

# Multiscale Statistical Process Control with Multiresolution Data

Marco S. Reis and Pedro M. Saraiva

GEPSI-PSE Group, Dept. of Chemical Engineering, University of Coimbra, Pólo II –Pinhal de Marrocos, 3030–290 Coimbra, Portugal

DOI 10.1002/aic.10805

Published online March 13, 2006 in Wiley InterScience (www.interscience.wiley.com).

*An approach is presented for conducting multiscale statistical process control that adequately integrates data at different resolutions (multiresolution data), called MR-MSSPC. Its general structure is based on Bakshi's MSSPC framework designed to handle data at a single resolution. Significant modifications were introduced in order to process multiresolution information. The main MR-MSSPC features are presented and illustrated through three examples. Issues related to real world implementations and with the interpretation of the multiscale covariance structure are addressed in a fourth example, where a CSTR system under feedback control is simulated. Our approach proved to be able to provide a clearer definition of the regions where significant events occur and a more sensitive response when the process is brought back to normal operation, when it is compared with previous approaches based on single resolution data.* © 2006 American Institute of Chemical Engineers *AIChE J*, 52: 2107–2119, 2006

*Keywords:* multiscale statistical process control, multiresolution data, wavelets

## Introduction

Data generated in chemical process plants arise from many sources, such as on-line and off-line process sensors, laboratory tests, readings made by operators, raw materials specifications, and weather conditions, to name just a few. To such a variety of origins are usually associated complex data structures, due to a diversity both in time acquisition and missing data patterns among the different variables (which give rise to sparse tables), as well as in variable resolutions, since the values from different variables may carry information over different time ranges (multiresolution data). In spite of several developments that have already been proposed to address the sparsity problem created by multirate<sup>1–3</sup> and missing data<sup>4–7</sup> problems, the issue of adequately handling multiresolution process data remains, to a large extent, unexplored, with developments mainly centered around signal and image processing problems.<sup>8–10</sup>

In a multiresolution data structure, we can find variables

whose values are collected punctually (high time resolution) at every node of a fine grid whose spacing is established by their (also higher) acquisition rates, and variables that represent averages over larger time ranges (that is, over several nodes of this grid), to which we will refer as “lower resolution variables.” (The term “averaging support,” AS, will also be used to address the range of time or number of nodes over which averages are computed.)

In industrial applications, multiresolution data structures usually arise when process sensor information is combined with data taken from other sources, such as the following: averages made by operators from several readings taken from process measurement devices during their shifts, which are then annotated in daily operation reports or introduced manually in a computer connected to the central data storage unit; measurements from pools of raw material or products accumulated during a certain period of time and mixed thoroughly before testing; averages of process variables over a period of time, which are computed automatically by local DCS computers (such as on an hourly basis); and aggregated measurements from an entire batch.

On the other hand, processes going on in chemical plants are

Correspondence concerning this article should be addressed to M. S. Reis at marco@eq.uc.pt.

themselves typically quite complex, and this complexity is also reflected in collected data, which contain the cumulative effect of many underlying phenomena and disturbances, with different location and localization patterns in the time/frequency plane. This means that not only the overall *system* has a multiscale nature, since it is composed of processing units that span different time scales and frequency bands, but also that the *inputs* (manipulation actions, disturbances, faults) can present a variety of features with distinct time/frequency characteristics. For such reasons, multiscale approaches, designed to handle and take advantage of the information contained at different scales, have been developed for addressing different tasks,<sup>11,12</sup> namely, process monitoring, which will be the focus of this article. In this regard, multiscale monitoring approaches provide an adequate basis to develop a (multiscale) process monitoring framework that integrates information with different resolutions, as the concept of resolution (or scale) is already present in their algorithmic structures, by design, in particular for those based on the wavelet transform as a tool to separate dynamic features contained at different scales. Therefore, the structure underlying Bakshi's<sup>13</sup> MSSPC (for data at a single resolution) was adopted in this work as an adequate starting point to integrate data with different resolutions, a research topic covered by a number of authors, as we will see in the next paragraph.

Kosanovich and Piovoso<sup>14</sup> presented an approach where the Haar wavelet transform coefficients from filtered data (with a finite impulse response median hybrid filter) were used to estimate a PCA model, which is then applied for monitoring purposes, but it was with the works of Bakshi et al.<sup>11,15</sup> that the first univariate and multivariate MSSPC methodologies were established. Several modifications and improvements were then introduced to the base methodology, such as its integration with monitoring methods specialized in detecting changes over data distribution features<sup>16</sup> or the use of variable grouping and the analysis of contribution plots when an event is detected in the control charts at any scale, in order to monitor the process and simultaneously perform early fault diagnosis.<sup>17</sup>

Other contributions include the development of nonlinear process monitoring approaches, similar to MSPCA, but with a nonlinear modeling step, namely, using neural networks.<sup>18,19</sup> Rosen<sup>20</sup> presented an alternative methodology, where the components at different scales in the original time domain (that is, not in the wavelet transform domain) are combined using background knowledge about the process, in order to reduce the number of monitoring statistics available when all the scales are monitored separately and to provide physical insight to the scales under analysis. In this approach, there is no reconstruction stage, as in Bakshi's MSSPC, and the coarser scale coefficients are omitted in the monitoring procedure, allowing for adaptation (in the mean) to nonstationary data. Yoon and MacGregor<sup>21</sup> also developed an approach based upon a multiscale representation of data in the original time domain that encompasses the successive extraction of principal components for the extended set (all variables represented at all scales), according to the decreasing magnitude of the associated eigenvalues. It turns out that, owing to the orthogonal properties of the wavelet transform, the loadings only contain non-zero entries for variables at the same scale, and therefore each extracted component strictly conveys information regarding a specific scale. As such, results are not very different from

what can be achieved with traditional MSSPC for the same number of principal components. However, this approach does allow for a ranking of the relevant structures underlying data variability, regarding the contributions from the variables covariance at different scales.

Trygg et al.<sup>22</sup> applied a 2-D wavelet transformation to compress data from NIR (near-infrared) spectra collected over time and estimated a PCA model for this 2-D compressed matrix. Such a model was then used to check whether new incoming spectra deviate from those collected during normal operation using PCA tools. In the context of a PLS based monitoring scheme, Teppola and Minkkinen<sup>23</sup> used a multiscale decomposition in order to remove seasonal and low-frequency contributions to variation. A scale (or frequency band) selection methodology is also adopted in Luo et al.<sup>24</sup> to select the frequency bands where data analysis should be focused when trying to detect sensor faults. Sun et al.<sup>25</sup> developed a framework for detecting abnormal situations, using the so-called wavelet-domain hidden Markov models (HMM). For further references on multiscale monitoring, we refer the interested reader to the review article by Ganesan et al.<sup>26</sup>

The remaining parts of this article are organized as follows: In the next section, we present the essential background material that supports the methodology here presented, and then introduce our MSSPC approach that integrates multiresolution data (MR-MSSPC). In the following section, several examples illustrate its improved effectiveness in identifying the region in the time domain where a fault occurs, and its promptness in detecting transition points, when compared with other alternatives based on single resolution data structures. A last example addresses the case of monitoring a non-linear multivariate dynamic process using the MR-MSSPC methodology, where several important practical issues regarding its real world implementation are addressed, as well as some extensions, namely, the possible definition of an adequate resolution for each variable being monitored. We finalize with a conclusions section that summarizes the main results presented in this article, and addresses some ongoing and future work to be pursued in this field.

## Theoretical Background

### *Principal components analysis (PCA) and PCA-SPC*

PCA<sup>27-29</sup> is a well known multivariate data analysis technique that addresses the problem of finding a reduced ( $p$ -dimensional) set of new variables, the principal components, which are linear combinations of the original ( $m$ ) variables, with the ability of explaining most of their variability ( $m \geq p$ ). Such linear combinations are those that successively present maximal (residual) variability after the portion explained by former components has been removed. The solution of such an optimization formulation can be reduced to an eigenvalue problem,<sup>30</sup> where the optimal linear combinations (loadings) are given by the successive normalized eigenvectors of the data covariance matrix, associated with the eigenvalues sorted in a decreasing ordering of magnitude. Applying PCA to the original data matrix, we transform a set of correlated variables into a smaller, decorrelated one (that is, with a diagonal covariance matrix), which quite often still explains a large amount of the structure and variability present in the original dataset.

PCA has been applied to Quality Control in multivariate

contexts,<sup>31-33</sup> where it provides an adequate alternative to the monitoring procedure based upon the Hotelling- $T^2$  statistic.<sup>34</sup> In fact, such a procedure involves an inversion of the covariance matrix, an operation that can be quite unstable from the numerical point of view in situations where the system is of low rank, or approximately so, as frequently happens in industrial contexts due to, for instance, dependencies caused by underlying phenomena and conservation laws (mass and energy), presence of control loops, use of redundant instrumentation, and the bare nature of some measuring devices employed, that leads to highly redundant sets of variables (such as spectrometers). In this context, process monitoring through PCA modeling circumvents these problems, and consists of monitoring process variability in a reduced space, the PCA subspace, where normal process variation is mostly concentrated, using a Hotelling- $T^2$  statistic. This is complemented with another control chart, which follows variability around the estimated PCA subspace through a residual statistic (often called  $Q$  or  $SPE$ , standing for Squared Prediction Error) given by the squared Euclidean distance of each multivariate observation to the PCA subspace. Thus, through PCA-SPC, a process with dozens or hundreds of variables can be adequately followed using just these two monitoring statistics. More details about this procedure, and the associated monitoring statistics and control limits, can be found in the literature.<sup>32,35,36</sup>

### Wavelets

Wavelets correspond to particular types of functions that are localized both in the time and frequency domains. Under certain circumstances, collections of this type of functions (wavelet families), whose elements differ only in the values attributed to their two parameters, form basis sets. The two parameters to be defined are the scale parameter, indexed by  $j$  (which stretches or compresses the wavelet function and, therefore, defines its shape on the time-frequency plane), and the time-translation or shift parameter,  $k$  (which positions the function in a given location in the time domain).

