Using warping information for batch process monitoring and fault classification

J.M. González-Martínez[a,*], J. A. Westerhuis[b], A. Ferrer[a]

[a]*Departamento de Estadística e Investigación Operativa Aplicadas y Calidad, Universidad Politécnica de Valencia, Camino de Vera s/n, 46022, Valencia, Spain*
[b]*Biosystems Data Analysis, Swammerdam Institute for Life Sciences, University of Amsterdam, 1090 GE Amsterdam, The Netherlands*

Abstract

This paper discusses how to use the warping information obtained after batch synchronization for process monitoring and fault classification. The warping information can be used for i) building unsupervised control charts or ii) fault classification when a rich faulty batches database is available. Data from realistic simulations of a fermentation process of the *Saccharomyces cerevisiae* cultivation are used to illustrate the proposal.

Keywords: Fault classification, Batch synchronization, Warping information, Relaxed Greedy Time Warping, Batch Multivariate Statistical Process Control.

1. Introduction

The application of latent structures-based bilinear models for batch process understanding, troubleshooting and monitoring has been widely studied and discussed in the literature [1, 2]. Also, over the last decade, there has been an increasing interest from industry to apply this methodology to their processes for a successful enhancement of the final product quality, both in off-line and on-line applications.

In batch processes, measurements belonging to J process variables are collected at K different sampling points over I batches. These data can be used to design monitoring systems using projection to latent structures-based methods, like Principal Component Analysis (PCA) and Partial Least Squares (PLS). Commonly, batches have different durations since time is not a crucial factor for the completion of a batch. Instead other criteria, such as the achievement of a considered amount of product or a temperature in a process phase or stage are used for completion. Due to the variability in the chemical composition of the raw material, variation in environmental conditions and/or cleanness of equipment, among others, these criteria are reached at different times across batches. In order to ensure the correspondence of all the process variables at any point throughout one batch to those at the same state in other batches (i.e. the synchronization of the key process events), the synchronization prior to bilinear modeling is required.

The approaches for synchronizing batch data can be roughly classified into three categories: i) methods based on compressing/expanding the raw trajectories using linear interpolation either in the batch time dimension [3, 4] or in an indicator variable dimension (the so-called Indicator Variable-based synchronization, IV) [5]; ii) methods based on features extraction [6, 7, 8, 9]; and iii) methods based on stretching, compressing and translating pieces of trajectories, such as Dynamic Time Warping (DTW) [10] and Relaxed Greedy Time Warping (RGTW) [11].

Once batch data are synchronized by using one of the synchronization methods mentioned previously, a 3-way data matrix $\mathbf{X}$ ($I \times J \times K$) is available for the subsequent multivariate analysis. Before batch

---
*Corresponding author
Email address: `jogonmar@gmail.com` (J.M. González-Martínez)

modeling, some issues must be addressed due to the nature of batch data, i.e. nonlinear and time-varying dynamics, and complex variable correlations structure varying throughout the batch run. After batch data have been synchronized and modeled, a monitoring scheme is built. Typically, two Shewhart control charts based on the Hotelling-T^2 and Squared Prediction Error (SPE) statistics are designed. Their control limits (thresholds) are estimated from normal operating conditions (NOC) process data and later adjusted using cross-validation techniques for an imposed significance level (ISL) [12, 13]. Once the scheme is designed, new measurements from a new batch data can be projected onto the data model, yielding to the aforementioned multivariate statistics, to check for the correct performance of the process.

The set of warping time profiles obtained from the batch synchronization are seldom used in the monitoring scheme. These warping functions provide valuable information about the process pace of each of the batches in the study (the so-called warping information). In particular, these profiles give information on the performance of the different stages or phases of the process throughout the batch run. This may be related to the appearance of faults during the process and may have a direct effect in product quality. Hence, a study of the set of warping profiles obtained from the off-line and on-line synchronization is highly desired. Some authors have emphasized the importance of not discarding the information derived from the synchronization [10] and others have used this warping information as an extra variable in the multivariate analysis [11, 14]. Nevertheless, there is no sound study on the use of the warping information for: i) unsupervised (i.e. requiring no a priori knowledge about the type of faults) process monitoring, and ii) supervised (i.e. incorporating prior knowledge from a data base of historical faults) fault classification. This is the main goal of this paper.

Section 2 explains how to obtain the warping information from the RGTW-based synchronization and introduces a novel NOC warping information-based control chart as a complementary tool for unsupervised end-of-batch and real-time process monitoring. Section 3 is devoted to illustrate how to use the warping information for end-of-batch supervised fault classification using different methods. Section 4 illustrates the application of the different approaches using data from a biofermentation simulator of the *Saccharomyces cerevisiae* cultivation. Finally, some conclusions are drawn in Section 5.

2. NOC warping information-based control chart for process monitoring

The RGTW algorithm [11] is used in this paper for batch synchronization due to its versatility both for end-of-batch and real-time applications. This algorithm synchronizes trajectories of a test batch $\mathbf{B}_{new}(K_{new} \times J)$ against a reference batch $\mathbf{B}_{ref}(K_{ref} \times J)$, by finding a minimum cost function (or warping path) $\mathbf{f}_{new}^T = \{w(1), w(k), ..., w(K_{w_{new}})\}$. Here each $w(k)$ is an ordered pair $[i(k), j(k)]$ indicating that the i-th and j-th sampling point belonging to $\mathbf{B}_{ref}$ and $\mathbf{B}_{new}$, respectively, are synchronized. The synchronization is assessed with a weighted local cost function $d(i, j)$, represented as a $K_{ref} \times K_{new}$ local distance matrix, which assigns a matching cost for synchronizing each possible pair sampling points from the reference and test batches. Several constraints are used to restrict the search of the warping path, namely a band fit to the batch variability that constrains the search space of such path, and local constraints (or predecessors) which define the warping function as monotonic and continuous. Additionally, a cumulative weighted distance matrix $\mathbf{D}(f_{new})$ is assessed by estimating the cumulative matching costs of each of the allowed paths f_{new}. The optimal warping path $\mathbf{f}_{new}^*$ (also called warping profile, or warping information) is assessed by obtaining the path that minimizes the cumulative distance. This synchronization procedure is carried out within a moving window ζ with a defined width, which is optimized by cross-validation [11].

