Multiscale SPC Using Wavelets: Theoretical Analysis and Properties

Hrishikesh B. Aradhye, Bhavik R. Bakshi, Ramon A. Strauss, and James F. Davis
Dept. of Chemical Engineering, The Ohio State University, Columbus, OH 43210

Most practical process data contain contributions at multiple scales in time and frequency, but most existing SPC methods are best for detecting events at only one scale. For example, Shewhart charts are best for detecting large, localized changes, while EWMA and CUSUM charts are best for detecting small changes at coarse scales. A multiscale approach for SPC, adaptable to the scale of relevant signal features and developed based on wavelet analysis, detects abnormal events at multiple scales as relatively large wavelet coefficients. Univariate and multivariate multiscale SPC (MSSPC) for detecting abnormal operation are theoretically analyzed, and their properties are compared with existing SPC methods based on their average run lengths. SPC methods are best for detecting features over a narrow range of scales. Their performance can deteriorate rapidly if abnormal features lie outside this limited range. Since in most industrial processes, the nature of abnormal features is not known a priori, MSSPC performs better on average due to its adaptability to the scale of the features and for monitoring autocorrelated measurements since dyadic wavelets decorrelate most stochastic processes. MSSPC with dyadic discretization is appropriate for SPC of highly autocorrelated or nonstationary stochastic processes. If normal measurements are uncorrelated or contain only mild autocorrelation, it is better to use MSSPC with integer or uniformly discretized wavelets. Many existing methods such as MA, EWMA, CUSUM, Shewhart, batch means charts, and their multivariate extensions are special cases of MSSPC.

Introduction

Statistical process control (SPC) has been an active area of research for many decades. A broad spectrum of methods have been developed, including methods for univariate SPC such as Shewhart, moving-average (MA), exponentially weighted moving-average (EWMA), and cumulative-sum (CUSUM) charts. Methods for multivariate SPC include multivariate extensions of univariate methods, and methods that monitor latent variables obtained by combining the measured variables with a lower dimension of space. Popular methods for reducing the dimensionality of the measured data include principal-component analysis (PCA) and partial least-square regression (PLS). Many extensions and applications of these have been developed (Kresta et al., 1991; Ku et al., 1995; MacGregor, 1994).

Correspondence concerning this article should be addressed to B. R. Bakshi.
Current addresses of: H. B. Aradhye, SRI International, Menlo Park, CA; R. A. Strauss, ExxonMobil, Fairfax, VA. J. F. Davis, University of California, Los Angeles, CA.

Most existing univariate and multivariate SPC methods operate at a fixed scale, and are best for detecting changes at a single scale. For example, Shewhart charts analyze the raw measurements at the scale of the sampling interval or the finest scale, and are best for detecting large, localized changes. In contrast, MA, EWMA, and CUSUM charts inherently filter the data, and, therefore, process measurements at a coarser scale. They are best for detecting small shifts or features at coarse scales. Tuning parameters such as window length or filter constant determine the scale at which the measurements are represented.

In contrast to the single-scale nature of SPC methods, data from most practical processes are inherently multiscale due to events occurring with different localizations in time, space, and frequency. A typical example of such data from a petrochemical process is shown in Figure 1. Figure 1a shows data during normal operation, while Figure 1b represents unusual operation due to a drier cooling event. In Figure 1b, the pro-

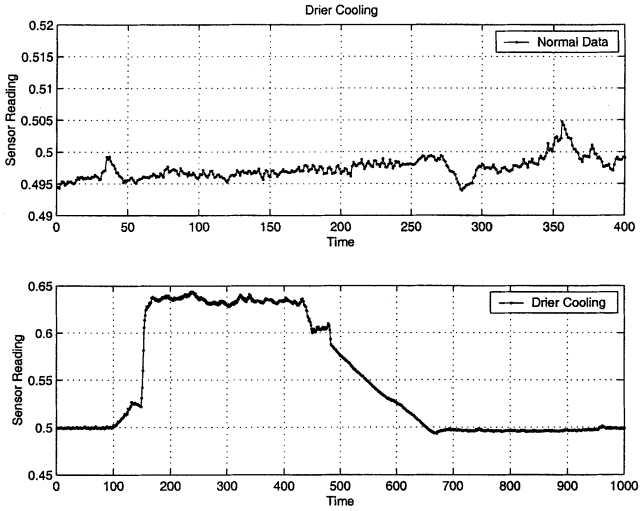

Figure 1. Data from a petrochemical process.

(a) Normal operation; (b) abnormal operation representing the drier cooling event.

cess change at approximately 150 time units is at a very fine scale and localized in time, but spans a wide range of frequencies. The steady portions of the signal are at coarse scales and span a wide temporal range. Finally, the change between 425 and 675 time units consists of a small sharp change followed by a short steady section and a slow ramp at an intermediate scale. Ideally, techniques for detecting changes at different scales, such as those shown in Figure 1b, should adapt automatically to the scale of the features. In response to this need, many heuristic or ad hoc techniques have been proposed for overcoming the single-scale nature of SPC charts. These include the Western Electric rules (Western Electric, 1956), useful for identifying patterns in data, and combined Shewhart and CUSUM charts (Lucas, 1982) for identifying large and small shifts. Other methods, such as CUSCORE charts (Box and Ramirez, 1992), may be specially designed to detect abnormal features if their character is known before they occur. Unfortunately, for most practical processes, such information about the nature of features representing abnormal operation is not known beforehand. The multiscale SPC method studied in this article can adapt to any type of signal change, and provides better average performance than existing methods for a range of changes.

Many existing SPC methods assume uncorrelated measurements, whereas, in practice, autocorrelated measurements are extremely common. A common approach for SPC of autocorrelated measurements is to decorrelate them by fitting a time-series model, and monitor the residual error. However, this approach is often not practical for industrial use, particularly for multivariate processes, due to their high dimensionality. Other univariate approaches for decorrelating autocorrelated measurements without time-series modeling include taking the batch means (Runger and Willemain, 1995) and finding the residuals between the measurements and their one-step ahead prediction by a moving center-line EWMA (MCEWMA) model (Mastrangelo and Montgomery, 1995). Batch means control charts often result in a significant delay in detecting large shifts, while MCEWMA is best only for integrated moving-average (IMA) stochastic processes. Consequently, neither approach is broadly applicable to different types of stochastic processes and process shifts, and they lack multivariate generalizations. For multivariate SPC, the dynamics are captured by a linear time-series model by methods such as dynamic PCA (DPCA) (Ku et al., 1995) or subspace identification (Negiz and Cinar, 1997). However, the autocorrelation is often still present in the measurements. Traditional T^2 and Q control charts are then applied in the space of the selected latent variables and their residuals. These approaches often work better than SPC by steady-state PCA or PLS for autocorrelated data.

In recent years, wavelets have been popular for analyzing multiscale or autocorrelated measurements due to their ability to compress multiscale features and approximately decorrelate many autocorrelated stochastic processes (Donoho et al., 1995; Miller and Willsky, 1995). Thus, wavelet coefficients provide compact information about a signal at different localizations in time and frequency. Also, the wavelet coefficients of many stochastic processes are approximately uncorrelated, since wavelets are approximate eigenfunctions of many mathematical operators (Beylkin et al., 1991). Wavelet-based multiscale methods have been developed for improved solution of many tasks including data compression, estimation, feature extraction, and filtering (Donoho et al., 1995; Mallat, 1989; Miller and Willsky, 1995; Nounou and Bakshi, 1999; Sadler and Swami, 1999).

The benefits of multiscale representation using wavelets also have been extended to change detection and process monitoring. Methods for Bayesian multiscale change detection require prior knowledge about the nature of the process change (Ogden and Lynch, 1999). Such methods can perform well if the necessary information is available, but are not popular in industrial monitoring applications due to a lack of the necessary information for the prior, and a lack of familiarity with Bayesian methods (Stoumbos et al., 2000). Another approach combines wavelets and multivariate SPC based on PCA (Bakshi, 1998). This approach, called multiscale SPC (MSSPC), consists of decomposing each measured variable to multiple scales by using a selected family of wavelet basis functions. The decomposition permits identification of signal features at various scales as relatively large coefficients in uncorrelated data. The characteristics of this multiscale approach have been studied and illustrated by many researchers, and some variations have also been suggested via examples, but without any statistically rigorous simulation or theoretical analysis (Teppola and Minkkinen, 2000; Rosen and Lennox, 2001; Yoon, 2001).

This article presents the theoretical analysis and properties of univariate and multivariate MSSPC proposed by Bakshi (1998). The performance of MSSPC and existing methods is compared based on the average run lengths (ARL) for detecting shifts of different sizes. The ARL is determined both empirically by Monte Carlo simulation, and theoretically for uncorrelated as well as stationary and nonstationary autocorrelated measurements. For autocorrelated univariate measurements, MSSPC is compared with residuals, weighted batch means (Runger and Willemain, 1995), and MCEWMA (Mastrangelo and Montgomery, 1995) charts. The performance of multivariate MSSPC using PCA is compared with steady-state and dynamic PCA. To permit a fair comparison

between popular existing methods and MSSPC, the ARL analysis presented in this article focuses mainly on detecting mean shifts. However, MSSPC is suitable for detecting a much broader variety of changes, as demonstrated by the industrial applications in this article and others (Bakshi, 1998; Aradhye et al., 2000, 2001; Kano et al., 2002). Furthermore, MSSPC subsumes existing methods, such as Shewhart, MA, EWMA, and CUSUM charts, depending on the nature of the selected wavelet. Thus, MSSPC can adapt the data filter and the detection limits according to the nature of the process change, and can specialize to existing methods if necessary.

Theoretical analysis and application to industrial data indicate that MSSPC is a good *general* method for SPC of processes containing features of different types and sizes and at different scales, from uncorrelated or autocorrelated measurements. These results indicate that if the size and shape of the features representing abnormal operation are known *a priori*, then one or more of the conventional SPC methods can be tailored to outperform MSSPC. For instance, if the abnormal deviations are known *a priori* to always be very large mean shifts and lack autocorrelation, it may be best to use the Shewhart chart instead of MSSPC. Similarly, if the abnormal deviations are known *a priori* to always be very small shifts and lack autocorrelation, the EWMA chart is most likely better. For detecting other types of features, special charts such as CUSCORE (Box and Ramirez, 1992) and spectral PCA can be designed. For many industrial applications, however, the exact nature of abnormal deviations (such as scale, shape, magnitude, autocorrelation) *cannot be anticipated a priori*. Furthermore, little or no annotated historical data may be available for many abnormal conditions. In such situations, the performance of existing SPC methods often deteriorates rapidly for changes outside their limited range of specialization. Consequently, it is necessary to have robust SPC methods that can do well on the average for detecting *any* type of change, even though such methods may not be the best for any single type of change. In light of these practical and industrially relevant facts, MSSPC is demonstrated to be a better and more general method for SPC of measurements with unknown and different types of changes.

The remainder of this article is organized as follows. A brief introduction to the relevant notation and properties of wavelets is provided in the next section. After that, the methodology of MSSPC and its relationship to existing SPC methods are described, followed by the ARL analysis of univariate MSSPC. The ARL analysis includes a comparison of the theoretically and empirically derived ARL for MSSPC, and a comparison with the ARLs of existing methods. ARL analysis for multivariate SPC and a comparison with SPC by PCA and dynamic PCA are presented next. Subsequently, the properties of MSSPC and a benefits over existing methods are delineated via case studies on data from industrial processes. Finally, a summary of this work and projected directions for future research are presented.