Any finite-energy signal can be expanded in terms of wavelet basis functions (which carry information about all the scales between the finest,  $j = 0$ , and the coarsest, given by the maximum decomposition depth selected for the wavelet representation of the signal,  $j = J_{dec}$ ) and the associated scaling functions (which contain all the contributions arising from scales  $j > J_{dec}$ ). These expansion coefficients are called details coefficients,  $d_k^i$  ( $i = 1, \dots, J_{dec}; k \in \mathbb{Z}$ ) (coefficients of the wavelet functions), and approximation coefficients,  $a_k^{J_{dec}}$  ( $k \in \mathbb{Z}$ ) (multipliers of the scaling functions), and are usually referred to as the orthogonal wavelet transform or wavelet coefficients.

A very efficient recursive scheme was developed by S. Mallat<sup>37</sup> for computing wavelet coefficients, with computational complexity  $O(n)$ , once projection at the finest scale,  $a_k^0$  ( $k \in \mathbb{Z}$ ), becomes available. It essentially consists of implementing a pyramidal algorithm based upon convolution with quadrature mirror filters. In practice, the finest scale approximation coefficients,  $a_k^0$  ( $k \in \mathbb{Z}$ ), take the values from the original data set,<sup>13</sup> and the calculation of the wavelet transform coefficients proceeds by implementing Mallat's efficient algorithm (more elaborate initialization strategies are discussed elsewhere<sup>38</sup>). Therefore, in practice, we usually apply the anal-

ysis and reconstruction quadrature mirror filters associated to a given wavelet without using any wavelet function explicitly.

Besides such a computational efficiency, wavelet transforms can be very effective in representing signals with localized features in the time/frequency plane, due to the nature of their basis functions. They are, for this reason, very promising tools for analyzing non-stationary data or data with localized regularity characteristics.

### Multiscale statistical process control (MSSPC)

MSSPC explores some key properties of the wavelet transform, namely, those regarding the ability to extract deterministic features in a few wavelet coefficients (*energy compaction*), whereas stochastic processes spread their energy across all the coefficients. Furthermore, they approximately decorrelate signals, that is, the autocorrelation matrix of such signals is approximately diagonalized (*decorrelation ability*).<sup>13,39,40</sup> This allows for handling contexts where signals monitored present autocorrelation and features with widely different time/frequency characteristics. More specifically, MSSPC is based upon multiscale principal component analysis (MSPCA), which combines the decorrelation ability of PCA, regarding cross-correlations among variables, with that of the wavelet transform for any potential autocorrelated behavior in each variable, as well as the deterministic/stochastic separation power characteristic of this type of transform. It was developed by Bakshi,<sup>13</sup> who then proposed its application to the context of statistical process monitoring (through MSSPC).

The MSSPC procedure consists of computing independent principal components models and control limits for PCA-SPC control charts at each scale, using data collected from a process running under normal conditions. Then, as new data are acquired, wavelet coefficients are calculated at each scale, according to the chosen discretization scheme, and control chart procedures implemented separately at each scale. If any significant activity is detected at any scale, the signal is reconstructed back to the time domain, using only those coefficients arising from the significant scales. The covariance matrix at the finest scale is also computed using information related to the significant scales, in order to implement the statistical tests at the finest scale ( $T^2$  and  $Q$  control charts) that will produce the final outcome of the MSSPC method, stating whether the process can be considered to be operating under normal conditions, or if a special event has occurred.

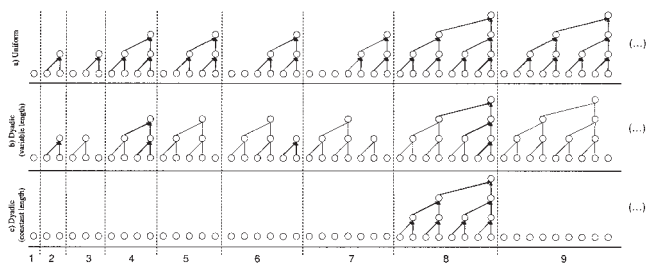
### MR-MSSPC

This section describes our MR-MSSPC approach, including the adopted discretization strategies underlying its application.

#### Discretization strategies

One of the differentiating features regarding alternative implementations of MSSPC, and an important one, regards the type of data windows over which the wavelet transform is applied, and based upon which the subsequent data analysis is carried out.

In one extreme, we have the moving window of constant dyadic length, used by Bakshi,<sup>13</sup> which consists of translating in time a window of length  $2^{J_{dec}}$  (where  $J_{dec}$  stands for the decomposition depth of the wavelet transform used in the



**Figure 1. Discretization strategies used in MSSPC, for  $J_{dec} = 3$ .**

(a) Overlapping moving windows of constant dyadic length (uniform discretization), (b) Dyadic moving windows for orthogonal wavelet transform calculations (variable window length dyadic discretization), (c) Non-overlapping moving windows of constant dyadic length (constant window length dyadic discretization).

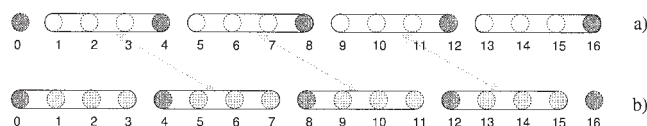
multiscale analysis), so that the last vector of observations is always included in the window, after an initial phase that goes from observation number 1 to observation number  $2^{J_{dec}}$ . Its length increases in such a way that a dyadic window is always used, until the maximum length is reached (Figure 1a).

Concerning now procedures that involve a time delay, we have the moving window of variable dyadic length that enables a successive calculation of the coefficients regarding an orthonormal wavelet transformation (Figure 1b), in opposition to the coefficients of its undecimated counterpart, also known as translation invariant wavelet transform, which is calculated using the first type of moving window. The former procedure corresponds to a uniform discretization of the wavelet translation parameter, while the second one implements a dyadic discretization strategy.<sup>41</sup> As can be seen from Figure 1b, the length of the window is not constant along time, and therefore not all the wavelet coefficients are used for monitoring at each stage.

Finally, we have the non-overlapping moving windows of constant dyadic length ( $2^{J_{dec}}$ ), over which all the relevant orthogonal wavelet coefficients can be calculated using batches of data collected every ( $2^{J_{dec}}$ )<sup>th</sup> observation (Figure 1c). This strategy also corresponds to a dyadic discretization of the wavelet translation parameter, but now all the coefficients for the selected decomposition depth are calculated simultaneously.

Let us now consider the situation where, among the collected data set, there are variables whose values regard averages over different time supports (multiresolution data). These values become available at the end of such time periods, at which they are recorded in the data storing units. The traditional way for incorporating this type of data in monitoring procedures designed to analyze information available at a single resolution usually consists of holding the last available value constant during the time steps at the finest resolution where no new information is collected, until a new average value become available (zero-order hold).

This strategy creates a mismatch between the time support where the averages were calculated, and the one attributed to the average values. To illustrate this point, let us consider the situation where a variable corresponding to averages over four successive observations at the finest resolution is being acquired. Figure 2a illustrates the time ranges along which the



**Figure 2. Time ranges over which the average values are actually calculated (a) and those where the values are held constant in a conventional strategy to incorporate multiresolution data in single resolution methodologies (b).**

average values are calculated, while Figure 2b depicts the ranges where the values are held constant with such a procedure. These are quite different, and in fact only have an interception point in the discretization grid at the finest resolution.

From the different discretization approaches above described, the one that we have found to be most adequate for setting up a multiresolution MSSPC procedure (MR-MSSPC) is the variable window length, dyadic discretization. As happens with its constant window length dyadic counterpart, this strategy has the important property of allowing low resolution measurements to maintain their effective time supports as represented in Figure 2a. This derives from the compact support associated with the calculations of the approximation coefficients, but without introducing large time delays in the monitoring procedure. The uniform procedure was designed to handle on-line MSSPC tasks in situations where all the variables have the same resolution (single resolution data), where it is quite effective, but requires, for this same reason, a data structure of the type represented in Figure 2b.

### Description of the MR-MSSPC methodology

Our MR-MSSPC methodology begins with a specification of the resolution relative to the values collected for each variable. Quite often there is a finest resolution, corresponding to variables collected at higher sampling rates, that is used to set the finest grid (scale index  $j = 0$ ). If variable  $X_i$  corresponds to averages over time supports of length  $2^{J_i}$  times that of the finest resolution, then its scale index or resolution level is set to  $J_i$ . A variable at a resolution  $J_i$  will only be decomposed to scales coarser (that is, higher) than  $J_i$ , and therefore it does not contribute to the monitoring procedures implemented at finer scales ( $j \leq J_i$ ). This attribution is straightforward in situations where the low resolution variables represent averages over dyadic supports. In case this does not happen, we propose setting  $J_i$  as the immediately next coarser scale, that is,  $J_i = \lceil \log_2(AS) \rceil$  where  $AS$  is the averaging time support,  $\lceil \log_2(AS) \rceil$  standing for the smallest integer  $n \geq \log_2(AS)$ , and project data onto this scale using a weighted averaging procedure that will be described later. The decomposition depth to be used in the wavelet transformation phase,  $J_{dec}$ , is another parameter to be set before implementation of the methodology. It must be higher than  $\max\{J_i\}_{i=1:m}$  (usually  $J_{dec} \geq \max\{J_i\}_{i=1:m} + 2$  for reasons related to the ability to reconstruct behavior of past events). A summary of the whole procedure is presented in Table 1.

