Multiscale Statistical Process Control Using Wavelet Packets

Marco S. Reis and Pedro M. Saraiva

Dept. of Chemical Engineering, University of Coimbra, Polo II, Rua Sílvio Lima, 3030-790, Coimbra, Portugal

Bhavik R. Bakshi

Dept. of Chemical and Biomolecular Engineering, Ohio State University, 125 Koffolt Laboratories, 140 West 19th Ave., Columbus, OH 43210

An approach is presented for conducting multiscale statistical process control (MSSPC), based on a library of basis functions provided by wavelet packets. The proposed approach explores the improved ability of wavelet packets in extracting features with arbitrary locations, and having different localizations in the time-frequency domain, in order to improve the detection performances achieved with wavelet-based MSSPC. A novel approach is also developed for adaptively selecting the best decomposition depth. Such an approach is described in detail and tested using artificial simulated signals, employed to compare average run length (ARL) performance against other SPC methodologies. Furthermore, its performance under real world situations is also assessed, for two industrial case studies using datasets containing process upsets, through the construction of receiver operating characteristic (ROC) curves. Both univariate and multivariate cases are covered. ARL results for a step perturbation show that the proposed methodology presents a steady good performance for all shift magnitudes, without significantly changing its relative scores, as happens with other current methods, whose relative performance depends on the shift magnitude being tested. For artificial disturbances, with features localized in the time/frequency domain, multiscale methods do present the best performance, and for the particular case of detecting a decrease in autocorrelation they are the only ones that can detect such a perturbation. In the examples using industrial datasets, where disturbances exhibit more complex patterns, multiscale approaches also present the best results, in particular in the range of low false alarms, where monitoring methods are aimed to operate. © 2008 American Institute of Chemical Engineers *AIChE J,* 54: 2366–2378, 2008
Keywords: multiscale statistical process control, wavelet-packets, wavelets, statistical process control

Introduction

Most processes in modern chemical plants are typically complex, and such a complexity appears 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. Usually, not only the overall *systems* have a multiscale nature, since they are composed of processing units that span different timescales and frequency bands, but their *inputs* (manipulation actions, disturbances, faults) can also present a

---

wide variety of features, with distinct time/frequency characteristics. This fact implies that even quite simple systems cannot be effectively treated under conventional methodologies, designed to operate only at a particular scale (single scale). 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,[1,2] as alternatives to the traditional single-scale methodologies, whose optimal performance is restricted just to a subset of all the relevant scales. They have been applied, among other situations, for signal and image denoising[3,4] and compression,[5] process monitoring,[6,7] system identification,[8,9] time series analysis,[10] regression, classification and clustering,[11,12] and numerical analysis.[13]

With respect to process monitoring approaches, Shewhart control charts and multivariate statistical process control based on principal components (PCA-MSPC) have proven to be adequate for the detection of large deviations occurring at the finest scales, whereas moving-average (MA), exponentially weighted moving-average (EWMA), and cumulative-sum (CUSUM) charts (as well as their multivariate counterparts) have provided very good results regarding the detection of small shifts at coarser scales, whose particular range depends on the tuning parameters adopted (window length or filter constants).

On the other hand, it is known that current approaches face problems when they have to deal with the rather common situation of analyzing autocorrelated data, arising from a dynamic process, as they assume collected measurements to be uncorrelated. Therefore, to be appropriately implemented under such situations, they require complementary procedures, such as fitting of a time-series model followed by monitoring of the decorrelated residuals, an approach that is not practical for multivariate processes of high-dimensionality. The residuals to be monitored can also be generated using the one-step-ahead prediction from a moving centerline EWMA model (MCEWMA), together with the collected measurements,[14] a procedure which is best suited for integrated moving-average processes (IMA), therefore, lacking the generalization ability to other cases. Another approach, also found in the literature, called dynamic PCA, consists of expanding the measurement matrix with time-lagged variables, in order to capture autocorrelation in the PCA modeling stage.[15,16] However, as in all of the aforementioned techniques, one still ends up focused at a single-scale, since the basis functions used in such data representations maintain the same support in the time-frequency plane at all locations.

To overcome such limitations, multiscale statistical process control (MSSPC)[7] provides a valid alternative process monitoring methodology, with the highly desirable features of being sensitive to a wide range of faulty patterns (with different magnitudes and localizations in the time-frequency plane), arising from a variety of processes (either uncorrelated or with autocorrelation). In the multivariate case, it combines the ability of PCA, for decorrelating the variables' covariance, with that of wavelets, for (approximately) decorrelating the serial dependencies (autocorrelation) of signals. Process monitoring, is, therefore, conveniently performed in a space where variability can be described in simpler terms, without loosing effectiveness, through a procedure that comprises two main stages (Figure 1).
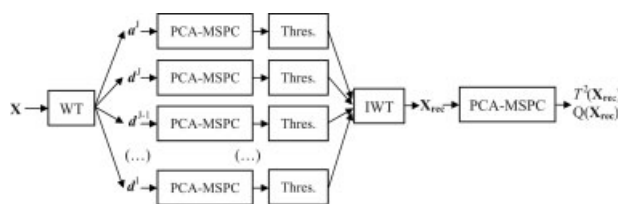

Figure 1. Wavelet-based multiscale statistical process control (MSSPC).

In the first stage, a preliminary monitoring task is conducted over the space of the PCA scores of wavelet coefficients at each scale, where the transformed variables appear now decorrelated, either in the variable direction (by PCA), or in the time direction (by the wavelet transformation). In this stage, significant events at each scale are detected, and those scales where such events do occur are selected to be used in the reconstruction of the signal at the finest scale, which can be considered as a feature extraction operation for a potential fault pattern. Then, in a second stage, over such a reconstructed signal a final statistical test is performed, in order to decide about the state of operation (normal or abnormal). This final stage of monitoring, performed in the original time domain, aims to avoid false alarms when the process returns back to normal operation, something that would otherwise occur due to the existence of significant events in the detail coefficients for finer scales (that are only sensitive to changes in the signals), as well as to allow for the continuous detection of a sustained shift, after it occurs, which may not be detected at the finer scale detail levels, because no change is occurring after an initial period of time has gone by.

Several modifications have been introduced to this basic MSSPC methodology, such as its integration with monitoring methods specialized in detecting changes over the data statistical distribution,[17] 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.[18] Other contributions include the development of nonlinear process monitoring approaches, similar to MSPCA, but with a nonlinear modeling step, by using neural networks.[19,20] Rosen[21] presented an alternative methodology, where the components at different scales in the original time 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 with regard to the scales under analysis. Yoon and MacGregor[22] have also developed an approach, based on 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 eigenvalues, thus, leading to a rank of the relevant structures underlying data variability, regarding the contributions from the variables covariance at different scales.

The added detection flexibility of MSSPC, in order to adapt to different types of disturbances, effectively extracting
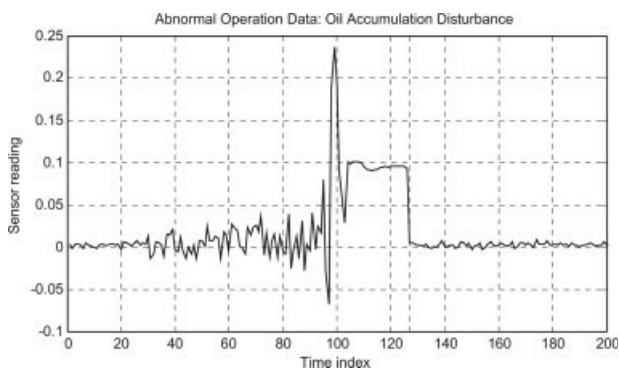
Figure 2. Process disturbance with multiscale features embedded (a sensor failure due to an oil accumulation event).

Vertical lines indicate the onset and the end times of the disturbance.

their patterns by analyzing the significant scales involved, arises from a different way of representing data. Such a representation consists of using basis functions, called wavelets that do span the time-frequency plane in different ways, allowing for adaptation to the signal's energy distribution in this domain (the time-frequency domain). These basis functions present good time resolution for high-frequency events and good frequency resolution for low-frequency patterns, in which case they are associated with longer time resolutions, since the mathematical inequality $\sigma(\psi) \cdot \sigma(\hat{\psi}) \geq 1/2$ ($\hat{\psi}$ representing the Fourier transform of ψ), has to be satisfied (this is also the key inequality of the Heisenberg uncertainty principle[23,24]). This type of coverage for the time-frequency plane is indeed adequate in a large class of signals, making of wavelets effective tools for analyzing a wide variety of datasets, particularly regarding those situations that present features at distinct locations and with different localizations in the time-frequency plane, or regarding nonstationary phenomena. One example of such a signal is presented in Figure 2, which corresponds to a sensor failure due to oil accumulation (it represents the difference between the readings of two redundant sensors, one of which is faulty).