Wavelets

A family of wavelet basis functions can be represented as

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

where s and u represent the dilation and translation parameters, respectively, and $\psi(t)$ is the mother wavelet. Application to discrete signals requires discretization of the translation and dilation parameters. It is common to discretize these parameters dyadically as $s = 2^m$, $u = 2^m k$, $(m, k) \in Z^2$. These choices of s and u allow the wavelets to be orthonormal (Daubechies, 1988), but cause a *downsampling* (dyadic reduction or decimation) in the number of wavelet coefficients from one scale to the next. Any signal can be decomposed into its contributions from multiple scales as a weighted sum of dyadically discretized orthonormal basis functions

$$y(t) = \sum_{m=1}^{L}\sum_{k=1}^{N} d_{mk}\psi_{mk}(t) + \sum_{k=1}^{N} a_{Lk}\phi_{Lk}(t) \tag{2}$$

where $\boldsymbol{y}$ is the vector of measurements, d_{mk} represents the wavelet or detail signal coefficient at scale m and location k, and a_{Lk} represent the scaled signal or scaling function coefficient of $\phi_{Lk}(t)$ at the coarsest scale L and location k. The scaling function or father wavelet, ϕ_{mk}, captures the low-frequency content of the original signal that is not captured by wavelets at the corresponding or finer scales. An example of a wavelet decomposition is shown in Figure 2. In this case, the signal is decomposed to the coarsest scale represented by $L = 3$. This decomposition is based on the Haar wavelet and corresponding scaling function, which are defined as

$$\psi(t) = \begin{cases} 1, & 0 \le t < 1/2 \\ -1, & 1/2 \le t < 1 \\ 0, & \text{otherwise} \end{cases}$$

$$\phi(t) = \begin{cases} 1, & 0 \le t < 1 \\ 0, & \text{otherwise}. \end{cases} \tag{3}$$

Efficient methods have been developed for decomposing a signal on a family of wavelet basis functions based on convolution with the corresponding filters (Mallat, 1989). The filter coefficients for Haar scaling functions and wavelets are $\boldsymbol{H} = [1/\sqrt{2}\ \ 1/\sqrt{2}]$ and $\boldsymbol{G} = [1/\sqrt{2}\ \ -1/\sqrt{2}]$, respectively. Filters for other wavelets have also been derived (Daubechies, 1988). The vector of coefficients at scale, m, are computed recursively as, $\boldsymbol{d_m} = \boldsymbol{G}\boldsymbol{d_{m-1}}$ and $\boldsymbol{a_m} = \boldsymbol{H}\boldsymbol{a_{m-1}}$, with $\boldsymbol{a_0} = \boldsymbol{y}$. Thus, wavelet and scaling function coefficients at coarser scales may be computed by applying the $\boldsymbol{H}$ and $\boldsymbol{G}$ filters to the scaling function coefficients at a finer scale. These recursions can also be written as (Bakshi, 1998)

$$\boldsymbol{a_m} = \boldsymbol{H_m}\boldsymbol{y}$$
$$\boldsymbol{d_m} = \boldsymbol{G_m}\boldsymbol{y} \tag{4}$$

where $\boldsymbol{H_m} = \Pi_m \boldsymbol{H}$ and $\boldsymbol{G_m} = \Pi_{m-1}\boldsymbol{H}$.

Orthonormal wavelets have found application in many data analysis and modeling tasks due to their computational efficiency, ability to compress deterministic features in a small number of relatively large wavelet coefficients, and ability to approximately diagonalize a variety of mathematical operators. These properties are illustrated in Figures 2 and 3. Figure 2 shows the wavelet decomposition of the signal representing the drier cooling event in Figure 1b. The fine-scale

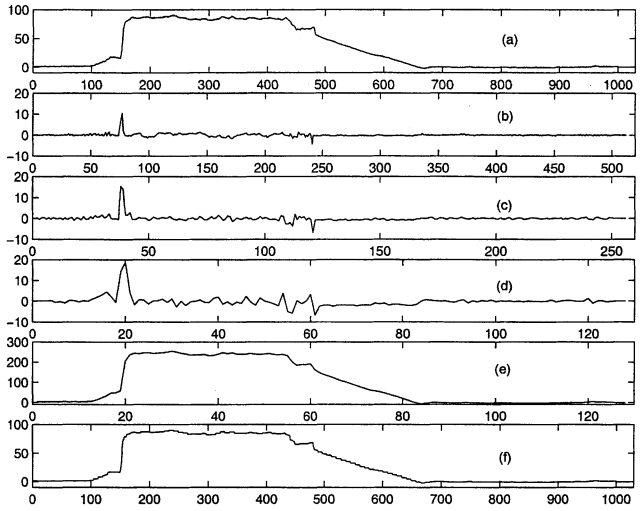

Figure 2. Extraction of deterministic changes by wavelet decomposition.

(a) Original signal, $m = 0$; (b) wavelet coefficients at $m = 1$; (c) wavelet coefficients at $m = 2$; (d) wavelet coefficients at $m = 3$; (e) last scaled signal coefficients at $m = 3$; (f) reconstructed signal after wavelet thresholding (Donoho et al., 1995).

features are captured by the large wavelet coefficients in Figures 2b, 2c, and 2d, and the remaining coarse scale features by the last scaled signal in Figure 2e. Eliminating small coefficients using thresholds and reconstructing the signal results in a signal with less random variation, as shown in Figure 2f. The properties of such a wavelet thresholding approach have been studied extensively (Donoho et al., 1995).

The ability of wavelets to decorrelate the data is depicted in Figure 3. The original signal, a simulated nonstationary stochastic process, and its autocorrelation function are shown in Figure 3a. The wavelet coefficients and corresponding autocorrelation functions shown in Figures 3b, 3c, and 3d indicate that the coefficients are approximately uncorrelated. The last scaled signal in Figure 3e contains residual correlation,

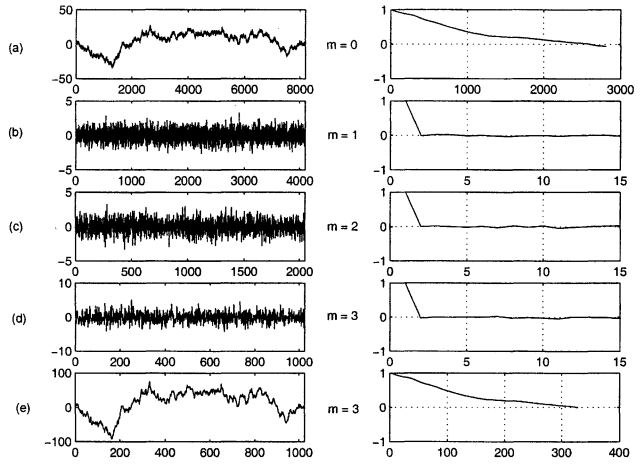

Figure 3. Approximate decorrelation due to dyadic wavelet transform.

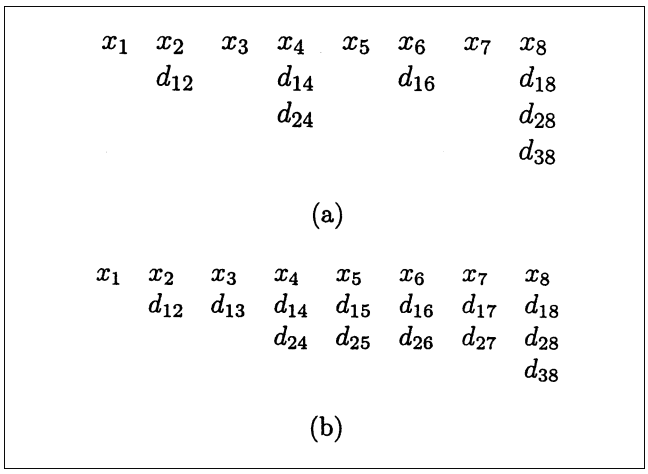

Figure 4. Wavelet decomposition with (a) dyadic, and (b) integer discretization of translation parameter.

which may be further reduced by decomposing to coarser scales. The variance of the wavelet coefficients at each scale varies according to the power spectrum of the original signal. Thus, the variance of the wavelet coefficients and the last scaled signal in Figures 3b, 3c, 3d, 3e, increases with decreasing frequency. These properties form the basis of the MSSPC method described in this article.

The wavelet decomposition in Figures 2 and 3 downsamples the coefficients at coarser scales due to the dyadically discretized translation parameter, as shown in Figure 4a. A disadvantage of using dyadically discretized wavelets is that measurements at nondyadic locations are decomposed after a time delay. This delay increases at coarser scales. Thus, as shown in Figure 4a, the fifth measurement, x_5, cannot be decomposed to obtain d_{16} until after after x_6 is obtained, and to d_{28} only after x_8 is obtained. This delay in decomposing the measurements causes a corresponding delay in interpreting the measurements and extracting features. For example, if a mean shift occurs at x_5, it will not be detected at scale $m = 1$ without a delay of one sample (x_5 to x_6), and at scale $m = 2$ without a delay of three samples (x_5 to x_8). This time delay also means that the dyadically discretized wavelet coefficients of translated signals are not related with each other in the same way as the translated signals (Mallat, 1989; Mallat and Zhong, 1992). Such a time delay and lack of translational invariance are unacceptable for many pattern-recognition, filtering, and feature-extraction tasks (Coifman and Donoho, 1995; Mallat and Zhong, 1992; Nounou and Bakshi, 1999).

A popular approach for eliminating this time delay is by using wavelets with a uniformly discretized translation parameter, $u = k$. This value of u results in wavelet decomposition with uniform or *integer* discretization or without downsampling, as shown in Figure 4b. The wavelet coefficients are no longer orthonormal to each other. Consequently, the decorrelation ability is lost, but the ability to compress changes in a small number of large wavelet coefficients is retained. This overcomplete wavelet representation is also referred to as the stationary wavelet transform (Nason and

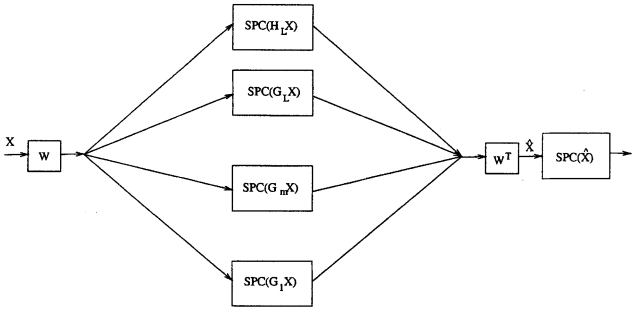

Figure 5. MSSPC methodology (Bakshi, 1998).

Silverman, 1995) or the translationally invariant wavelet transform (Coifman and Donoho, 1995). These wavelets are convenient for pattern recognition and for on-line multiscale methods, including the on-line MSSPC used in this article.

Multiscale Statistical Process Control

Methodology

The methodology for univariate and multivariate SPC is depicted in Figure 5 and Table 1. The measurements for each variable are decomposed on the selected family of wavelets. The resulting coefficients at each scale are monitored by the selected control chart. Thus, for univariate SPC, the coefficients at each scale are subjected to a Shewhart chart. For multivariate SPC, the coefficient matrix at each scale is subjected to the selected multivariate monitoring method such as PCA or PLS. The detection limits at each scale are determined from the coefficients at that scale of data representing normal operation. Applying separate control charts at each scale permits identification and selection of the scales or frequency bands that contain the features representing the current abnormal operation. The signal is reconstructed from the wavelet and scaling function coefficients larger than the limits at each scale, and the monitoring method is applied again to the reconstructed signal. If linear methods are used for monitoring, the control limits for the reconstructed signal can be determined from the limits or variance of the selected scales (Bakshi, 1998). This is because the wavelet decomposition decomposes the variance of the data to multiple scales as

$$X^TX = (H_LX)^T(H_LX) + (G_LX)^T(G_LX) + \cdots + (G_mX)^T(G_mX) + \cdots + (G_1X)^T(G_1X) \quad (5)$$

If a nonlinear method is used for SPC, the detection criterion for the reconstructed signal need not be a linear combination of the detection criteria at each scale. Consequently, it may be necessary to determine the detection criterion for the signal reconstructed from all combinations of scales from data representing normal operation, and stored for later use (Aradhye et al., 2000, 2001; Kano et al., 2002).

The state of the process is determined by applying the adjusted detection limit to the reconstructed signal. This last step of monitoring the reconstructed signal is essential for extracting the relevant features and for a quicker detection of process changes. Without reconstructing the signal, it is difficult to ascertain whether a coefficient outside its detection limits is due to an outlier, a sustained shift, return to normal operation, or a new shift. If the process only experiences mean shifts then, like Shewhart charts, MSSPC also can be used to detect a return to normal operation. In this special case, the reconstruction step can expedite detection of return-to-normal operation. However, MSSPC is meant to detect a much wider variety of changes, and its use for detecting a return-to-normal operation from any change is not statistically valid. Most SPC charts, including MSSPC, are hypothesis tests designed to reject the null hypothesis (the normal state) based on the observations. Strictly speaking, for a theoretically valid detection of the return-to-normal state, it is necessary to design control charts with each of the possible abnormal states as the null hypothesis, which can then detect the return to the normal process condition as a rejection of these null hypotheses. For most practical implementations, however, the *a priori* knowledge of all possible deviations and their exact nature is not available from the normal process conditions. Often, very few annotated historical data for abnormal situations are available. It is, therefore, a common industrial practice to consider the process to be "not abnormal" for points that do not violate the detection limit. Such points may be due to false negatives, return-to-normal operation, or a change to a new process state.

