Multiscale PCA with Application to Multivariate Statistical Process Monitoring

Bhavik R. Bakshi
Dept. of Chemical Engineering, The Ohio State University, Columbus, OH 43210

Multiscale principal-component analysis (MSPCA) combines the ability of PCA to decorrelate the variables by extracting a linear relationship with that of wavelet analysis to extract deterministic features and approximately decorrelate autocorrelated measurements. MSPCA computes the PCA of wavelet coefficients at each scale and then combines the results at relevant scales. Due to its multiscale nature, MSPCA is appropriate for the modeling of data containing contributions from events whose behavior changes over time and frequency. Process monitoring by MSPCA involves combining only those scales where significant events are detected, and is equivalent to adaptively filtering the scores and residuals, and adjusting the detection limits for easiest detection of deterministic changes in the measurements. Approximate decorrelation of wavelet coefficients also makes MSPCA effective for monitoring autocorrelated measurements without matrix augmentation or time-series modeling. In addition to improving the ability to detect deterministic changes, monitoring by MSPCA also simultaneously extracts those features that represent abnormal operation. The superior performance of MSPCA for process monitoring is illustrated by several examples.

Introduction

Principal-component analysis (PCA) is among the most popular methods for extracting information from data, which has been applied in a wide range of disciplines. In chemical process operation and control, PCA is used to solve several tasks including data rectification (Kramer and Mah, 1994), gross-error detection (Tong and Crowe, 1995), disturbance detection and isolation (Ku et al., 1995), statistical process monitoring (Kresta et al., 1991; Wise et al., 1990), and fault diagnosis (MacGregor et al., 1994; Dunia et al., 1996). PCA transforms the data matrix in a statistically optimal manner by diagonalizing the covariance matrix by extracting the cross-correlation or relationship between the variables in the data matrix. If the measured variables are linearly related and contaminated by errors, the first few components capture the relationship between the variables, and the remaining components are composed only of the error. Thus, eliminating the less important components reduces the contribution of errors in the measured data and represents it in a compact manner. Applications of PCA rely on its ability to reduce the dimensionality of the data matrix while capturing the underlying variation and relationship between the variables.

PCA is popular for process monitoring since it allows extension of the principles of univariate statistical process monitoring (SPM) to monitoring of multivariate processes (Jackson, 1980; Kresta et al., 1991). Abnormal operation is detected if the measurements deviate from the region of normal operation in the space of the retained principal-component scores, or the space of the residual error. The scores and residuals are usually plotted as a Shewhart chart at the scale of the measurements, but may also be plotted as a cumulative sum (CUSUM) or an exponentially weighted moving average (EWMA). Shewhart charts are quick at detecting large changes, but slow at detecting small shifts. In contrast, CUSUM and EWMA with properly selected filter parameters are better at detecting small changes, but slow at detecting large changes. Consequently, the best type of control chart depends on the nature of the change in the signal.

Conventional PCA is best for analyzing a two-dimensional matrix of data collected from a steady-state process, containing linear relationships between the variables. Since these conditions are often not satisfied in practice, several extensions of PCA have been developed. Multiway PCA allows the analysis of a multidimensional matrix (Nomikos and MacGregor, 1994). Hierarchical or multiblock PCA permits easier modeling and interpretation of a large matrix by decomposing it into smaller matrices or blocks (Wold et al., 1996; Mac-

Gregor et al., 1994). Dynamic PCA extracts time-dependent relationships in the measurements by augmenting the data matrix by time-lagged variables (Kresta et al., 1991; Ku et al., 1995). Nonlinear PCA (Kramer, 1991; Hastie and Stuetzle, 1989; Dong and McAvoy, 1996; Tan and Mavrovouniotis, 1995) extends PCA to extracting nonlinear relationships between the variables. On-line adaptive PCA updates the model parameters continuously by exponential smoothing (Wold, 1994).

Modeling by PCA and its extensions is done at a single scale, that is, the model relates data represented on basis functions with the same time-frequency localization at all locations. For example, PCA of a time series of measurements is a single-scale model since it relates variables only at the scale of the sampling interval. Such a single-scale modeling approach is appropriate if the data contains contributions at just one scale. Unfortunately, data from almost all practical processes are multiscale in nature due to

- Events occurring at different locations and with different localization in time and frequency.
- Stochastic processes whose energy or power spectrum changes with time and/or frequency.
- Variables measured at different sampling rates or containing missing data.

Consequently, conventional PCA is not ideally suited for modeling of most process data. Techniques have been developed for PCA of some types of multiscale data such as missing data, but the single-scale approach forces data at all scales to be represented at the finest scale, resulting in increased computational requirements.

Another shortcoming of conventional PCA and its extensions is that its ability to reduce the error by eliminating some components is limited, since an imbedded error of magnitude proportional to the number of selected components will always contaminate the PCA model (Malinowski, 1991). This limited ability of PCA to remove the error deteriorates the quality of the underlying model captured by the retained components, and adversely affects the performance of PCA in a variety of applications. For example, in process monitoring by PCA, due to the presence of errors, detection of small deviations may not be possible and that of larger deviations may be delayed. Similarly, contamination by the imbedded error also deteriorates the quality of the gross-error detection and estimation of missing data. Consequently, the performance of PCA may be improved by methods that allow better separation of the errors from the underlying signal.

A popular approach for improving the separation between the errors and the underlying signal is to pretreat the measurements for each variable by an appropriate filter. In the chemical industry, univariate linear filters such as exponential smoothing and mean filtering (Tham and Parr, 1994), plus univariate nonlinear filters, such as finite impulse-response median-hybrid (FMH) filters (Heinonen and Neuvo, 1987; Piovoso et al., 1992), are most popular. Linear filters represent data at a single scale, and suffer from the disadvantages of single-scale PCA. Nonlinear filters are multiscale in nature, and cause less distortion of the retained features, but perform best for piecewise constant or slowly varying signals, and are often restricted to off-line use. The recent development of wavelet-based filtering methods (Donoho et al., 1995) overcomes the disadvantages of other nonlinear filters, and can be used on-line for all types of signals (Nounou and Bakshi, 1998). Despite these advances in filtering methods, preprocessing of the measured variables is still not a good idea, since it usually destroys the multivariate nature of the process data, which is essential for multivariate SPM and other operation tasks (MacGregor, 1994).

For reaping the benefits of reducing errors by filtering to improve process monitoring, it is necessary to develop an integrated approach to both these tasks (Kramer and Mah, 1994; Venkatasubramanian and Stanley, 1994; Bakshi, 1995). Such an approach is developed in this article by combining the ability of PCA to extract the relationship between the variables and decorrelate the cross-correlation with the ability of wavelets to extract features in the measurements and approximately decorrelate the autocorrelation. The development of wavelets has sparked a flurry of research activity over the last decade, but most work has focused on using wavelets for multiscale data analysis. Multiscale modeling has received surprisingly little attention, despite the fact that most existing modeling methods are inherently single scale in nature, whereas most data contain contributions at multiple scales. Research on multiscale modeling to date has focused on multiscale state-space modeling and estimation (Chou et al., 1994; Stephanopoulos et al., 1997). This article develops a multiscale approach for modeling by PCA that can be generalized to transform other single-scale empirical modeling methods to multiscale modeling.

The MSPCA methodology consists of decomposing each variable on a selected family of wavelets. The PCA model is then determined independently for the coefficients at each scale. The models at important scales are then combined in an efficient scale-recursive manner to yield the model for all scales together. For multivariate SPM by MSPCA, the region of normal operation is determined at each scale from data representing normal operation. For new data, the important scales are determined as those where the current coefficient violates the detection limits. The actual state of the process is confirmed by checking whether the signal reconstructed from the selected coefficients violates the detection limits of the PCA model for the significant scales. This approach is equivalent to adaptively filtering each value of the scores and residuals by a filter of dyadic length that is best suited for separating the deterministic change from the normal process variation. The detection limits for the scores and residuals also adapt to the nature of the signal. Thus, monitoring by MSPCA can automatically specialize to Shewhart, CUSUM, or EWMA of the scores and residuals with filters of dyadic length and parameters depending on the nature of the measurements, and the selected wavelet. Furthermore, retaining only those coefficients that violate the detection limits at each scale integrates the tasks of process monitoring and extraction of features relevant to abnormal operation, without any prefiltering. The MSPCA approach is analogous to multiblock PCA (Wold et al., 1996), with the subblocks being defined by the wavelet coefficients at each scale, and the superblock by the selected scales together. It is also similar to the algorithm for multiscale Kalman filtering (Chou et al., 1994), and multigrid methods (Briggs, 1987).

Wavelets and PCA have also been combined by other researchers, but the resulting techniques are different from the MSPCA method proposed in this article. The fast approxi-

mate PCA of Wickerhauser (1994) determines the approximate eigenvectors by selecting them from a library of wavelet packet basis functions. Since this library can be searched very efficiently, the approximate PCA algorithm takes $O(np \log p)$ computations vs. $O(p^3)$ computations for the exact PCA of an $n \times p$ data matrix. The local PCA approach of Coifman and Saito (1996) aims at overcoming the global nature of the basis functions determined by PCA. It decomposes the data matrix by local cosine transforms, and selects those temporally localized basis functions that explain the maximum variance. The approach of computing the PCA of the wavelet coefficients instead of the time-domain data, and its application to process monitoring, has also been suggested by Kosanovich and Piovoso (1997). Their approach preprocesses the data by the univariate FMH filter and then transforms it to the wavelet domain before applying PCA to the coefficients. This approach does not fully exploit the benefits of multiscale modeling, and the univariate filtering is not integrated with PCA. Furthermore, monitoring a process based only on its wavelet decomposition will result in too many false alarms after a process returns to normal operation.

The rest of this article is organized as follows. Conventional PCA is introduced in the next section, followed by a brief introduction to wavelet analysis. The methodology for MSPCA is developed in detail, and its relation to multiblock PCA and multiscale Kalman filtering is described. The MSPCA method is applied to multivariate statistical process monitoring when the abnormal operation is due to a deterministic change in the variables. The superiority of process monitoring by MSPCA over conventional PCA is illustrated by several examples, followed by a discussion of the results and future work.