The PCA models developed in the initial stage, involving wavelet coefficients calculated from reference data (I), are not only for the detail coefficients at each scale ( $0 < j \leq J_{dec}$ ) and

**Table 1. Summary of Our MR-MSSPC Methodology**

- 
- I. Calculate PCA models at each scale using reference data.
- For each variable ( $X_i, i = 1:m$ ), perform the wavelet decomposition from  $J_i + 1$  to  $J_{dec}$ .
  - Calculate the mean vectors and covariance matrices for all scales.
  - Select the number of PC and calculate PCA models for all scales.
- II. Implement the MR-MSSPC methodology.
- For each observation index,  $k$ , multiple of  $2^{J_{min}+1}$  ( $J_{min} = \min\{J_i\}_{i=1:m}$ ):
    - Get dyadic window corresponding to current observation (length equal to  $2^{J_{max}(k)}$ ).
    - Decompose those variables  $X_i$  for which  $J_i < J_{max}(k)$ , from  $J_i + 1$  to  $J_{max}(k)$ .
    - Implement PCA-based SPC at each scale where coefficients are available, using Hotelling's  $T^2$  and  $Q$  statistics, and select the scales where significant events are detected from the standpoint of these statistics.
    - Using the scales where significant events are detected, reconstruct data at the resolution levels between the coarser scale where a significant event was detected,  $J^*$ , and  $J_{min}$ , i.e., scales  $j$  that satisfy the condition  $j: J_{min} < j < J^* \wedge j \in \{J_i\}_{i=1:m}$ . Do the same for the mean vectors and covariance matrices associated with the selected scales.
  - Using the reconstructed data, recombined covariance matrices, and mean vectors, calculate  $T^2$  and  $Q$  statistics at each intermediate resolution, and look for significant events in these charts. If none detects a significant event, consider the process to be operating under normal conditions; if any of them does show an abnormal value, then trigger an alarm, and study the contribution plots for the reconstructed statistics at the scale where the signal occurs. The plots of the tests performed at each scale also contain information about the frequency ranges involved in the process perturbation identified.
- 

for the approximation coefficients at scale  $J_{dec}$ , as happens with MSSPC with uniform discretization, but also for the approximation coefficients at scales  $0 < j \leq J_{dec}$ . Thus, we often do have available approximation coefficients for  $j < J_{dec}$ , and, in practice, we found that they can play indeed an important role in the early detection of sustained shifts.

Furthermore, PCA models at different scales may have a different number of variables associated with them. The models for the finest scales have fewer variables than those for the coarser scales, as these also integrate lower resolution variables. For this reason, the number of components for the models at each scale has to be chosen individually. A rule must also be defined to specify the number of PC to be used in the PCA model for reconstructed data at the finest scale. We will use the minimum from all the scales involved in the reconstruction, so that the number of PC will always be smaller than the number of reconstructed variables.

Stage II.iv corresponds to an extension of the MSSPC's reconstruction phase using information from scales where significant events were detected, back to the original time domain, when multiresolution data are present. As we are now dealing with variables having different resolutions, we test the statistics derived from reconstructed data at these intermediate resolutions, besides  $j = 0$  (now converted to the more general  $J_{min}$ ), provided that they stay below the coarser scale where significant activity was detected. Therefore, we may end up with more than one plot for the reconstructed  $T^2$  and  $Q$  statistics (one per resolution satisfying the conditions mentioned). Thus,

in order to maintain the overall significance level of the SPC procedure adopted in the confirmation phase for each of the two statistics, control limits are adjusted using a correction factor applied over significance level ( $\alpha$ ), derived from the Bonferroni inequality:  $\alpha/n_{charts}$ , where  $n_{charts}$  is the number of charts used simultaneously for each statistic.

Another relevant issue regards the wavelet decomposition of variables available at coarser resolutions. Filtering operations, followed by dyadic down-sampling at each stage of the wavelet decomposition, encompass scaling operations that assure energy conservation for the orthonormal transformation (Parseval relation). As the coarser resolution variables have fewer decomposition stages than the other finer resolution variables, scaling operations that might have been made initially to the whole data set would now be distorted at each scale if no additional scaling is imposed to the coarser resolution variables. We have considered this issue in our computations, preventing such distortion from occurring in the implementation of MR-MSSPC.

### Illustrative Examples of MR-MSSPC Application

In this section, we will illustrate the main features of the proposed MR-MSSPC methodology through its application to a number of case studies. The good properties of MSSPC methodologies in the monitoring of systems exhibiting autocorrelation were already widely explored in the literature.<sup>13,16,17,20,41</sup> Such properties are inherited by the proposed MR-MSSPC. In fact, since our methodology is based upon a dyadic discretization strategy, the decorrelation ability of the multiresolution decomposition is even higher than the one associated with a uniform discretization scheme. It is therefore expected to be even more suited to address highly correlated and nonstationary processes.<sup>41</sup> For this reason, we will focus our application studies mainly over stationary uncorrelated systems, where the main features of the method can be more clearly illustrated. However, we also present an example regarding a more complex dynamic system (CSTR under feedback control), where several interesting features connected to real world implementations of the methodology are also addressed.

The following latent variable model was adopted for data generation in the first three studies presented below,<sup>13</sup> since this kind of model structure is representative of data collected from many real world industrial processes:<sup>42,43</sup>

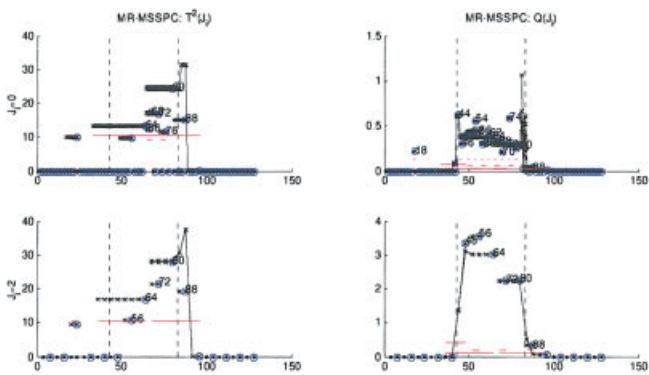
$$X(k) = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ 1 & -1 \end{bmatrix} \cdot L(k) + \varepsilon(k) \quad (1)$$

$$L(k) \sim iid N(0, \Sigma_l), \Sigma_l = \mathbf{I}_2$$

$$\varepsilon(k) \sim iid N(0, \Sigma_e), \Sigma_e = 0.2 \cdot \mathbf{I}_4$$

where  $\mathbf{I}_m$  is the identity matrix with dimension  $m$ .

For the purposes of illustrating our MR-MSSPC framework, variable  $X_4$  contains coarser resolution information, given as the successive averages over nonoverlapping windows of length  $AS$  (to be defined in each example), while the remaining variables are all available at the finest resolution. Therefore, variable  $X_4$  will only be acquired at the end of each period of



**Figure 3. Plots of the  $T^2$  and  $Q$  statistics at the two resolutions available in the data set,  $J_i = \{0,2\}$ , using data reconstructed from significant scales.**

Control limits are set for a confidence level of 99% (horizontal line segments). Legend:  $\ominus$  - signals effective plotting times (“current times”);  $\times$  - appears if the statistic is significant at “current time,” in which case its values in the same dyadic window are also represented (the “current time” index also appears next to the corresponding circle); — control limit for the statistic, which is represented every time a significant event is detected at some scale relevant for the control chart;  $\bullet$  - indicates a “common cause” observation (event is not significant). [Color figure can be viewed in the online issue, which is available at [www.interscience.wiley.com](http://www.interscience.wiley.com).]

AS consecutive observations, representing the mean of its values over that period of time.

In the first example, presented next, we illustrate the situation where the averaging window length, AS, is a dyadic number ( $2^j$ ), leaving for later the situation where such a support is nondyadic.

**Example 1. MR-MSSPC for multiresolution data with dyadic supports**

A reference set with 4096 observations was generated using the latent variable model.<sup>1</sup> Variables  $\{X_1, X_2, X_3\}$  are available at the finest scale ( $J_1 = J_2 = J_3 = 0$ ), while variable  $X_4$  represents averages over windows with length 4 ( $J_4 = 2$ ). To test the monitoring features of MR-MSSPC, 128 observations are generated and a shift of magnitude +1 is imposed to all variables between observations 43 and 83 (included). The fact that transition times do not fall in the dyadic grid at a boundary between two averaging windows is intentional, in order to see how the method behaves in such less favorable conditions.