Table 1. Pseudocode for MSSPC

[Class] = MSSPC(X,covw,meanw,MaxDepth,numPC)
1. FOR $j = 1$: *SignalLength*,
   (a) Extract matrix, $X_2 = X(j - WinLength : j, :)$, where, $WinLength \le 2^{MaxDepth}$
   (b) Determine limits for normal data at each scale from *covw* and *meanw*
   (c) Decompose each column of X_2 with selected wavelet $= X_2w$
   (d) FOR $i = 1$: *MaxDepth*
      i. Transform X_2w by applying selected modeling method to coefficients at each scale
      ii. Select coefficients from transformed data that violate detection limits for transformed data
      iii. Determine coefficients from X_2w corresponding to limit violation in Step (d)(ii) $= X_2wSelect$
   (e) END FOR
   (f) Repeat Steps (d), (i)–(iii) for last scaled signal
   (g) Reconstruct $X_2wSelect$ using selected wavelet $= X_2wR$
   (h) Determine mean, covariance and detection limit for reconstructed signal from selected scales and *covw, meanw*
   (i) IF $X_2wR(j) > limit$
      i. *Class = Abnormal*
   (j) ELSE
      i. *Class = Normal*
   (k) END IF
2. END FOR

Adaptive nature of MSSPC

The MSSPC methodology and its features are illustrated by the simple univariate example shown in Figures 6 and 7. These figures use the Haar wavelet. The data representing normal operation are independent, identically distributed (IID) Gaussian with unit variance. The detection limits at each scale are determined from the normal data. In this case,

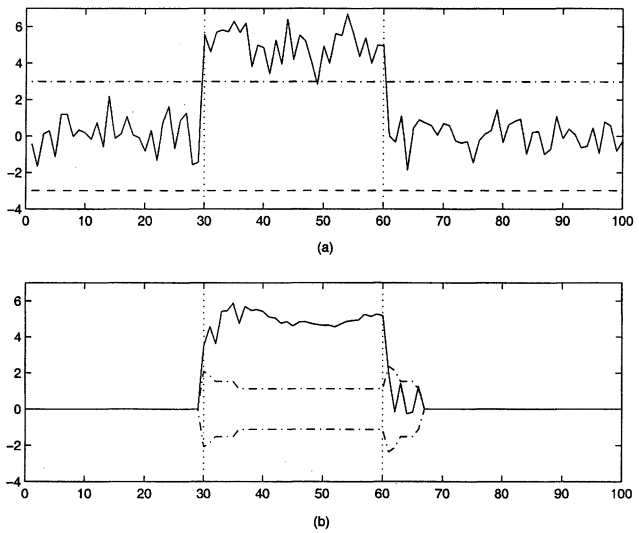

Figure 6. (a) Shewhart chart, (b) MSSPC.

Detailed figures for obtaining each point on the reconstructed signal and limits are shown in Figure 7.

the detection limits at each scale are equal for dyadically discretized wavelets, since the coefficients are also uncorrelated with approximately unit variance. For uniformly discretized wavelets, the limits need to be adjusted to account for the lack of downsampling and autocorrelation in the coefficients, as discussed in sections "Integer vs. Dyadic Discretization" and "Uncorrelated Gaussian Processes." Abnormal operation is indicated in Figure 6a by a mean shift of magnitude 5 between samples 30 and 60. The results for the Shewhart chart and MSSPC for this case are shown in Figures 6a and 6b, respectively. MSSPC results in a filtered signal that mainly contains the feature representing abnormal operation. Furthermore, the MSSPC detection limits change according to the nature of the signal and the scales at which the features are present.

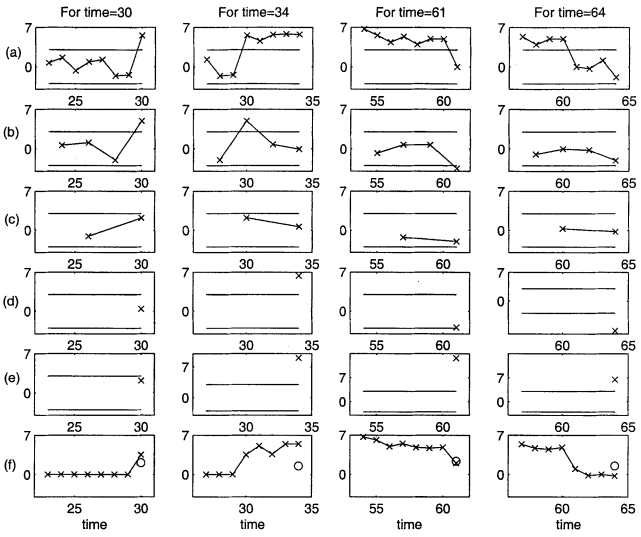

Figure 7. MSSPC methodology.

(a) Original signal used for monitoring measurement at selected time; (b), (c), (d) wavelet coefficients at $m = 1, 2, 3$; (e) coefficients for last scaled signal, $m = 3$; (f) reconstructed signal ($\times$) and detection limit ($\circ$) corresponding to selected time, and plotted in Figure 6b.

The detailed steps in MSSPC at four time steps are shown in Figure 7. When the shift first occurs at $t = 30$, it is detected only at the finest scale, as shown in Figure 7b for $t = 30$. At each point in time, scales at which the wavelet coefficients are outside their respective limits are *selected* for signal reconstruction. The detection limit for the reconstructed signal is calculated based on the variance of normal data (zero-mean IID Gaussian in this example) when reconstructed in the exact same set of selected scales. Since, at time $t = 30$, the shift is detected only at the finest scale, only the finest-scale coefficient is used to compute the reconstructed signal and detection limit for the reconstructed signal. Variance of normal data, subjected to wavelet decomposition and subsequent reconstruction with only the finest scale, is used to set the detection limit at the current time ($t = 30$). The reconstructed signal and the detection limit are shown in Figure 7f, indicating that MSSPC detects the shift. The values of the limit and reconstructed signal at $t = 30$ in Figure 7f form the corresponding points in Figure 6b. The plots for $t = 34$ depict the behavior of MSSPC after the shift has persisted for some time. Now, the shift is detected by the wavelet and scaling-function coefficients at the coarsest scale in Figures 7d and 7e. The detection limit at $t = 34$ is depicted by the circle in Figure 7f, and is determined based on the variance of the normal data at the selected scales ($m = 3$). As the shift persists, it is detected only by the last scaled signal. The behavior of MSSPC when the process returns to normal operation is shown by the plots at $t = 61$ and $t = 64$. At $t = 61$, the last scaled signal in Figure 7e continues to violate the detection limit, but the change is picked up at the finest scale, as shown in Figure 7b. The reconstructed signal and limit at $t = 61$ confirm that the process has changed and may have returned to normal operation. As the return to normal persists, the last scaled signal continues to be outside the limit, but wavelet coefficients at other scales help in keeping the reconstructed signal within its limits. Finally, the last scaled signal also stops violating the detection limit. These figures demonstrate the adaptive nature of MSSPC, the need for the final wavelet reconstruction, and final monitoring step to expedite detection of changes and adaptation, as discussed earlier in this section. The large library of filters available to MSSPC and their relationship with existing methods are discussed later in this section.

Integer vs. dyadic discretization

As discussed in the second section, the wavelet translation parameter may be discretized dyadically or uniformly. The type of discretization may be selected according to the nature of the measured data, and the objective of the SPC problem. Dyadic discretization permits the use of orthonormal wavelets and approximate decorrelation of autocorrelated measurements, but introduces a time delay in detecting changes due to a lag in the computation of the wavelet coefficients. Uniform discretization does not downsample the wavelet coefficients and permits truly on-line SPC, but the decorrelation ability and orthonormality are lost. These properties and the average run-length analysis in this article indicate that, in general, wavelets with uniform discretization are best for on-

line SPC of uncorrelated or moderately autocorrelated measurements, particularly for detecting large shifts. If the measurements are highly correlated or nonstationary, it is best to use dyadically discretized wavelets. For small shifts, the performance with dyadic or integer discretization is quite similar.

For *dyadic* discretization, if the measurements representing normal process operation are uncorrelated and Gaussian, the coefficients at each scale will also be uncorrelated and Gaussian with almost equal variance. If the normal data are autocorrelated, then the coefficients of an orthonormal wavelet decomposition at each scale will be approximately uncorrelated Gaussian, with the variance proportional to the power spectrum of the measurements in the corresponding frequency band. Due to these properties, the limits at each scale for MSSPC with dyadically discretized wavelets can be determined directly from the wavelet decomposition or power spectrum of normal data.

For *integer* discretization, the variance of the coefficients at each scale is still proportional to the power spectrum of the measured data, but the detection limits need to be adjusted to account for the increase in the number of coefficients and for the autocorrelation between them. Since the total number of wavelet coefficients is $O[N \times (L+1)]$ for a decomposition depth of L and $O(N)$ data points, the integer or uniformly discretized wavelet transform may appear to achieve data expansion rather than data compression. However, since the number of *significant* coefficients for detecting a change remains small, data compression is still achieved. As discussed in the previous section, this overcomplete or redundant wavelet decomposition is popular for many tasks (Lang et al., 1996; Malfait and Roose, 1997; Nason and Silverman, 1995; Pesquet et al., 1996).

The increase in the number of coefficients in the decomposed signal requires the limit at each scale to be larger than that for the original data to maintain equal false alarm rates. Different approaches may be used to determine the detection limits for the wavelet and scaling function coefficients and the reconstructed signal to obtain the same false-alarm rate. For example, the wavelet and scaling function coefficients can be subjected to smaller confidence limits to be more sensitive to changes in the signal, followed by a higher confidence limit for the reconstructed signal. Based on Bonferroni's inequality, the relationship between the desired confidence limit, C (such as 95%) for the reconstructed signal, and the confidence limit applied to the coefficients at each scale, C_L, for a decomposition of depth L can be written as (Bakshi, 1998)

$$C_L = 100 - \frac{1}{L+1}(100 - C) \tag{6}$$

This approach is used for the examples illustrating the ARL analysis of multivariate processes. The autocorrelation in the coefficients requires a further adjustment of the detection limits to obtain the desired in-control run lengths. This adjustment is similar to that used by SPC methods that filter the measurements such as MA, EWMA, and CUSUM (Montgomery, 1996). For the examples in this article, this adjusted detection limit is determined empirically, as discussed in the section, "Uncorrelated Gaussian Process."

Relation with existing methods

The characteristics of MSSPC can be appreciated by comparing the filters used by MSSPC with those used by existing SPC methods. The filters used by the simplest forms of Shewhart, EWMA, MA, and CUSUM charts, shown in Figure 8 (Hunter, 1986), indicate that these methods differ in the scale at which they filter the measurements, and the nature of the filter. Shewhart, MA, and CUSUM charts filter the data at increasingly coarse scales. The scale of the EWMA filter is determined by the value of the filter parameter, and that of the MA chart by the length of the filter window. These methods represent the measurements at a single fixed scale, since the size of the filter does not vary, making them best for detecting changes that occur at a single scale. Consequently, these methods are best for detecting changes that span only one scale or range of temporal and frequency localization, since the filter does not adapt to the scale of the signal features. Thus, Shewhart charts are best for detecting large and localized changes, while CUSUM charts are best for detecting smaller coarse changes. More advanced versions of these charts aim to improve their performance over a broader range of scales by using nonlinear filters. For example, combined Shewhart–CUSUM charts (Lucas, 1982) represent the measurements at two scales—finest and coarsest—while tabular and algorithmic CUSUM adjust the filter length based on the measurements.