Principal-Component Analysis

PCA transforms an $n \times p$ data matrix, X, by combining the variables as a linear weighted sum as

$$X = TP^T, \tag{1}$$

where P are the principal-component loadings, T are the principal-component scores, and n and p are the number of measurements and variables, respectively. The principal-component loadings denote the direction of the hyperplane that captures the maximum possible residual variance in the measured variables, while maintaining orthonormality with the other loading vectors. The eigenvectors of the covariance matrix of X are the loadings, and the eigenvalues indicate the variance captured by the corresponding eigenvector. The data matrix may also be decomposed by singular-value decomposition as,

$$X = U\Lambda^{1/2}V,$$

where Λ is a diagonal matrix of the eigenvalues $P^T = V$ and $T = U\Lambda^{1/2}$.

If the measured variables are redundant or correlated, the number of nonzero eigenvalues is equal to the rank of the matrix, and the data matrix may be reproduced exactly without considering the loadings and scores corresponding to the zero eigenvalues. Thus, PCA reduces the dimensionality of the data matrix by extracting the linear relationship between the variables and decorrelating the cross-correlation in the data matrix. In practice, the measured variables are usually contaminated by errors, and none of the eigenvalues are exactly zero, but the loadings and scores corresponding to small eigenvalues are composed of the errors only. Thus, the contribution of the errors in the data matrix may be decreased by eliminating the loadings and scores corresponding to the small eigenvalues, and reconstructing the filtered or rectified matrix as

$$\hat{X} = \hat{T}\hat{P}^T, \tag{2}$$

where $\hat{T}$ and $\hat{P}$ represent the selected scores and loadings, respectively.

An important decision in PCA is to select the appropriate number of principal components that capture the underlying relationship, while eliminating the errors. Several techniques are available for this task, and are reviewed by Jackson (1991) and Malinowski (1991). If an estimate of the error is not available, methods such as the scree test, parallel analysis, or cross-validation may be used. The scree test plots the eigenvalues and relies on the less important ones to be much smaller than the others and to lie on a straight line. The test looks for a break in the steep slope of the relevant eigenvalues and the almost flat slope of the eigenvalues representing the errors. The subjective nature of this test is reduced by parallel analysis, where another matrix with uncorrelated, normally distributed variables is constructed to be of the same size as the matrix being analyzed. The eigenvalues of both matrices are plotted, and the intersection of the two plots is used to separate the relevant from the irrelevant components. Selecting the components by cross-validation relies on determining the PCA model from training data and selecting the appropriate number of components as those that minimize the error on testing data.

PCA is best suited for the analysis of steady-state data with uncorrelated measurements. In practice, the dynamics of a typical chemical or manufacturing process cause the measurements to be autocorrelated, that is, the data matrix will have both cross-correlation and autocorrelation. The steady-state PCA approach may be extended to modeling and monitoring of dynamic systems by augmenting the data matrix by including lagged variables as,

$$X = [x_1(t)\ x_1(t-1),\ \ldots,\ x_2(t)\ x_2(t-1),\ \ldots].$$

The augmented data matrix permits implicit time-series modeling to decorrelate the relationship between the measurements, or autocorrelation. This approach has been widely used for system identification, model reduction, and disturbance detection of dynamic processes (Box and Tiao, 1977; Ljung, 1987; MacGregor, 1994; Ku et al., 1995; Cheng and McAvoy, 1997). It requires prior knowledge or assumptions about the order of the dynamics, and increases the computational complexity of the modeling.

For the monitoring of large-scale and complex processes, a modification of PCA called hierarchical or multiblock PCA has been developed (Wold et al., 1996). This approach divides the large data matrix into multiple blocks of variables, or measurements, or both. The PCA model is then developed

for the data in each block, as well as for multiple blocks together to capture the relationship between the subblocks. The blocks are usually defined based on knowledge about the system being modeled, such as variables measured on distinct pieces of equipment, or corresponding to different regimes of operation. This approach can provide greater insight into the data than conventional PCA, which treats the entire data matrix as just one block, and has been applied to monitoring and diagnosis of large continuous and batch processes (MacGregor et al., 1994; Kourti et al., 1995; Wold et al., 1996; Cheng and McAvoy, 1997).

Wavelets

Wavelets are a family of basis functions that are localized in both time and frequency, and may be represented as

$$\psi_{su}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-u}{s}\right),$$

where s and u represent the dilation and translation parameters, respectively. For most practical applications to measured data, the wavelet dilation and translation parameters are discretized dyadically, and the family of wavelets is represented as

$$\psi_{mk}(t) = 2^{-m/2}\psi(2^{-m}t - k), \tag{3}$$

where $\psi(t)$ is the mother wavelet, and m and k are the dilation and translation parameters, respectively. The translation parameter determines the location of the wavelet in the time domain, while the dilation parameter determines the location in the frequency domain as well as the scale or extent of the time-frequency localization. The wavelets represented by Eq. 3 may be designed to be orthonormal to each other (Daubechies, 1988).

Any signal may be decomposed into its contributions in different regions of the time-frequency space by projection on the corresponding wavelet basis function, as depicted in Figure 1. The lowest frequency content of the signal is represented on a set of scaling functions, as depicted in Figure 1f. The number of wavelet and scaling function coefficients decreases dyadically at coarser scales due to the dyadic discretization of the dilation and translation parameters. Fast algorithms for computing the wavelet decomposition are based on representing the projection of the signal on the corresponding basis function as a filtering operation (Mallat, 1989). Convolution with a filter, $\boldsymbol{H}$, represents projection on the scaling function, and convolution with a filter, $\boldsymbol{G}$, represents projection on a wavelet. Thus, the coefficients at different scales may be obtained as

$$\boldsymbol{a}_m = \boldsymbol{H}\boldsymbol{a}_{m-1}, \qquad \boldsymbol{d}_m = \boldsymbol{G}\boldsymbol{a}_{m-1}, \tag{4}$$

where $\boldsymbol{d}_m$ is the vector of wavelet coefficients at scale m, and $\boldsymbol{a}_m$ is the vector of scaling function coefficients. The original data are considered to be the scaling function coefficients at the finest scale, that is, $\boldsymbol{x} = \boldsymbol{a}_0$. Equation 4 can also be represented in terms of the original measured data vector, $\boldsymbol{x}$, as

$$\boldsymbol{a}_m = \boldsymbol{H}_m\boldsymbol{x}, \qquad \boldsymbol{d}_m = \boldsymbol{G}_m\boldsymbol{x}, \tag{5}$$

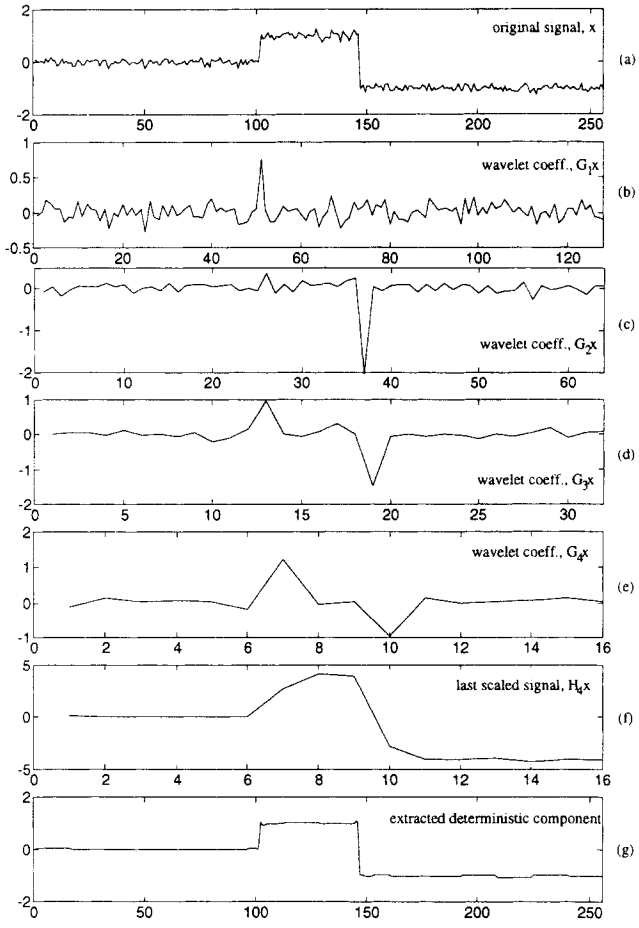

Figure 1. Wavelet decomposition and separation of stochastic and deterministic components.

(a) Original signal; (b)–(e) wavelet coefficients ($m = 1, \ldots, 4$); (f) scaling function coefficients ($L = 4$); (g) extracted deterministic component.

where $\boldsymbol{H}_m$ is obtained by applying the $\boldsymbol{H}$ filter m times, and $\boldsymbol{G}_m$ is obtained by applying the $\boldsymbol{H}$ filter $(m-1)$ times and the $\boldsymbol{G}$ filter once. The original data may be reconstructed exactly from its wavelet coefficients at all scales, d_m for $m = 1, 2, \ldots, L$, and scaling function coefficients at the coarsest scale, a_L.

Wavelets have found wide use for signal analysis and noise removal in a variety of fields due to their ability to represent deterministic features in terms of a small number of relatively large coefficients, while stochastic processes contaminate all the wavelet coefficients according to their power spectrum. This property is due to the time-frequency localization and can be seen in Figure 1, where the wavelet coefficients corresponding to the two step changes are larger than the coefficients corresponding to the uncorrelated stochastic process. The deterministic and stochastic components of the signal may be separated by an appropriate threshold, resulting in the deterministic component shown in Figure 1g. The statistical properties of this wavelet thresholding approach have been studied by Donoho et al. (1995), and several techniques have been developed for selecting the threshold (Nason, 1996).

Another useful property of wavelets is that although they

are not known to be the exact eigenfunctions or principal components of any operators, they are the approximate eigenfunctions of a large variety of operators (Wornell, 1990; Dijkerman and Majumdar, 1994). Consequently, the wavelet coefficients of most stochastic processes are approximately decorrelated. The variance of the wavelet coefficients at different scales represents the energy of the stochastic process in the corresponding range of frequencies, and corresponds to its power spectrum. Thus, for an uncorrelated Gaussian stochastic process or white noise, the variance of the wavelet coefficients is constant at all scales, whereas for colored noise, the variance decreases at finer scales.