The warping profiles obtained from the RGTW-based synchronization are composed of a set of different transitions at each sampling point, i.e. vertical, horizontal and diagonal steps. Based on the number of the different transitions the warping path contains, conclusions regarding the good performance of the different process stages can be drawn. Let us assume that the test and reference batch are located on the x-axis and y-axis, respectively. In this scenario, an excessive number of vertical or horizontal transitions (data expansion or compression, respectively) in the warping profiles belonging to the calibration batches in relation to the reference batch, means that the process has needed less or more time, respectively, to release the product. Hence, the use of this information in batch process monitoring may be valuable.
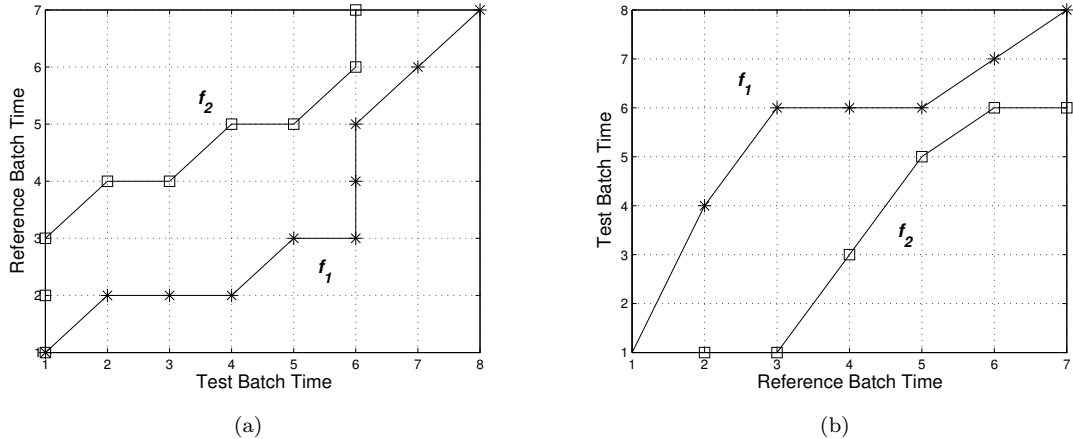
Figure 1: Warping profiles f_1 and f_2 belonging to two different trajectories obtained from the RGTW-based synchronization (a) and expressed as a function of the reference time (b).

RGTW-based synchronization provides not only the synchronized multivariate measurements, but also the optimal warping functions $\mathbf{f}_i$ derived for each of I batches. For each of the test batches synchronized against the reference batch, a warping function with length equal to K_{w_i} is obtained. The length of the warping information among batches is different. In order to use this information for building the NOC warping information-based control chart (WICC) and for designing the fault classifiers, all the warping profiles need to have the same length. Hence, the warping profiles must be expressed as function of the reference batch to have equal length. This transformation is performed as follows (see Figure 1). The test batch sampling point that matches with each one of the reference batch sampling points is estimated. In the case that a set of consecutive horizontal transitions are present, i.e. n test batch sampling points are matched with the k_{ref}-th reference batch time point, the last time point of this set is taken as the matched point. Note that the interpretation of this consecutive test batch sampling points matching a certain reference batch sampling point is different in the inner RGTW algorithm. In that case, an average of the values belonging to the multivariate batch trajectories is calculated and matched with a defined reference batch sampling point. Hence, warping profiles transformation is not suitable for interpreting synchronization outcomes since it may cause misleadings. In the example in Figure 1, 2nd-4th sampling points of the test batch are matched with the 2nd sampling point of the reference batch (see warping profile f_1 in Figure 1(a), line depicted with asterisks), hence, the 4th sampling point of the test batch is matched with the 2nd sampling point of the reference batch (see warping profile f_1 in Figure 1(b), line depicted with asterisks). This procedure is repeated for each one of the K_{ref} reference batch sampling points over all I_{NOC} test batches. At the end of the execution, a set of warping profiles with equal length is available (see Figure 1(b)). Once this transformation has been performed, the matrix $\mathbf{F}_{NOC}$ $(I_{NOC} \times K_{ref})$ containing the I_{NOC} warping profiles expressed as a function of the reference batch is obtained. Data containing this matrix define the consistent and normal processing pace through the batch time. For the sake of interpretability, the monotonic increasing behavior of the warping profiles is removed by subtracting the average values of each one of the reference batch points (columns of matrix $\mathbf{F}_{NOC}$). From this centered $\mathbf{F}_{NOC}$ matrix in this paper we propose to build the NOC-WICC. This is a complementary tool to Hotelling-T^2 and SPE control charts for end-of-batch process monitoring. The corresponding control limits at 99% confidence level can be assessed by estimating the percentile 0.5 and 99.5 at each reference batch time (columns of matrix $\mathbf{F}_{NOC}$).

In real-time applications, NOC-WICC can be also used for unsupervised process monitoring. In this context, a new point from the warping information is available when a set of time points from an ongoing batch are matched with the next k_{ref}-th time point from the reference batch. Hence, the real-time monitoring of the warping profile would have a certain delay from the original batch time. This is necessary in order

3

to ensure that the ongoing warping profile has the same length as those corresponding to NOC and can be monitored with NOC-WICC.

3. Fault classification procedures

Let us assume that a set of warping profiles derived from the RGTW-based synchronization of batch trajectories belonging to historical faulty batches with different types of faults l (l=1, 2, ... , L) are collected in the matrices $\mathbf{F}_1$, $\mathbf{F}_2$, ..., $\mathbf{F}_L$. This warping information can also be used for fault classification. Once the monitoring system has detected an out-of-control signal, supervised methods can be used to classify the type of fault that occurred in end-of-batch applications.