The filters available to MSSPC are of many different shapes, and include the filters used by many existing methods. The filters available to MSSPC with Haar wavelets for a decomposition depth of $L = 3$ are shown in Figure 9. The filters are labeled by a 4-bit binary string to indicate the coefficients retained for reconstruction. For example, the combination 0100 corresponds to reconstruction after retaining only the wavelet coefficient at $m = 2$, and the combination 1101 corresponds to the selection of wavelet coefficients at scales $m = 1,2$ and the scaling coefficient at $m = 3$. Comparison with Figure 8 reveals that these MSSPC filters subsume the filters

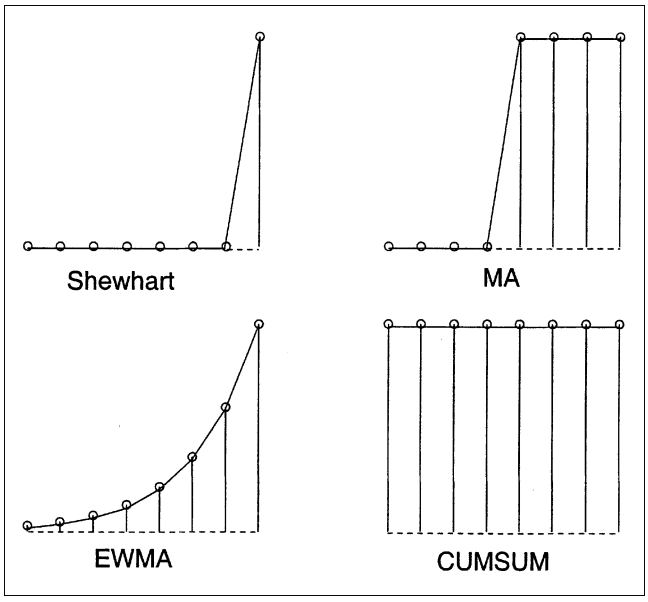

Figure 8. Commonly used SPC filters (Hunter, 1986).

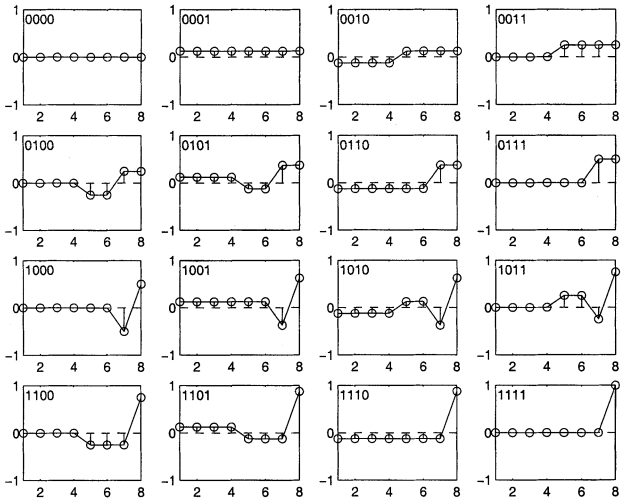

Figure 9. Filters available in MSSPC using Haar wavelets for $L = 3$.

used by Shewhart, MA, and CUSUM charts. For example, the combination 1111 corresponds to a Shewhart chart, the combination 0011 corresponds to an MA filter of window size 4, while the combination 0001 corresponds to a CUSUM window for a signal of length 8. The use of smoother boundary-corrected wavelets, such as Daubechies wavelets, approximates an EWMA filter as one of the possible cases (Nounou and Bakshi, 1999). Thus, MSSPC can automatically specialize to a range of existing methods, depending on the nature of the measurements and abnormal features. As described earlier, the example in Figure 7 shows that when the shift is first detected at scale $t = 30$, only the finest scale was chosen for reconstruction. This selection of scales corresponds to the filter represented by the binary string 1,000 in Figure 9. The scales selected at $t = 34$ correspond to the filter 0011, which is a mean filter of length 4, and to 1,001 at $t = 61$. Thus, the filter selected in MSSPC adapts to the features in the developing signal.

MSSPC with dyadic discretization is also related to the approach of decorrelating the measurements by taking their weighted or unweighted batch means (Runger and Willemain, 1995). Batch-means control charts select the window size according to the nature of the autocorrelation, thus requiring a longer window with increasing correlation. Since the windows do not overlap, this approach cannot detect a shift sooner than the window length. Unweighted batch means with a window of dyadic length is equivalent to the last scaled signal from a Haar decomposition at a depth equal to the base-two logarithm of the batch means window length. Thus, MSSPC using Haar wavelets can subsume batch means of dyadic lengths. Unlike batch means, MSSPC also has the wavelet coefficients at finer scales available, permitting a quicker detection of changes than a batch-means control chart, as illustrated in the next section.

Average Run Length Analysis of Univariate MSSPC

This section compares the performance of univariate MSSPC with existing univariate SPC methods for detecting a mean shift in uncorrelated and autocorrelated measurements. The ARL is the average number of samples required to detect a shift, and is determined both theoretically and by Monte Carlo simulation. The theoretical derivation of ARL for wavelets with integer and dyadic discretization is presented in the Appendix. When the magnitude of the shift is zero, the corresponding ARL value indicates the probability of false alarms, and is referred to as the in-control run length. For the same in-control run length, it is desirable to have the lowest possible ARL values for nonzero mean shifts or other abnormal events. This approach provides a standard way of comparing the relative performance of different SPC techniques. When plotted against the magnitude of the shift, the ARL curve is expected to be nonincreasing and typically converges to the location of the mean shift, as the magnitude of shift tends to infinity. For the examples in this article, the mean shift is located at the first measurement. The Monte Carlo simulation results are generated from at least one thousand realizations.

Uncorrelated Gaussian process

In this analysis, the measurements are considered to be IID Gaussian with unit variance

$$x(t) \sim N(0, 1) \tag{7}$$

The performance of Shewhart and MA charts is compared with MSSPC with dyadic or integer discretization. In each case, the parameters are selected to maintain approximately equal in-control run lengths of 370 samples. MA charts are selected for comparison with MSSPC using Haar wavelets, since both methods use filters of the same shape.

For *dyadically* discretized wavelets, the ARL curves obtained from theory and simulation for different depths of decomposition are shown in Figure 10. The ARL for a Shewhart chart and a moving average chart with a window of size 16 are also shown in Figure 10 for comparison. The theoretical run lengths are determined based on the assumption that

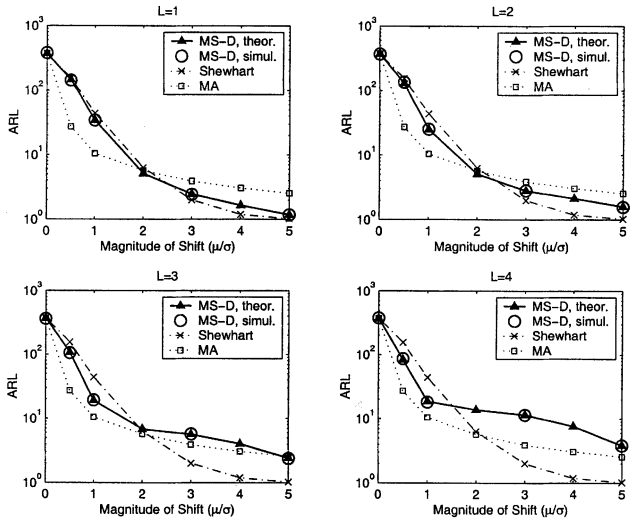

Figure 10. ARL of MSSPC-dyadic simulated, MSSPC-dyadic theoretical, Shewhart, and MA charts applied to univariate IID Gaussian process for different depths of decomposition.

the downsampled coefficients are uncorrelated and Gaussian with variance equal to that of the original measurements. The close match between the theoretical and simulated curves confirms the validity of the approach. Detailed derivation of the theoretical approach is provided in Appendix A1. With increasing depth of the wavelet decomposition, the ability of MSSPC dyadic to detect large shifts deteriorates due to the increasing time delay in obtaining the wavelet and scaling function coefficients at coarser scales, as discussed in the previous section and illustrated in Figure 4a. The ability to detect small shifts improves at greater depths due to a larger signal-to-noise ratio caused by a greater separation between the stochastic variation and deterministic mean shift at coarser scales. Depending on the selected depth of decomposition, the performance of MSSPC tends to be better than that of Shewhart charts for small shifts, and better than MA charts for large shifts.

The ARL plots for *integer* discretization shown in Figure 11 indicate better performance than that with dyadic discretization for detecting large shifts. The improvement is particularly significant for larger depths of decomposition, since the time delay caused by dyadic discretization causes more harm in detecting larger shifts. The ARL curves for a Shewhart chart and MA chart with a window of 16 samples are also plotted in Figure 11. The theoretically computed ARL for MSSPC with integer discretization is shown in Figure 12 for a depth of $L = 1$, and indicates a good match with the empirical values. The theoretical analysis of Appendix A1 cannot be accurately applied to MSSPC with integer discretization, since the assumption of uncorrelated coefficients is violated. However, this correlation can be modeled as a Markov chain, and a theoretical method for ARL can be derived as shown in Appendix A2. This approach uses numerical integration to compute the probability of detection in terms of the magnitude of the shift and the number of time steps passed since the introduction of the shift. These probabilities are then used to derive the value of the average run length. Detailed derivation of ARL values for a two-scale wavelet decomposition with Haar wavelets for this signal is provided in Appendix A2. The procedure is general and can be extended to multiple scales and other wavelets. The computation cost for numerical integration, however, increases exponentially with the depth of wavelet decomposition.

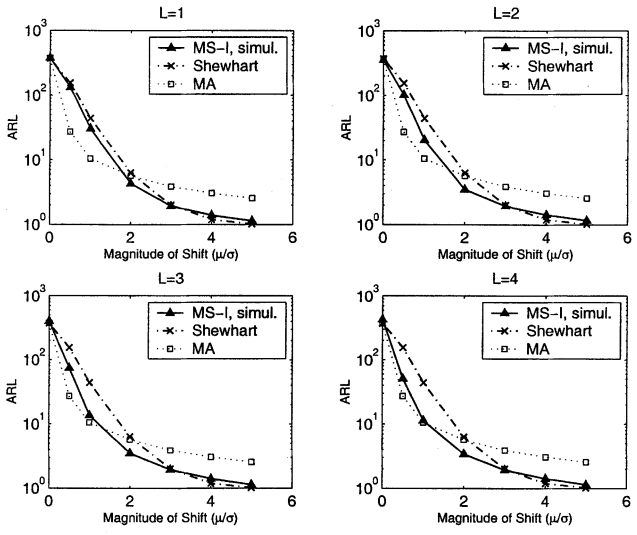

Figure 11. ARL of MSSPC-integer simulated, Shewhart, and MA charts applied to univariate IID Gaussian process for different depths of decomposition.

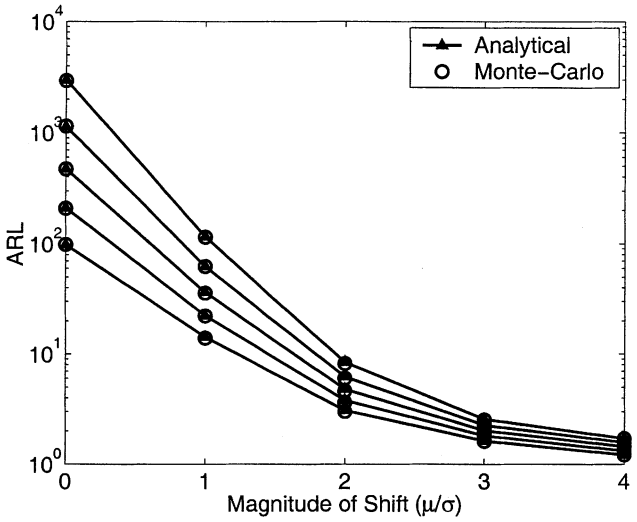

Figure 12. Theoretical vs. simulated ARL values for different detection thresholds with MSSPC-integer for $L = 1$ applied to IID Gaussian process.

Since the number of coefficients from the uniformly discretized wavelet decomposition is more than the number of samples in the original signal, the confidence limit at each scale is increased by using Eq. 6. The lack of downsampling also results in autocorrelated coefficients at each scale, requiring further adjustment in the limits to obtain the desired in-control run lengths. The adjusted limits are determined by simulating the run length for different limits at each scale. The limit corresponding to the desired run length is then determined by interpolation between the simulated values. The resulting limits at each scale for different decomposition depths are shown in Figure 13. These limits provide an in-control run length of 370 for the final reconstructed signal.

Comparison of the ARL curves for MSSPC-dyadic, MSSPC-integer, MA, and Shewhart charts in Figures 10 and 11 indicates that no single method performs best for all shift sizes. Thus, if the objective of SPC is to detect only small shifts, it is best to use an MA control chart. Similarly, if the objective is to detect only large shifts, it is best to use a Shewhart chart. Other methods, such as CUSCORE charts (Box and Ramirez, 1992), can be tailored to detect specific changes whose nature is known beforehand. We note again that in most practical problems, it is impossible to predict the type or extent of change due to abnormal operation, hence, our interest in MSSPC. For such problems, MSSPC is more appropriate due to its ability to adapt for detecting a wide variety of shifts and signal features.