The conventional algorithm for wavelet decomposition is restricted to off-line use since downsampling of the wavelet coefficients causes a time delay in the computation. This limitation may be overcome by decomposing the data in a moving window of dyadic length, with the most recent sample included in the window (Nounou and Bakshi, 1998). The wavelet coefficients obtained by this approach are identical to those obtained by the stationary or nondecimated wavelet transform (Mallat, 1991; Nason and Silverman, 1995), but the orthonormality of the basis functions and approximate decorrelation of the measurements in each window are maintained. Averaging the signal reconstructed from each moving window decreases the spurious features created near sudden changes and improves the accuracy and smoothness of the extracted features.

Multiscale Principal-Component Analysis

MSPCA combines the ability of PCA to extract the cross-correlation or relationship between the variables with that of orthonormal wavelets to separate deterministic features from stochastic processes and approximately decorrelate the auto-correlation among the measurements. The steps in the MSPCA methodology are shown in Figure 2 and the following algorithm:

1. for each column in data matrix,
2. compute wavelet decomposition
3. end
4. for each scale,
5. compute covariance matrix of wavelet coefficients at selected scale
6. compute PCA loadings and scores of wavelet coefficients
7. select appropriate number of loadings
8. select wavelet coefficients larger than appropriate threshold
9. end
10. for all scales together,
11. compute PCA by including scales with significant events
12. reconstruct approximate data matrix from the selected and thresholded scores at each scale
13. end

To combine the benefits of PCA and wavelets, the measurements for each variable (column) are decomposed to the column's wavelet coefficients using the same orthonormal wavelet for each variable. This results in transformation of the data matrix, $\boldsymbol{X}$, into a matrix, $\boldsymbol{WX}$, where $\boldsymbol{W}$ is an $n \times n$ orthonormal matrix representing the orthonormal wavelet transformation operator containing the filter coefficients,

$$\boldsymbol{W} = \begin{bmatrix} h_{L,1} & h_{L,2} & \cdots & & & & & h_{L,N} \\ g_{L,1} & g_{L,2} & \cdots & & & & & g_{L,N} \\ g_{L-1,1} & \cdots & & g_{L-1,\frac{N}{2}} & 0 & \cdots & & 0 \\ 0 & \cdots & & 0 & g_{L-1,\frac{N}{2}+1} & \cdots & & g_{L-1,N} \\ \vdots & & & & & & & \vdots \\ g_{1,1} & g_{1,2} & 0 & \cdots & & & & 0 \\ \vdots & & & & & & & \vdots \\ 0 & 0 & \cdots & & & 0 & g_{1,N-1} & g_{1,N} \end{bmatrix} = \begin{bmatrix} \boldsymbol{H}_L \\ \boldsymbol{G}_L \\ \boldsymbol{G}_{L-1} \\ \vdots \\ \boldsymbol{G}_m \\ \vdots \\ \boldsymbol{G}_1 \end{bmatrix}, \tag{1}$$

where $\boldsymbol{G}_m$ is the $2^{\log_2 n - m} \times n$ matrix containing wavelet filter coefficients corresponding to scale $m = 1, 2, \ldots, L$, and $\boldsymbol{H}_L$ is the matrix of scaling-function filter coefficients at the coarsest scale. The matrix, $\boldsymbol{WX}$, is the same size as the original data matrix, $\boldsymbol{X}$, but due to the wavelet decomposition, the deterministic component in each variable in $\boldsymbol{X}$ is concentrated in a relatively small number of coefficients in $\boldsymbol{WX}$, while the stochastic component in each variable is approximately decorrelated in $\boldsymbol{WX}$, and is spread over all components ac-

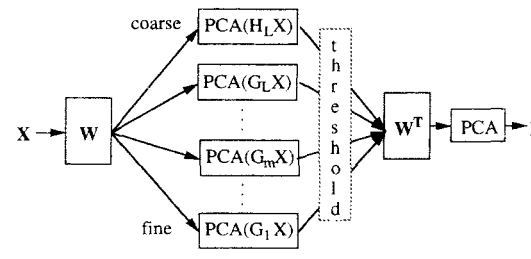

Figure 2. Methodology for multiscale principal-component analysis.

cording to its power spectrum. The relation between the PCA of X and WX is given by the following theorem.

Theorem 1. The principal-component loadings obtained by the PCA of X and WX are identical, whereas the principal-component scores of WX are the wavelet transform of the scores of X.

Proof. Since each column of X is transformed by the same orthonormal transformation matrix, W, the relationship between the columns in WX is the same as that between the columns in X, that is, the nature of the cross-correlation is unchanged. For a large number of samples and orthonormal wavelets, the mean of each column of X is identical to that of the corresponding column in WX, and the covariance matrix of X is the same as that of WX, since

$$(WX)^T(WX) = X^T W^T W X = X^T X. \tag{6}$$

Equation 6 also implies that the principal-component loadings of X and WX are identical. Since the PCA of X is given by Eq. 1, the PCA of the wavelet transformed matrix, WX, is given as

$$WX = (WT)P^T,$$

which proves that the scores of WX are the wavelet transform of the scores of X.

The covariance of the wavelet transformed matrix, and equivalently of the original data matrix, may be written in terms of the contribution at multiple scales as

$$(WX)^T(WX) = (H_L X)^T(H_L X) + (G_L X)^T(G_L X) + \cdots + (G_m X)^T(G_m X) + \cdots + (G_1 X)^T(G_1 X). \tag{7}$$

To exploit the multiscale properties of the data, the PCA of the covariance matrix of the coefficients at each scale is computed independently of the other scales. The resulting scores at each scale are not cross-correlated due to PCA, and their autocorrelation is approximately decorrelated due to the wavelet decomposition. Depending on the nature of the application, a smaller subset of the principal-component scores and wavelet coefficients may be selected at each scale. The number of principal components to be retained at each scale are not changed by the wavelet decomposition since it does not affect the underlying relationship between the variables at any scale. Consequently, existing methods such as cross-validation, the scree test, or parallel analysis may be applied to the data matrix in the time domain or to all the wavelet coefficients to select the relevant number of components. For selecting the relevant wavelet coefficients, different techniques may be devised depending on the application. For separating the stochastic and deterministic components of a signal, thresholding criteria such as the median absolute deviation (Donoho et al., 1995) or cross-validation (Nason, 1996) or the T^2 and Q detection limits (Jackson, 1991) may be applied to the scores at each scale. Applying separate thresholds at each scale allows MSPCA to be more sensitive to scale-varying signal features such as autocorrelated measurements. Thresholding of the coefficients at each scale identifies the region of the time-frequency space and scales where there is significant contribution from the deterministic features in the signal.

The covariance matrix for computing the loadings and scores for all scales together is computed by combining the covariance matrices of the scales at which the coefficients violate the threshold. This final covariance matrix may be computed efficiently by refining the covariance matrix at a coarser scale in a scale-recursive manner by incorporating the covariance matrix of the wavelet coefficients at the same scale as

$$(H_{m-1}X)^T(H_{m-1}X) = (H_m X)^T(H_m X) + \gamma(G_m X)^T(G_m X), \tag{8}$$

where

$$\gamma = \begin{cases} 1 & \text{if the PCA at scale } m \text{ has significant events} \\ 0 & \text{otherwise.} \end{cases}$$

Equation 8 is identical to that for time-recursive computation of the covariance matrix (Ljung, 1987; Dayal and MacGregor, 1997), except that the time parameter is replaced by scale, and the parameter, γ, is analogous to the forgetting factor. This MSPCA modeling method represents one way of using the PCA models at multiple scales, and other approaches may be devised, depending on the application.

MSPCA is related to several existing methods, such as multiblock or hierarchical PCA (Wold et al., 1996), multiscale Kalman filtering (Chou et al., 1994), and multigrid methods for solving differential equations (Briggs, 1987). Multiblock PCA divides a large data matrix from a complex process into multiple blocks defined in terms of the variables, measurements, or both. In contrast, in MSPCA, each variable is decomposed to multiple scales, and the matrix of coefficients at each scale are treated as a separate block. As in the multiblock PCA algorithm, the MSPCA algorithm also involves determining the PCA of each subblock or scale, of the superblock, or of all scales together. The PCA of each subblock is useful for identifying "interesting" blocks, and that of the superblock is useful for determining the relationship between the selected subblocks. Furthermore, like multiblock PCA, multiscale PCA also reduces to conventional PCA, as stated by the following theorem.

Theorem 2. MSPCA reduces to conventional PCA if neither the principal components nor the wavelet coefficients at any scale are eliminated.

Proof. If neither the principal components nor the wavelet coefficients at any scale are eliminated, then the covariance matrix at all scales computed by Eq. 8 will have $\gamma = 1$ for all scales, making it identical to Eq. 7, which is the covariance matrix of the original data. Consequently, the results of MSPCA and PCA will be identical.

MSPCA is also related to multiscale Kalman filtering (Chou et al., 1994), and the model determined by both methods may be represented in terms of a binary tree whose nodes represent wavelet coefficients. The methodology for both multiscale Kalman filtering and MSPCA starts with a fine-to-coarse scale sweep of the tree. The next MSPCA step of determining the PCA model at each level of the tree and selecting the relevant principal components and wavelet coefficients is not present in the multiscale Kalman filtering method. Finally, both MSPCA and multiscale Kalman filtering fuse the mod-

els at different scales recursively in a coarse-to-fine scale sweep of the tree. Like multiscale Kalman filtering, the MSPCA algorithm can also be implemented efficiently in parallel since the computation at each scale can be done independently, and only communication between neighboring scales is needed for the scale-recursive fusion of covariance matrices. This should result in a logarithmic improvement in the computational complexity, that is, an algorithm requiring $O(n)$ computations will require only $O(\log n)$ computations when implemented in parallel.