Figure 3 presents the results obtained regarding control charts for the  $T^2$  and  $Q$  statistics at the two resolutions available in the data set, that is, at  $J_i = \{0, 2\}$ . In our MR-MSSPC charts, circles are used to indicate that the respective statistic’s abscissa corresponds to the time where the last value of the method’s analyzing dyadic window is acquired, and where calculations are actually made and results plotted. For instance, in Figure 3, circles appear every 2<sup>nd</sup> observation in the plots for  $J_i = 0$ , and every 4<sup>th</sup> observation in the plots for  $J_i = 2$ , since only at such time instants do new values become available at these resolutions. Therefore, a decision about the state of the process is only taken at times corresponding to observations signaled by circles.

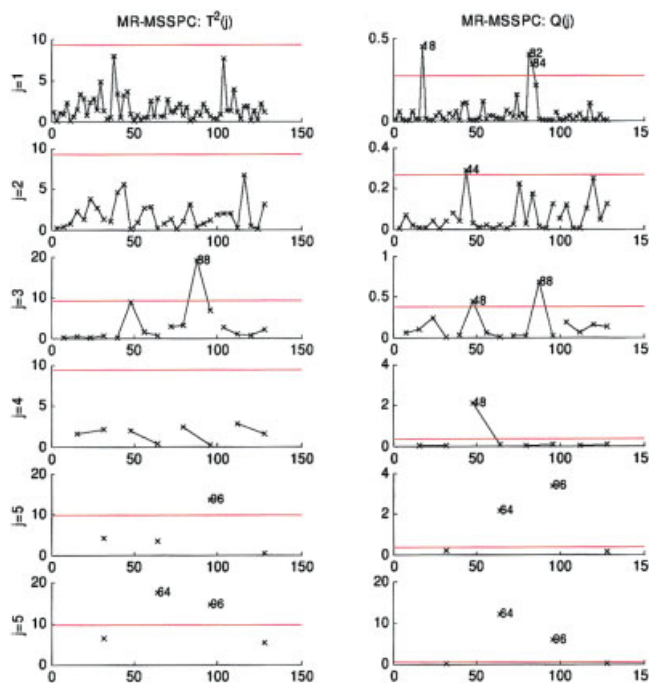
If a statistic ( $T^2$  or  $Q$ ) signaled with a circle falls above

control limits, its observation number also appears next to it. That being the case, all the values of this statistic regarding the same dyadic window are also presented (as well as the associated control limits). This plotting feature enables a more accurate reconstruction of the time at which a special cause occurred, even if it is detected at a later stage.

We also represent the points of the statistics and control limits even if the last observation is not significant, provided that there is at least one scale where a significant event was detected (with no number associated with it in the plots). This allows us to visualize more clearly when the process returns back to normal operation, as well as to see imminent abnormal situations in their early stages.

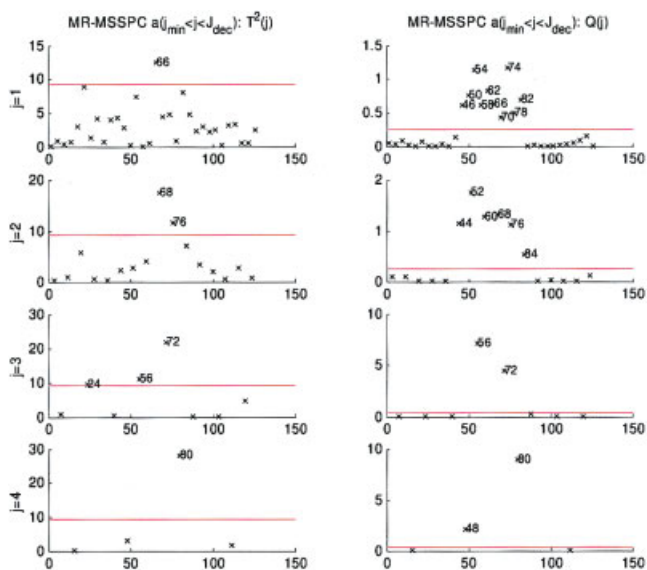
When no significant event is detected at any scale, a “zero” point is plotted. From Figure 3, we can see that  $Q$  charts are more sensitive than  $T^2$  charts for the type of fault analyzed in this example. The  $Q$  statistic at  $J_i = 0$  clearly indicates that an abnormal observation has occurred in the immediate past neighborhood of observation 44, and that the process has returned back to normal shortly after observation 80. A mild spurious observation is also detected, again at time 88, but in general the procedure was able to reconstruct quite clearly what has occurred when the process returned back to normality.

Figure 4 illustrates the underlying MR-MSSPC monitoring tasks conducted at each scale over detail coefficients for  $0 < j \leq J_{dec}$  and on the approximation coefficients at scale  $J_{dec}$ , while Figure 5 regards the ones involving approximation coefficients for scales  $0 < j \leq J_{dec}$ . As can be seen from these plots, detail coefficients play an important role in the detection of transition times, while approximation coefficients have the



**Figure 4. Plots of the  $T^2$  and  $Q$  statistics for detail coefficients at each scale ( $J_{dec}$ ) and for approximation coefficients at scale ( $0 < j \leq J_{dec}$ ), with control limits set for a confidence level of 99%.**

[Color figure can be viewed in the online issue, which is available at [www.interscience.wiley.com](http://www.interscience.wiley.com).]



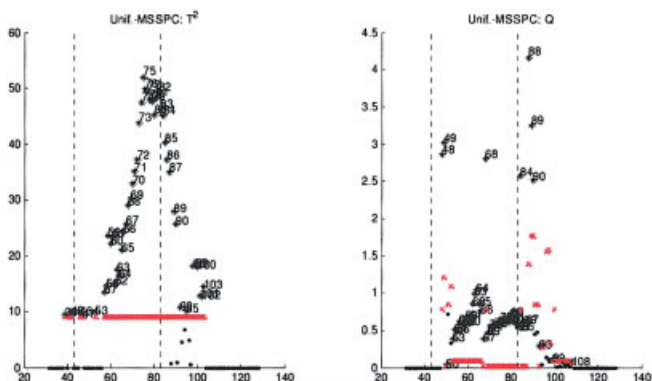
**Figure 5. Plots of the  $T^2$  and  $Q$  statistics for the approximation coefficients at scales  $0 < j < J_{dec}$ , with control limits set for a confidence level of 99%.**

[Color figure can be viewed in the online issue, which is available at [www.interscience.wiley.com](http://www.interscience.wiley.com).]

complementary role of signaling abnormalities during the duration of a sustained shift.

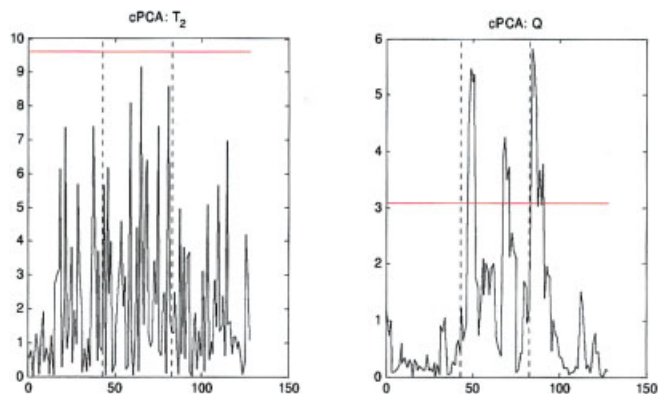
We now analyze the same situation, but using techniques designed to handle data at a single resolution that adopt the procedure for handling multiresolution data represented in Figure 2b. Results regarding the  $T^2$  and  $Q$  statistics for the MSSPC methodology with uniform discretization (Unif.-MSSPC) are presented in Figure 6. Once again, the number of the observation appears as a label when it is significant from the standpoint of the control chart statistic. One can see that control charts detect the shift quite promptly, but the definition of the region where the shift occurs is distorted due to the way values for lower resolution variables are handled.

Figure 7 presents results regarding the use of PCA-SPC. We



**Figure 6. Results for MSSPC with uniform discretization: plots of the  $T^2$  and  $Q$  statistics for reconstructed data.**

Control limits are set for a confidence level of 99% (represented by symbol  $\times$ ). [Color figure can be viewed in the online issue, which is available at [www.interscience.wiley.com](http://www.interscience.wiley.com).]



**Figure 7. Results for cPCA-SPC: plots of the  $T^2$  and  $Q$  statistics, with control limits set for a confidence level of 99% (cPCA stands for “classic” PCA, also referred to here simply as PCA).**

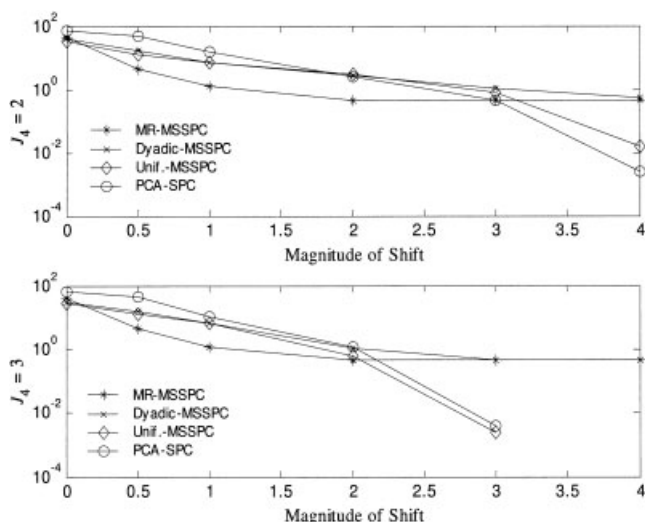
[Color figure can be viewed in the online issue, which is available at [www.interscience.wiley.com](http://www.interscience.wiley.com).]

can see that in this case only the  $Q$  statistic detects significant abnormal activity going on during the duration of the shift, even though its detection rate is not as high as that exhibited by multiscale methods. This difference derives from the increased sensitivity of MSSPC methods, which have the ability of zooming into process behavior at different scales (octave frequency bands) in the search for changes in normal variability patterns.