The capacity of MSSPC to adjust to a given situation is illustrated in Figure 14, which shows the mean-square error

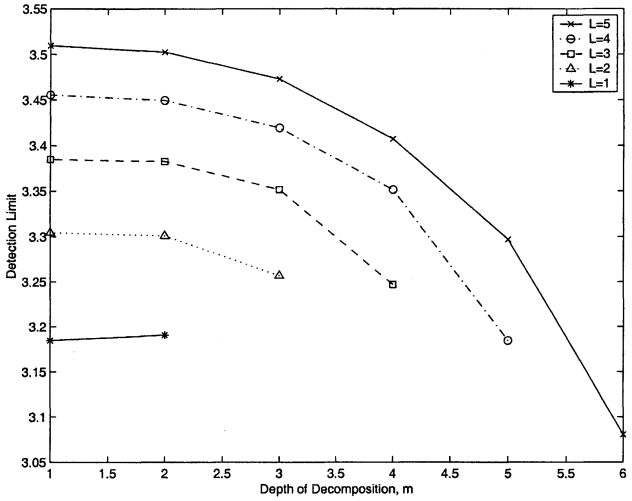

Figure 13. Adjusted detection limits to obtain confidence, C_L, computed by Eq. 6, for different decomposition depths.

These limits approximate in-control run length of 370 for MSSPC-integer of IID Gaussian data. Limit at largest m for each curve is for the last scaled signal.

(MSE) for classifying a mean shift of different sizes. The shift persists for 50 samples, and is surrounded on both sides by 50 samples each of normal data, as shown in Figure 14a. Similar to the ARL analysis, the MSE analysis from Figure 14b shows the MSSPC approach to be robust over a wide range of shift magnitudes. Average MSE values for all shifts in Figure 14 are the smallest for MSSPC integer over the range of mean shifts considered.

Stationary autocorrelated process

Many methods for SPC of autocorrelated measurements have been devised (Montgomery, 1996; Wardell et al., 1994). If a time-series model of the measurements is available, it can be used to decorrelate the measurements. The residuals will be approximately uncorrelated, and may be monitored by existing SPC methods (Harris and Ross, 1991). Since time-series models may not be readily available and are often not practical to use, other methods that do not require an explicit time-series model of the measurements have been developed for decorrelating the measurements. These methods include batch-means control charts (Runger and Willemain, 1995) and MCEWMA (Mastrangelo and Montgomery, 1995). Batch-means control charts decorrelate the data by taking the average of the measurements in nonoverlapping windows. The window size is selected such that the means in each window are uncorrelated. Thus, more correlation will require a longer window size. Weighted batch means (WBM) determine the weights based on knowledge of the autocorrelation, while unweighted batch means use equal weights regardless of the type of correlation. Significant limitations of this approach are that it cannot detect a shift sooner than the window length, and the window size may have to be determined empirically. As given by Eq. 4, the wavelet decomposition approach also takes a weighted mean of the measurements. However, MSSPC does better than the batch-means approach since MSSPC filters the data over multiple windows of dyadic length, while batch means uses a filter of a fixed length. This multiscale nature of MSSPC allows it to detect shifts of different sizes more effectively than the batch-means approach. MCEWMA fits an EWMA to the measurements to minimize the one-step ahead prediction error, and is ideal for decorrelating IMA stochastic processes. It has also been applied to other types of autocorrelated processes (Mastrangelo and Montgomery, 1995). As illustrated by the examples in this section, MCEWMA does not perform as well as MSSPC for other types of stochastic processes.

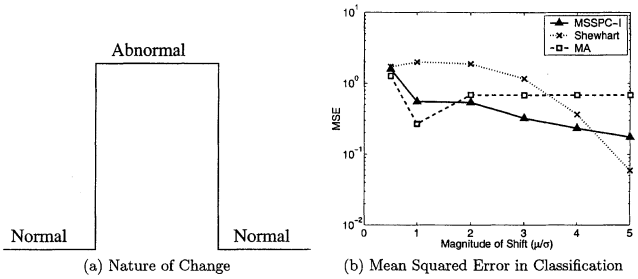

Figure 14. Mean-square error for classification by MSSPC-integer, Shewhart, and MA charts for IID Gaussian process with mean shift.

Total MSE for all shifts is 7.1722×10^{-3} for Shewhart, 4.2808×10^{-3} for MA, and 3.4099×10^{-3} for MSSPC-integer.

The ARLs for the following AR(1) process

$$x(t) = 0.5x(t-1) + \epsilon(t) \tag{8}$$

for different depths are shown in Figure 15 based on Monte Carlo simulation and the derivation presented in Appendix A1. The x-axis plots the ratio of the magnitude of the mean shift, μ, to the standard deviation, σ, of the noise, $\epsilon(t)$. The match between the analytical and Monte Carlo results is not as good as that for MSSPC of uncorrelated data shown in Figure 10. This is because the assumption of uncorrelated coefficients at each scale is not as valid as for uncorrelated measurements. The match between theory and simulation is likely to be better for smoother wavelets due to their better ability to decorrelate the measurements. Furthermore, the match between theory and experiment deteriorates with increasing depth, since the assumption of uncorrelated coefficients is less accurate at coarser scales. Figure 15 also shows the ARL of SPC residuals and weighted batch means charts (Runger and Willemain, 1995). These results indicate that WBM does well only for small shifts, while residual charts work well only for very large shifts. On the average, MSSPC with dyadic discretization works quite well, particularly for small L.

Figure 16 depicts the ARL for an AR(1) process given by

$$x(t) = 0.9x(t-1) + \epsilon(t) \tag{9}$$

The high degree of autocorrelation in this stochastic process makes it more difficult to detect shifts. This figure compares the ARL of MSSPC-dyadic with that of WBM, and

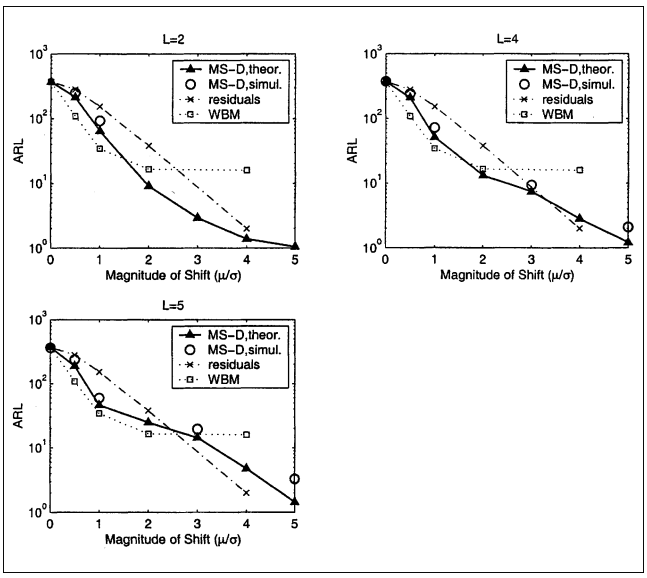

Figure 15. ARL curves for MSSPC-dyadic simulated, MSSPC-dyadic theoretical, residuals, and weighted batch means charts for AR(1) process given by Eq. 8.

Please note that σ refers to the standard deviation of the white generating noise $\epsilon(t)$.

MCEWMA control charts. WBM does well for detecting small shifts, but deteriorates quickly for larger shifts. MCEWMA does better than WBM only for the largest shift. MSSPC-dyadic does worse than WBM for small shifts, but does significantly better than the other two approaches for larger shift sizes. The time delay due to downsampling is made up to a certain extent by the easier detection of shifts in uncorrelated coefficients.

Nonstationary process

Nonstationary stochastic processes present special challenges for SPC, since their mean tends to change over time. The ARL performance of MSSPC and MCEWMA are compared in Figure 17. In this case, the stochastic process is IMA(1,1), given by

$$x(t) = x(t-1) + \epsilon(t) - 0.5\epsilon(t-1) \qquad (10)$$

which can be modeled optimally by EWMA. Using wavelets without downsampling is not feasible for SPC of such nonstationary measurements, since the high autocorrelation in the nondownsampled wavelet coefficients increases the rate of false alarms for the same fault-detection ability.

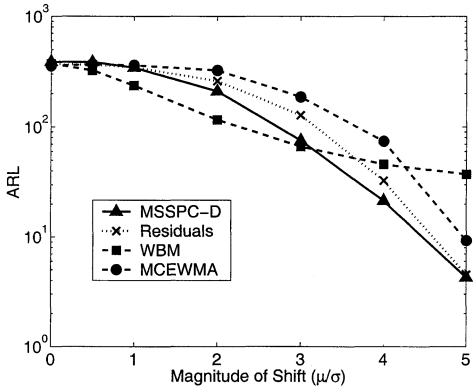

Figure 16. ARL curves for MSSPC-dyadic, WBM, and MCEWMA charts for an AR(1) process given by Eq. 9.

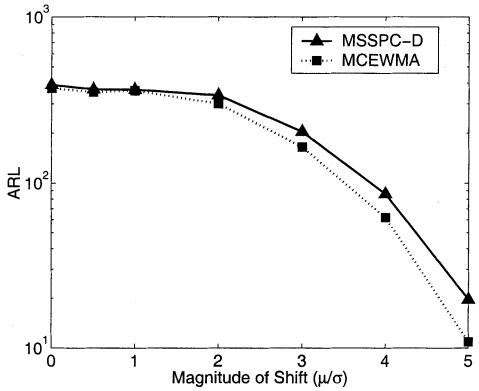

Figure 17. ARL of MSSPC-dyadic and MCEWMA for an IMA(1,1) process given by Eq. 10.

Figure 17 shows that for small shifts, the performance of MCEWMA and MSSPC-dyadic are equivalent, but MCEWMA performs better for detecting large shifts. The superiority of MCEWMA for this type of stochastic process is due to it being the optimal approach for decorrelating an IMA(1,1) time series. A mean shift in an IMA(1,1) process appears as a spike of very small duration in the decorrelated residuals (Wiel, 1996). The MSSPC-dyadic approach may easily miss this spike, even when it is large, due to the downsampling of the wavelet coefficients. Furthermore, the mean shift cannot be detected in the coefficients of the last scaled signal due to the residual autocorrelation and extremely large detection limits. The extremely localized nature of the spike indicates that a Shewhart chart on uncorrelated data should work best, which is essentially the approach used by MCEWMA. ARL for other approaches such as CUSUM, EWMA, and Shewhart charts on residuals are presented in Wiel (1996). These are not compared in Figure 17, since MCEWMA is the best approach for the IMA(1,1) process.

Average Run-Length Analysis of Multivariate MSSPC

The general framework of MSSPC shown in Figure 5 can be used to transform any existing multivariate SPC method into a multiscale approach. This section focuses on multivariate SPC using PCA and dynamic PCA (DPCA). PCA has been popular for process monitoring, since it allows extension of the principles of univariate statistical process monitoring to multivariate processes by capturing a large number of related measured variables in a small number of uncorrelated principal-component scores (Jackson, 1990; Kresta et al., 1991). The T^2 plot monitors the space of selected principal components,

whereas the Q plot monitors the residual space. The detection limits are based on the assumption of IID Gaussian data, or are found empirically.

Multivariate monitoring by MSPCA operates on the coefficient matrix at each scale using conventional PCA (Bakshi, 1998). Each column of the data matrix is decomposed by the selected wavelet, and T^2 and Q plots are developed for the coefficients at each scale. Since the wavelet transform is a linear operation, MSPCA preserves the modeling qualities of PCA and does not affect the eigenvectors or eigenvalues of the original matrix. The covariance of the data matrix is decomposed according to Eq. 5.

Many extensions and variations of PCA have been developed for dealing with various practical situations. A common approach for monitoring autocorrelated measurements is by DPCA (Ku et al., 1995; MacGregor, 1994). DPCA augments the data matrix by lagged variables to implicitly find the time-series model between the lagged and other variables. However, the autocorrelation within each variable still remains. Multivariate EWMA has also been applied to the selected principal-component scores to benefit from the dimensionality reduction and to improve the detection of small shifts (Scranton et al., 1996). A rigorous ARL analysis of PCA and its many variations is usually not available. This section studies the properties of PCA, DPCA, and MSPCA based on simulated uncorrelated and autocorrelated data.