Instead of the MSCPA methodology described earlier, some of the benefits of the wavelet representation may be reaped by just transforming the measured data on a selected wavelet basis and computing the PCA of $\boldsymbol{WX}$ instead of $\boldsymbol{X}$. PCA of $\boldsymbol{WX}$ will make it easier to separate deterministic features in a stochastic process, but this approach will be restricted to off-line use and will not fully exploit the benefits of the multiscale representation, since it will implicitly assume that the nature of the data does not change with scale, or that the stochastic process is uncorrelated and that the contribution of the deterministic features is equal at all scales. This assumption will cause too many false alarms for autocorrelated measurements, as compared to the MSPCA approach that accounts for the scale-dependent power spectrum. False alarms will also be created for process monitoring based on the scores of $\boldsymbol{WX}$ after a process returns to normal operation, as discussed in the next section.

Multivariate Statistical Process Monitoring by MSPCA

PCA is popular for multivariate SPM (Jackson, 1980; Kresta et al., 1991) since it does not encounter problems due to ill-conditioning faced by other multivariate monitoring methods such as multivariate CUSUM (Crosier, 1988) and EWMA (Lowry et al., 1992). The principal-component loadings and detection limits for the scores and residuals are computed from data representing normal operation. For new data, the score space is monitored by computing the T^2 value as the sum of the squares of the selected scores scaled by the respective eigenvalue computed from data representing normal operation as

$$T_i^2 = \sum_{j=1}^{\hat{p}} \frac{t_{ij}^2}{\lambda_j},$$

where T_i^2 is the T^2 value for the ith row of measurements, $\hat{p}$ is the number of scores selected, and λ_j is the eigenvalue of the jth score. For Gaussian distributed measurements, the T^2 detection limits may be calculated based on the F-distribution (Jackson, 1991). For monitoring in the residual space, the squared prediction error is computed as

$$Q_i = \sum_{j=1}^{\hat{p}} (x_{ij} - \hat{x}_{ij})^2$$

and the detection limits are determined from the eigenvalues of the components not included in the PCA model (Jackson, 1980). Changes in the process that affect all the variables while satisfying the PCA model are detected in the score space, whereas changes that violate the PCA model are detected in the residual space.

The score and residual spaces are usually monitored by plotting the time series of T^2 and Q values without any filtering. This Shewhart-type control chart can detect large changes in the mean quickly, but is slow in detecting small changes. The ability to detect small changes may be improved by plotting the CUSUM or EWMA of T^2 and Q, but it requires proper selection of the filter parameters, and detection of large changes may be delayed (Scranton et al., 1996). These popular charting methods differ in the scale at which they represent the data, since Shewhart charts plot the data at the finest scale, CUSUM is a representation at the coarsest scale, and the scale of EWMA depends on the value of the filter parameter. As described in this section, monitoring by MSPCA automatically represents the T^2 and Q values at the scale that is best for detecting the abnormal operation.

Process monitoring by MSPCA involves computing independent principal-component loadings and detection limits for the scores and residuals at each scale from data representing normal operation. For new data, a statistically significant change is indicated if the scores or residuals based on wavelet coefficients computed from the most recent measurements violate the detection limits at any scale. Since the wavelet coefficients are sensitive only to changes, if a variable goes outside the region of normal operation and stays there, the wavelet coefficients will be statistically significant only when the change first occurs. The change is first detected at the finest scale that covers the frequencies present in the feature representing abnormal operation. As the change persists, it is detected by wavelet coefficients at coarser scales present in the abnormal feature. If the change persists, the most recent scaling-function coefficients are the last ones to violate the detection limit, and continue to do so for as long as the process remains in an abnormal state. Similarly, when a process returns from abnormal to normal operation, the change will be detected by the wavelet coefficients, but the scaling function coefficients will continue to indicate abnormal operation for several samples due to the coarser scale of representation. Thus, process monitoring based only on PCA of the wavelet and scaling-function coefficients will not permit quick and continuous detection of a shift, and may create false alarms after a process has returned to normal operation, as illustrated by the first example in the next section. This time delay in detecting a shift at coarser scales is because the longer filter dilates the deterministic change, and delays its detection. A similar time delay is also observed in CUSUM and EWMA charts (Lowry et al., 1992).

Fortunately, the last MSPCA steps of selecting the scales that indicate significant events, reconstructing the signal to the time domain, and computing the scores and residuals for the reconstructed signal, improve the speed of detecting abnormal operation and eliminate false alarms after a process returns to normal operation. For SPM, those scales are selected at which the wavelet coefficient corresponding to the current measurements violates the detection limits. Thus, the definition of the parameter, γ, in Eq. 8 is modified to

$$\gamma = \begin{cases} 1 & \text{if the current coefficient at scale } m \text{ violates the detection limit} \\ 0 & \text{otherwise.} \end{cases}$$

Since the reconstructed signal in the time domain is generated from the large wavelet coefficients, MSPCA integrates the task of monitoring with that of extracting the signal features representing abnormal operation, with minimum distortion and time delay. Consequently, there is no need for a separate step for prefiltering the measured variables. Furthermore, since the covariance matrix computed by Eq. 8 depends on the scales at which the current scores or residuals violate the detection limits, the final detection limits for confirming the state of the process also adapt to the nature of the signal features.

For on-line monitoring, the MSPCA algorithm is applied to measurements in a moving window of dyadic length. This approach is based on the on-line multiscale rectification method of Nounou and Bakshi (1998), except that instead of rectifying each measured variable, on-line MSPCA rectifies the T^2 and Q values computed from the PCA at each scale. Consequently, as shown for on-line multiscale rectification, on-line monitoring by MSPCA automatically specializes to a Shewhart, CUSUM, or EWMA chart applied to the T^2 and Q values for each measurement, as described below:

- If none of the current coefficients at any scale are eliminated, the reconstructed value and corresponding detection limits will be identical to those for the original measurement, making MSPCA for that measurement identical to a Shewhart chart.
- If the signal is decomposed using Haar wavelets to the maximum depth (coarsest scale), and only the current coefficient of the last scaled signal is selected, then MSPCA will be similar to a CUSUM of the maximum possible dyadic length of measurements.
- If current coefficients for the last scaled signal and detail signals at consecutively finer scales are selected, then MSPCA will be equivalent to a mean filter at the finest scale of the selected coefficients for Haar wavelets, or similar to an EWMA filter at the finest scale of the selected coefficients for smoother boundary-corrected wavelets.

The use of a moving window makes the on-line wavelet decomposition algorithm equivalent to wavelet decomposition without downsampling, causing a signal of length n to result in a total of $n(L+1)$ coefficients, where L is the depth of the wavelet decomposition. This increase in the number of coefficients requires on-line monitoring by MSPCA to increase the detection limits at each scale to maintain the desired confidence limit for the reconstructed signal. For example, for normally distributed uncorrelated measurements in a window of length 128, approximately one sample will lie outside the 99% limits. The off-line wavelet transform will also result in 128 uncorrelated coefficients, and approximately one coefficient will violate the 99% limits. In contrast, the on-line wavelet transform of these data will result in 128 coefficients at each scale, and approximately one coefficient will violate the 99% detection limits at *each* scale. Thus, if the signal is decomposed to four detail signals and one scaled signal, that is, for $L=4$, application of the 99% limit at each scale will result in an effective confidence of only 95% for the reconstructed signal, since the coefficients violating the detection limits at each scale need not be at the same location. Consequently, the detection limits at each scale for on-line monitoring by MSPCA need to be adjusted to account for the overcompleteness of the on-line wavelet decomposition by the following equation:

$$C_L = 100 - \frac{1}{L+1}(100 - C), \tag{9}$$

where C is the desired overall confidence limit, C_L is the adjusted confidence limit at each scale in percent, and L is the number of scales to which the signal is decomposed, resulting in L detail signals and one scaled signal.

Another effect of the on-line wavelet decomposition approach is that the wavelet coefficients retain more of the autocorrelation in the signal due to the use of overlapping windows for the decomposition. Fortunately, as shown by the illustrative examples, the performance of monitoring by MSPCA is not adversely affected by the autocorrelated coefficients in adjacent windows, since the confidence limits at each scale are increased by Eq. 9, and even relatively small deterministic features are captured by large wavelet coefficients.

An important practical decision in the MSPCA methodology is that of the depth or number of scales of the wavelet decomposition. Ideally, the depth should be selected to provide maximum separation between the stochastic and deterministic components of a signal. If the depth is too small, then the last scaled signal will have a significant amount of noise that will be retained in the result of MSPCA, but if the depth is too large, the matrix of coefficients at coarser scales will have very few rows due to the dyadic downsampling, and will affect the accuracy of the PCA at that scale. The reduction in the number of rows is not a limitation for on-line or nondecimated wavelet decomposition. The depth may be determined by cross-validation, but for the examples in this article, a heuristic maximum depth that results in 32 coefficients at the coarsest scale, that is, $L = \log_2 n - 5$ is found to be a good choice.

Illustrative Examples

The MSPCA methodology for on-line SPM is implemented in Matlab based on the wavelet analysis functions available in WaveLab (Buckheit and Donoho, 1995).

Mean shift in uncorrelated measurements

This example consists of four variables, with an underlying dimensionality of two. Two variables are represented as uncorrelated Gaussian measurements of zero mean and unit variance, as shown in Eqs. 10a and 10b. The remaining two variables are formed by adding and subtracting the first two variables, respectively, as shown in Eqs. 10c and 10d:

$$\tilde{x}_1(t) = N(0,1) \tag{10a}$$

$$\tilde{x}_2(t) = N(0,1) \tag{10b}$$

$$\tilde{x}_3(t) = \tilde{x}_1(t) + \tilde{x}_2(t) \tag{10c}$$

$$\tilde{x}_4(t) = \tilde{x}_1(t) - \tilde{x}_2(t). \tag{10d}$$

The measured data matrix, X, consists of these four variables contaminated by uncorrelated Gaussian errors of zero mean and standard deviation of 0.2 as,

$$X(t) = \tilde{X}(t) + 0.2N(0,I).$$

The data representing normal operation consist of 256 equally spaced samples. The PCA and MSPCA models are developed from these data by selecting the first two principal components. Since the measurements are uncorrelated, the variances at each scale for MSPCA are almost equal. Abnormal operation is indicated by a step change in the mean of all four variables between samples 176 and 225. The performance of MSPCA is illustrated and compared with that of PCA for two different magnitudes of the mean shift.