This abnormal situation comprises an initial stage, where the disturbance does have a relatively small magnitude and presents an oscillatory pattern (usually difficult to be detected by current methodologies), whose magnitude grows along time, until the fault fully manifests itself as an high-magnitude steady change, after some spikes, also of high-magnitude. Clearly, such a pattern contains multiscale features, since its spectrum spans quite different regions of the frequency domain along time.

Even though wavelets already offer an effective (i.e., compact) representation, for most complex signals, it is possible to develop new basis sets, that can provide even more flexible descriptions of such patterns, and may be tailored for achieving optimal data representations (optimality being assumed here in a given sense, to be specified *a priori*), with the wavelet basis being just a particular case of such a variety. Broader basis sets can be obtained under the context of wavelet packets, which are basically libraries of functions, from which it is possible to extract a number of basis sets

with functions that can have a greater variety of covertures over the time-frequency plane. Furthermore, it is possible to choose the most adequate basis set for a given application in an efficient way, using specialized search algorithms (such as the "best basis" algorithm[25]), associated with a particular cost function that provides the necessary quality measure to drive the optimization search among all the possible sets (e.g., entropy, threshold or l^p norm cost[24,26]). For instance, regarding the signal presented in Figure 2, it is possible to come up with a basis set that provides a more compact representation for a given number of retained expansion coefficients, i.e., whose retained energy (squared norm of signal vector), for a given number of retained coefficients, is the highest (Figure 3). The reason why the difference, in this particular case, is not very significant, arises from the fact that the best basis derived for this signal is quite close to the wavelet basis.

In the context of chemical engineering and related fields of knowledge, wavelet packets have been mostly used in compression and denoising applications,[27–31] following the main stream of applications for wavelet-based approaches. However, other developments have also been achieved, that explore the possibility of choosing the most adequate basis set for a given application, namely in pattern recognition and classification analysis,[12,32] as well as in the sometimes related problem of fault detection, as happens for instance with the detection of partial discharges in gas-insulated substations,[33] or regarding gap abrasion of main bearing in diesel engines, where a pattern matching methodology was adopted using images of the phase space (time-frequency plane) obtained through the application of wavelet packets to
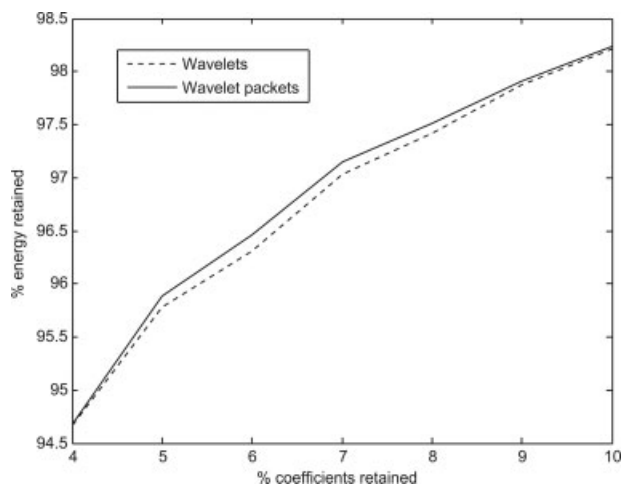
Figure 3. Curves of percentage of coefficients retained vs. percentage of signal's energy retained, using a wavelet basis (Daubechies-4 with five decomposition levels), and a basis set chosen from the corresponding wavelet packet decomposition, selected by the best basis algorithm implemented with an entropy criteria.

Calculations were conducted in the Matlab platform (The MathWorks, Inc.) using the WavLab package (available at http://www-stat.stanford.edu/~wavelab) and code written by the authors.

noisy measurements, adequately taken at the exterior of the engine.[34]

With regard to process monitoring, Li and Qian[35] presented an approach that involves a preliminary wavelet packet analysis stage, followed by application of PCA to extract the linear relationships present in data and of a non-linear mapping technique (input-training neural network, IT-net) for modeling any nonlinearities that might also be present. However, these authors do not provide any comparative results with alternative methodologies, regarding the performance of their approach. Developments concerning condition monitoring of vibration in industrial machinery have also been reported by Yen and Lin,[36] where wavelet packets were adopted to provide the means for generating a rich prediction space for classification of operation status, which is then submitted to a dimension reduction operation, so that best features for analysis are identified, and later on selected.

In this article, we present a multiscale statistical process control procedure that uses a wavelet packet analysis of successively collected measurements, in order to effectively extract faulty patterns that may be immersed in such stochastic signals. In the next section, we provide a brief introduction to wavelet analysis and explain how wavelet packets emerge from it, exploring their connection and addressing the potential advantages of using such an alternative data representation. Then, in the Multiscale Statistical Process Control Using Wavelet Packets (WP-MSSPC) section, we describe the proposed multiscale statistical process control approach based on wavelet packets (WP-MSSPC). Next, results obtained for two sets of comparative studies, encompassing WP-MSSPC and other approaches available in the literature, are presented and discussed. The first one involves an analysis of their performances under simulated scenarios (average run length study, ARL), while the second is devoted to their application in real world industrial situations (based on the receiver-operating characteristic, ROC, curves). We finalize with a summary of the main conclusions reached.

Theoretical background

Wavelets. Wavelets correspond to particular types of functions, whose location and localization characteristics in the time/frequency plane are ruled by two parameters: both the *localization* in this plane, and *location* in the frequency domain are determined by the scale parameter s, whereas *location* in the time domain is controlled by the time translation parameter b. Each wavelet $\psi_{s,b}(t)$, can be obtained from the so called "mother wavelet" $\psi(t)$, through a scaling operation that "stretches" or "compresses" the original function, establishing its form, and a translation operation, that controls its positioning in the time axis:

$$\psi_{s,b}(t) = \frac{1}{\sqrt{|s|}}\psi\left(\frac{t-b}{s}\right) \qquad (1)$$

The shape of the mother wavelet is such that it has an equal area above and below the time axis. As a result, besides having a compact localization in this axis, wavelets do also oscillate around it. In the continuous wavelet transform (CWT), scale and translation parameters vary continuously, constituting a redundant transformation. Therefore, in order

to construct a basis set, these parameters should be appropriately sampled, so that the set of wavelet functions parameterized by the new indices (scale index j, and translation index k) does cover the time-frequency plane in a nonredundant way. Such a sampling strategy consists of applying a dyadic grid, in which b is sampled more frequently for lower values of s, and where s grows exponentially with the power of 2 (dyadic discretization):

$$\psi_{j,k}(t) = \psi_{s,b}(t)\Big|_{\substack{s=2^j \\ b=k\cdot 2^j}} = \frac{1}{2^{j/2}}\psi\left(\frac{t-k\cdot 2^j}{2^j}\right) = \frac{1}{2^{j/2}}\psi\left(\frac{t}{2^j}-k\right)$$

$$(2)$$

Within the scope of a multiresolution (or multiscale) decomposition framework, developed by Mallat,[24,37,38] it is possible to decompose any signal x, into its contributions at different scales, ranging from the finest ones (low-values for scale index j, corresponding to high-frequency contents) to the coarsest ones (higher-scale index values, associated with low-frequency components). This can be done by expanding the signals using wavelet functions for the intermediate scales (i.e., between the finest and coarsest scales of analysis) and scaling functions for the coarsest scale (that conveys information for all scales higher than the maximum scale used in the analysis), which, altogether, form a basis set. The expansion coefficients can be computed very efficiently, with computational complexity $O(n)$, through a recursive procedure proposed by Mallat,[38] that essentially consists on the successive application of quadrature mirror filters followed by dyadic downsampling. There are two types of filters involved in this analysis: one providing the wavelet transform coefficients relative to the lower frequency band of the signal $\mathbf{H}_j$, while the other leads to the coefficients corresponding to the higher frequency band of the signal $\mathbf{G}_j$. In the wavelet transform, only the coefficients for lower frequency bands are successively decomposed, as indicated in Eq. 3 (where matrices that implement such filtering schemes are used), and illustrated in Figure 4.
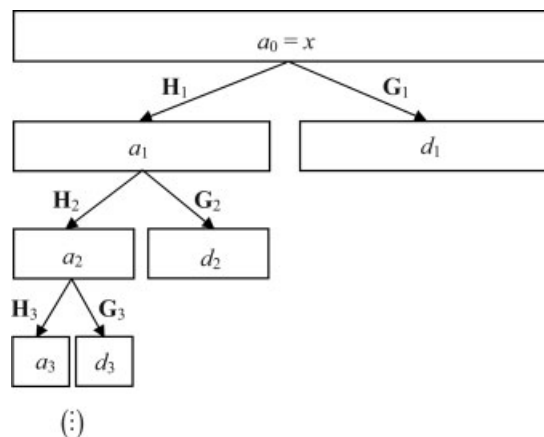
Figure 4. Wavelet decomposition procedure, where only the low-frequency coefficients (vectors a_j) are successively subjected to filtering operations.

$$a_0 = x \text{ (original signal)}$$
$$a_1 = \mathbf{H}_1 a_0, \ d_1 = \mathbf{G}_1 a_0$$
$$a_2 = \mathbf{H}_2 a_1, \ d_2 = \mathbf{G}_2 a_1$$
$$a_3 = \mathbf{H}_3 a_2, \ d_3 = \mathbf{G}_3 a_2$$
$$\vdots$$
$$a_J = \mathbf{H}_J a_{J-1}, \ d_J = \mathbf{G}_J a_{J-1} \tag{3}$$

The low-frequency coefficients are the so-called approximation coefficients (a's), while those for the high-frequency bands are the detail coefficients (d's). They are the multipliers of the basis functions: approximation coefficients are associated with the scaling functions, and detail coefficients with wavelets.