In this paper, three different supervised procedures are compared. In the first procedure, warping information is used to build supervised faulty warping information-based control charts (faulty WICC). In the other two procedures, warping information is used to fit classification models using standard chemometrics tools, such as Partial Least Squares-Discriminant Analysis (PLS-DA)[15] [15] or Soft Independent Modelling of Class Analogy (SIMCA) [16]. For the design of these classifiers, each one of matrices $\mathbf{F}_l$ ($I_l \times K_{ref}$) is split up into two different data sets, a training and test data set containing warping profiles of $I_{training}$ and I_{test} faulty-l batches, respectively. Using the training data set a model/classifier is developed and optimized after outliers have been removed. The test data set is used to estimate a different classification index for each classifier: membership probability to fault l in faulty WICC, predictions in PLS-DA and Squared Prediction Error (SPE) in SIMCA.

In order to assess the quality of classification using a defined threshold, measures derived from the confusion table are used [17]. To establish the best classification threshold, the Matthews Correlation Coefficient (MCC) [18] is estimated. That threshold whose MCC value is the closest to 1 will be selected as the classification threshold. For the sake of comparison among classifiers, the Receiver Operator Characteristic (ROC) curve will be used. In order to assess the accuracy of the classifiers, the area under the ROC curve (the so-called AUROC) are calculated as an index measuring the goodness of the classifier.

3.1. Supervised faulty warping information-based control charts

The idea is to build a control chart from the warping information contained in each one of the matrices $\mathbf{F}_l$ corresponding to the training data set. Firstly, each warping profile is centered to the NOC average warping profile. Afterwards, the control limits at 99% confidence level are assessed by estimating the percentile 0.5 and 99.5 at each reference batch time (columns of the matrices $\mathbf{F}_l$). The test data set of each one of the l types of faults is then plotted onto the faulty-l WICC and the percentage of points falling within the control limits is calculated. This is an index of the membership probability to fault l. This procedure is repeated for each one of matrices $\mathbf{F}_1$, $\mathbf{F}_2$, ..., $\mathbf{F}_L$, i.e., for each one of known faults. Finally, the faulty WICC-based classifier is built as explained above. A new complete faulty batch can be classified into fault class l, if its percentage of points falling within the control limits of the fault-l WICC is larger than the corresponding classification threshold.

3.2. Classifiers based on Partial Least Squares Discriminant Analysis (PLS-DA)

The following approach is based on the fit of a PLS-DA model from the matrix $\mathbf{F}$, which was obtained after arranging the matrices $\mathbf{F}_1$, $\mathbf{F}_2$, ..., $\mathbf{F}_L$ corresponding to the training data set, one below the other. The response $\mathbf{Y}$ matrix is defined by dummy variables denoting the type of faulty batches (i.e. classes). In this type of analysis, the elements of the column vector $\mathbf{y}_c$ are one for batches belonging to class c (fault-1, fault-2, ..., fault-l, ..., fault-L) or zero otherwise. Both matrices $\mathbf{F}$ and $\mathbf{Y}$ are autoscaled. Once the PLS-DA model is fitted, the warping profiles of each type of fault belonging to the test data set are preprocessed and projected onto the PLSDA model, yielding the fault class predictions. After checking that their Euclidean distances to the latent model (SPE) are lower than the 99% confidence level (control limits estimated using the approximation method by Jackson and Mudholkar [19]), the best classification threshold is calculated by following the procedure explained in Section 3. When a new complete warping profile $\mathbf{f}_{new}$ corresponding to a faulty batch is available, it can be classified. For this purpose, $\mathbf{f}_{new}$ is centered and scaled using the

mean and variance estimated from the training data matrix $\mathbf{F}$, and its prediction vector $\hat{\mathbf{y}}_{new}$ is predicted. If the prediction $\hat{\mathbf{y}}_{new}$ is above the classification threshold of fault-l classifier, this new faulty batch can be classified as fault-l.

3.3. Classifiers based on the SIMCA approach

Soft Independent Modeling of Class Analogy (SIMCA), is a well-known classification method. This approach consists of the fit of a PCA model from each autoscaled matrix $\mathbf{F}_l$ containing the warping information of the faulty-l batch ($l = 1, ..., L$), corresponding to the training data set. Once the PCA models have been fitted, the warping profiles of faulty batches belonging to the test data set are preprocessed and projected onto the latent space, obtaining the squared Euclidean distances to the latent model (SPE). In case these values do not exceed the control limits at 99% confidence level (control limits estimated from theoretical results [19]), they are used to estimate the best classification threshold by following the procedure explained in Section 3. When a new complete warping profile $\mathbf{f}_{new}$ corresponding to a faulty batch is available, this can be classified. For this purpose, $\mathbf{f}_{new}$ is centered and scaled using the mean and variance estimated from the training data matrix $\mathbf{F}$, and then its SPE value is estimated. If this value is above the classification threshold of fault-l classifier, this new faulty batch can be classified as fault-l.

4. Results and discussion

Process data based on the biological model of the aerobic growth of *S. cerevisiae* on glucose limited medium [20] was generated using Simulink for Matlab release 2007b$^{\text{®}}$ (©The MathWorks, Inc). In order to simulate the physical uncertainty caused by the biological variability, slightly modified values of constants of the first principles model were introduced into the parametric space that defines the Simulink scheme for simulation. Also, Gaussian noise of low magnitude in the initial conditions (10%) and measurements (5%) to simulate the typical errors produced by sensors were added. Furthermore, the simulation achieved here took into account the biological variability of yeasts. In fermentation processes, characterized by a duration of several days, some microorganisms may have different generation times, having a significant influence on biomass growth and quality features, yielding batches with different duration.

A set of 85 unsynchronized batches was simulated under normal conditions following the assumptions commented previously. Two additional sets of 44 faulty batches each with two different kinds of abnormalities due to abnormal operating conditions were simulated.