Case 1. This case considers a mean shift of unity, which is large enough for easy detection by both PCA and MSPCA. The T^2 and Q charts for PCA for the abnormal operation are shown in Figure 3. The Q chart easily and unambiguously identifies the mean shift between the samples where it was introduced, but some measurements do go below the 95% detection limit.

The T^2 and Q charts for on-line MSPCA are constructed as illustrated in Figures 4, 5, and 6. These figures contribute to the final results shown in Figures 7 and 8. The Q chart at each scale for the wavelet decomposition of data in a window from sample number 49 to 176 is shown in Figure 4a. Since a step change in the mean contributes to a wide range of frequencies, it is first detected by the Q chart at the finest scale, $m = 1$, as shown in Figure 4a. The reconstructed time-domain signal based on the coefficient that violates the 99% limit is shown in Figure 4b. The 99% detection limit for the most recent measurement is computed from the variance of the coefficients representing normal operation at the finest scale by applying Eq. 8, and is shown as an $\times$ in Figure 4b. The Q chart for on-line monitoring in Figure 8b is constructed by including only the current value of the reconstructed signal and its detection limit at the sample number 176. Since only the last point is used for on-line monitoring, detection limits are not shown for any other measurements in Figure 4b.

Detection of the mean shift after it persists for a few samples is illustrated in Figure 5 for data in a window ending at

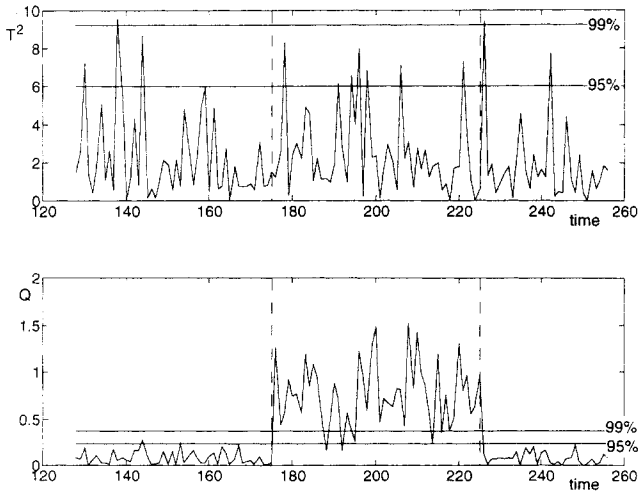

Figure 3. Performance of PCA for detecting mean shift of 1 between samples [176, 225] in uncorrelated measurements.

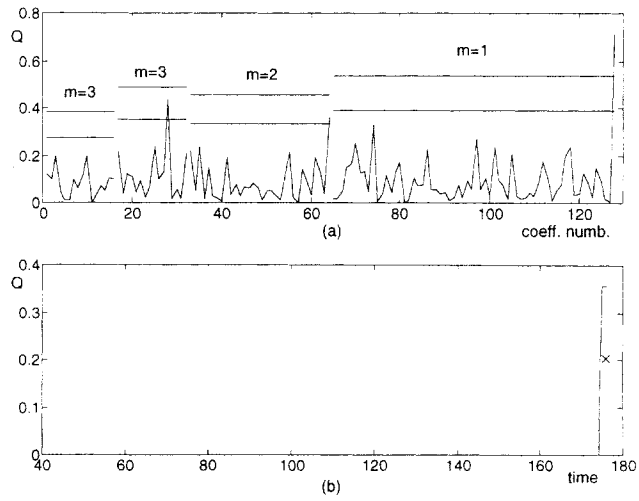

Figure 4. Multiscale detection of the unit mean shift at 175 in uncorrelated measurements for data in window [49, 176].

(a) Q chart for wavelet decomposition at each scale with 95 and 99% limits. Mean shift is detected with 99% confidence at finest scale, $m = 1$. (b) Q chart for reconstructed signal based on the coefficients outside the 99% limit in (a). The 99% detection limit for current measurement is computed from the variance of the normal data at scale $m = 1$, and indicated by $\times$. The values of the signal and detection limit at 176 contribute to Figure 8b.

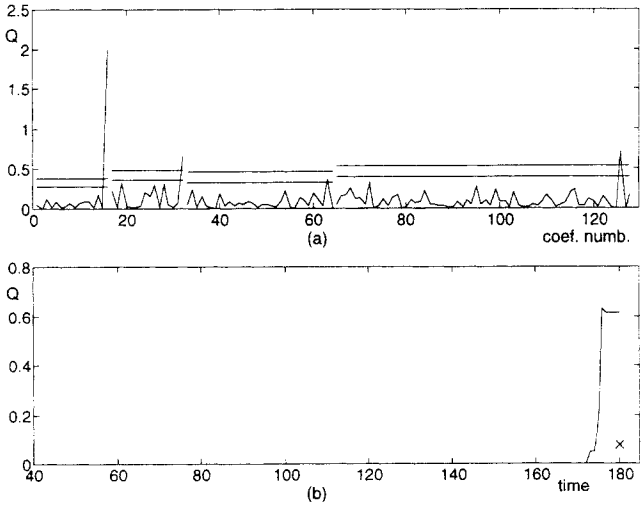

Figure 5. Detection of the unit mean shift in uncorrelated measurements for data in window [53, 180].

(a) Q chart for wavelet decomposition at each scale with 95 and 99% limits. Mean shift is detected at coarser scales for wavelet and scaling function coefficients at $m = 3$. (b) Q chart for reconstructed signal based on the coefficients outside the 99% limit in (a). The 99% detection limit for current measurement is computed from the variance of the scaling function and wavelet coefficients for normal data at scale $m = 3$, and indicated by $\times$. The values of the signal and detection limit at 180 contribute to Figure 8b.

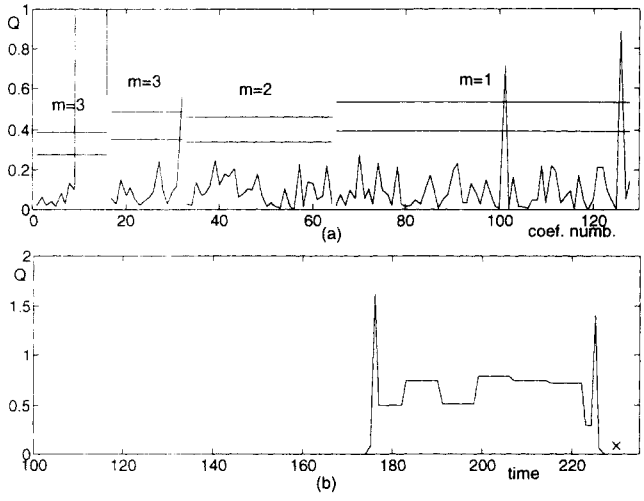

Figure 6. Detection of the end of the mean shift in uncorrelated measurements for data in window [103, 230].

(a) Q chart for wavelet decomposition at each scale with 95 and 99% limits. End of mean shift is detected at $m = 1$, but coefficients at coarser scales continue to violate the detection limit due to shift at 175. (b) Q chart for reconstructed signal based on the coefficients outside the 99% limit in (a). The 99% detection limit for current measurement is computed from the variance of the scaling function and wavelet coefficients for normal data at scale $m = 3$, and indicated by $\times$. The return of the process to normal operation is clearly indicated by the reconstructed signal and detection limit at 230.

sample 180. The change is no longer detected by the current coefficients at the finest scale, but is detected by the current coefficients at coarser scales, at $m = 3$, as shown in Figure 5a. The reconstructed signal based on the coefficients that violate the 99% detection limit in Figure 5a, and the detection limit for these important scales, are shown in Figure 5b, and contribute to the signal in Figure 8b at sample number

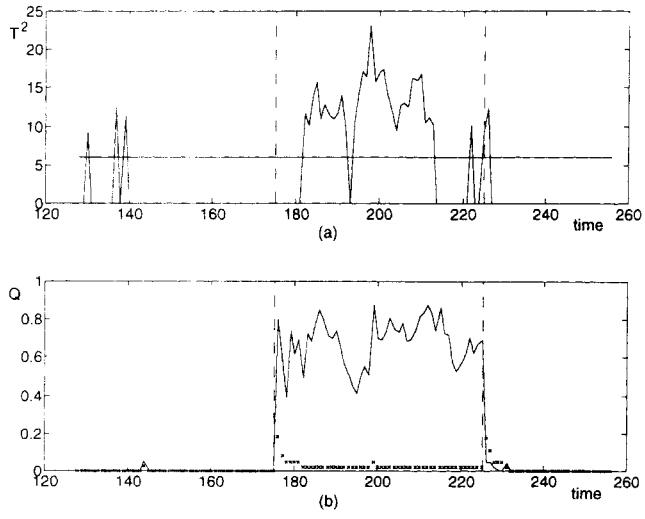

Figure 7. Monitoring by MSPCA of the uncorrelated measurements for the mean shift of 1.

(a) T^2 plot with 95% confidence limit. (b) Q plot with 95% confidence limit indicated by $\times$.

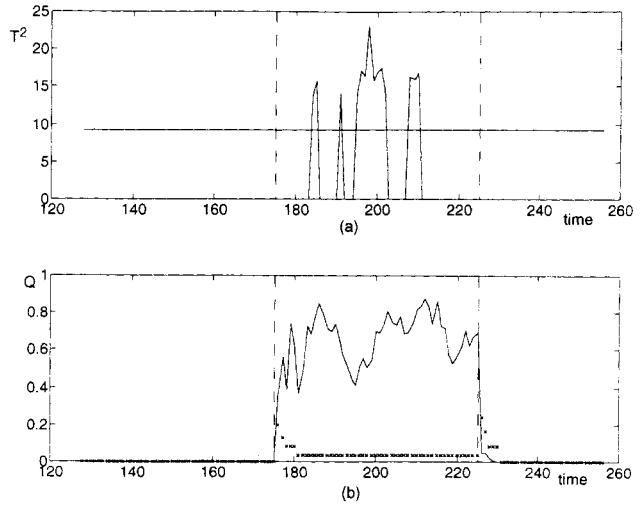

Figure 8. Monitoring by MSPCA of the uncorrelated measurements for mean shift of 1.

(a) T^2 plot with 99% confidence limit. (b) Q plot with 99% confidence limit indicated by $\times$.