Wavelet Packets. In the wavelet decomposition, only the approximation coefficients are successively decomposed by the application of quadrature mirror filters $\mathbf{H}_j$ and $\mathbf{G}_j$. In terms of vector spaces, this amounts to say that only the approximation spaces are decomposed into new coarser approximation and details spaces, which are orthogonal to each other. In terms of the basis functions for these spaces, namely for the detail spaces, this implies that they lose time resolution as the decomposition depth increases, gaining, on the other hand, frequency definition, since such a decomposition scheme originates wavelet basis with constant relative bandwidth (the so-called constant-Q scheme). Introduced by Coifman, Meyer and Wickerhauser,[24] wavelet packets result from the decomposition of not only the successively coarser approximation spaces, but also of the associated detail spaces. In practice, this corresponds to the successive application of quadrature mirror filters to both approximation and detail coefficients (Eq. 4), as illustrated in Figure 5.

$$w_0^0 = x \text{ (original signal)}$$
$$w_0^1 = \mathbf{H}_1 w_0^0, \ w_1^1 = \mathbf{G}_1 w_0^0$$
$$w_0^2 = \mathbf{H}_2 w_0^1, \ w_1^2 = \mathbf{G}_2 w_0^1, \ w_2^2 = \mathbf{H}_2 w_1^1, \ w_3^2 = \mathbf{G}_2 w_1^1$$
$$w_0^3 = \mathbf{H}_3 w_0^2, \ w_1^3 = \mathbf{G}_3 w_0^2, \ w_2^3 = \mathbf{H}_3 w_1^2, \ w_3^3 = \mathbf{G}_3 w_1^2,$$
$$w_4^3 = \mathbf{H}_3 w_2^2, \ w_5^3 = \mathbf{G}_3 w_2^2, \ w_6^3 = \mathbf{H}_3 w_3^2, \ w_7^3 = \mathbf{G}_3 w_3^2$$
$$\vdots \tag{4}$$

The further decomposition of detail spaces, now undertaken, originates new basis functions for such subspaces that tile the time-frequency plane differently. It also leads to a situation where we have a number of possible basis functions sets available for representing the signal (more than $2^{2^{J-1}}$, for a decomposition depth of J[24]). Therefore, it is possible to choose the one that best suits our data analysis purposes, be it compression, classification, or, as in this situation, fault pattern extraction for process monitoring. Following the decomposition scheme presented in Figure 5, we will refer to each vector of wavelet packets coefficients by the nodes of such a binary tree, which can be indexed using the pair (j, p), where index j stands for the decomposition depth ($j = 0$: J_{dec}, with 0 regarding the original signal), and p is the horizontal position given by the number of nodes that are to the left for depth j ($p = 0$: $2^j - 1$). One should notice that index
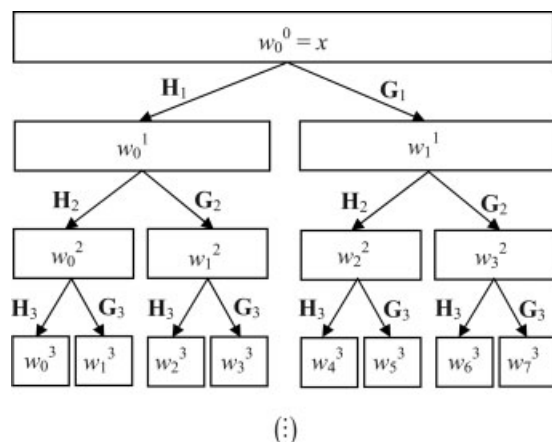
Figure 5. Illustration of the wavelet packet decomposition tree, where both the low-frequency and high-frequency components of each signal at a finer scale (vectors w_n^j) are successively subjected to the quadrature mirror filtering operations.

p does not order the basis functions, at a given j, in the sense of covering increasing frequency regions (Paley ordering), something that would require swapping the branches from some nodes.[24,36]

Figure 6 presents the basis functions waveforms obtained for the Haar case (also known as Walsh wavelet packets), from which we can see that a wider variety of shapes can be indeed more effectively represented through their adequate combination. As an illustrative example of such a capability, we present the results derived from analyzing an artificial signal, with 1,024 observations consisting of i.i.d. Gaussian white noise with a standard deviation of 2, where, after observation 512, a sinusoidal component, with unit magnitude and frequency 0.7813π rad, begins to develop and lasts until the end of the time series. The full wavelet packet table of coefficients, resulting from the decomposition (using a Daubechies-6 filter) of this signal, is presented in Figure 7. The original signal lies at the top, where a quick visual inspection only barely indicates the onset of the sinusoidal perturbation. Also signaled in this figure are the coefficients for the standard wavelet basis (boxes with thick lines), where, once again, it is not very easy to spot the time instant where the change has occurred. This implies that conventional MSSPC may not be able to appropriately detect this sinusoidal change. However, there are other nodes in the wavelet packet tree whose coefficients are quite sensitive to this disturbance, so that the transition time can be easily pointed out, as happens, in this case, for the signal in the dashed box. In fact, the best basis tree for this example, using an entropy criterion for conducting the best basis algorithm, is quite different from the standard wavelet basis, as depicted in Figure 8, where the binary trees, associated with each basis, are presented. We can see that more wavelet packet coefficients, relative to nodes closer to the dashed boxes, are picked up, being also responsible for strong improvements in the assumed cost criteria.
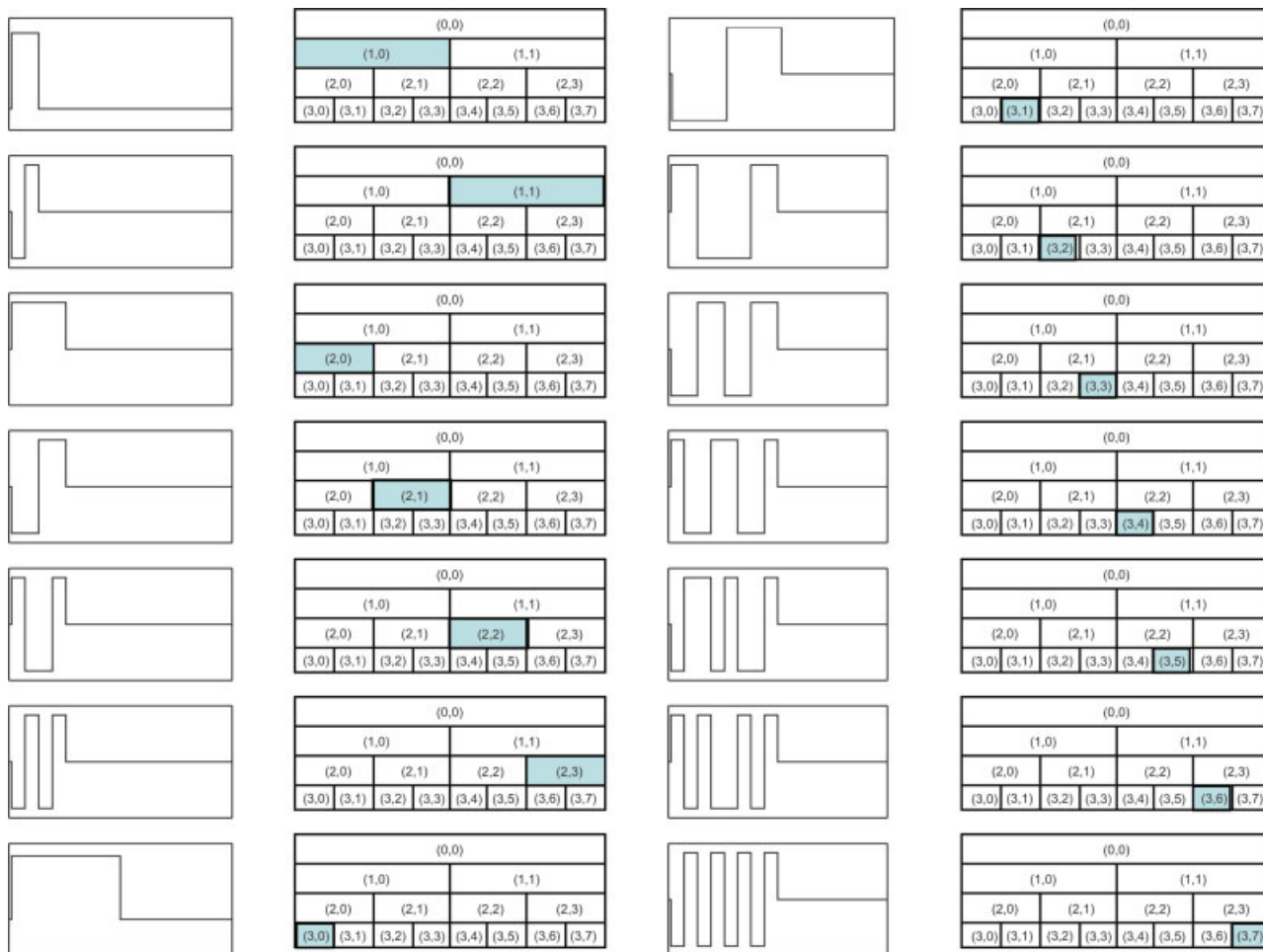
Figure 6. Waveforms for the basis corresponding to each node of the wavelet packet tree (there are $n/2^j$ translated versions at each node, but only the first one is presented for illustration purposes).