Namely, interference processes have been simulated, i.e. factors that directly influence the maximum reaction rate (Vmax, k1L in the model) of the lumped biochemical reaction considered in the model. This Vmax represents in which way the substrate is processed by the yeast S. cerevisiae in glucose limited media. Although biochemically based (highly efficient strains will be able to consume glucose more quickly, showing higher intrinsic Vmax values), this parameter may also be influenced by processes such a diffusion. For example, if the bioreactor is not correctly stirred or the viscosity of the mixture is too high and hinders nutrient diffusion, substrates may not be accessible for the microorganism, resulting in low consumption rates. When these operating conditions are overcome, a better material transport is expected, and hence, a higher maximum reaction rate (Vmax). To simulate these scenarios, the values of the kinetic constants k1l (associated to the reaction describing the glucose uptake system and the glycolytic pathway) and k6 (associated to the reaction describing the formation of ethanol from acetaldehyde) were modified in the stoichiometric equations. In particular, these values were increased in comparison to the nominal values indicated in [20]. Modifying the constants, the consumption of glucose is higher than in normal operating conditions, causing an excess of glucose in the microorganism (the so-called metabolic overflow). In this scenario, the rate of glycolysis exceeds a critical value resulting in by-product formation (ethanol, acetaldehyde, acetate) from pyruvate and ethanol (activation of the fermentation pathway). Consequently, the amount of carbon dioxide is also higher in media than in normal operating conditions. This has a direct effect on the duration of the second stage of the fermentation (from the 50th sampling point -i.e. after 20h the batch started approximately- onwards), which takes longer than usual to reduce the amount of these products.

Ten variables were measured every sampling time over all batches: concentrations (glucose, pyruvate, acetaldehyde, acetate, ethanol and biomass), active cell material, acetaldehyde dehydrogenase (proportional

to the measured activity), specific oxygen uptake rate and specific carbon dioxide evolution rate. The simulated data were split up in a calibration and a test data set. 60 NOC batches and 20 batches from each of the simulated abnormalities were randomly selected to form the training data set, and the remaining batches were used to arrange the corresponding test data sets (25 NOC, 24 faulty-1 and 24 faulty-2 batches). Before the multivariate modeling is carried out, the synchronization of NOC batches must be performed. In order to establish the proper parameters of the RGTW algorithm, the cross-validation procedure proposed in [11] was run.
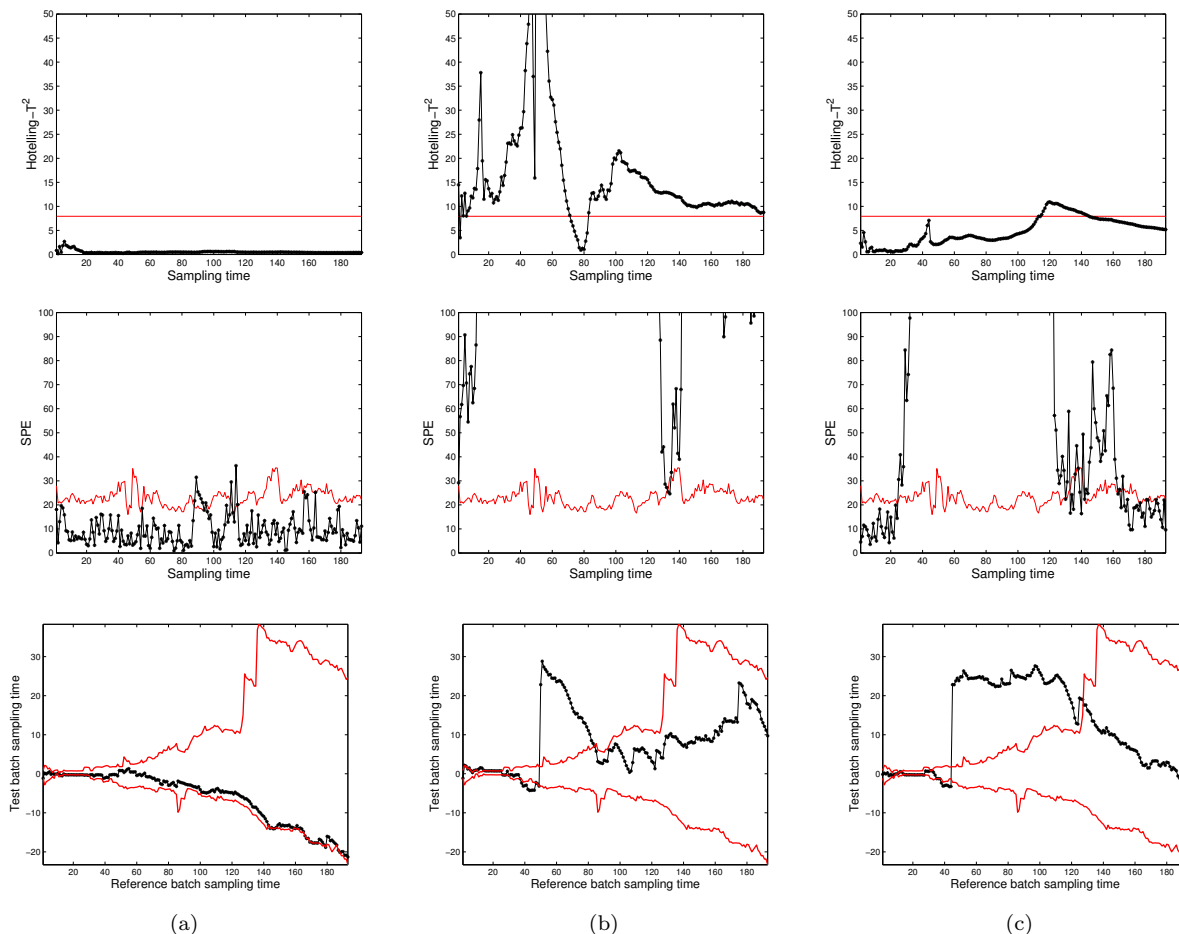

Figure 2: Hotelling-T^2, SPE and NOC warping information-based control charts monitoring a NOC (a), faulty-1 (b) and faulty-2 test batch (c). Cross-validated control limits for a 99% confidence level (solid red line) in the Hotelling-T^2 and SPE control charts are shown. Also, the upper and lower control limits of the NOC warping information-based control chart established at the percentile 0.5 and 99.5 are denoted by solid red line.