Uncorrelated measurements

This example considers the linear uncorrelated multivariate process modeled as (Bakshi, 1998)

$$x_1(t) \sim N(0, 1)$$

$$x_2(t) \sim N(0, 1)$$

$$x_3(t) = \frac{x_1(t) + x_2(t)}{\sqrt{2}}$$

$$x_4(t) = \frac{x_1(t) - x_2(t)}{\sqrt{2}}$$

$$\boldsymbol{Y}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \\ x_4(t) \end{bmatrix} + 0.2\epsilon(t) \qquad (11)$$

The data matrix, $\boldsymbol{Y}$, for normal operation is generated as in Eq. 11, while that for abnormal operation is generated by adding a mean shift of various sizes to $x_i(t)$, $i = 1, \ldots, 4$. Both, PCA and MSPCA select two principal components for the T^2 plots. A run is terminated if there is a violation of the limits in either of the T^2 or Q plots. The ARL curves for both methods shown in Figure 18 indicate that the behavior of PCA and MSPCA-integer is analogous to that of the Shewhart chart and MSSPC-integer shown in Figure 11. Similar to the univariate uncorrelated case, MSPCA-integer shows improvement over PCA for this multivariate process, particularly in detecting small shifts.

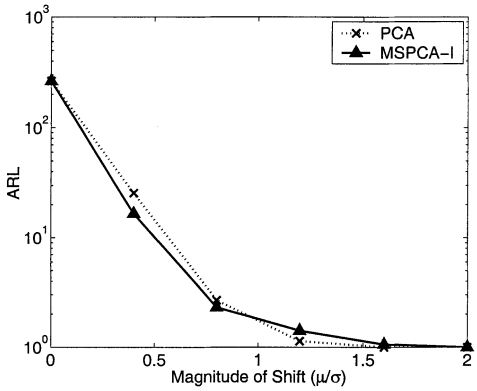

Figure 18. ARL of PCA and MSPCA-integer for a multivariate uncorrelated process given by Eqs. 11.

Autocorrelated measurements

This section presents ARL analysis of the following multivariate time-series model (Ku et al., 1995)

$$z(t) = \begin{bmatrix} 0.118 & -0.191 \\ 0.847 & 0.264 \end{bmatrix} z(t-1) + \begin{bmatrix} 1 & 2 \\ 3 & -4 \end{bmatrix} u(t-1)$$

$$y(t) = z(t) + v(t)$$

$$u(t) = \begin{bmatrix} 0.811 & -0.226 \\ 0.477 & 0.415 \end{bmatrix} u(t-1) + \begin{bmatrix} 0.193 & 0.689 \\ -0.320 & -0.749 \end{bmatrix} w(t-1) \qquad (12)$$

The data matrix for steady-state PCA is $[u(t)\ y(t)]$, while that for dynamic PCA is $[u(t)u(t-1)y(t-1)]$. Following Ku et al., two principal components are selected for steady-state PCA, and seven for DPCA. Data representing abnormal operation are generated by introducing a shift in u. The ARL analysis for this process using MSPCA-dyadic is shown in Figure 19 and using MSPCA-integer in Figure 20. Both figures include ARL plots for PCA and DPCA. For all methods, a run is

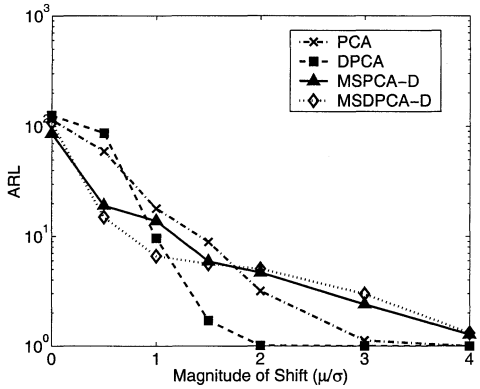

Figure 19. ARL of PCA, DPCA, MSPCA-dyadic, and MSDPCA-dyadic for a multivariate correlated time series given by Eq. 12.

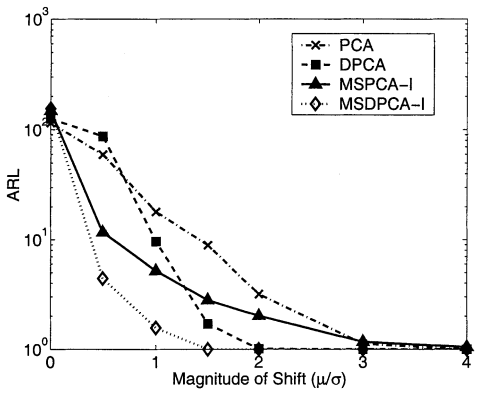

Figure 20. ARL of PCA, DPCA, MSPCA-integer, and MSDPCA-integer for a multivariate correlated time series given by Eq. 12.

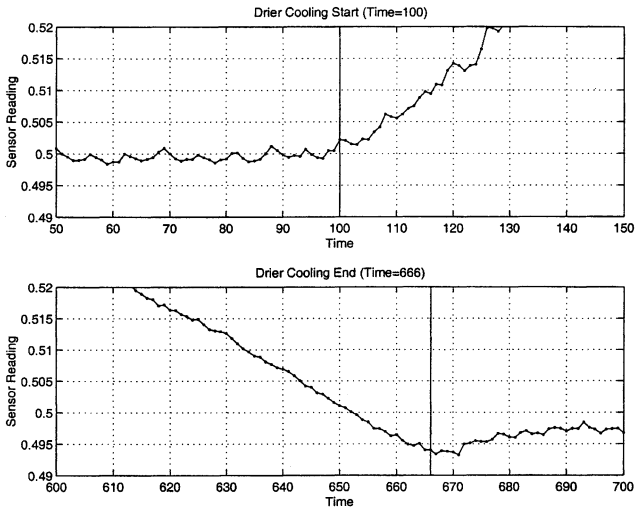

Figure 21. Onset and end of the drier cooling event.

terminated when either the T^2 or Q plot violates its detection limit.

These figures indicate that DPCA does perform better than PCA, particularly for detecting large shifts in the mean. Both, MSPCA and MSDPCA with dyadic discretization do well for detecting small shifts, but not as well as PCA or DPCA for detecting large shifts. Their performance for detecting large shifts deteriorates due to the delay introduced by downsampling. These results are comparable to those for univariate MSSPC by Shewhart chart and MSSPC-dyadic shown in Figure 10. MSPCA and MSDPCA with integer discretization do better for small shifts than their single-scale counterparts. In fact, MSDPCA integer shows the best performance for *all* shift sizes, since it benefits from extraction of the dynamic model by DPCA, and quicker detection by the multiscale approach.

Industrial Case Studies

This section demonstrates the application and performance of the SPC methods studied in this article to univariate and multivariate data obtained from a petrochemical process. The data are provided by the Abnormal Situation Management Consortium. Since it is not possible to perform ARL analysis on industrial data, the performance of various methods is compared by plotting the empirical receiver–operating characteristic (ROC) curves for each example. This is a plot of the true-positive rate (TPR) vs. the false-positive rate (FPR), and is commonly used to compare the performance of classification methods. For a given set of data, the TPR and FPR may be obtained empirically by applying each method to the data for different sets of thresholds and other parameters. As is common in industrial practice, TPR is the fraction of points for which the detection limits are violated in the region where abnormal operation is known to be present. FPR is the fraction of points for which the limits are violated in the region where normal operation is known to be present. For each example, the region of abnormal operation is determined via operator input.

Univariate processes

The univariate measurements analyzed in this section represent malfunctions in the charge gas drier unit and due to oil accumulation. Both events exhibit very small levels of noise relative to the magnitudes of the events. This facilitates localizing the onset and end times for faulty behavior by operator annotation and manual inspection of data. Unlike the previous examples of simulated data containing step changes only, this industrial data set contains other types of changes in the measurements. As illustrated by these examples, since practical data usually contain changes of different sizes and shapes, MSSPC is ideally suited for detecting abnormal operation from such data.

Charge Gas Drier Cooling. Measurements under normal and drier cooling conditions are shown in Figure 1. Since the normal data have much less variation than the abnormal measurements, it is possible for plant personnel to identify the onset and end of the drier cooling event at 100 and 666 time units, respectively, as shown in Figure 21. The thresholds for the Shewhart, MA, and MSSPC charts were determined empirically to obtain approximately the same rate of false alarms of 1% on the normal data. The large sudden change at the start of the abnormal operation is ideally suited to detection by a Shewhart chart. As shown in Figure 22, a Shewhart chart performs slightly better than MSSPC, which does better than MA for identifying the starting point of the fault. In contrast, it is more difficult for the Shewhart chart to detect the slow ramp near the end of the abnormal event. As shown in Figure 22, MSSPC is the best method for detecting abnormal behavior during the slow ramp event, with the Shewhart chart providing the worst performance. Since the change is slow and occurs at a coarse scale, it is difficult to detect only from the magnitude of the measurements, but is easier to detect based on the slope of the signal. MSSPC is particularly good for detecting such changes, since the wavelet coefficients correspond to the local slope of the measurements, and capture the behavior at multiple scales. Consequently, MSSPC is the only method that detects abnormal operation between samples at 655 and 666, as shown in Figure 22.

The overall performance of Shewhart, MA, and MSSPC over the entire signal are compared by plotting the ROC curve for each method, as shown in Figure 23. The points represent

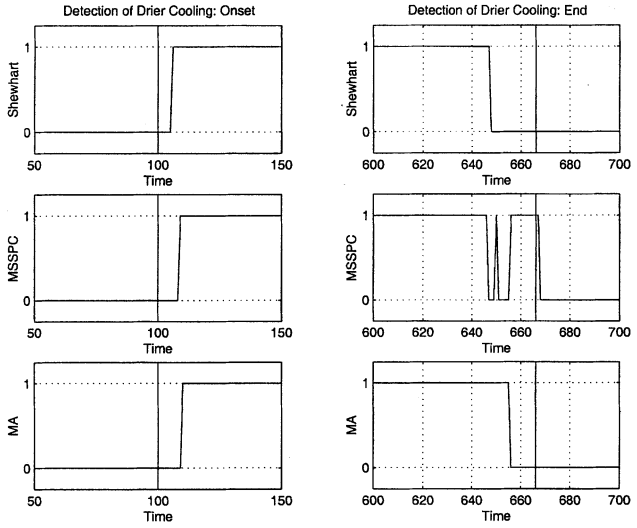

Figure 22. Detection of drier cooling event by Shewhart, MSSP-integer, and MA charts.

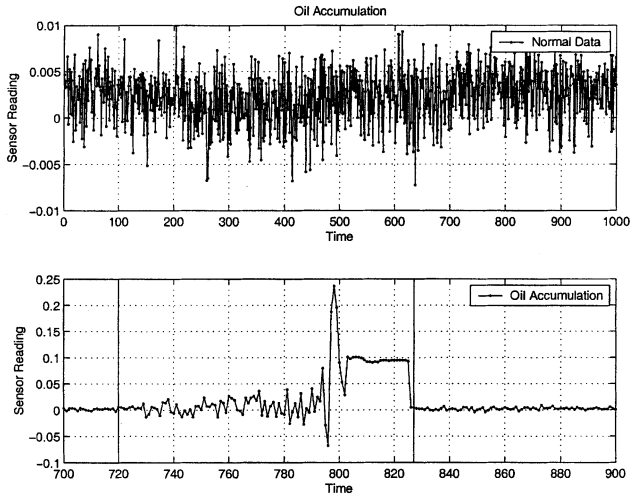

Figure 24. Data for normal operation and the oil accumulation event.

empirically determined values, while the curves are their least-square polynomial fits. These graphs clearly show the overall superiority of MSSPC for detecting the abnormal event, since the MSSPC curve shows a higher rate of detecting true positives for the same rate as false positives.

Sensor Malfunction Due to Oil Accumulation. This example analyzes a sensor failure due to oil accumulation. The difference between the readings of a faulty sensor and a coupled redundant sensor is shown in Figure 24. The abnormal event is indicated by a change in the frequency and amplitude of the measurements between 720 and 827 time points. The small oscillations make it difficult to detect the beginning of this fault by any of the methods used in this article. Once again, the rate of false alarms was maintained equal for all methods to approximately 1%. As illustrated in Figure 25, the MA chart results in the worst performance for detecting this event. MSSPC and Shewhart charts show a similar performance. The ROC curves for this event, shown in Figure 26, confirm that MA charts have the worst performance, with Shewhart and MSSPC charts performing similarly.