The associated indices (j, p) are indicated at the corresponding diagrams on the left of the waveforms (the signal's length is $n = 16$). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]
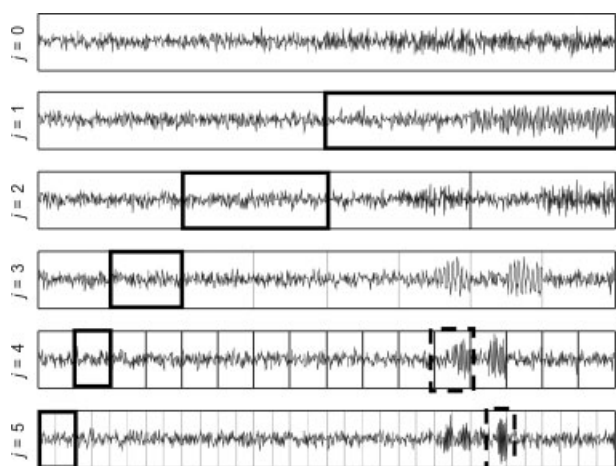
Figure 7. Wavelet packet decomposition of an artificial signal consisting of white noise with a sinusoidal disturbance superimposed during the second half of the interval (a Daubechies-6 filter was used with $J_{dec} = 5$).
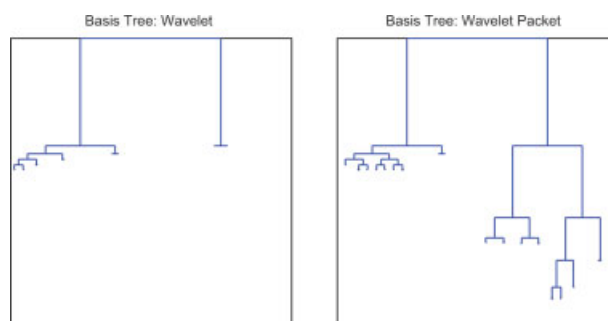
Figure 8. Binary trees representing the nodes whose basis functions participate in the (overall) basis set, for the standard wavelet base (a), and for the best basis of the wavelet packet tree (b), using an entropy criterion, for the signal analyzed in Figure 7.

The height of each branch is proportional to the improvement in the cost function (entropy) obtained after each splitting. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Table 1. Pseudocode for WP-MSSPC (with Dyadic Discretization)

Initialization
1. Define wavelet packet filter (*Wave*);
2. Define maximum decomposition depth (J_{dec});
3. Define level of significance (α);
4. Compute the wavelet packet decomposition of the training data set, and calculate the mean and variance (univariate case) or variance-covariance matrices (multivariate case) for data blocks at each node;
5. In the multivariate case, select also the number of principal components to retain in the PCA models at each node.

Implementation
1. For each new point with an even time index ($2k, k \in \square$):
   a. Select data contained in a window of maximal dyadic length ($2^{J_{wind}(k)}$), such that $J_{wind}(k) \leq J_{dec}$: $X_{wind}(2k)$;
   b. Calculate the wavelet packet table for $X_{wind}(2k)$: $WPX_{wind}(2k)$;
   c. Get current coefficients from the blocks of data corresponding to each node of $WPX_{wind}(2k)$: $Coef_{curr}^{wind}(2k)$;
   d. Select, among the current coefficients, $Coef_{curr}^{wind}(2k)$, those that are relative to significant events, at the predefined level of significance α: $^*Coef_{curr}^{wind}(2k)$;
   e. Construct a new wavelet packet table containing only the current significant coefficients, $^*Coef_{curr}^{wind}(2k)$: $^*WPX_{wind}(2k)$;
   f. Reconstruct data using a basis set compatible with $^*WPX_{wind}(2k)$: $Xrec_{wind}(2k)$;
   g. Combine variance (univariate case) or variance-covariance matrices (multivariate case) relative to significant events and test whether the reconstructed signal is significant or not at the level of significance α, using univariate SPC (Shewhart) or multivariate SPC (PCA-MSPC);
   h. If the reconstructed signal is significant, signal it as a special event;
2. $k \leftarrow k + 1$, GO TO 1.

## Multiscale Statistical Process Control Using Wavelet Packets (WP-MSSPC)

The proposed WP-MSSPC approach follows the basic sequence of steps that have been proven to be successful for conducting wavelet-based MSSPC (Figure 1), but with major modifications introduced in the multiscale decomposition stage. Instead of applying a wavelet orthogonal transformation, we introduce a wavelet packet redundant decomposition. This also implies redesigning all the subsequent stages, namely regarding the selection of "significant scales", thus giving rise to the selection of a "significant and proper basis" for describing the pattern of a potential disturbance.

The overall procedure can be conceptually divided into an initialization stage and an implementation stage (Table 1). In the preliminary *initialization stage*, user-defined features and parameters of the method are selected, namely the type of wavelet packet filter to use, the maximum possible decomposition depth to be considered (unlike existing MSSPC methods, this approach can select the actual depth adaptively for each window of data, as discussed later), the level of significance to be used in statistical tests (probabilistic control limits), and the number of principal components to be adopted (for the multivariate case). Furthermore, using historical data from normal operation, the normal operation condition (NOC) regions, at each wavelet-packet node, are also defined, using the transformed data and significance levels previously defined.

In the *implementation stage*, as new data gets acquired (with dyadic length, to allow the proper computation of wavelet packet coefficients), wavelet packet coefficients are computed, with the last coefficient at each node being retained for analysis of its significance, through the nodes' specific SPC charts. If no significant activity is registered at any node (all current coefficients fall within statistical control limits), the method proceeds to the next data window. However, if some coefficients are found to be significant, our method picks up only those and adopts the corresponding wavelet packet basis, which represents the potential abnormal patterns more efficiently. It is this flexible basis selection strategy that allows for the adaptive selection of decomposi-

tion depths. In fact, the decomposition depth effectively used in the analysis is the one that corresponds to the coarsest significant wavelet packet coefficient, and not the maximal decomposition depth defined initially. As a result, decomposition proceeds deeper (increasing decomposition depth) if no significant coefficients are identified, until the maximum specified depth is reached. On the other hand, if significant coefficients are identified before reaching the maximum depth, the decomposition does not need to go any deeper. The advantageous consequences of this strategy are discussed more thoroughly later. Then, using only the significant coefficients, our algorithm computes the corresponding pattern in the original time domain and the statistical limits of the SPC chart to be used for performing the final and confirmatory significance tests. These limits are computed from knowledge about the wavelet packet basis adopted and the statistical quantities estimated from normal operation data at each node, in the initialization stage. If the reconstructed trend is found to be significant (falls outside statistical control limits) then this pattern is confirmed to be abnormal. On the other side, if it is not statistically significant, no signaling is produced, and process monitoring moves to the next data window. More details of the proposed approach can be found in its pseudocode, presented in Table 1.