Firstly, the basic parameters of the off-line DTW algorithm were assessed. The batch whose duration was the closest to the median length of historical NOC batches was selected as the reference batch. In this case, batch #30 was chosen with a duration of 193 sampling points. The weights were assessed as the geometric average of the weights estimated by the iterative procedure of Kassidas *et al.* [10] and Ramaker *et al.* [21] approaches. Secondly, the proper window width ζ was estimated. For this purpose, the performance of the RGTW algorithm varying the window width between 1 and 5 units were studied. The RGTW-based synchronization using a window width equal to 3 units was finally selected and the bands were calculated based on the warping information derived from the latter synchronization.

Once the set of 60 calibration NOC batches were synchronized, the slices containing information of all process variables at the sampling time k were arranged side by side in a two-way data matrix $\mathbf{X}$(60 batches × (10 variables × 193 sampling points)). After batch data were autoscaled, i.e. the mean trajectory was subtracted and all process variables at every sampling time were scaled to unit variance, a PCA model was fitted. The selection of the optimum number of principal components (PCs) was carried out based on the PRESS (PREdicted Residual Sum of Squares) function derived from the cross-validation procedure and the results obtained from the study of the performance of the PCA model using the Overall Type I (OTI) and Type II (OTII)[1] risks [12].

For performance evaluation of the PCA model, the first ten PCs were taken into consideration in the study. For each of the PCs, a monitoring system was built by designing two multivariate Shewhart control charts based on Hotelling-T^2 and SPE statistics. Their control limits were estimated from NOC process data and later readjusted using cross-validation techniques for an imposed significance level (ISL). The NOC and faulty test sets composed by 25 and 48 batches (24 batches for each of the two abnormalities), respectively, were projected onto the model and the OTI and OTII values for both statistics were calculated. Once the complete procedure was repeated for each of the PCs considered, the values of both indices as a function of the number of PCs were studied jointly with the PRESS. In this example, two principal components were finally extracted since the corresponding model had better performance in relation to the aforementioned parameters.

4.1. NOC warping information-based control charts

The warping information obtained from the RGTW-based synchronization of the batch trajectories corresponding to the set of 60 calibration NOC batches were used to build the NOC warping information-based control chart introduced in Section 2, where an upper and lower control limit established at percentile 0.5 and 99.5, respectively.

Three different batches, one NOC, one faulty-1 and one faulty-2 batches were randomly chosen from the test data set to be synchronized and monitored, yielding the warping information, Hotelling-T^2 and SPE statistics throughout the batch run (see Figure 2). In the case of the NOC test batch, no clear out-of-control signal is detected in any of the three control charts (see Figure 2(a)). Regarding the end-of-batch monitoring results of the two different faults, the monitoring system has correctly detected the abnormalities through the SPE control charts (see Figures 2(b) and 2(c)). It is worth noticing that the statistics-based control charts detect the fault earlier than the warping information-based control charts, in particular in the case of the faulty-1 batch. Nevertheless the use of the latter control chart provides a good insight about the process performance and also a valuable complementary tool for fault classification. This will be explained in Section 4.2. The remaining test batches also showed the same behavior as the three selected batches in Figure 2 (results not shown).

4.2. Supervised warping information-based control charts

Using the 20 batches for each abnormality belonging to the training data set, the faulty WICCs were built by following the procedure explained in Section 3.1. Differences among the different classes can be found by looking at the control limits of the faulty WICCs shown in Figure 3. From the control limits of the faulty-1 WICC (Figure 3(a)) and NOC WICC (Figure 2, bottom), one can check that these batches required more time than the NOC batches to achieve the stage limited from the 45th to the 50th reference time point. In particular, a large amount of vertical transitions are shown at this time interval, hence, the batch trajectories were compressed by the RGTW algorithm to synchronize the process events. From the 50th to the 120th sampling point, a larger number of horizontal than vertical transitions are shown. Consequently, faulty-1 batches needed less time to reach the end of such process stage with respect to NOC batches. Regarding the faulty-2 batches, the control limits of the faulty-2 WICC (see Figure 3(b)) show that from the 40th to 45th reference sampling point, the RGTW algorithm compressed the batch trajectories,

---

[1]The OTII values are calculated by following $OTII = 100 \cdot \frac{nnf}{I_{ab} \cdot k}\%$, where nnf is the number of non-detected faults, I_{ab} is the number of faulty batches and k is the length of the faulty period.
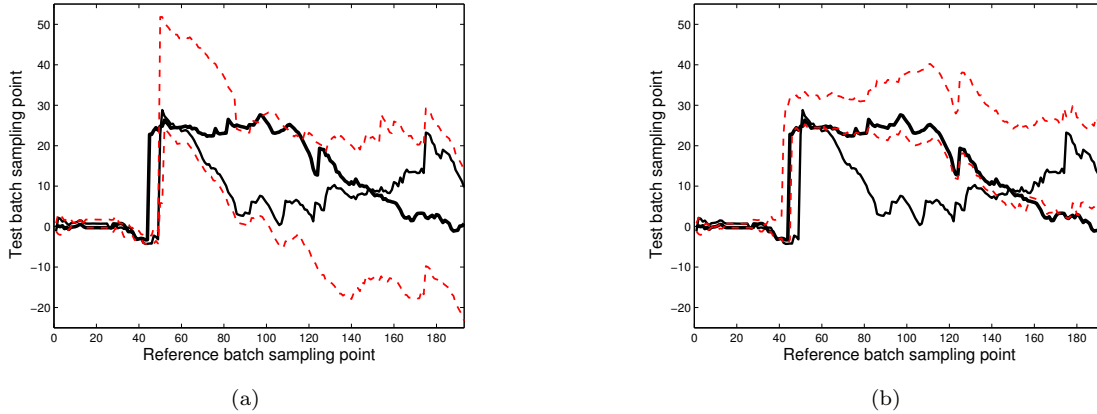
Figure 3: Faulty-1(a) and faulty-2 (b) WICC with control limits defined at percentiles 0.5 and 99.5 (dashed lines) are shown. Warping information belonging to faulty-1 (medium line-width) and faulty-2 (thick line-width) batches is plotted.