180. The contribution of the change is present at the finer scales, but in older coefficients. While the change continues, it will not be detected by any current wavelet coefficient, but will be detected by the current coefficients in the last scaled signal. If the signal is decomposed to a depth such that the last scaled signal contains little contribution from the stochastic process variation, then the current coefficients in the last scaled signal will continue to violate the detection limits after an initial time delay. If the process returns to normal operation, the change will again be detected first at the finest scales, but the coefficients of the last scaled signal will continue to violate the detection limits for several samples after the process has returned to normal, as shown in Figure 6a. Thus, any monitoring scheme based only on the wavelet decomposition, without the last MSPCA step of monitoring based on the time-domain reconstructed signal, may not be able to detect a return to normal operation as soon as it occurs. Figures 4b, 5b, and 6b demonstrate the ability of MSPCA to extract the feature representing abnormal operation and to adjust the detection limits, depending on the scales at which the abnormal operation is detected.

The T^2 and Q charts for on-line MSPCA are shown in Figure 7 for a confidence limit of 95%, and in Figure 8 for a confidence of 99%. Comparison with Figure 3 shows that the Q chart for both confidence limits detects the mean shift at least as accurately as conventional PCA, and all the measurements are well above the confidence limits. The original measured variables and features extracted by MSPCA based on coefficients that violate the 99% limit on the Q chart are shown in Figure 9. These signals are the mean of the reconstructed signals like the ones in Figures 4b, 5b, and 6b. The spikes in the extracted feature of Figure 9 near the end of the mean shift may be decreased by soft thresholding, that is, by subtracting the detection limit from the retained coefficients (Nounou and Bakshi, 1998).

Case 2. In this case, the magnitude of the mean shift is 0.3, which is only 30% of the standard deviation of the measurements. The T^2 and Q charts in Figure 10 show that the

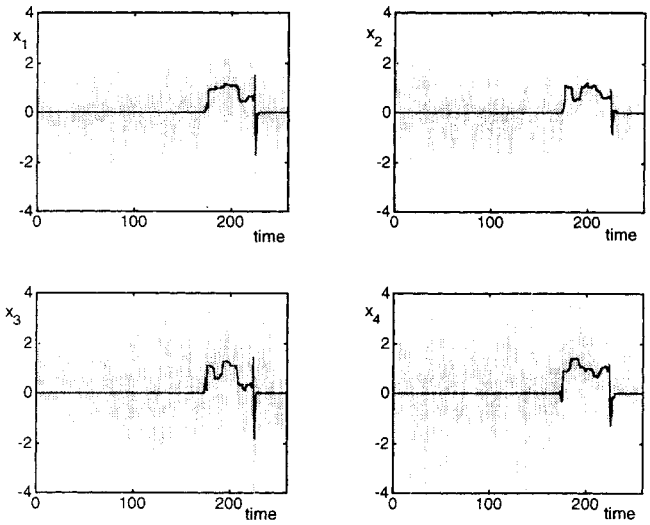

Figure 9. Features relevant to abnormal operation extracted from each variable by MSPCA residuals plot for 99% confidence shown in Figure 8.

Plot in lighter shade is actual data, darker shade is extracted feature. True feature is a unit mean shift between samples 176 and 225.

Q values for the abnormal operation are a little larger in the region of the mean shift, but do not consistently violate the confidence limits. Thus, conventional PCA is unable to detect this change. The T^2 and Q charts for on-line MSPCA demonstrate the superiority of monitoring by PCA, as shown in Figures 11 and 12. The Q charts for 95% and 99% confidence clearly identify the mean shift over most of the range of abnormal operation, which is significantly better than monitoring by PCA.

Monitoring of autocorrelated measurements

The data for this example are generated from a model suggested by Ku et al. (1995). The input variables, z and u, are given by

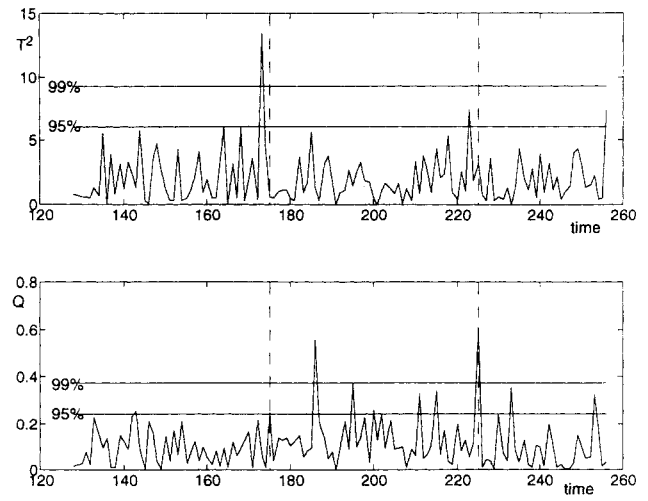

Figure 10. Performance of monitoring uncorrelated measurements with mean shift of 0.3 by PCA.

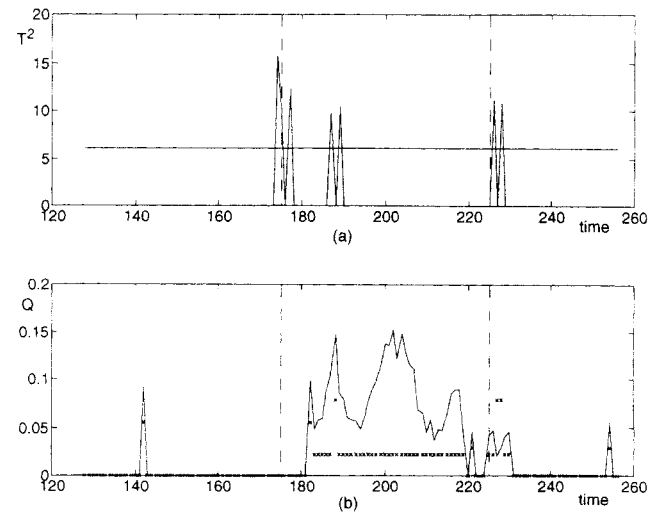

Figure 11. Monitoring uncorrelated measurements with mean shift of 0.3 by MSPCA.

(a) T^2 plot with 95% confidence limit. (b) Q plot with 95% confidence limit indicated by $\times$.

$$z(k) = 0.8z(k-1) + u(k-1)$$
$$u(k) = 0.7u(k-1) + w(k-1),$$

where $w(k)$ is white noise with unit variance. The measured data are corrupted by white noise of standard deviation 0.2. The data matrix used for steady-state PCA is given by

$$X(k) = [z(k)\ u(k)] + 0.2N(0, I).$$

The matrix representing normal operation consists of 1,024 measurements. Parallel analysis results in the selection of one principal component. Since the measurements are correlated, the eigenvalues for MSPCA vary with scale according to their power spectrum.

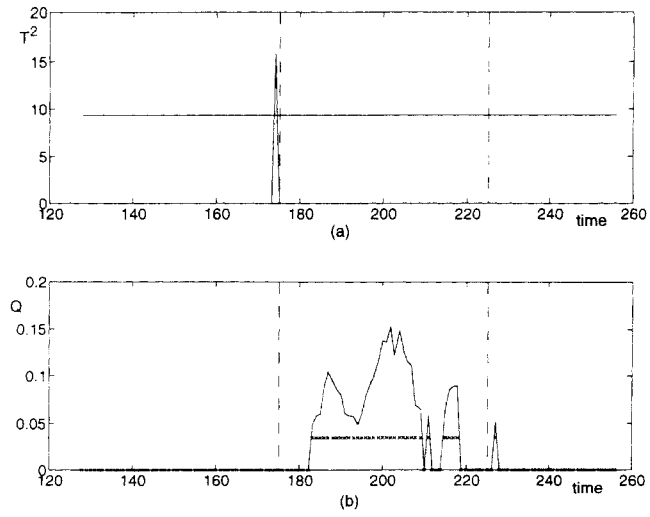

Figure 12. Monitoring uncorrelated measurements with mean shift of 0.3 by MSPCA.

(a) T^2 plot with 99% confidence limit. (b) Q plot with 99% confidence limit indicated by $\times$.

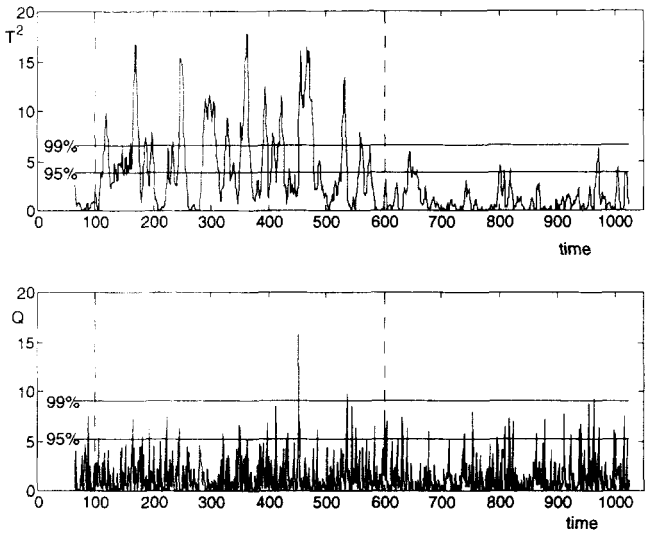

Figure 13. Monitoring of autocorrelated measurements with mean shift of 0.5 by steady-state PCA.

The testing data are generated by introducing a disturbance in the form of a mean shift of magnitude 0.5 in the input variable, u, between samples 101 and 600. The T^2 and Q charts for PCA are shown in Figure 13. The disturbance does affect the T^2 values, but cannot be detected with reasonable confidence from this chart. The T^2 and Q charts for MSPCA are shown in Figures 14 and 15. Clearly, the T^2 chart for MSPCA is significantly better at identifying the disturbance. Figures 14 and 15 also show that steady-state MSPCA for autocorrelated measurements does not result in too many false alarms.