The implementation described in Table 1 relates to a dyadic discretization scheme for building the current data windows, over which the wavelet packet analysis and subsequent monitoring procedures are implemented, at each step. Dyadic and integer discretization are the two modes currently used for implementing MSSPC approaches. The first one is based on a nonredundant (orthogonal) representation of data along time, which allows for an approximate decorrelation of autocorrelated measurements, at the expense of introducing some time delay in the analysis. The second one uses a redundant representation, due to the implicit suppression of downsampling, leading to a decreased decorrelation ability, but with the advantage of enabling on-line real-time implementations. These two discretization procedures appear illustrated in Figure 9. The pseudocode of Table 1 is also valid for the integer discretization mode, after replacing all code instructions containing $2k$ by k.
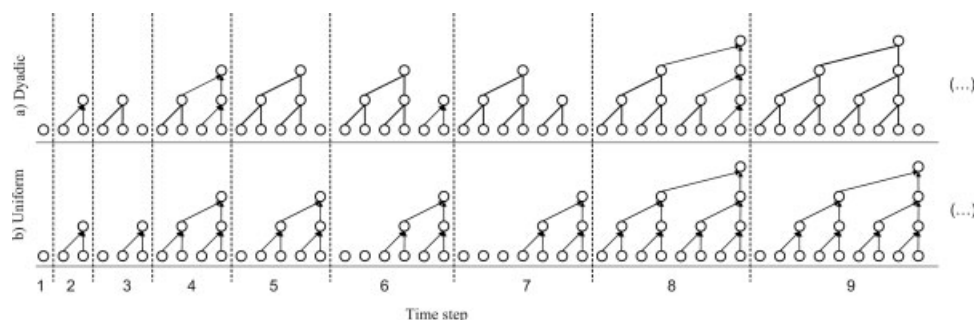
Figure 9. Discretization strategies used in MSSPC methodologies: (a) dyadic discretization, and (b) integer or uniform discretization.

(Example for the Haar wavelet with $J_{dec} = 3$). Circles represent wavelet coefficients at each scale (observations are acquired at scale $j = 0$, and used to compute the coefficients at scale $j = 1$, represented immediately above, and these then used to compute coefficients at scale $j = 2$, represented in the next level, and so on, until the decomposition depth, $j = 3 = J_{dec}$ is attained). Arrows indicate the coefficients that are computed at each time step. It is clear that in the dyadic case, there are time steps where no coefficients are computed (at odd values for the time index), and for the coefficients at coarser scales the delay involved is even larger.

One important feature of WP-MSSPC is its ability to automatically select the adequate decomposition depth for representing each significant pattern, a desirable property not shared by wavelet-based MSSPC approaches. For such approaches, a maximum decomposition depth is always performed. This raises the issue that one needs to find out the most adequate decomposition depth for performing wavelet-based MSSPC, since a too fine decomposition depth leads to a quite noisy last scaled signal (a_j), and a limited decorrelation ability, whereas a too coarse decomposition depth leads to very few coefficients being used for estimating the NOC statistics, therefore affecting the accuracy of SPC at coarser scales, which also somewhat distorts the reconstructed extracted features. Such a rather problematic trade-off, is, therefore, overtaken by our proposed adaptive depth selection strategy. The selected basis, from the wavelet packet tree, depends strictly on the coefficients found to be significant. Furthermore, theoretically, WP-MSSPC subsumes wavelet-based MSSPC, as the data decomposition in which this method is based corresponds to one out of the many possible basis sets available in the wavelet packet library of WP-MSSPC.

Average Run Length (ARL) Analysis of WP-MSSPC

In this section, we compare the univariate versions of WP-MSSPC (dyadic and integer) with other existing SPC methodologies in their capability to detect several perturbations (step change, sinusoidal perturbation, appearance of autocorrelation) in an uncorrelated homogeneous process with unit variance $x_i \sim i.i.d.N(0,1)$, $i = 0,1,2,\ldots$ (where *i.i.d.* stands for an "independent and identically distributed" random variable following, in this case, a zero mean Gaussian distribution with unit variance), as well as a decrease in the autocorrelation parameter of an AR(1) dynamic process, by computing their average run length (ARL) performance over various perturbation magnitudes. ARL is the average number of observations required to detect a shift after it begins. In this study it was determined through Monte Carlo simulations. In order to make all methods comparable in a fair

way, control limits were previously adjusted, so that the different techniques present the same ARL(0) performance, i.e., the same ARL scores in the absence of any process shift. The SPC methods analyzed here were the following: Shewhart,[39] moving-average (MA, with an averaging time support of $2^{J_{dec}}$ observations, to maintain results comparable with MSSPC approaches),[40] and exponentially-weighted moving average control charts (EWMA, with $\lambda = 0.2$),[41] MSSPC[7] (based on wavelets, in the dyadic and integer discretization modes), and WP-MSSPC (also implemented in both the dyadic and integer discretization modes). All decompositions were based on the Haar wavelet filter (with $J_{dec} = 3$).
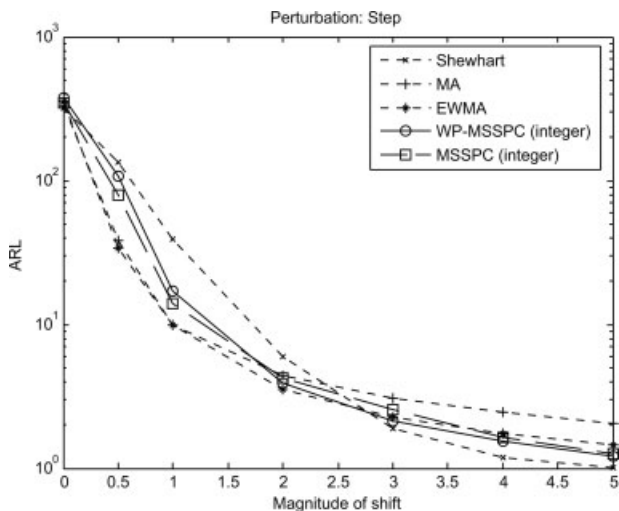
Step perturbation

ARL results for mean shifts of increasing magnitude, following a step perturbation, are presented in Figure 10, where
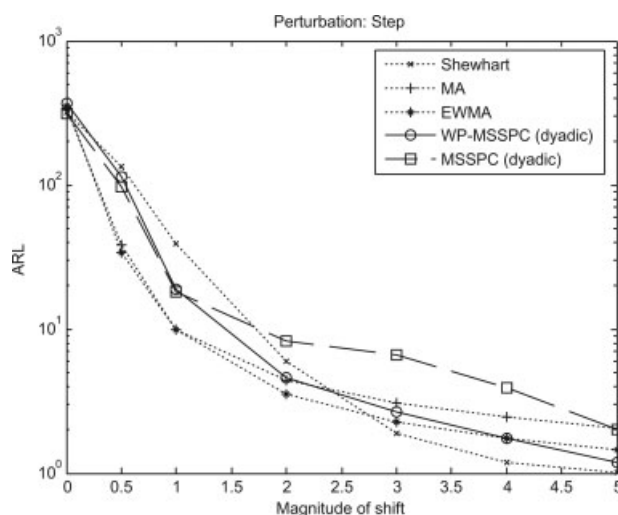
Figure 10. ARL curves for a step perturbation in the mean, where all MSSPC methods presented are implemented with a dyadic discretization scheme.

Figure 11. ARL curves for a step perturbation in the mean (MSSPC methods implemented in the integer discretization mode).

only MSSPC methods with *dyadic* discretization are plotted, for the sake of discerning all trends involved. Those relative to the integer discretization mode are presented in Figure 11. In the dyadic implementation (Figure 10) it is possible to see that WP-MSSPC considerably enhances the detection ability of wavelet-based MSSPC for shifts of magnitude higher than 1. The best performances are achieved by MA and EWMA for shifts of small magnitude, and Shewhart control charts for high-magnitude shifts. This relative performance of conventional methods is well known, and the observed behavior of the multiscale methods is due to the simplicity of the perturbation adopted, which presents a clear location in the time-frequency domain, therefore, creating the necessary conditions for single-scale approaches to perform well, as they were designed to operate optimally only at a certain scale. However, one should notice that, as soon as the perturbation magnitude leaves the optimal range, where each conventional method is designed to operate well, their relative performance, when compared with other methodologies, degrades considerably, something that does not happen with WP-MSSPC, which keeps its ARL difference to the best approaches approximately constant over the whole spectra of shift magnitudes, due to its adaptation capabilities to the characteristic scale of the perturbation taking place. The adaptive decomposition depth and library of basis functions available in WP-MSSPC make it possible that, even with dyadic discretization, one does not have to wait until enough coefficients are obtained to detect a shift. In contrast, conventional MSSPC, despite using windows of data with the same length as WP-MSSPC, does not perform as well in the fault detection, due to the adoption of a fixed set of basis functions (only one of the many representatives contained in the library explored by WP-MSSPC), thus preventing an efficient extraction of features from shorter time windows. The same comments can be made regarding the results shown in Figure 11, but where the observed performances for MSSPC and WP-MSSPC are now quite similar, in the integer implementation mode, mainly because the effect of time delay in

obtaining the scaled signal coefficients in MSSPC is eliminated under this mode. However, integer discretization requires additional computation and benefits of the decorrelation ability associated with orthonormal wavelets may become more reduced.
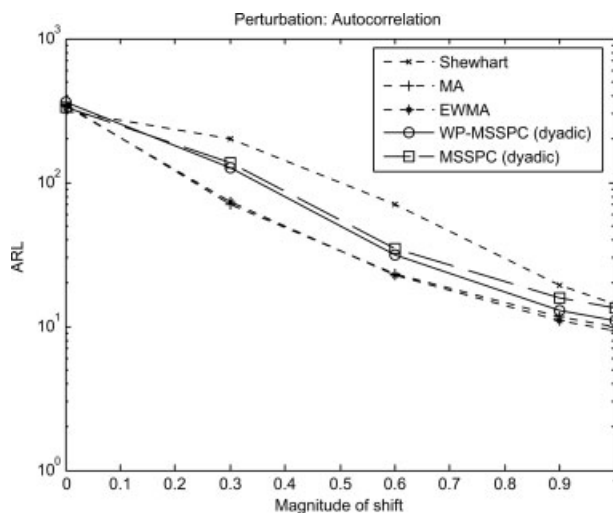
Increase in autocorrelation

Analyzing now what happens when an uncorrelated process suddenly acquires some autocorrelation (characterized by the autocorrelation parameter $\phi \in [0,1]$, Eq. 5),

$$
\begin{aligned}
X(k) &= \phi \cdot X(k-1) + \varepsilon(k) \\
\varepsilon(k) &\sim i.i.d.N(0,1) \\
\phi &= 0(k = 1 : K) \rightarrow \phi = \Delta(k = K+1 : end)
\end{aligned}
\tag{5}
$$

one can observe that multiscale methods present the same kind of intermediate performance behavior as before. WP-MSSPC is slightly better than its wavelet-based counterpart in the dyadic situation (Figure 12).