The main feature of the MR-MSSPC methodology illustrated in this example is its ability to define well the duration of abnormalities when multiresolution data are present. It is quite sensitive in detecting its beginning, but, even more than this, effective in the detection of its end, due to a consistent use of the time supports regarding low resolution values achieved through the implementation of an orthogonal wavelet transformation over a variable dyadic length window. In the next example, we consolidate these features through an extended Monte Carlo simulation study.

### Example 2. MR-MSSPC: extended simulation study

The previous example illustrated several interesting properties of MR-MSSPC. In order to consolidate such properties and show that they are not connected with the particular dataset analyzed, an extended Monte Carlo simulation study was also conducted, where several shifts were tested, together with different resolution levels associated to variable  $X_4$  (or, stated equivalently, with averaging windows of different lengths used for variable  $X_4$ ). Resolution levels tested were  $J_4 = \{2, 3\}$ , and the shift magnitudes analyzed were as follows:  $shifts = \{0, 0.5, 1, 2, 3, 4\}$ . For each resolution level, a reference data set composed of 2048 observations was created in order to estimate the models underlying each of the tested methodologies, after which a test set with 256 observations was generated with a shift introduced between observation number  $n_i$  and observation number  $n_f$ . To avoid biases due to shift location,  $n_i$  is randomly extracted from a uniform distribution:  $n_i \sim U(40, 50)$ , while the duration of the shift was kept constant, corresponding to the next 40 observations ( $n_f = n_i + 40$ ). The



**Figure 8. ARL results for the different methodologies, using shifts of different magnitude and two levels of resolution associated with variable  $X_4$ .**

generation of the test set and shift location was repeated 2000 times for each shift magnitude.

The methods tested and compared here are the following: MR-MSSPC, Dyadic-MSSPC (MR-MSSPC with dyadic discretization, but using only data at a single resolution), Unif.-MSSPC, and PCA-SPC.

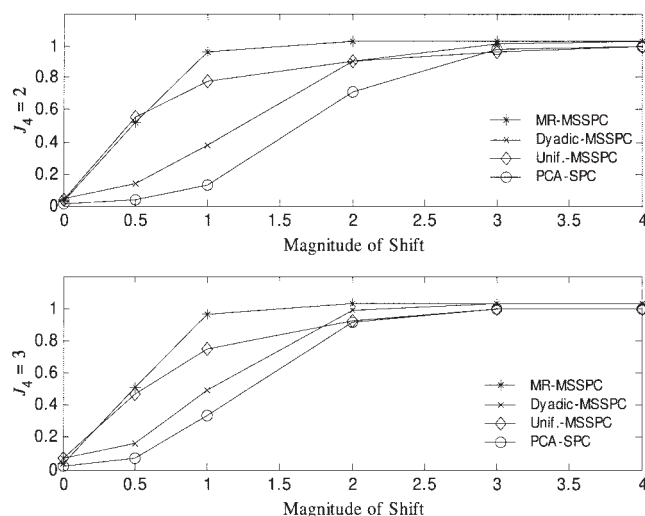
The detection metrics that will provide the ground for our comparisons are:

- Average run length (ARL), calculated considering the first occurrence of a significant event either in the  $T^2$  or  $Q$  control charts;
- True Positive Rate (TPR), which in this work corresponds to the fraction of significant events detected during the duration of the shift (between  $n_i$  and  $n_f$ ), relative to the maximum possible amount of detections that could be achieved with each methodology (that is, if all the statistics' values computed during this interval were assumed to be significant);
- False Positive Rate (FPR), here defined as representing the fraction of false alarms detected right after the process returns to normality, in a range of time with the same amplitude as the one used for calculating TPR (between  $n_f + 1$  and  $n_f + 41$ ), once again relative to the maximum possible amount of detections that could be achieved with each technique.

The number of selected principal components was kept constant at 2,  $J_{dec}$  set to 5, and the wavelet transform used was the Haar transform. Our significance levels, adopted for each method, were also adjusted in order to obtain similar ARL(0) performances (average run length obtained under the absence of any shift).

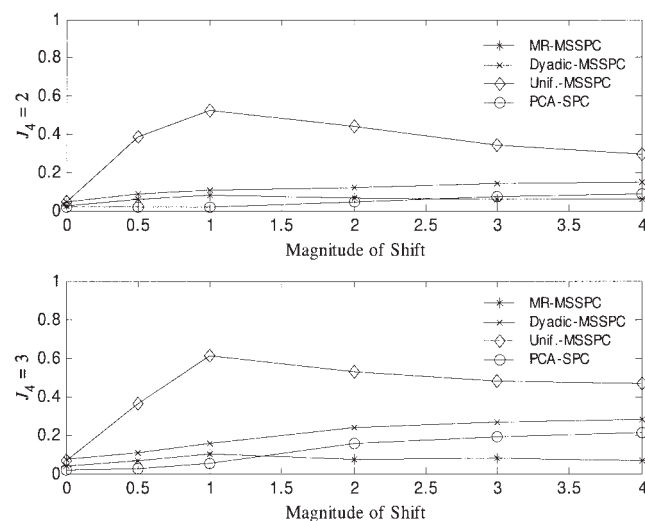
Figure 8 compares ARL performances obtained for the various methods. The time delay associated with MR-MSSPC only becomes an issue for shifts of magnitude greater than 2, after which it stabilizes at around 0.5, which seems acceptable for most applications. Thus, even though speed of detection was not a specific goal motivating the conception of our framework, it ends up also performing well in this regard.

Results regarding TPR are shown in Figure 9, where we can



**Figure 9. TPR results for the different methodologies, using shifts of different magnitude and two levels of resolution associated with variable  $X_4$ .**

see that MR-MSSPC performs better than its alternatives. It also does quite well when the process goes back to normal operation (Figure 10), with a low false alarm rate, which is only sometimes overtaken by PCA-SPC. However, in this situation, one must not forget that such a technique presents lower true positive detection metrics (TPR), as shown in Figure 9. Therefore, these results point towards an improved overall performance achieved by MR-MSPSC regarding the duration of the fault (higher TPR), quick detection of its beginning (low ARL), and effective delimitation of its end (low FPR).



**Figure 10. FPR results for the different methodologies, using shifts of different magnitude and two levels of resolution associated with variable  $X_4$ .**



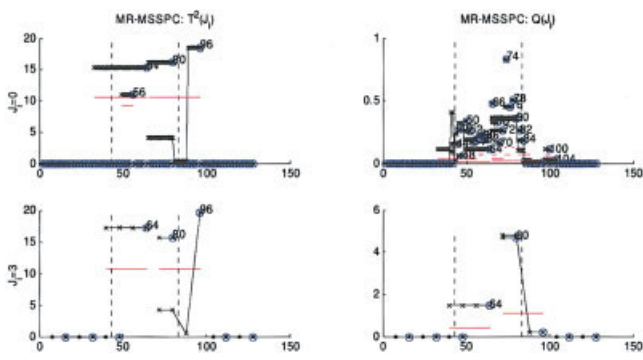
**Table 2. Selection of Resolution Index ( $J_i$ ) When the Averaging Support for a Lower Resolution Variable Is Not Dyadic**

- I. Set  $J_i$  as the index for the next coarser scale, i.e.,  $J_i = \lceil \log_2(AS) \rceil$ , where  $AS$  is the averaging time support and  $\lceil \log_2(AS) \rceil$  stands for the smallest integer  $n \geq \log_2(AS)$ .
- II. Project data onto scale  $J_i$  using the following weighted averaging procedure:
  - a. FOR each window of dyadic length under analysis (length  $2^{J_{max}(K)}$ ), divide it into sub-regions of length  $2^{J_i}$  ( $K$  sub-regions).
    1. FOR each sub-region ( $l = 1:K$ ):
      - i. Collect lower resolution values ( $x_j$ ) and calculate the fractions of their averaging support contained in the sub-region  $l$  under analysis ( $w_j$ ).
      - ii. Calculate the weighted average of the collected values:  $X_i(l) = \sum_j w_j x_j / \sum_j w_j$ .

**Example 3. MR-MSSPC for multiresolution data with nondyadic supports**

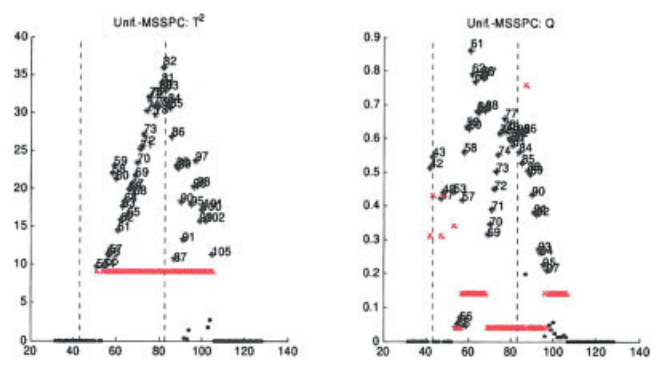
When the values of a lower resolution variable represent the mean values over a non-dyadic time support, the attribution of its scale index is not straightforward. In this example we show how this issue can be effectively dealt with within the scope of our MR-MSSPC framework. The proposed solution consists of implementing the sequence of steps summarized in Table 2.