Multivariate process

This case study applies PCA and MSPCA-integer to detecting a disturbance in furnace feeds extracted from a real petrochemical process. Normalized sensor readings for 10 furnace-feed sensors are used as test data, and are presented in Figure 27. The event began at approximately $t = 108$. The detection performance presented in Figure 28 shows that MSPCA detects the event slightly earlier than PCA and continues to detect it with its Q value well above the detection

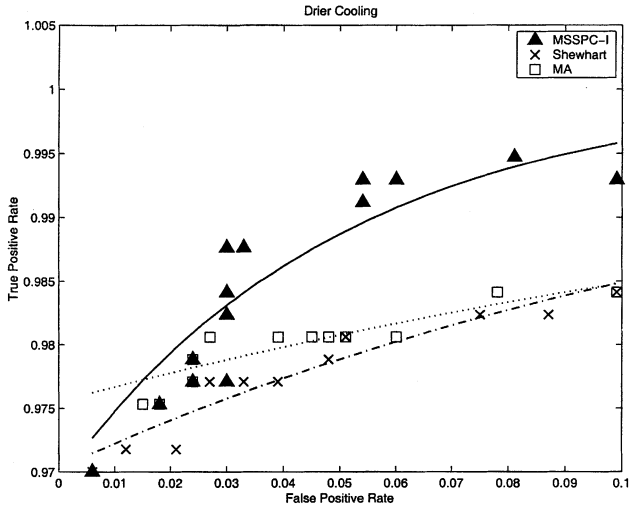

Figure 23. Receiver-operating characteristic curves for the drier cooling event using Shewhart, MA, and MSSPC charts.

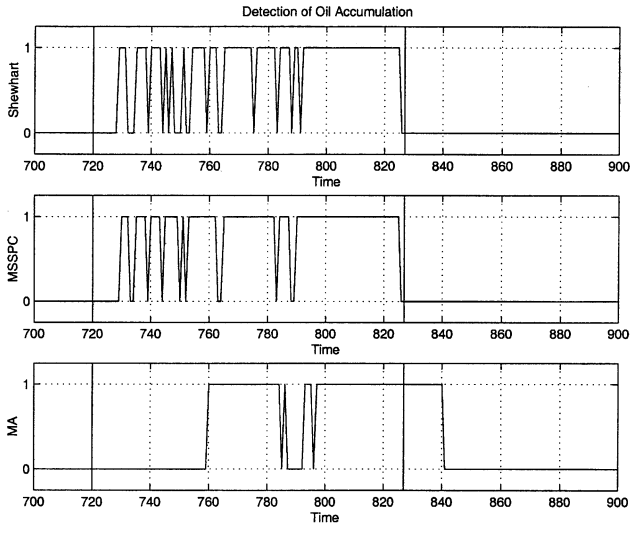

Figure 25. Performance of Shewhart, MSSPC-integer, and MA charts for detection of oil accumulation event.

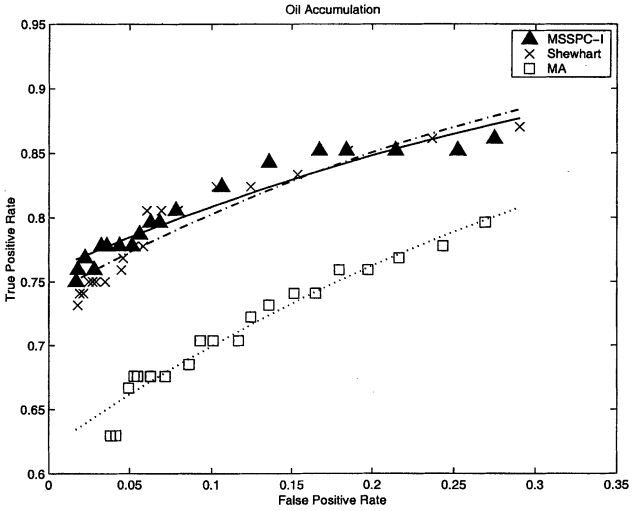

Figure 26. Receiver–operating characteristic curves for the oil accumulation event using Shewhart, MA, and MSSPC charts.

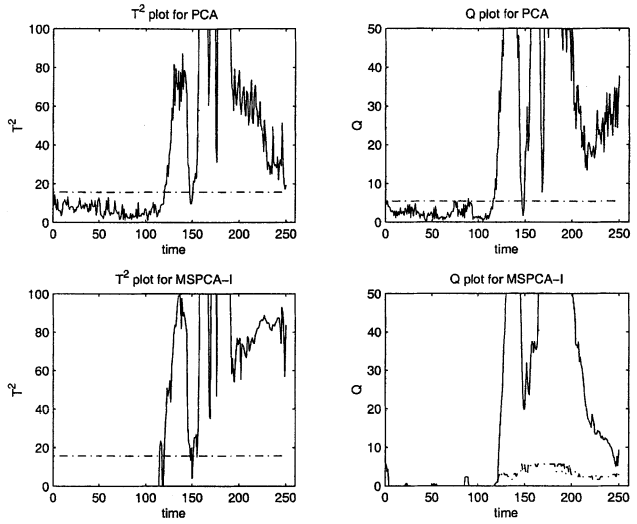

Figure 28. Performance of T^2 and Q plots of PCA and MSPCA-integer for monitoring multivariate industrial data.

limits. The ROC curves for this fault, shown in Figure 29, indicate that MSPCA performs better than PCA, particularly for small false-positive rates. For larger false-positive rates, MSPCA is slightly better or comparable. Additional applications to simulated and industrial data are described in Bakshi (1998), Aradhye et al. (2000, 2001), and Kano et al. (2002).

Conclusions and Discussion

The multiscale approach studied in this article is ideal for SPC of measurements that contain features with different localization in time or frequency and where the nature of the features is not known *a priori*. Examples of such features include deterministic changes with different temporal duration, and stochastic variation with changing intensity over frequency or time. MSSPC exploits the ability of wavelets to compress deterministic features at all scales to a small number of relatively large coefficients. It adapts to features at different scales by focusing only on those scales that contain coefficients outside the detection limits at each scale. The value of the detection limits at each scale depends on the nature of the temporal correlation in measurements representing normal operation. MSSPC adjusts the nature of the filter applied to the measurements as well as the detection limits for a selected confidence limit according to the scale of the abnormal feature. Thus, MSSPC can specialize to many existing SPC methods, including Shewhart, MA, CUSUM, and EWMA charts, depending on the scale of the abnormal feature and the selected wavelet.

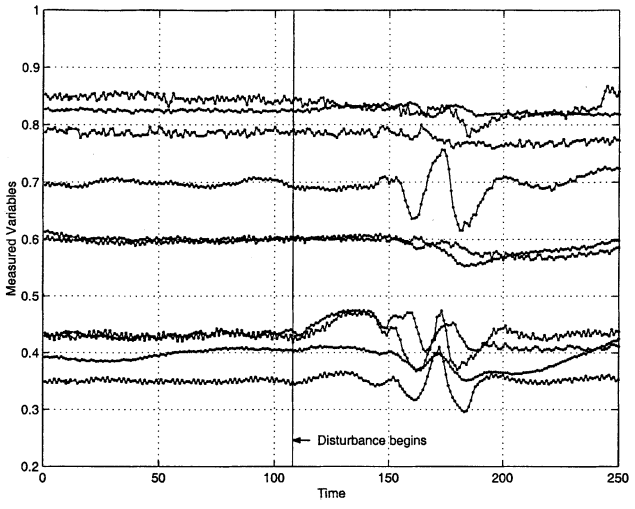

Figure 27. Test data for furnace feed disturbance detection.

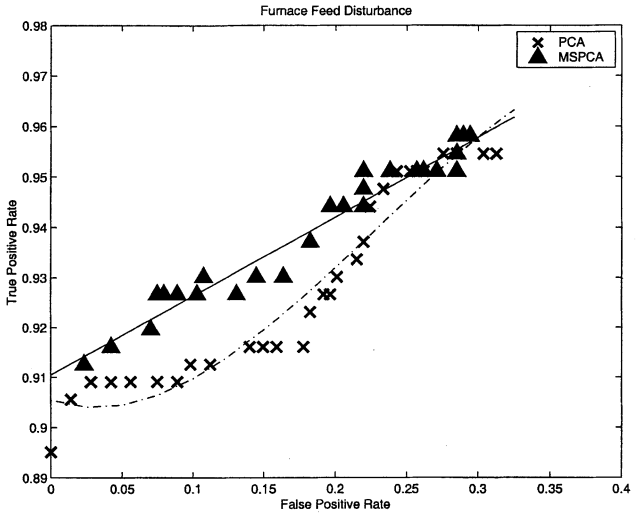

Figure 29. Receiver–operating characteristic curves for the furnace feed disturbance event using PCA and MSPCA-integer; maximum decomposition depth = 2.

Two variations of MSSPC are studied in this article, depending on how the wavelet translation parameter is discretized. If the translation parameter is discretized dyadically, then the coefficients are downsampled at each scale. This permits the use of orthonormal wavelets, and approximate decorrelation of autocorrelated measurements, but introduces a time delay in obtaining the coefficients and detecting some changes. If the translation parameter is discretized uniformly, then there is no downsampling of the coefficients. This permits on-line decomposition without delay while retaining the compression ability, but the wavelets are not orthonormal, and the coefficients are autocorrelated. ARL analysis shows that MSSPC with dyadic discretization performs well for detection of shifts from highly autocorrelated or nonstationary measurements, since the benefits of decorrelation outweigh the delay due to downsampling. If the measurements are uncorrelated or only moderately correlated, the benefits of decorrelation are not as significant, and it is better to decrease the time delay in detecting a change. In such situations, it is better to use MSSPC with uniform or integer discretization of the wavelet translation parameter.

Comparison of the average run lengths with existing methods for univariate and multivariate SPC indicates that MSSPC does not perform better than methods specially designed to detect specific types of changes. For example, Shewhart charts are best for detecting large shifts, while MA or EWMA charts are best for detecting small shifts of a certain size. For changes other than mean shifts such as spectral changes, Shewhart, MA, or CUSUM charts may not perform as well as MSSPC. However, other methods such as CUSCORE (Box and Ramirez, 1992) and spectral PCA can be tailor-made for detecting such changes. These methods are not adaptive or general, and do very well *only* for the type and magnitude of change they are designed to detect. In contrast, MSSPC is a general method that exhibits better average performance in detecting a range of changes in different types of measurements. Since the nature of the change is unknown for most industrial processes, MSSPC seems to be ideal for such systems as demonstrated by the industrial applications in this article and others (Aradhye et al., 2001). Furthermore, MSSPC can easily deal with autocorrelated measurements and multivariate problems. The MSSPC methodology can also be used to transform any single-scale process monitoring method to its multiscale version (Aradhye et al., 2000; Kano et al., 2002).

In addition to developing multiscale forms of other process monitoring methods, there are many other areas of ongoing and future work. The tuning parameters in MSSPC include the depth of decomposition and type of wavelet. The depth is selected manually, but it is expected that this selection may be automated. This article has focused only on the use of Haar wavelets, but the approach can easily be used with other types of wavelets. The use of smoother, boundary-corrected wavelets may provide better performance than Haar wavelets due to better feature extraction and decorrelation abilities. Furthermore, extension to libraries of basis functions such as, wavelet packets may permit MSSPC to automatically select the best family of basis functions from the library (Bakshi, 2002). MSSPC can also be extended to identification of the root causes of the abnormal operation. Since process operation involves making decisions at different time scales, the multiscale approach may also permit integration of various operation tasks. Research on these and related topics will be the subject of future publications.

Acknowledgments

We gratefully acknowledge partial financial support from the National Science Foundation through CAREER award CTS-9733627, the Presidential Fellowship Program of the Ohio State University, and the Abnormal Situation Management Consortium.

Literature Cited

Aradhye, H. B., B. R. Bakshi, J. F. Davis, and S. C. Ahalt, "Clustering in Wavelet Domain—A Multiresolution ART Network for Anomaly Detection," Tech. Rep., Dept. of Chemical Engineering, Ohio State Univ., Columbus (2000).

Aradhye, H. B., J. F. Davis, B. R. Bakshi, and S. C. Ahalt, "ART-2 and Multiscale ART-2 for On-Line Process Fault Diagnosis—Validation via Industrial Case Studies and Monte Carlo Simulation," *Proc. IFAC Workshop on On-Line Fault Detection and Supervision in the Chemical Process Industries*, Jejudo Island, Korea, published by Elsevier (2001).