Sine perturbation

In this case, our perturbation consists of a sine wave with frequency $\omega = 0.7813\pi$ rad (as the one considered in Figure 7), whose magnitude increases from zero, under normal operation conditions to $\Delta = \{0,0.5,1,2,3\}$. Figure 13 presents the results obtained when all MSSPC methods are implemented with dyadic discretization. We can see that MSSPC methods (especially WP-MSSPC) tend now to perform better over the whole range of magnitudes, in particular regarding the small amplitudes regions, where fault detection is particularly difficult. The same general conclusions hold for the integer implementation mode, as can be seen from the results presented in Figure 14. This example illustrates that it is possible, in general, to encounter simple perturbations for which some method performs better than the others. In this case,

Figure 12. ARL curves for an increase in the autocorrelation parameter, from $\phi = 0$ (uncorrelated process) to $\phi = \Delta = \{0.3,0.6,0.9,1\}$ (MSSPC methods implemented in the dyadic discretization mode).
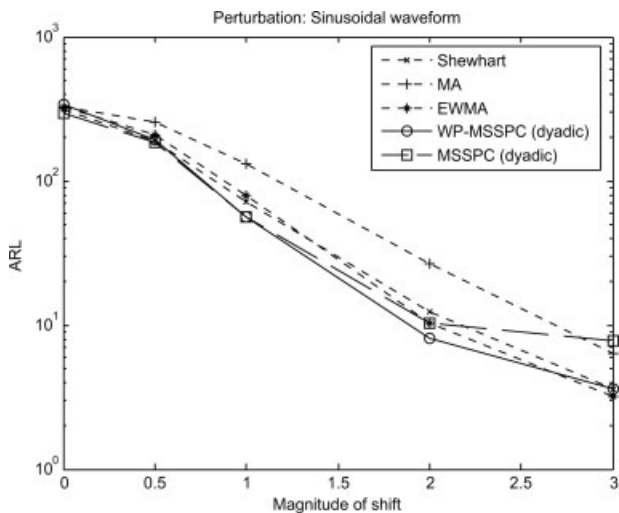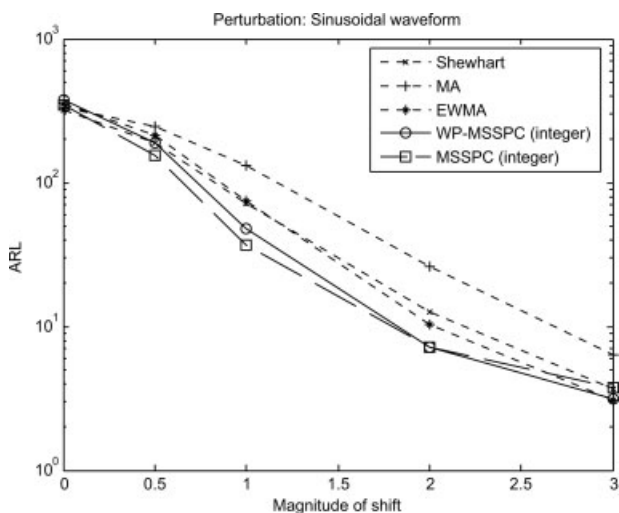
Figure 13. ARL curves for the sinusoidal perturbation (MSSPC methods implemented in the dyadic discretization mode).

Wave frequency is 0.7813π rad and amplitudes studied are {0,0.5,1, 2,3}.

for instance, the multiscale methods take advantage of their better detection sensitivity for perturbations well localized in the frequency domain, due to the way they represent data. Therefore, the type of knowledge acquired through this reasoning (methods performance evaluation under simple perturbation scenarios), while being certainly valuable for getting acquainted about the strengths and weaknesses of each alternative, also presents limitations, due to its case-based nature.

Decrease in autocorrelation

To finalize this section, regarding the comparison of methods' performances using Monte Carlo simulations, we analyze a situation where the system under normal operation is described by the following AR(1) process: $X(k) = \phi_{NOC} \cdot X(k-1) + \varepsilon(k)$, with $\phi_{NOC} = 0.5$, and $\varepsilon(k) \sim iid\ N(0,1)$. In this example, we tested all the different methods in the detection of a decrease in the autocorrelation parameter, from the normal operation conditions value of $\phi_{NOC} = 0.5$, to 0.5 (no change), 0.3, 0.1 and 0, i.e., with shifts of magnitude: $\phi_{NOC} - \phi_i$ {0,0.2,0.4,0.5}. Furthermore, in each trial, the variance of $\varepsilon(k)$ was also adjusted, so that the overall variance of $X(k)$ was kept constant over the whole time series, in order to make detections more difficult. Figures 15 and 16 present the results obtained. We can see that, in this case, the only methodologies that are able to detect such a perturbation are indeed the multiscale methods. In fact, all the remaining methodologies worsen their performance as the magnitude of the change increases, even though other techniques, not included in this study, might also have potential to detect such a type of perturbation.[17] These results are essentially explained by the good frequency resolution of the multiscale monitoring methods, arising from the nature of their basis functions, which allows for an effective monitoring of the signal's power spectra. This makes them able to detect changes in energy from one band to the other, something that single-scale methodologies are not always able to achieve (as demonstrated in this example), since they center monitoring efforts at a single, loosely defined, frequency band.

Industrial Case Studies

The artificial perturbations used in the previous section, to evaluate the methods' performances through Monte Carlo simulations, are quite simple in nature, even though useful to
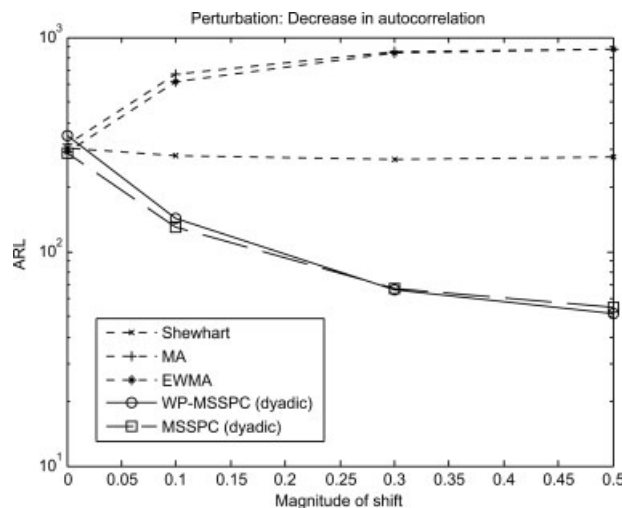

Figure 14. ARL curves for the sinusoidal perturbation (MSSPC methods implemented in the integer discretization mode).

Wave frequency is 0.7813π rad and amplitudes studied are {0,0.5,1, 2,3}.

Figure 15. ARL curves for a decrease in the autocorrelation parameter of an AR(1) process: $\Delta_i = \phi_{NOC} - \phi_i$, being the normal operation condition value of the parameter, $\phi_{NOC} = 0.5$ (variance of $\varepsilon(k)$ was adjusted so that the overall variance of $X(k)$ is kept constant).