yielding to vertical transitions. Again, the first process stage lasted longer than NOC batches. Unlike the faulty-1 batches, the faulty-2 batches showed a similar behavior as NOC batches from the 50th reference time point onwards.
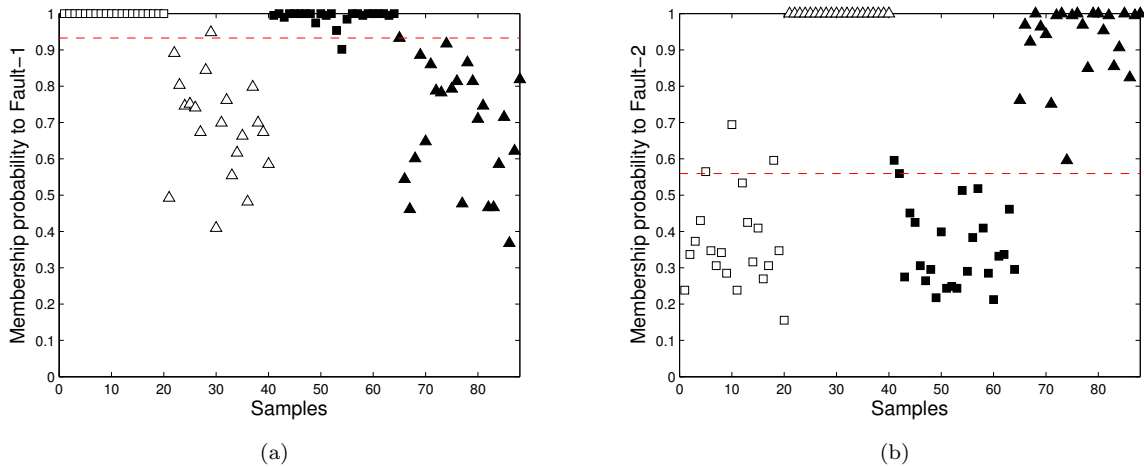

Figure 4: Classification of warping profiles for the Fault-1 (a) and Fault-2 classes using faulty WICC. Faulty-1 (empty squares) and faulty-2 batches (empty triangles) corresponding to the training data set are shown. Filled squares and triangles denote the faulty-1 and faulty-2 batches from the test data set, respectively. The dashed lines are the thresholds that yielded the highest MCC on the corresponding faulty classifier. Recall that only the batches corresponding to the training data set (those represented with empty symbols) were used to assess the classification thresholds.

To illustrate the performance of the control charts proposed for end-of-batch fault classification, only the two faulty batches selected in Section 4.1 (one faulty-1 and one faulty-2 batch) were plotted in the faulty-1 and faulty-2 WICCs, respectively (see Figure 3). The membership probability (MP) of the two selected faulty test batches to fault-1 and fault-2 classes was calculated. In the case of the faulty-1 test batch, the corresponding warping information (medium line-width) falls fully inside the control limits of the faulty-1 WICC (193 out of 193 points), yielding a $MP_{l=1} = 100\%$ (see Figure 3(a)). For the faulty-2 test batch the $MP_{l=1} = 86.53\%$ (167 out of 193 points). Concerning the faulty-2 WICC (see Figure 3(b)),

8

164 out of 193 points of the faulty-2 batch were inside the control limits ($MP_{l=2} = 84.97\%$) while 100 out of 193 points belonging to the fault-1 batch falls within the control limits of the mentioned control chart ($MP_{l=2} = 51.81\%$) (see Figure 3(b)).
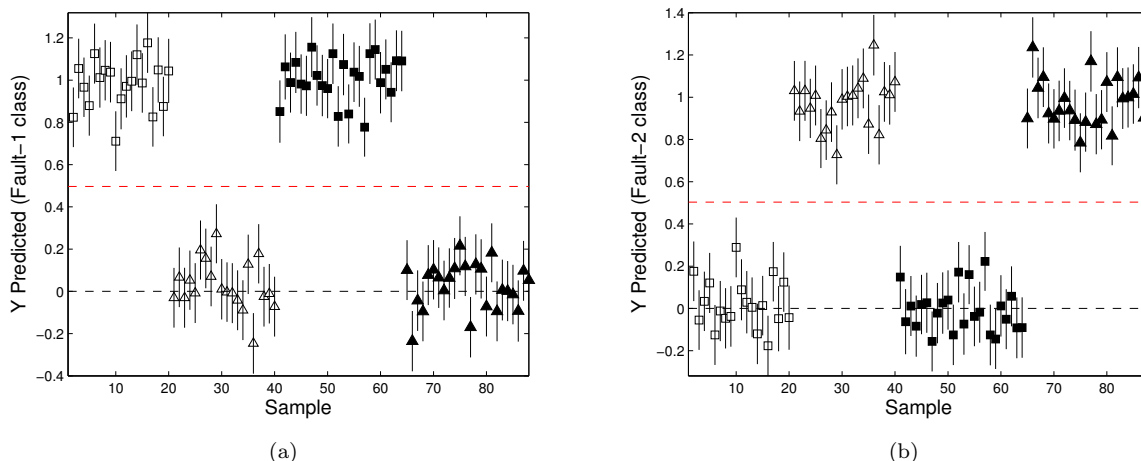

Figure 5: PLS-DA two-LV model predictions for the Fault-1 (a) and Fault-2 (b) class. Faulty-1 (empty squares) and faulty-2 batches (empty triangles) corresponding to the train data set are shown. Filled squares and triangles denote the faulty-1 and faulty-2 batches from the validation data set, respectively. The red dashed lines are the classification thresholds whose MCC value associated is the highest on the corresponding faulty data set.

Following the above procedure, the warping profiles corresponding to the 24 test batches of each type of fault were used to estimate the membership probability to the known fault. Once these probabilities were obtained (see Figure 4), a threshold per faulty WICC was calculated by following the procedure explained in Section 3.1 (classification thresholds for Fault-1 and Fault-2 classes are 0.93 and 0.56, respectively). Note that these thresholds notably differ from each other, mainly due to the different patterns found in the warping profiles of faulty-1 and faulty-2 batches. As can be appreciated in Figure 4(a), most of the points corresponding to the warping profile of the selected faulty-2 batch fall inside the limits in the faulty-1 WICC, except in the time intervals [8,14], [45,50], [86,98] and [110,117]. The rest of the faulty-2 batches also showed the same pattern in the faulty-1 WICC (results not shown). Hence, the membership probabilities to Fault-1 class estimated for both faulty-1 and faulty-2 batches are expected to be high, being for the formers slightly higher than for the latter. It causes that the classification threshold has a high value, close to 1. In contrast, a larger amount of points corresponding to the warping profile of the selected faulty-1 batch fall outside the control limits in the faulty-2 WICC, in particular, in the time intervals [45,50] and [60,143] (see Figure 4(b)(b)). Again, this behavior is also observed in the rest of the faulty-1 batches (results not shown). Hence, the membership probabilities to Fault-2 for faulty-1 and faulty-2 batches differ considerably, yielding to a lower threshold for classifier of Fault-2 class than for classifier of Fault-1 class.

