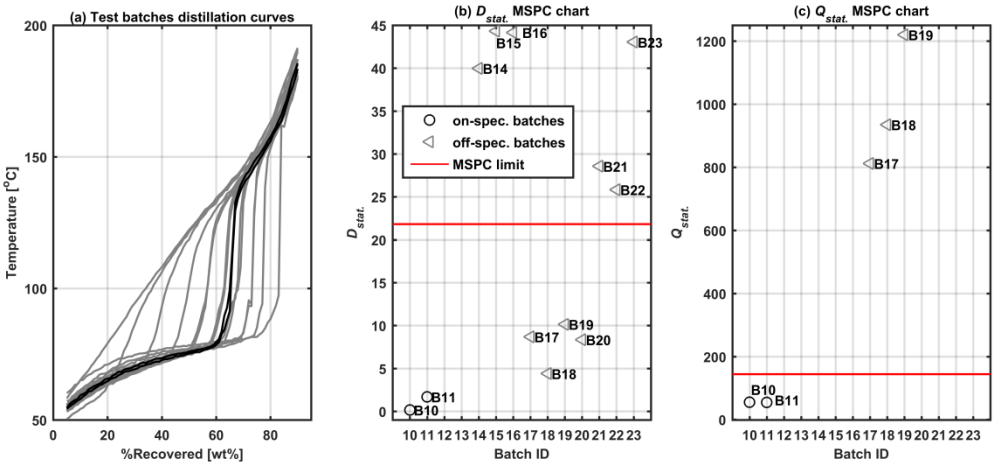

Figure 7 (a) Full distillation curves used to test PCA-based off-line batch MSPC model, on-specification in black and off-specification in gray, (b)$D_{stat.}$ and (c) $Q_{stat.}$ MSPC charts. Some batches show higher $D_{stat.}$ and $Q_{stat.}$ values and are not shown for better visualization of the control limits.

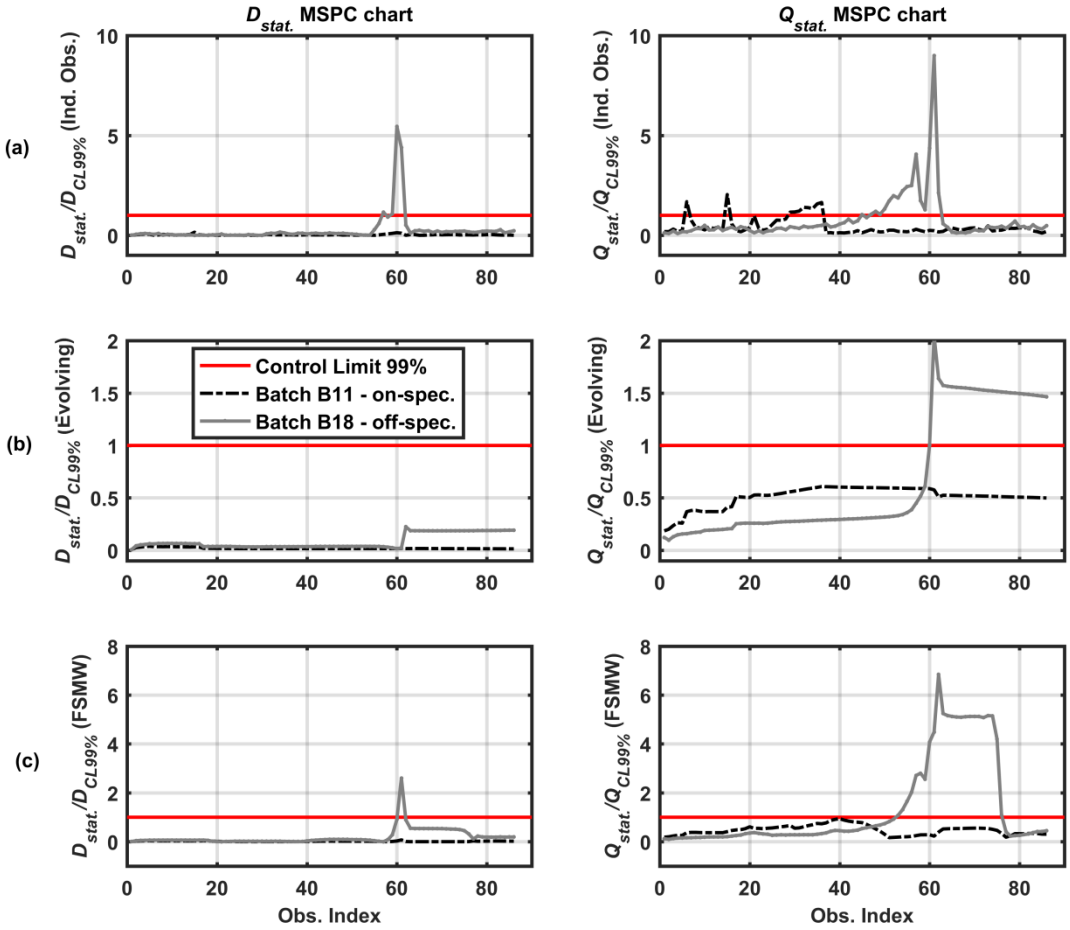

Figure 8 On-line MSPC charts for batches B11 (on-specification) and B18 (off-specification) for the a) Individual process observation models, b) Evolving MSPC strategies and c). FSMW evolving MSPC.