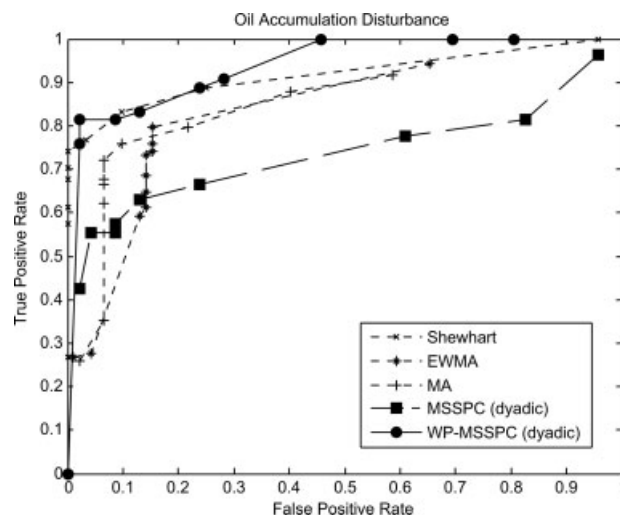All MSSPC methods were implemented with a dyadic scheme.
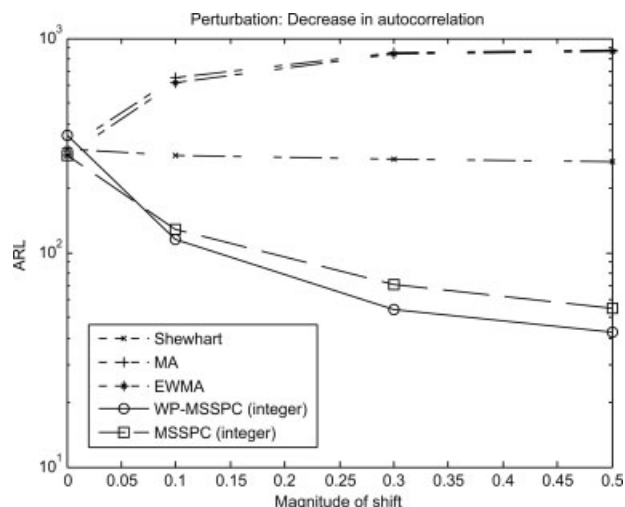
Figure 16. ARL curves for a decrease in the autocorrelation parameter of an AR(1) process: $\Delta_i = \phi_{NOC} - \phi_i$, being the normal operation condition value of the parameter, $\phi_{NOC} = 0.5$ (variance of $\varepsilon(k)$ was adjusted so that the overall variance of $X(k)$ is kept constant).

All MSSPC methods were implemented with an integer scheme.

Figure 17. ROC curves for the "oil accumulation event" (MSSPC methods were implemented with a dyadic discretization scheme).

Sensor malfunction due to an oil accumulation event

This case study involves a sensor malfunction. The corresponding dataset is the result of taking the difference between the readings from two redundant sensors, when one of them undergoes a faulty situation due to oil accumulation (Figure 2). The overall performance of the methods (considering only the dyadic implementation of MSSPC approaches), over the entire signal, is given in terms of their empirical ROC curves, reported in Figure 17. We can see that WP-MSSPC presents the best performance, as it leads, almost over the entire plot, to the highest TPR scores (higher sensitivity) for any given FPR value. Focusing now only on the MSSPC methods (Figure 18), we can also see that the two WP-MSSPC implementations outperform their MSSPC counterparts, even though MSSPC performance is also quite good in the integer discretization mode. In summary, WP-

provide reference scenarios for a preliminary comparative assessment. However, in industrial environments, faults tend to manifest themselves through much more complex patterns, presenting features that spread out through several time-frequency regions. Therefore, it is also important to test and evaluate the performances of the different methods under such circumstances.
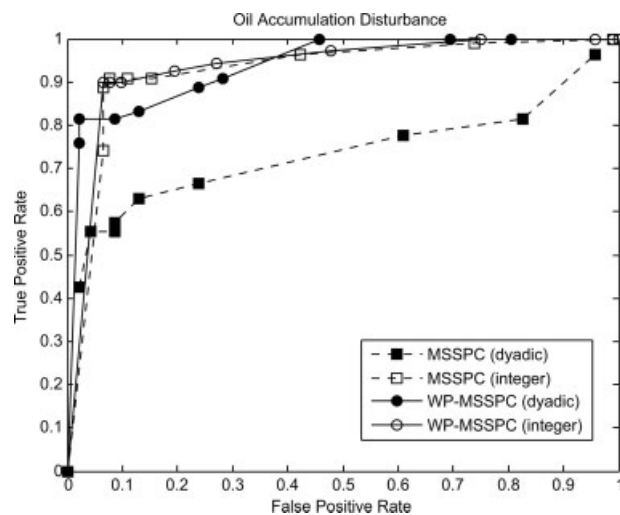
In this section, we present two real world case studies and analyze the performance of several monitoring methodologies in the detection of abnormal situations occurring in each dataset (provided by the Abnormal Situation Management Consortium). As no ARL study can be conducted in this situation, the performances achieved by the various methodologies are compared by plotting their empirical receiver-operating characteristic (ROC) curves, a common approach used for assessing classification algorithms in two-class problems. They correspond to a plot of the "true positive rate" (TPR) versus "false positive rate" (FPR). TPR corresponds to the fraction of points for which the detection limits are violated in the region where abnormal operation is known to be present. FPR, on the other hand, is the fraction of points for which detection limits are violated in the region where normal operation is known to be present. ROC curves are drawn by varying the methods thresholds (control limits) T, and computing the two detection rates: FPR (T_i) vs. TPR (T_i), $T_i \in [T_{min}, T_{max}]$. TPR and FPR are, thus, related to the notions of sensitivity and specificity.[42,43] Sensitivity is the probability of detecting the abnormality, given the true state to be abnormal (TPR), while specificity is the probability of not detecting the abnormality, given the true state to be of normal operation $(1 - \text{FPR})$. In each of the following case studies, the region of abnormal operation was determined from process operator inputs and specific knowledge available.

Figure 18. ROC curves for the "oil accumulation event": comparison of MSSPC approaches.

MSSPC does not perform worse then MSSPC, and often does much better, due to the increased sensitivity provided by the nature of the basis functions adopted in WP-MSSPC (derived from the wavelet packet library), which makes it able to extract more effectively, and, therefore, detect in an easier way any abnormal patterns present in the signal.

Furnace feed event

This case study addresses a multivariate situation, where measurements were collected from 10 different sensors located in a furnace from a real petrochemical process. A disturbance has occurred in the furnace feed, and the faulty operation measurements, for the different variables, can be seen in Figure 19.

These measurements were autoscaled (centered to zero mean and normalized to unit variance, using the mean and standard deviation computed from normal operation data) and the multivariate versions of the several MSSPC methodologies applied, as well as PCA-MSPC, for comparison purposes. Since there are now two monitoring statistics (T^2 and Q or SPE), a significant event is considered to occur when at least one of them exceeds the corresponding statistical control limits. Five principal components were retained in all PCA models. The results obtained are presented in Figure 20, where both WP-MSSPC approaches (dyadic and integer) present the best overall performance (the dyadic implementation of MSSPC, once again, faces some problems, which do not fall in the region of the ROC curve analyzed).

Conclusions

In this article, we presented an approach for conducting MSSPC using a library of basis functions provided by the wavelet packet transformation. The use of such a library has the potential of enhancing the extraction of patterns from abnormal events, making the method more sensitive to a wider class of faults, with arbitrary localization and developing at distinct locations in the time-frequency plane. Our
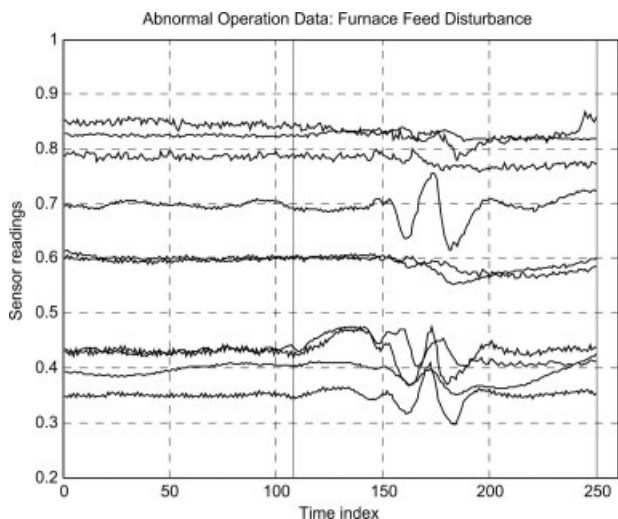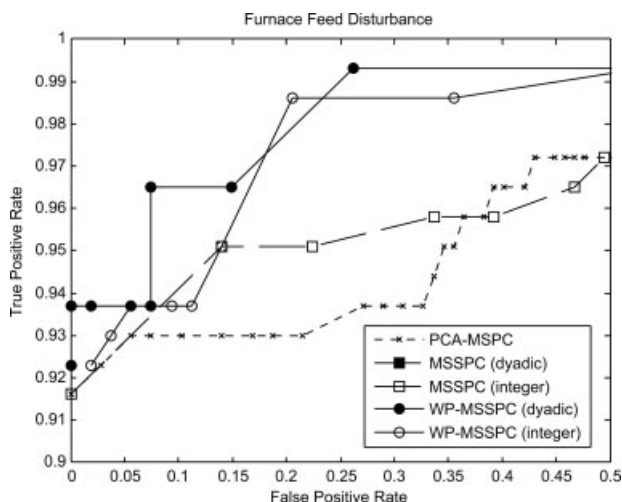
Figure 20. ROC curves for the "furnace feed event".

method also presents the desirable feature of autonomously selecting the adequate decomposition depth to represent each significant pattern, a property not shared by wavelet-based MSSPC approaches.