When the averaging supports have dyadic length, the above procedure provides the same values for  $X_i$  as the standard procedure. When they do not have such a property, it balances the contribution from each value within each sub-region of length  $2^{J_i}$ . To illustrate the application of this strategy, we will now consider variable  $X_4$  in model (1) to represent the average taken over a window of 5 successive values. Thus, according to step I in Table 2,  $J_4 = \lceil \log_2(5) \rceil = 3$ . The data set is also processed in order to be used with approaches based on single resolution data, by holding average values constant until a new mean value becomes available (Figure 2b). The results obtained for MR-MSSPC, Unif.-MSSPC, and PCA-MSSPC, when a shift of magnitude 1 is introduced between observation 43 and 83 (included), are presented in Figures 11 to 13. An improvement in the definition of the faulty region, as well as in



**Figure 11. Plots of the  $T^2$  and  $Q$  statistics at the two resolutions available in the data set,  $J_i = \{0,3\}$ , using data reconstructed from significant scales.**

Control limits are set for a confidence level of 99%. [Color figure can be viewed in the online issue, which is available at [www.interscience.wiley.com](http://www.interscience.wiley.com).]



**Figure 12. Results of MSSPC with uniform discretization: plots of the  $T^2$  and  $Q$  statistics for the reconstructed data.**

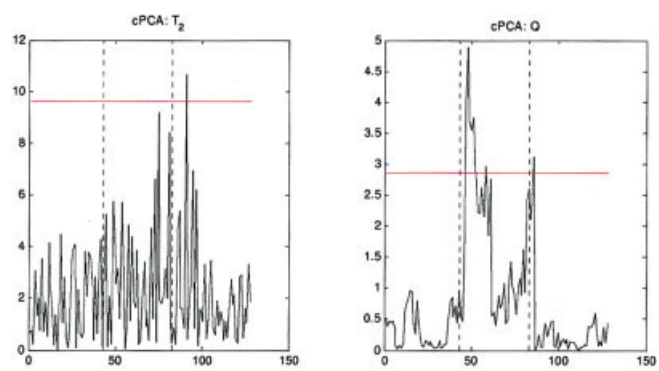
Control limits are set for a confidence level of 99% (now represented by symbol  $\times$ ). [Color figure can be viewed in the online issue, which is available at [www.interscience.wiley.com](http://www.interscience.wiley.com).]

the detection of returns back to normality, is obtained through MR-MSSPC, even in this situation where the averaging window does not have a dyadic support.

**Example 4. MR-MSSPC applied to a CSTR under feedback control**

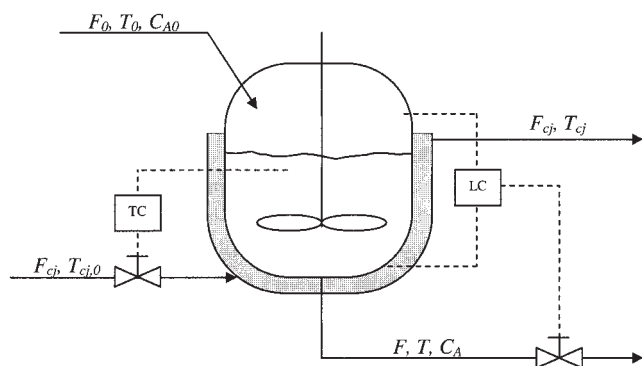
This final example aims to illustrate several interesting features that may arise in real world process monitoring implementations of MR-MSSPC, under the presence of time dynamics and non-linearities, besides variables collinearity. In particular, we will analyze how the number of components for the PCA models at each scale can be selected from an analysis of the multiscale covariance structure relative to normal operation conditions, and show how our method performs in the detection of a bias in the temperature sensor for the inlet stream. Furthermore, we also see how such information can be applied in order to set the resolution needed for some variables to higher scale indices, without compromising, in general, monitoring performance.

The system considered here consists of a simulated industrial nonisothermal CSTR, where an irreversible and exothermic



**Figure 13. Results for cPCA-SPC: plots of the  $T^2$  and  $Q$  statistics.**

Control limits are set for a confidence level of 99%. [Color figure can be viewed in the online issue, which is available at [www.interscience.wiley.com](http://www.interscience.wiley.com).]



**Figure 14. Representation of CSTR with level and temperature control.**

first order reaction ( $A \rightarrow B$ ) takes place.<sup>44</sup> This reactor is equipped with a water jacket that removes excess heat released and two control loops (proportional action) that act upon two manipulated variables, that is, outlet flow rate ( $F$ ) and flow rate through the jacket ( $F_{c_j}$ ), in order to control the process variables volume,  $V$ , and reactor temperature,  $T$ , respectively. Figure 14 illustrates this system, whereas more details about its mathematical model, parameters, and operating conditions can be found in the Appendix.

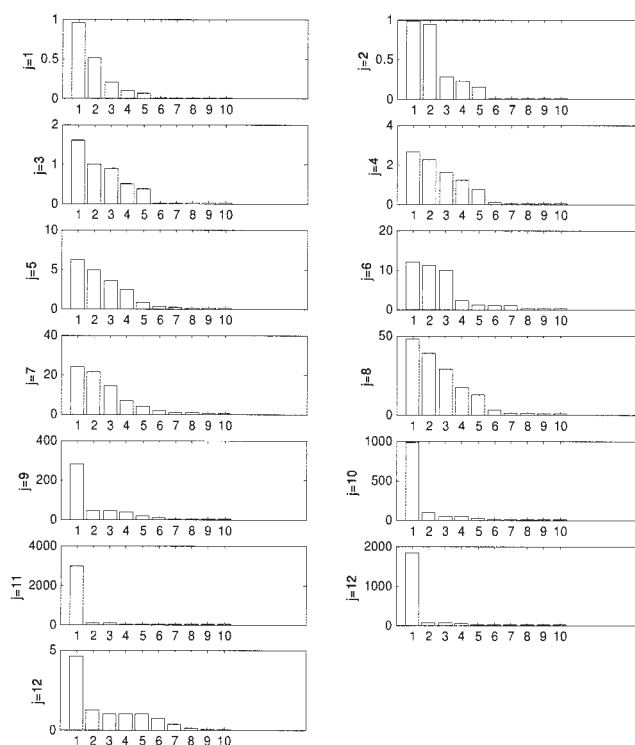
Normal operating conditions variability was generated by considering randomness associated with variables  $F_0$ ,  $T_0$ ,  $T_{c_j,0}$  and  $C_{A0}$ . For the first three variables, they were assumed to be of the autoregressive type (first order, Eq. 2), with parameters as presented in Table 3:

$$X(k) = \phi \cdot X(k-1) + \varepsilon(k), \quad \varepsilon(k) \sim N(0, \sigma_x^2) \quad (2)$$

As for  $C_{A0}$ , we assume that the reactant is fed to the CSTR from successive tanks, with  $2 \text{ m}^3$  each, for which concentration shows little variation (homogenous mixture within each tank), but changing from tank to tank, according to  $C_{A0} \sim N(0.5, 0.1^2)$ . A reference data set was generated, representing 364 hours of normal operation, during which 10 variables,  $\{X_i\}_{i=1:10} = \{V, T, T_{c_j}, C_A, F_0, C_{A0}, T_0, T_{c_j,0}, F, F_{c_j}\}$ , were collected every 10 s, and analyzed in order to set monitoring parameters, estimate models at each scale, and gain insight regarding multiscale features. In this preliminary analysis stage, a decomposition depth of  $J_{dec} = 12$  was chosen, which is high enough to characterize all the phenomena going on in the range of scales relevant for this process. Figure 15 presents the eigenvalues profiles for covariance matrices at each scale. This plot clearly illustrates that the dimension of the relevant PCA subspaces for process monitoring purposes in dynamic non-linear systems is, in general, a function of scale, according to the power spectra of the variables involved and

**Table 3. Parameters of Autoregressive Models Used for Simulating Normal Operation Regarding Variables  $F_0$ ,  $T_0$ , and  $T_{c_j,0}$**

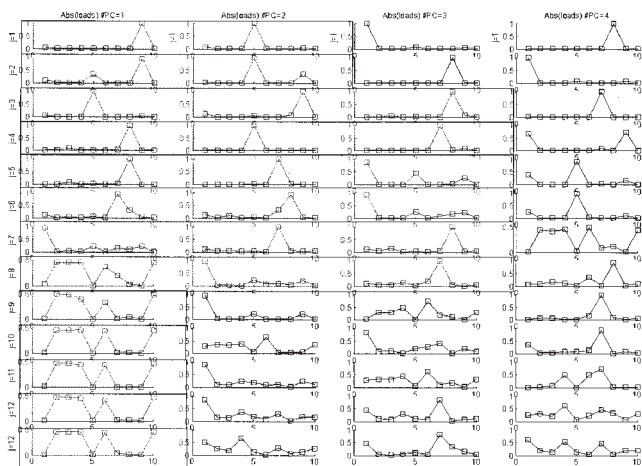
Variable	$\phi$	$\sigma_x^2$
$F_0$	0.5	0.750
$T_0$	0.95	0.878
$T_{c_j,0}$	0.9	1.72



**Figure 15. Eigenvalues plots for the covariance matrices regarding variables wavelet detail coefficients at each scale ( $j = 1:12$ ) and for the wavelet approximation coefficients at the coarsest scale ( $j = 12$ , last plot at the bottom).**

the correlation they present in the frequency bands corresponding to different scales. Using the information conveyed by such plots, we can also choose the number of principal components to be adopted for PCA models of the wavelet coefficients at each scale, as well as get important clues regarding the decomposition depths that should be used in order to capture the system's main dynamic features. In this particular case, we did set  $J_{dec} = 9$  because above this scale the behavior of the correlation structure does not seem to change significantly, thus meaning that relevant dynamic features of the system are expressed at lower scales.