Accuracy for faulty-1 and faulty-2 classes is measured by the area under the ROC curve (AUROC), leading to a value of 0.9911 and 0.9951, respectively, which indicates a good performance of the faulty WICC-based classifier. This is also illustrated in Figure 4(a) (membership probability to Fault-1 class) and Figure 4(b) (membership probability to Fault-2 class), where almost all the test warping profiles were correctly classified. Note that the membership probabilities belonging to the training data set are plotted for visualization purpose as well.

4.3. PLS-DA-based classifier

Using the 20 batches for each abnormality belonging to the training data set, a PLS-DA model was fitted. The resulting PLS-DA cross-validated model yielded two latent variables, with R^2X, R^2Y and Q^2

values of 74.9%, 94.4% and 92.8%, respectively. The faulty batches from the test data set (24 batches for each of the faults) were projected onto the PLS-DA model as external validation. All SPE values were inside the corresponding 99% confidence limits. The predictions for class fault-1 and class fault-2 models are shown in Figure 5. A threshold per each class was estimated by following the approach explained in Section 3.1 (classification thresholds for Fault-1 and Fault-2 classes are 0.49 and 0.50, respectively). At this point it is worth noticing that these thresholds are almost equal. This is because in both cases, the prediction distribution generated for Fault-1 and Fault-2 classes do not overlap each other, meaning that the classes are well separated. This interpretation can be observed in Figure 5, where all the test faulty batches were correctly classified, both in Faulty-1 and Faulty-2 classes.

In order to check for model consistency, a random permutation test was performed to study the model consistency. The permutation test aims at comparing both goodness of fit and goodness of prediction of the original model with the values estimated after class randomization [22]. The R^2Y and Q^2 values found (result not shown) using real class labels were clearly outside the distributions of those statistics found when employing random class labels, which indicates a clear distinction between the permutated-classification and the original classification. From these results, we can conclude that the PLS models are statistically significant ($P < 0.001$).
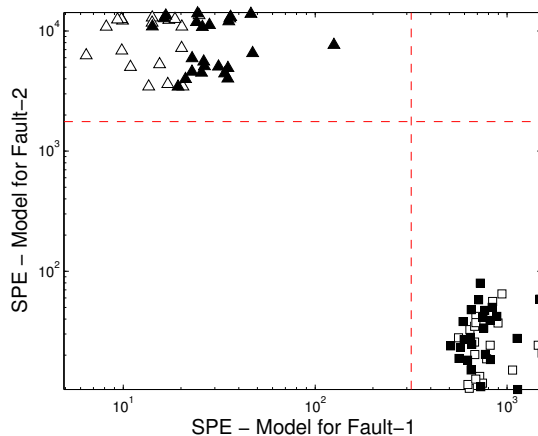

Figure 6: Cooman's plot for the faulty-1 and faulty-2 batches model. Faulty-1 (empty squares) and faulty-2 (empty triangles) batches corresponding to the training data set are plotted. Filled squares and triangles denote the faulty-1 and faulty-2 batches from the test data set, respectively. The red dashed lines are the classification thresholds selected as those with the highest MCC.

In this case, the AUROC value corresponding to the ROC curves for faulty-1 and faulty-2 classes is 0.9946, showing similar performance than the faulty WICC-based classifier.

4.4. SIMCA-based classifier

A cross-validated PCA model was fitted from each of the aforementioned faulty-1 and faulty-2 training data set. The first PCA model yielded 3 PCs, with R^2 and Q^2 values of 84.8% and 72.2%, respectively. The second one was defined by 4 PCs, a goodness of fit equal to 84.8% and a goodness of prediction equal to 67.6%.

The faulty batches from the test data set (24 batches for each of the faults) were projected onto the PCA models to classify them into their correct model. A Cooman's plot has been done (Figure 6), representing the distances of the different training and test sets to the Fault-1 and Fault-2 models. Once these distances were obtained, a threshold per PCA model was calculated by following the procedure explained in Section 3 (classification thresholds for Fault-1 and Fault-2 PCA models are 316.45 and 1763.3, respectively). Note that as happened in PLSDA-based classifier, both models are able to clearly distinguish faulty-1 and faulty-2

batches, both those corresponding to the training and the test data sets, since their corresponding distances to the model of different fault classes are large.

Again, accuracy was measured estimating the area under the ROC curve designed for each one of classes. The AUROC value of the ROC curves associated to faulty-1 and faulty-2 class is 0.9946, denoting a good performance for classification.

5. Conclusions

In this paper, the use of the warping information obtained from the RGTW-based batch synchronization both for process monitoring and for supervised fault classification is addressed.

A unsupervised control chart based on the warping profiles from NOC batches (NOC-WICC) is proposed as a complementary tool to the Hotelling-T^2 and SPE control charts for end-of-batch and real-time batch process monitoring. In the case of process faults are fingerprinted in the warping profiles, this chart can be useful to detect their occurrence in the process. Nevertheless, the NOC-WICC may not notably improve the performance of the traditional multivariate Shewhart control charts. This improvement is subject to different factors, such as the nature of the process or the influence of the fault in the process phases, among others. For subtle changes (ramps, small step changes, etc.) detection, memory control charts, such as EWMA or CUSUM, should be used.