ARL results, derived from Monte Carlo simulations with simple fault patterns (step, sinusoidal increase and decrease in autocorrelation), demonstrate the adaptive ability of MSSPC methods over different shift magnitudes. Despite the fact that WP-MSSPC does not constitute the best approach for some simple perturbations (e.g., step perturbation), it is still able to keep a well balanced performance, without presenting any significant degradation of its relative performance regarding all other tested methods, at any particular range of scales, as happens with single-scale approaches, once they move away from their optimal operation regions. However, one should also notice that, as simple perturbations are concerned, MSSPC methods were found to perform better in detecting an hidden sinusoidal perturbation, and were the only ones able to detect a decrease in autocorrelation for an AR(1) process.

The analysis of real world case studies underlines the true potential of MSSPC approaches and illustrates the added value that WP-MSSPC can bring to process monitoring, given its increased flexibility to cope with the more complex shapes that real process disturbances usually tend to present. Thus, WP-MSSPC seems to be the most appropriate method for performing SPC on processes where the nature of the data patterns due to abnormal situations is not known beforehand. This is likely to be the case for most complex processes and practical applications.

Future work can focus on exploring other types of wavelet filters for the wavelet packet decompositions, with the aim of improving detection capabilities. The application of this methodology to other types of processes, namely batch processes, is also envisioned as an interesting area for additional research activities to be carried out.

Figure 19. Disturbance for the "furnace feed event".

Vertical lines indicate the onset and the end of the disturbance.

Literature Cited

1. Bakshi BR. Multiscale analysis and modeling using wavelets. *J Chemomet*. 1999;13:415–434.

2. Motard RL, Joseph B, eds. *Wavelet Applications in Chemical Engineering*. Boston: Kluwer; 1994.

3. Donoho DL, Johnstone IM. *Ideal Spatial Adaptation by Wavelet Shrinkage: Department of Statistics*, Stanford University; 1992.

4. Jansen M. *Noise Reduction by Wavelet Thresholding*. Vol 161. NY: Springer; 2001.

5. Vetterli M, Kovačević J. *Wavelets and Subband Coding*. Upper Saddle River, NJ: Prentice Hall; 1995.

6. Aradhye HB, Bakshi BR, Strauss R, Davis JF. Multiscale SPC using wavelets: theoretical analysis and properties. *AIChE J*. 2003;49(4):939–958.

7. Bakshi BR. Multiscale PCA with application to multivariate statistical process control. *AIChE J*. 1998;44(7):1596–1610.

8. Carrier JF, Stephanopoulos G. Wavelet-based modulation in control-relevant process identification. *AIChE J*. 1998;44(2):341–360.

9. Nikolaou M, Vuthandam P. FIR model identification: parsimony through kernel compression with wavelets. *AIChE J*. 1998;44(1):141–150.

10. Percival DB, Walden AT. *Wavelets Methods for Time Series Analysis*. Cambridge, U.K: Cambridge University Press; 2000.

11. Alsberg BK, Woodward AM, Winson MK, Rowland JJ, Kell DB. Variable selection in wavelet regression models. *Anal Chim Acta*. 1998;368:29–44.

12. Walczak B, van den Bogaert V, Massart DL. Application of wavelet packet transform in pattern recognition of near-IR Data. *Anal Chem*. 1996;68:1742–1747.

13. Beylkin G, Coifman R, Rokhlin V. Fast wavelet transforms and numerical algorithms I. *Commun Pure Appl Math*. 1991;XLIV:141–183.

14. Mastrangelo CM, Montgomery DC. SPC with correlated observations for the chemical and process industries. *Qual Reliab Eng Intl*. 1995;11(79).

15. Kresta JV, MacGregor JF, Marlin TE. Multivariate statistical monitoring of process operating performance. *Can J Chem Eng*. 1991;69:35–47.

16. Ku W, Storer RH, Georgakis C. Disturbance detection and isolation by dynamic principal component analysis. *Chemom Intell Lab Syst*. 1995;30:179–196.

17. Kano M, Nagao K, Hasebe S, et al. Comparison of multivariate statistical process monitoring methods with applications to the eastman challenge problem. *Comp Chem Eng*. 2002;26:161–174.

18. Misra M, Yue HH, Qin SJ, Ling C. Multivariate process monitoring and fault diagnosis by multi-scale PCA. *Comp Chem. Eng*. 2002;26:1281–1293.

19. Fourie SH, de Vaal P. Advanced process monitoring using an online non-linear multiscale principal component analysis methodology. *Comp Chem Eng*. 2000;24:755–760.

20. Shao R, Jia F, Martin EB, Morris AJ. Wavelets and non-linear principal components analysis for process monitoring. *Control Eng Pract*. 1999;7:865–879.

21. Rosen C, Lennox JA. Multivariate and multiscale monitoring of wastewater treatment operation. *Water Res*. 2001;35(14):3402–3410.

22. 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; 4–6 June, 2001*; Korea.

23. Kaiser G. *A Friendly Guide to Wavelets*. Boston: Birkhäuser; 1994.

24. Mallat S. *A Wavelet Tour of Signal Processing*. 2nd ed. San Diego: Academic Press; 1999.

25. Coifman RR, Wickerhauser MV. Entropy-based algorithms for best basis selection. *IEEE Trans Inf Theory*. 1992;38(2):713–718.

26. Mallet Y, Coomans D, de Vel O. Wavelet Packet Transforms and Best Basis Algorithms. In: Walczak B, ed. *Wavelets in Chemistry*. Vol 22. Amsterdam: Elsevier; 2000:151–164.

27. Walczak B, Massart DL. Noise Suppression and Signal Compression Using the Wavelet Packet Transform. *Chemom Intell Lab Syst*. 1997;36:81–94.

28. Leung AK-M, Chau F-T, Gao J-B, Shih T-M. Application of wavelet transform in infrared spectrometry: spectral compression and library search. *Chemom Intell Lab Syst*. 1998;43:69–88.

29. Walczak B, Massart DL. Wavelet packet transform applied to a set of signals: a new approach to the best-basis selection. *Chemom Intell Lab Syst*. 1997;38:39–50.

30. Cohen I, Raz S, Malah D. Translation-invariant denoising using the minimum description length criterion. *Sig Proc*. 1999;75:201–223.

31. Bakshi BR, Stephanopoulos G. Compression of Chemical Process Data by Functional Approximation and Feature Extraction. *AIChE J*. 1996;42(2):477–492.

32. Cocchi M, Seeber R, Ulrici A. WPTER: wavelet packet transform for efficient pattern recognition of signals. *Chemom Intell Lab Syst*. 2001;57:97–119.

33. Chang CS, Jin J, Chang C, Hoshino T, Hanai M, Kobayashi N. Separation of corona using wavelet packet transform and neural network for detection of partial discharge in gas-insulated substations. *IEEE Trans Power Delivery*. 2005;20(2):1363–1369.

34. Chen L, GuiZeng W, QuingGang Q.Wavelet Packet Images Matching Applied to Noise Faults Diagnosis. Paper presented at: *International Symposium on Advanced Control of Chemical Processes - ADCHEM 2003*; 11–14 January, 2004; Hong Kong.

35. Li X, Qian Y.Process Monitoring Based on Nonlinear Wavelet Packet Principal Component Analysis. Paper presented at: *ESCAPE-14, European Symposyum on Computer Aided Process Engineering, 2004*; Lisbon, Portugal.

36. Yen GG, Lin K-C. Wavelet Packet Feature Extraction for Vibration Monitoring. *IEEE Trans Ind Electron*. 2000;47(3):650–667.

37. Daubechies I. *Ten Lectures on Wavelets*. Vol 61. Philadelphia, Pennsylvania: SIAM; 1992.

38. Mallat S. A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. *IEEE Trans Pattern Anal Mach Intell*. 1989;11(7):674–693.

39. Shewhart WA. *Economic Control of Quality of Manufactured Product*. Vol Republished in 1980 as a 50th Anniversary Commemorative Reissue by ASQC Quality Press. New York: D. Van Nostrand Company, Inc.; 1931.

40. Montgomery DC. *Introduction to Statistical Quality Control*. 4th ed. New York: John Wiley &Sons Inc; 2001.

41. Hunter JS. The exponentially weighted moving average. *J Qual Technol*. 1986;18(4):203–210.

42. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning*. NY: Springer; 2001.

43. van der Heijden F, Duin RPW, de Ridder D, Tax DMJ. *Classification, Parameter Estimation and State Estimation*. Chichester U.K: John Wiley & Sons Inc; 2004.