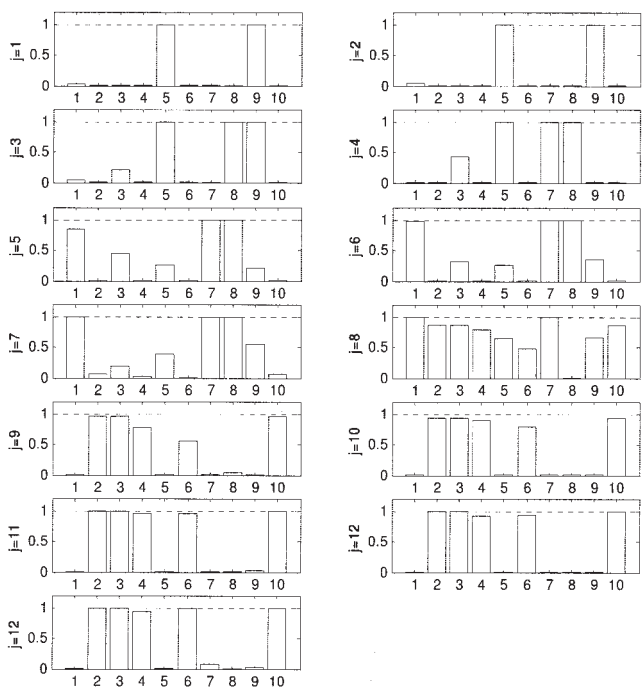
The absolute values of the load vectors for the selected principal components, at each scale, are presented in Figure 16 (shaded plots), where we can deepen the analysis of the correlation structure at different scales, looking for the main active relationships in each frequency band, and distinguish which variables are more significantly involved. This last point can be conducted more effectively by looking at the fraction of explained variance for each variable in the PCA model developed at each scale (Figure 17). From these two figures we can see that, although there is some overlapping due to the interception between frequency bands characteristic of some variables, in scales 1–3 the variables involved are mainly those with fast dynamics (notably, flow rates), whereas in the intermediate scales (3–8) we get those variables with slower dynamics (temperatures), as well as the attenuated effect of flow rate “filtered” by reactor capacity (volume). Finally, in scales 8–12, the slow mode variables ( $C_{A0}$ , the majority of the system outputs, and control variables) become relevant.



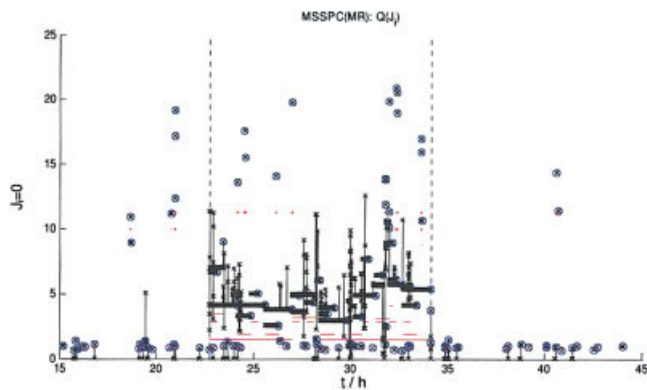
**Figure 16. Absolute value of the load vectors associated with the principal components selected at each scale (shaded graphs).**

After having estimated the MR-MSSPC monitoring parameters, a test data set containing about 45 hours of operation data was also generated, with a bias of 6 K introduced in  $T_0$  ( $\Delta/\sigma \approx 2$ ), between times 22h:46m and 34h:09m. The monitoring results obtained for the MR-MSSPC  $Q$  statistic are presented in Figure 18, showing that our method successfully detected this shift.

As Figure 17 points out, variable number 4,  $C_{AO}$ , only becomes relevant at coarser scales. This means that we can use a lower resolution to represent its behavior along time, without losing much detail but introducing a time delay related with



**Figure 17. Fraction of explained variance for each variable in the PCA model developed at each scale.**



**Figure 18. Plot of the  $Q$  statistic for MR-MSSPC applied over the test data set, with all variables available at the finest scale  $J_i = 0$  ( $i = 1:10$ ) (control limits defined for a 99% confidence level).**

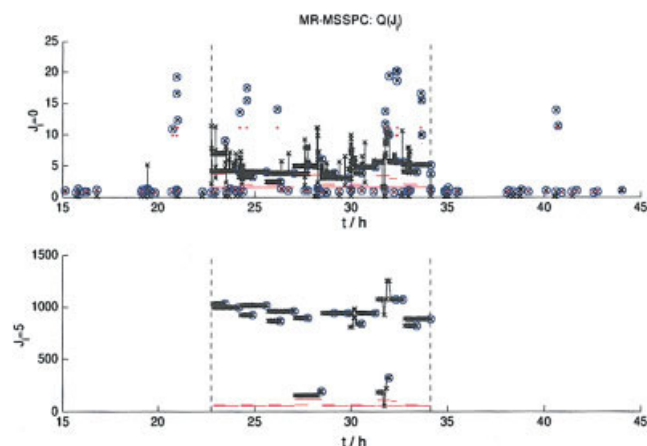
[Color figure can be viewed in the online issue, which is available at [www.interscience.wiley.com](http://www.interscience.wiley.com).]

the process of average computation. Figure 19 illustrates what happens when we set the resolution of  $C_{AO}$  at  $J_4 = 5$ , and conduct MR-MSSPC over the same test data set. Detection results do not change significantly, but the location of the fault becomes even more evident in the representation for  $J_i = 5$ .

In Figure 19 we are not handling existing multiresolution data, but are actually creating a multiresolution data structure, after analyzing the multiscale characteristics of the system operating under normal conditions. The coarser scale selected represents a trade-off between the adequate scale to express a certain variable and the time delay involved in the computation of its mean values.

## Conclusions

We presented an approach for conducting MSSPC that adequately integrates data with different resolutions (multireso-



**Figure 19. Plot of the  $Q$  statistic for MR-MSSPC applied over the test data set, with all variables available at the finest scale, except  $C_{AO}$ , which is available at  $J_4 = 5$  (control limits defined for a 99% confidence level).**

[Color figure can be viewed in the online issue, which is available at [www.interscience.wiley.com](http://www.interscience.wiley.com).]

lution data). Such an approach was then tested under four different scenarios, in order to illustrate and establish its main features. The first three examples underline a consistent use of the time supports regarding lower resolution variables, thus enabling a clearer definition of the regions where significant events occur and a more sensitive response when the process is brought back to normal operation. They also show that, as long as the fault does not happen exclusively in the lower resolution variables, no significant time delay is introduced by the proposed methodology. A final example brings into discussion both interesting and important issues regarding practical applications involving dynamic systems with non-linearities, such as the interpretation of their multiscale covariance structure, selection of monitoring parameters, and corresponding scales. Ongoing and future work address an extension of these ideas to the development of a multiresolution dynamic modeling framework, which encompasses the definition of a multiscale model structure, its identification, and posterior use in other tasks, such as multiresolution optimal estimation and control. At this stage, we are also examining some applications of our data analysis framework to sets of industrial data obtained from chemical plants.

## Acknowledgments

The authors would like to acknowledge financial support from the Portuguese Foundation for Science and Technology (project POCTI/EQU/47638/2002).