Since the measurements in this example are autocorrelated, the data are also modeled by dynamic PCA by augmenting the data matrix by lagged variables as

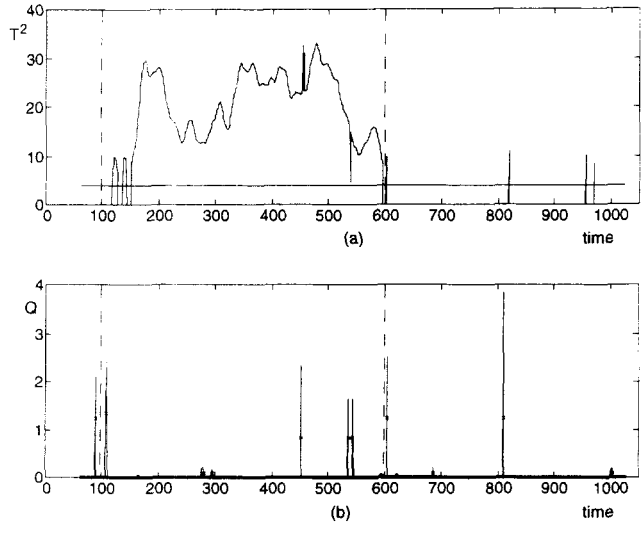

Figure 14. Monitoring of autocorrelated measurements by steady-state MSPCA.

(a) T^2 plot with 95% confidence limit. (b) Q plot with 95% confidence limit indicated by $\times$.

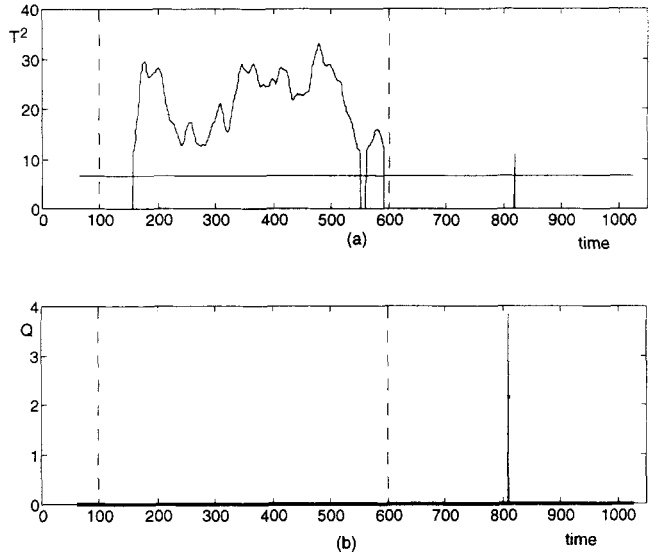

Figure 15. Monitoring of autocorrelated measurements by steady-state MSPCA.

(a) T^2 plot with 99% confidence limit. (b) Q plot with 99% confidence limit indicated by $\times$.

$$X(k) = [z(k)\ u(k)\ z(k-1)\ u(k-1)] + 0.2N(0, I).$$

Three components are retained by parallel analysis. As shown by the charts in Figure 16, dynamic PCA is also unable to detect the mean shift.

Fault detection in industrial fluidized catalytic cracker unit

This case study is provided by the Abnormal Situation Management (ASM) consortium to evaluate various fault-detection methodologies. The fluidized catalytic cracker unit (FCCU) consists of a stacked regenerator and disengager. The

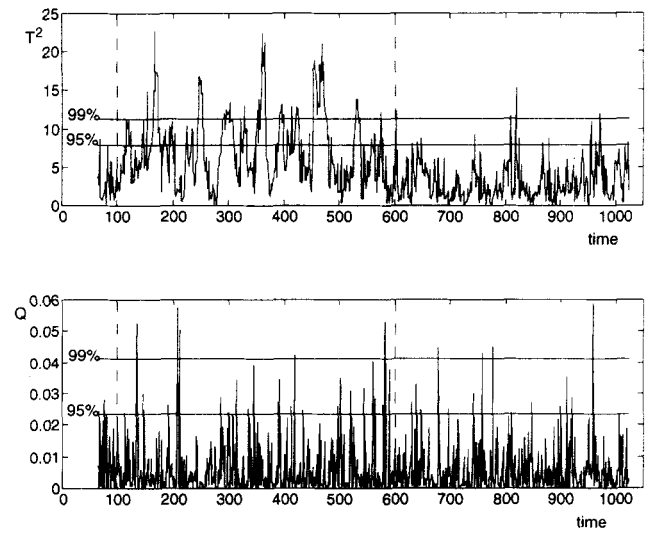

Figure 16. Monitoring of autocorrelated measurements with mean shift of 0.5 by dynamic PCA.

Table 1. Landmarks in the Simulation of Slurry Pumparound Fault as a Slow Ramp

Time	Event
0:00:00	Simulation begins
0:01:00	Spurious transmitter drift
0:05:00	Malfunction begins as a slow ramp
0:11:00	Spurious transmitter ends and holds at new value
1:05:00	Malfunction ends and controllers take over
1:15:00	Start of remedial action
1:26:00	End of simulation

Source: Ali (1997).

feed entering the unit is heated in the feed/slurry pump around exchanger followed by the feed preheater. The heated fresh feed is combined with steam and the catalyst and enters the riser. The catalyst and product are separated in the disengager by two cyclones. The catalyst is returned to the regenerator, while the product is sent to the bottom of the main fractionator. Additional details about the FCCU are provided by Ali (1997).

The FCCU simulator generates output from 110 sensors every 20 s. Data representing normal operation are available from a steady-state simulation of about 10 min and are corrupted by adding uncorrelated Gaussian errors of standard deviation 0.1. Three principal components are selected based on a parallel analysis of the steady-state data. The malfunction selected to demonstrate the features of monitoring by MSPCA is degradation of the slurry pumparound performance to 40% of the original value. This malfunction is simulated as a slow ramp that starts after 5 min and ends after 65 min. Other landmark events for this fault are in Table 1.

The ability of PCA and MSPCA to detect this malfunction is compared in Figure 17 based on the Q chart. The T^2 charts are not shown, since they are unable to detect the malfunction. For PCA, the squared prediction error continuously violates the 95% confidence limit after about 47 min of operation, and the 99% limit only after the malfunction is over and after the controllers take over. In contrast, for MSPCA the error violates the 95% limit continuously after 17 min, and the 99% limit after 24 min. Since the performance of adaptive resonance theory (ART) for detecting this fault is similar to that of PCA (Ali, 1997), monitoring by MSPCA is also likely to perform better than monitoring by ART.

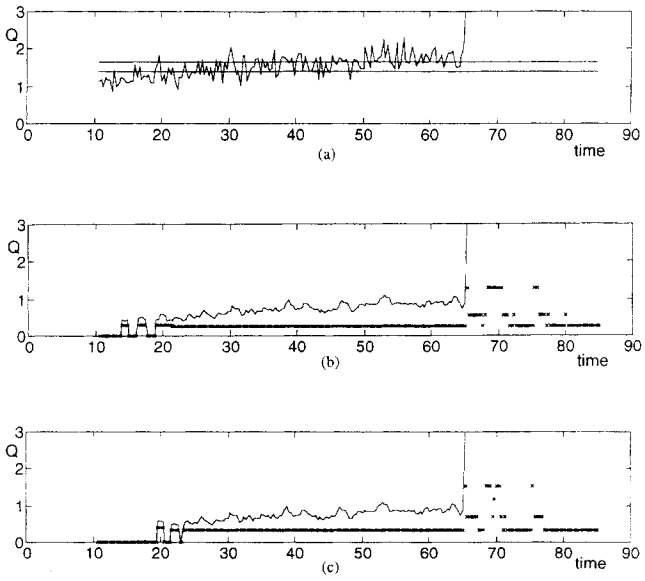

Figure 17. Q charts for slow ramp failure of slurry pumparound in FCCU.

(a) PCA with 95% and 99% confidence limits. (b) MSPCA for 95% confidence. (c) MSPCA for 99% confidence.

Conclusions

This article presents an approach for combining the attractive properties of PCA and wavelet analysis. The resulting MSPCA extracts relationships between the variables by PCA, and between the measurements by wavelet analysis. Consequently, MSPCA enables greater compression of the data matrix than either PCA or wavelet analysis alone. The MSPCA methodology determines separate PCA models at each scale to identify the scales where significant events occur. Models at these interesting scales are then fused to obtain a PCA model for all scales together. MSPCA transforms conventional single-scale PCA to a multiscale modeling method, which is better suited for modeling data containing contributions that change over time and frequency. Such multiscale data include nonwhite stochastic processes, and deterministic features localized in time and frequency.

Applying MSPCA to multivariate statistical process monitoring results in a method that automatically filters the scores and residuals at each scale for easiest detection of a deterministic change in the measurements. Since the popular control charts such as Shewhart, CUSUM, and EWMA of the scores and residuals only differ in scale at which they filter the data, monitoring by MSPCA can specialize to these charts, and automatically select the best among them and other filters for every measurement. The detection limits for each measurement also depend on the scales at which the change is detected, and adapt to the data. Since wavelets approximately decorrelate autocorrelated measurements, MSPCA can also detect changes in autocorrelated measurements without too many false alarms. Selection of the wavelet coefficients that violate the detection limits at each scale also extracts the features that are relevant to the abnormal operation. Thus, MSPCA is better at SPM than PCA, it integrates the tasks of feature extraction and process monitoring, and eliminates the need for prefiltering the data.

The MSPCA methodology may be extended in several different directions to further improve its performance. Robustness to outliers may be provided by combining multiscale median filtering with the wavelet decomposition (Bakshi et al., 1997; Nounou and Bakshi, 1998). The basis functions may be selected from a library such as wavelet packets to improve the separation of deterministic and stochastic features. Multidimensional wavelets may also be used for the decomposition. This article has focused on using MSPCA to detect deterministic changes. Since wavelets are also able to characterize and model stochastic processes, MSPCA may also be extended to detect changes in the stochastic behavior of the variables. MSPCA may also be used for replacing PCA in other tasks, such as gross-error detection, estimation of missing data, and fault diagnosis. The multiscale approach may

also be exploited for monitoring by methods other than PCA, such as adaptive resonance theory, to obtain similar improvement. Since most existing regression methods, including ordinary least-squares regression, partial least-squares regression, and artificial neural networks, are also single scale in nature, the insight provided by MSPCA may help in transforming these methods for multiscale modeling.