When a rich faulty database is available, warping information can be used to build the so-called supervised warping information-based control charts (faulty WICC) or to fit classification models using supervised chemometric tools. Although in this paper simple and widely used tools such as PLSDA and SIMCA have been used, other classification techniques, such as Support Vector Machines [23], *Linear and Quadratic Discriminant Analysis* (LDA & QDA) [24], and KNN [25], could be taken into consideration, among others. In this paper, the three approaches studied showed good classification performance in terms of the area under the ROC) curve (the so-called AUROC). The use of the faulty-WICC-based classifiers depends much on the type of fault -i.e. if faults have characteristic fingerprints in their corresponding warping profiles at specific time periods that are different from the rest. The more different the warping profiles from faulty batches, the better the accuracy of the classifier. In contrast, PLSDA and SIMCA-based classifiers are more accurate in fault classification when no clear differences among warping profiles are found.

In this study, using the warping profiles derived from the aRGTW-based synchronization has been enough to design the classifiers with good prediction performance. In cases that the warping information belonging to different faulty batches does not show clear different patterns, the use of the raw batch trajectories jointly with the warping profiles for fault classification is suggested.

References

[1] S. Wold, N. Kettaneh-Wold, J. McGregor, K. Dunn, Batch process modeling and MSPC, Comprehensive Chemometrics ñ Chemical and Biochemical Data Analysis, In: Brown S, Tauler R, Walczak R (eds.) 1 (2009) 163–195.

[2] T. Kourti, Multivariate statistical process control and processes control, using latent variables, Comprehensive Chemometrics ñ Chemical and Biochemical Data Analysis, In: Brown S, Tauler R, Walczak R (eds.) 4 (2009) 21–54.

[3] D. J. Louwerse, A. A. Tates, A. K. Smilde, G. L. M. Koot, H. Berndt, Pls discriminant analysis with contribution plots to determine differences between parallel batch reactors in the process industry, Chemometrics and Intelligent Laboratory Systems 46 (1999) 197 – 206.

[4] M. Zarzo, A. Ferrer, Batch process diagnosis: Pls with variable selection versus block-wise pcr, Chemometrics and Intelligent Laboratory Systems 73 (2004) 15–27.

[5] P. Nomikos, J. MacGregor, Monitoring batch processes using multiway principal components, AICHE Journal 40 (1994) 1361–1375.

[6] C. Ündey, A. Çinar, Statistical monitoring of multistage, multiphase batch processes, IEEE Control System Magazine 22 (2002) 40–52.

[7] N. Kaitsha, C. F. Moore, Extraction of event times in batch profiles for time synchronization and quality predictions, Industrial & Engineering Chemistry Research 40 (2001) 252–260.

[8] J. Ramsay, B. Silverman, Functional data analysis, New York:Springer-Verlag, 1997.

[9] S. W. Andersen, G. C. Runger, Automated feature extraction from profiles with application to a batch fermentation process, Journal of the Royal Statistical Society: Series C (Applied Statistics) 61 (2012) 327–344.

[10] A. Kassidas, J. MacGregor, P. Taylor, Synchronization of batch trajectories using dynamic time warping, AIChE Journal 44 (1998) 864–875.

[11] J. Gonzlez-Martnez, A. Ferrer, J. Westerhuis, Real-time synchronization of batch trajectories for on-line multivariate statistical process control using dynamic time warping, Chemometrics and Intelligent Laboratory Systems 105 (2011) 195–206.

[12] J. Camacho, J. Picó, A. Ferrer, On-line monitoring of batch processes based on pca: Does the modelling structure matteré, Analytica chimica acta 642 (2009) 59–69.

[13] H. Ramaker, E. van Sprang, J. A. Westerhuis, F. van der Meulen, Performance assessment and improvement of control charts for statistical batch process monitoring, Statistica Neerlandica 60 (2006) 339–360.

[14] S. Garca-Muoz, T. Kourti, J. MacGregor, Troubleshooting of an industrial batch process using multivariate methods, Industrial and Engineering Chemistry Research 42 (2003) 3592–3601.

[15] M. Sjöström, S. Wold, B. Söderström, PLS discriminant plots, in: Proceedings of PARC in Practice, Amsterdam, June 19-21, 1985 Elsevier Science Publishers B.V., North-Holland.

[16] S. Wold, C. Albano, W. Dunn, U. Edlund, K. Esbensen, P. Geladi, S. Hellberg, E. Johansoon, W. Lindberg, M. Sjöström, Multivariate data analysis in chemistry, B.R. Kowalski (Ed.), Chemometrics: Mathematics and Statistics in Chemistry, D. Reidel Publishing Company, Dordrecht, Holland, 1984.

[17] J. Westerhuis, H. C. Hoefsloot, S. Smit, D. J. Vis, A. K. Smilde, E. van Velzen, J. van Duijnhoven, F. van Dorsten, Assessment of plsda cross validation, Metabolomics 4 (2008) 81–89.

[18] B. Matthews, Comparison of the predicted and observed secondary structure of t4 phage lysozymes, Biochimica et Biophysica Acta - Protein Structure 405 (1975) 442–451.

[19] J. Jackson, G. Mudholkar, Control procedures for residuals associated with principical component analysis, Technometrics 21 (1979) 341.

[20] F. Lei, M. Rotbøll, S. Jørgensen, A biochemically structured model for saccharomyces cerevisiae, Journal of biotechnology 88 (2001) 205–221.

[21] H. Ramaker, E. van Sprang, J. Westerhuis, A. G. Smilde, Dynamic time warping of spectroscopic batch data, Analytica Chimica Acta 498 (2003) 133–153.

[22] F. Lindgren, B. Hansen, W. Karcher, W. Sjöström, L. Eriksson, Model validation by permutation tests: applications to variable selection, Journal of Chemometrics 10 (1996) 521–532.

[23] Y. Xu, S. Zomer, R. Brereton, Support vector machines: a recent method for classification in chemometrics, Critical Reviews in Analytical Chemistry 36 (2005) 177–188.

[24] G. McLachlan, Discriminant Analysis and Statistical Pattern Recognition, Wiley, New York, 1992.

[25] D. Coomans, D. L. Massart, Alternative k-nearest neighbour rules in supervised pattern recognition. part 1. k-nearest neighbour classification by using alternative voting rules, Analytica Chimica Acta, 136 (1982) 15–27.