## Literature Cited

- Izadi I, Zhao Q, Chen T. An optimal scheme for fast rate fault detection based on multirate sampled data. *J Process Control*. 2005; 15:307-319.
- Lu N, Yang Y, Gao F, Wang F. Multirate dynamic inferential modeling for multivariable processes. *Chem Eng Sci*. 2004;59:855-864.
- Tangirala AK. *Multirate Control and Multiscale Monitoring of Chemical Processes*, Ph.D. Thesis, University of Alberta; 2001.
- Arteaga F, Ferrer A. Dealing with missing data in MSPC: several methods, different interpretations, some examples. *J Chemometrics*. 2002;16:408-418.
- Little RJA, Rubin DB. *Statistical Analysis with Missing Data* (2nd ed.). Hoboken, NJ: Wiley; 2002.
- Nelson PRC, Taylor PA, MacGregor JF. Missing data methods in PCA and PLS: score calculations with incomplete observations. *Chemometrics and Intelligent Laboratory Systems*. 1996;35:45-65.
- Walczak B, Massart DL. Dealing with missing data. *Chemometrics and Intelligent Laboratory Systems*. 2001;58:Part I:15-27, Part II:29-42.
- Basseville M, Benveniste A, Chou KC, Golden SA, Nikoukhah R, Willsky AS. Modeling and estimation of multiresolution stochastic processes. *IEEE Trans Information Theory*. 1992;38(2):766-784.
- Chou KC, Willsky AS, Benveniste A. Multiscale recursive estimation, data fusion, and regularization. *IEEE Trans Automatic Control*. 1994; 39(3):464-478.
- Willsky AS. Multiresolution Markov models for signal and image processing. *Proceedings IEEE*. 2002;90(8):1396-1458.
- Bakshi BR. Multiscale analysis and modeling using wavelets. *J Chemometrics*. 1999;13:415-434.
- Motard RL, Joseph B, eds. *Wavelet Applications in Chemical Engineering*. Boston: Kluwer; 1994.
- Bakshi BR. Multiscale PCA with application to multivariate statistical process control. *AIChE J*. 1998;44(7):1596-1610.
- Kosanovich KA, Piovoso MJ. PCA of wavelet transformed process data for monitoring. *Intelligent Data Analysis*. 1997;1(1-4):85-99.
- Top S, Bakshi BR. Improved statistical process control using wavelets. Paper presented at Foundation of Computer Aided Process Operations FOCAPO 98, Snowbird, UT. 1998.
- Kano M, Nagao K, Hasebe S, et al. Comparison of multivariate statistical process monitoring methods with applications to the eastman challenge problem. *Computers Chem Eng*. 2002;26:161-174.
- Misra M, Yue HH, Qin SJ, Ling C. Multivariate process monitoring and fault diagnosis by multi-scale PCA. *Computers Chem Eng*. 2002; 26:1281-1293.
- Fourie SH, de Vaal P. Advanced process monitoring using an on-line non-linear multiscale principal component analysis methodology. *Computers Chem Eng*. 2000;24:755-760.
- Shao R, Jia F, Martin EB, Morris AJ. Wavelets and non-linear principal components analysis for process monitoring. *Control Engineering Practice*. 1999;7:865-879.
- Rosen C, Lennox JA. Multivariate and multiscale monitoring of wastewater treatment operation. *Water Research*. 2001;35(14):3402-3410.
- Yoon S, MacGregor JF. Unifying PCA and multiscale approaches to fault detection and isolation. Paper presented at 6th IFAC Symposium on Dynamics and Control of Process Systems, Korea, 4-6 June 2001.
- Trygg J, Kettaneh-Wold N, Wallbäcks L. 2D wavelet analysis and compression of on-line industrial process data. *J Chemometrics*. 2001; 15:299-319.
- Teppola P, Minkinen P. Wavelet-PLS regression models for both exploratory data analysis and process monitoring. *J Chemometrics*. 2000;14:383-399.
- Luo R, Misra M, Himmelblau DM. Sensor fault detection via multiscale analysis and dynamic PCA. *Industrial & Eng Chem Res*. 1999; 38:1489-1495.
- Sun W, Palazoğlu A, Romagnoli JA. Detecting abnormal process trends by wavelet-domain hidden Markov models. *AIChE J*. 2003; 49(1):140-150.
- Ganesan R, Das TK, Venkataraman V. Wavelet based multiscale process monitoring—a literature review. *IIE Trans on Quality and Reliability Eng*. 2004;36(9):787-806.
- Jackson JE. *A User's Guide to Principal Components*. New York: Wiley; 1991.
- Johnson RA, Wichern DW. *Applied Multivariate Statistical Analysis* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall; 1992.
- Martens H, Naes T. *Multivariate Calibration*. Chichester: Wiley; 1989.
- Johnson RA, Wichern DW. *Applied Multivariate Statistical Analysis* (5th ed.). Upper Saddle River, NJ: Prentice Hall; 2002.
- Jackson JE. Quality control methods for several related variables. *Technometrics*. 1959;1(4):359-377.
- Kresta JV, MacGregor JF, Marlin TE. Multivariate statistical monitoring of process operating performance. *Canadian J Chem Eng*. 1991;69:35-47.
- MacGregor JF, Kourti T. Statistical process control of multivariate processes. *Control Engineering Practice*. 1995;3(3):403-414.
- Montgomery DC. *Introduction to Statistical Quality Control* (4th ed.). New York: Wiley; 2001.
- Jackson JE, Mudholkar GS. Control procedures for residuals associated with principal component analysis. *Technometrics*. 1979;21(3): 341-349.
- Wise BW, Gallagher NB. The process chemometrics approach to process monitoring and fault detection. *J Process Control*. 1996;6(6): 329-348.
- Mallat S. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans Pattern Analysis Machine Intelligence*. 1989;11(7):674-693.
- Strang G, Nguyen T. *Wavelets and Filter Banks*. Wellesley, MA: Wellesley-Cambridge Press; 1997.
- Dijkerman RW, Mazumdar RR. Wavelet representations of stochastic processes and multiresolution stochastic models. *IEEE Trans Signal Processing*. 1994;42(7):1640-1652.
- Vetterli M, Kovachevic J. *Wavelets and Subband Coding*. Upper Saddle River, NJ: Prentice Hall; 1995.
- Aradhye HB, Bakshi BR, Strauss R, Davis JF. Multiscale SPC using wavelets: theoretical analysis and properties. *AIChE J*. 2003;49(4): 939-958.
- Burnham AJ, Macgregor JF, Viveros R. Latent variable multivariate regression modeling. *Chemometrics and Intelligent Laboratory Systems*. 1999;48:167-180.
- MacGregor JF, Kourti T. Multivariate statistical treatment of historical data for productivity and quality improvements. Paper presented at Foundation of Computer Aided Process Operations—FOCAPO 98, Snowbird, UT. 1998.

## Appendix

The CSTR mathematical model, supporting the simulations carried out in example 4, is shown below,<sup>44</sup> according to the nomenclature, steady state, and parameter values presented in Table A1.

Global mass balance to CSTR

$$\frac{dV}{dt} = F_0 - F \quad (\text{A1})$$

Partial mass balance to component A

$$\frac{dVC_A}{dt} = F_0C_{A0} - FC_A - k_0e^{-E/RT}C_AV \quad (\text{A2})$$

Global CSTR energy balance

$$\frac{dVT}{dt} = F_0T_0 - FT - \frac{\Delta H}{\rho C_p}k_0e^{-E/RT}C_AV - \frac{UA}{\rho C_p}(T - T_{cj}) \quad (\text{A3})$$

Global cooling jacket energy balance

$$\frac{dV_{cj}T_{cj}}{dt} = F_{cj}(T_{cj,0} - T_{cj}) + \frac{UA}{\rho_j C_{p,cj}}(T - T_{cj}) \quad (\text{A4})$$

Control of reacting mixture volume (reactor level) using outlet flow rate

$$F = F_{set} - K_{c2}(V_{set} - V) \quad (\text{A5})$$

Control of CSTR temperature using cooling water flow rate

$$F_{cj} = F_{cj,set} - K_{c1}(T_{set} - T) \quad (\text{A6})$$

**Table A1. Variables Used in the Mathematical Model and Their Steady State Values, Along with Model Parameter Values**

Variable/ Parameter	Description	Steady state value/ parameter value
$F$	Outlet flow rate	$1.133 \text{ m}^3 \cdot \text{h}^{-1}$
$V$	Reacting mixture volume	$1.359 \text{ m}^3$
$C_{A0}$	Concentration of reactant A in the inlet stream	$8.009 \text{ kmol} \cdot \text{m}^{-3}$
$C_A$	Concentration of reactant A in the CSTR	$3.924 \text{ kmol} \cdot \text{m}^{-3}$
$T$	Temperature in the CSTR	333.3 K
$T_{cj}$	Temperature in the cooling jacket	330.3 K
$F_{cj}$	Water flow in the cooling jacket	$1.413 \text{ m}^3 \cdot \text{h}^{-1}$
$T_0$	Temperature in the inlet stream	294.4 K
$V_{cj}$	Cooling jacket volume	$0.1090 \text{ m}^3$
$k_0$	Pre-exponential factor	$7.08 \times 10^{10} \text{ h}^{-1}$
$E$	Activation energy	$69780 \text{ kJ} \cdot \text{kmol}^{-1}$
$R$	Gas constant	8.332
$U$	Overall heat transfer coefficient	$0.8517 \text{ kW} \cdot \text{m}^{-2} \cdot \text{K}^{-1}$
$A$	Heat transfer area	$23.23 \text{ m}^2$
$T_{cj,0}$	Temperature in the cooling jacket's inlet stream	294.4 K
$\Delta H$	Heat of reaction	$-69780 \text{ kJ} \cdot \text{kmol}^{-1}$
$C_p$	Heat capacity of the mixture	3.140
$\rho$	Density of the mixture	$800.9 \text{ kg} \cdot \text{m}^{-3}$
$C_{p,cj}$	Heat capacity of the cooling liquid (water)	4.187
$\rho_{cj}$	Density of the cooling liquid (water)	$998.0 \text{ kg} \cdot \text{m}^{-3}$
$K_{ct}$	Tuning constant for the proportional action in the temperature control loop	$0.3568 \text{ m}^3 \cdot \text{h}^{-1} \cdot \text{K}^{-1}$
$K_{c2}$	Tuning constant for the proportional action in the level control loop	$10 \text{ h}^{-1}$
$F_{set}$	Set point for outlet reactor flow	$1.133 \text{ m}^3 \cdot \text{h}^{-1}$
$F_{cj,set}$	Set point for cooling jacket flow	$1.413 \text{ m}^3 \cdot \text{h}^{-1}$
$T_{set}$	Set point for reactor temperature	333.3 K
$V_{set}$	Set point for reacting mixture volume	$1.359 \text{ m}^3$

Manuscript received Oct. 10, 2005, and revision received Jan. 27, 2006.