One of the main contributions of the field of wavelets is that it has provided a general framework for studying existing methods in a variety of fields, leading to new insight and often improving the existing methods (Sweldens, 1996). This article and other related work (Bakshi et al., 1997; Nounou and Bakshi, 1998; Top and Bakshi, 1998) show how wavelets can also provide a general framework for various existing approaches to univariate and multivariate SPM, such as Shewhart, CUSUM, and EWMA charts, and can result in improved process monitoring.

Acknowledgments

The authors thank Profs. Prem Goel and Xiaotong Shen of the Department of Statistics at Ohio State for extensive discussions on the statistical aspects of this work; the Abnormal Situation Management consortium, Mr. Zafar Ali, and Prof. Jim Davis for providing data for the FCCU case study; and the authors of WaveLab (http://playfair.stanford.edu/~wavelab) for providing the basic wavelet functions used in MSPCA.

Literature Cited

Ali, Z., "On-Line State Estimation in Intelligent Diagnostic Decision Support Systems for Large Scale Process Operations," MS Thesis, Ohio State Univ., Columbus (1997).

Bakshi, B. R., "Towards Integration of Measured Data-Dependent Process Operation Tasks Using a Time-Frequency Framework," *Proc. Amer. Control Conf.*, Seattle, p. 1260 (1995).

Bakshi, B. R., P. Bansal, and M. Nounou, "Multiscale Rectification of Random Errors without Process Models," *Comput. Chem. Eng.*, **21**, S1167 (1997).

Box, G. E. P., and G. C. Tiao, "A Canonical Analysis of Multiple Time Series," *Biometrika*, **64**(2), 355 (1977).

Briggs, W. L., *A Multigrid Tutorial*, SIAM, Philadelphia (1987).

Buckheit, J. B., and D. L. Donoho, "WaveLab and Reproducible Research," *Wavelets and Statistics*, A. Antoniadis and G. Oppenheim, eds., Springer-Verlag, New York (1995).

Cheng, G., and T. J. McAvoy, "Multi-Block Predictive Monitoring of Continuous Processes," *Proc. IFAC ADCHEM*, Banff, Canada (1997).

Chou, K. C., A. S. Willsky, and A. Benveniste, "Multiscale Recursive Estimation, Data Fusion and Regularization," *IEEE Trans. Automat. Contr.*, **AC-39**(3), 464 (1994).

Coifman, R., and N. Saito, "The Local Karhunen-Loeve Basis," *IEEE Time-Frequency Time-Scale Symp.*, Paris (1996).

Crosier, R. B., "Multivariate Generalizations of Cumulative Sum Quality Control Schemes," *Technometrics*, **30**, 291 (1988).

Daubechies, I., "Orthonormal Bases of Compactly Supported Wavelets," *Commun. Pure Appl. Math.*, **XLI**, 909 (1988).

Dayal, B. S., and J. F. MacGregor, "Recursive Exponentially Weighted PLS and its Applications to Adaptive Control and Prediction," *J. Process Control*, **7**, 169 (1997).

Dijkerman, R. W., and R. R. Majumdar, "Wavelet Representations of Stochastic Processes and Multiresolution Stochastic Models," *IEEE Trans. Signal. Process.*, **42**, 1640 (1994).

Dong, D., and T. J. McAvoy, "Nonlinear Principal Component Analysis—Based on Principal Curves and Neural Networks," *Comput. Chem. Eng.*, **20**, 65 (1996).

Donoho, D. L., I. M. Johnstone, G. Kerkyacharian, and D. Picard, "Wavelet Shrinkage: Asymptopia?" *J. R. Stat. Soc. B*, **57**, 301 (1995).

Dunia, R., S. J. Qin, T. F. Edgar, and T. J. McAvoy, "Identification of Faulty Sensors Using Principal Component Analysis," *AIChE J.*, **42**, 2797 (1996).

Hastie, T. J., and W. Stuetzle, "Principal Curves," *J. Amer. Stat. Assoc.*, **84**, 505 (1989).

Heinonen, P., and Y. Neuvo, "FIR-Median Hybrid Filters," *IEEE Trans. Acoust. Speech, Signal Processing*, **ASSP-35**, 6 (1987).

Jackson, J. E., "Principal Components and Factor Analysis: I. Principal Components," *J. Qual. Technology*, **12**(4), 201 (1980).

Jackson, J. E., *A User's Guide to Principal Components*, Wiley, New York (1991).

Kosanovich, K. A., and M. J. Piovoso, "PCA of Wavelet Transformed Process Data for Monitoring," *Intell. Data Anal.*, http://www.elsevier.com/locate/ida, 1, 2 (1997).

Kourti, T., P. Nomikos, and J. F. MacGregor, "Analysis, Monitoring and Fault Diagnosis of Batch Processes Using Multiblock and Multiway PLS," *J. Proc. Control*, **5**, 277 (1995).

Kramer, M. A., and R. S. H. Mah, "Model-Based Monitoring," *Proc. Int. Conf. on Foundations of Computer Aided Process Operations*, D. Rippin, J. Hale, J. Davis, eds. CACHE, Austin, TX (1994).

Kramer, M. A., "Nonlinear Principal Component Analysis Using Autoassociative Neural Networks," *AIChE J.*, **37**, 33 (1991).

Kresta, J., J. F. MacGregor, and T. E. Marlin, "Multivariate Statistical Monitoring of Process Operating Performance," *Can. J. Chem. Eng.*, **69**, 35 (1991).

Ku, W., R. H. Storer, and C. Georgakis, "Disturbance Detection and Isolation by Dynamic Principal Component Analysis," *Chem. Intell. Lab. Syst.*, **30**, 179 (1995).

Ljung, L., *System Identification: Theory for the User*, Prentice Hall, Englewood Cliffs, NJ (1987).

Lowry, C. A., W. H. Woodall, C. W. Champ, and S. E. Rigdon, "A Multivariate Exponentially Weighted Moving Average Control Chart," *Technometrics*, **34**, 46 (1992).

MacGregor, J. F., "Statistical Process Control of Multivariate Processes," *IFAC ADCHEM*, Kyoto, Japan (1994).

MacGregor, J. F., C. Jaeckle, C. Kiparissides, and M. Koutoudi, "Process Monitoring and Diagnosis by Multiblock PLS Methods," *AIChE J.*, **40**, 827 (1994).

Malinowski, E. R., *Factor Analysis in Chemistry*, Wiley, New York (1991).

Mallat, S. G., "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, **PMAI-11**, 674 (1989).

Mallat, S. G., "Zero-Crossings of a Wavelet Transform," *IEEE Trans. Inform. Theory*, **IT-37**(4), 1019 (1991).

Nason, G. P., and B. W. Silverman, "The Stationary Wavelet Transform and Some Statistical Applications," *Wavelets and Statistics*, A. Antoniadis and G. Oppenheim, eds., Springer-Verlag, New York (1995).

Nason, G. P., "Wavelet Shrinkage Using Cross-Validation," *J. R. Stat. Soc. B*, **58**(2), 463 (1996).

Nomikos, P., and J. F. MacGregor, "Monitoring Batch Processes Using Multiway Principal Component Analysis," *AIChE J.*, **40**(8), 1361 (1994).

Nounou, M. N., and B. R. Bakshi, "On-Line Multiscale Rectification of Random and Gross Errors without Process Models," Technical Report, Dept. of Chemical Engineering, Ohio State Univ. (1998).

Piovoso, M. J., K. A. Kosanovich, and R. K. Pearson, "Monitoring Process Performance in Real-Time," *Proc. ACC*, **3**, 2359 (1992).

Scranton, R., G. C. Runger, J. B. Keats, and D. C. Montgomery, "Efficient Shift Detection Using Multivariate Exponentially-Weighted Moving Average Control Charts and Principal Components," *Qual. Relat. Eng. Int.*, **12**, 165 (1996).

Stephanopoulos, G., M. Dyer, and O. Karsligil, "Multi-Scale Modeling Estimation and Control of Processing Systems," *Comput. Chem. Eng.*, **21**, S797 (1997).

Sweldens, W., "Wavelets: What Next?" *Proc. IEEE*, **84**, 680 (1996).

Tan, S., and M. L. Mavrovouniotis, "Reducing Data Dimensionality through Optimizing Neural Network Inputs," *AIChE J.*, **41**, 1471 (1995).

Tham, M. T., and A. Parr, "Succeed at On-Line Validation and Reconstruction of Data," *Chem. Eng. Prog.*, **90**, 46 (1994).

Tong, H., and C. M. Crowe, "Detection of Gross Errors in Data Reconciliation by Principal Component Analysis," *AIChE J.*, **41**, 1712 (1995).

Top, S., and B. R. Bakshi, "A General Framework for Univariate

Statistical Process Control Methods Using Wavelets," *Found. Comput. Aided Process. Oper.*, Snowbird, UT (July 1998).
Venkatasubramanian, V., and G. M. Stanley, "Integration of Process Monitoring, Diagnosis and Control: Issues and Emerging Trends," *Proc. Int. Conf. on Foundations of Computer Aided Process Operations*, D. Rippin, J. Hale, J. Davis, eds., CACHE Austin, TX (1994).
Wickerhauser, M. V., *Adaped Wavelet Analysis from Theory to Software*, Peters, Wellesley, MA (1994).
Wise, B. M., N. L. Ricker, D. F. Veltkamp, and B. R. Kowalski, "A Theoretical Basis for the Use of Principal Component Models for Monitoring Multivariate Processes," *Process Control and Quality*, **1**, 41 (1990).
Wold, S., "Exponentially Weighted Moving Principal Component Analysis and Projection to Latent Structures," *Chem. Intell. Lab. Sys.*, **23**, 149 (1994).
Wold, S., N. Kettaneh, and K. Tjessem, "Hierarchical Multiblock PLS and PC Models for Easier Model Interpretation and as an Alternative to Variable Selection," *J. Chemometrics*, **10**, 463 (1996).
Wornell, G. W., "A Karhunen-Loeve-Like Expansion for 1/f Processes via Wavelets," *IEEE Trans. Inform. Theory*, **IT-36**(4), 859 (1990).

Manuscript received Jan. 15, 1998, and revision received Apr. 24, 1998.