Bakshi, B. R., "Multiscale PCA with Application to Multivariate Statistical Process Monitoring," *AIChE J.*, **44**, 1596 (1998).

Bakshi, B. R., "Improving Multiscale Statistical Process Control Using Libraries of Basis Functions," AIChE Meeting, Indianapolis, IN (2002).

Beylkin, G., R. Coifman, and V. Rokhlin, "Fast Wavelet Transforms and Numerical Algorithms I," *Commun. Pure Appl. Math.*, **XLIV**, 141 (1991).

Box, G., and J. Ramirez, "Cumulative Score Charts," *Qual. Reliab. Eng. Int.*, **8**, 17 (1992).

Coifman, R. R., and D. L. Donoho, "Translation-Invariant De-Noising," *Wavelets and Statistics*, Lecture Notes in Statistics, Springer-Verlag, New York (1995).

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

Donoho, D. L., I. M. Johnstone, G. Kerkyacharian, and D. Picard, "Wavelet Shrinkage: Asymptopia?" *J. Roy. Stat. Soc., Ser. B*, **57**, 41 (1995).

Harris, T. J., and W. H. Ross, "Statistical Process Control Procedures for Correlated Observations," *Can. J. Chem. Eng.*, **69**, 48 (1991).

Hunter, J. S., "The Exponentially Weighted Moving Average," *J. Qual. Technol.*, **18**, 203 (1986).

Jackson, J. E., "Principal Components and Factor Analysis: I. Principal Components," *J. Qual. Technol.*, **12**, 201 (1990).

Kano, M., K. Nagao, S. Hasebe, I. Hashimoto, H. Ohno, R. Strauss, and B. Bakshi, "Comparison of Statistical Process Monitoring Methods: Application to the Eastman Challenge Problem," *Comput. Chem. Eng.*, **26**, 161 (2002).

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," *Chemometrics Intelligent Lab. Syst.*, **30**, 179 (1995).

Lang, G., H. Guo, J. E. Odegard, C. S. Burrus, and R. O. Wells, "Noise Reduction Using an Undecimated Discrete Wavelet Transform," *IEEE Signal Process. Lett.*, **3**, 10 (1996).

Lucas, J. M., "Combined Shewhart-CUSUM Quality Control Schemes," *J. Qual. Technol.*, **14**, 51 (1982).

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

Malfait, M., and D. Roose, "Wavelet Based Image Denoising Using a Markov Random Field a Priori Model," *IEEE Trans. Image Process.*, **6**, 549 (1997).

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

Mallat, S. G., and S. Zhong, "Characterization of Signals from Multistage Edges," *IEEE Trans. Pattern Anal. Machine Intell.*, **14**, 710 (1992).
Mastrangelo, C. M., and D. C. Montgomery, "SPC with Correlated Observations for the Chemical and Process Industries," *Qual. Reliab. Eng. Int.*, **11**, 79 (1995).
Miller, E., and A. S. Willsky, "A Multiscale Approach to Sensor Fusion and the Solution of Linear Inverse Problems," *Appl. Comput. Harmonic Anal.*, **2**, 127 (1995).
Montgomery, D. C., *Introduction to Statistical Quality Control*, Wiley, New York (1996).
Nason, G. P., and B. W. Silverman, "The Stationary Wavelet Transform and Some Statistical Applications," *Wavelets and Statistics*, Lecture Notes in Statistics, Springer-Verlag, New York (1995).
Negiz, A., and A. Cinar, "Statistical Monitoring of Multivariable Dynamic Processes with State-Space Models," *AIChE J.*, **43**, 2002 (1997).
Nounou, M. N., and B. R. Bakshi, "Online Multiscale Filtering of Random and Gross Errors Without Process Models," *AIChE J.*, **45**, 1041 (1999).
Ogden, R. T., and J. D. Lynch, "Bayesian Analysis of Change-Point Models," *Bayesian Inference in Wavelet-Based Methods*, Springer-Verlag, New York (1999).
Pesquet, J., H. Krim, and H. Carfantan, "Time Invariant Orthonormal Wavelet Representations," *IEEE Trans. Signal Process.*, **44**, 1964 (1996).
Rosen, C., and J. A. Lennox, "Multivariate and Multiscale Monitoring of Wastewater Treatment Operation," *Water Res.*, **35**, 3402 (2001).
Runger, G. C., and T. R. Willemain, "Model-Based and Model-Free Control of Autocorrelated Processes," *J. Qual. Technol.*, **27**, 283 (1995).
Sadler, B. M., and A. Swami, "Analysis of Multiscale Products for Step Detection and Estimation," *IEEE Trans. Inf. Theory*, **45**, 1043 (1999).
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. Reliab. Eng. Int.*, **12**, 165 (1996).
Stoumbos, Z. G., M. R. Reynolds, T. P. Ryan, and W. H. Woodall, "The State of Statistical Process Control as We Proceed into the 21st Century," *J. Amer. Stat. Assoc.*, **95**, 992 (2000).
Teppola, P., and P. Minkkinen, "Wavelet-pls Regression Models for Both Exploratory Data Analysis and Process Monitoring," *J. Chemometrics*, **14**, 383 (2000).
Wardell, D. G., H. Moskowitz, and R. D. Plante, "Run-Length Distributions of Special-Cause Control Charts for Correlated Processes," *Technometrics*, **36**, 3 (1994).
Western Electric, *Statistical Quality Control Handbook*, Western Electric Corp., Indianapolis, IN (1956).
Wiel, S. A. V., "Monitoring Processes that Wander Using Integrated Moving Average Models," *Technometrics*, **38**, 139 (1996).
Yoon, S., *Using External Information for Statistical Process Control*, PhD Thesis, McMaster Univ., Hamilton, Ont., Canada (2001).

Appendix A: Analytical Derivation of ARL Curve

Consider an IID Gaussian process, X, with mean 0 and standard deviation σ. If it undergoes a mean-shift of magnitude δ at time $k = 1$, then the variables $X[k]$, $k = 2, 3, \ldots, \infty$ are IID Gaussian variables with mean δ and standard deviation σ. The probability of not detecting this shift with a Shewhart chart using limits $[-\eta\ \eta]$ is

$$\begin{aligned}\beta &= P\{-\eta \le X \le \eta\} \\ &= \int_{-\eta}^{\eta} f(x-\delta,\ \sigma)\ dx \\ &= \Phi(\eta-\delta) - \Phi(-\eta-\delta) \end{aligned} \tag{A1}$$

where

$$f(x,\ \sigma) = \frac{1}{\sigma\sqrt{2}} \exp^{-(1/2)(x/\sigma)^2}$$

and Φ is the standard normal cumulative distribution function. The ARL is then (Montgomery, 1996)

$$ARL = \sum_{k=1}^{\infty} k\beta^{k-1}(1-\beta) = \frac{1}{1-\beta} \tag{A2}$$

A1. ARL with dyadic discretization

This derivation is for MSSPC with Haar wavelets. It assumes that the wavelet and scaling function coefficients from a wavelet decomposition are completely uncorrelated. Since Haar wavelets and scaling functions have a support of 2^m, at scale m, a mean shift can appear only in the first wavelet coefficient at each scale after the shift. Subsequent wavelet coefficients are not affected by the shift, and are statistically similar to wavelet coefficients for IID Gaussian data. A mean shift of size δ located at the second measurement gets scaled to $\delta/2^{m/2}$ in the first wavelet coefficient at any scale. The first coefficient of the last scaled signal becomes $\delta(2^L - 1)/2^{L/2}$ and the other coefficients become $\delta 2^{L/2}$. This difference between the change in the first and other coefficients is sufficiently small to be ignored in the derivation.

The probability of not detecting the shift in the first wavelet coefficient at scale m may be written as

$$\beta_m = \Phi\left(\eta - \frac{\delta}{2^{m/2}}\right) - \Phi\left(-\eta - \frac{\delta}{2^{m/2}}\right) \tag{A3}$$

The probability of not detecting the shift in wavelet coefficients other than the first one is denoted as β_0, and is given by Eq. A1c, which is the same as the probability of not detecting a shift in IID Gaussian measurements.

For a decomposition of depth, L, the probability of not detecting a shift in any scaled signal coefficient is

$$\bar{\beta}_L = \Phi(\eta - \delta 2^{L/2}) - \Phi(-\eta - \delta 2^{L/2}) \tag{A4}$$

By using Eq. A4 for all the scaled signal coefficients, the slight difference between the size of the shift in the first coefficient vs. others is being ignored.

The ARL for MSSPC-dyadic with $L = 1$ is determined as follows

$$\begin{aligned} ARL_1 = &\left(1 - \beta_1 \bar{\beta}_1\right) \\ &+ 3\beta_1 \bar{\beta}_1\left(1 - \beta_0 \bar{\beta}_1\right) \\ &+ 5\beta_1 \beta_0 \bar{\beta}_1^2\left(1 - \beta_0 \bar{\beta}_1\right) \\ &+ \cdots \end{aligned} \tag{A5}$$

Equation A5 corresponds to Figure A1a. The first term in Eq. A5 is the probability of detecting the shift in the first wavelet or scaling-function coefficient. The second term rep-

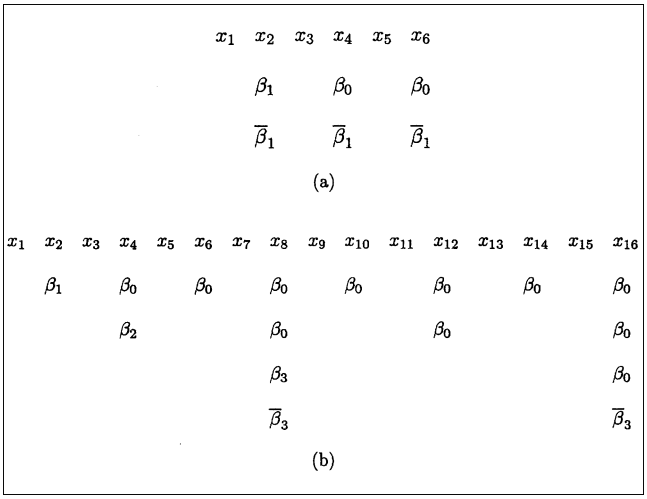

Figure A1. Dyadic discretization grid for ARL equations.

(a) Grid for obtaining Eq. A5 for $L = 1$; (b) grid for obtaining Eq. A9 for $L = 3$.

resents the probability of detecting the shift in the second wavelet or scaling-function coefficient, knowing that there was no detection in the first wavelet or scaling-function coefficients, and so on. Equation A5 can be written as

$$ARL_1 = (1 - \beta_1 \bar{\beta}_1) + (1 - \beta_0 \bar{\beta}_1)\beta_1 \bar{\beta}_1 \sum_{k=1}^{\infty} (2k+1)(\beta_0 \bar{\beta}_1)^{k-1} = 1 + \frac{2\beta_1 \bar{\beta}_1}{1 - \beta_0 \bar{\beta}_1} \quad \text{(A6)}$$

Similarly, for $L = 2$, the ARL can be written by referring to two scales in Figure 4a as

$$\begin{aligned} ARL_2 = {} & (1 - \beta_1) \\ & + 3\beta_1(1 - \beta_0 \beta_2 \bar{\beta}_2) \\ & + 5\beta_1 \beta_0 \beta_2 \bar{\beta}_2 (1 - \beta_0) \\ & + 7\beta_1 \beta_0^2 \beta_2 \bar{\beta}_2 (1 - \beta_0^2 \bar{\beta}_2) \\ & + 9\beta_1 \beta_0^4 \beta_2 \bar{\beta}_2^2 (1 - \beta_0) \\ & + 11\beta_1 \beta_0^5 \beta_2 \bar{\beta}_2^2 (1 - \beta_0^2 \bar{\beta}_2) + \cdots \end{aligned} \quad \text{(A7)}$$

which may be simplified to

$$ARL_2 = 1 + 2\beta_1 + \frac{2\beta_0 \beta_1 \beta_2 \bar{\beta}_2 (1 + \beta_0)}{1 - \beta_0^3 \bar{\beta}_2} \quad \text{(A8)}$$

Similarly, the ARL for $L = 3$ based on Figure A1b is

$$ARL_3 = 1 + 2\beta_1 + 2\beta_0 \beta_1 \beta_2 + 2\beta_0^2 \beta_1 \beta_2 + \frac{2\beta_0^4 \beta_1 \beta_2 \beta_3 \bar{\beta}_3 (1 + \beta_0 + \beta_0^3 + \beta_0^4)}{1 - \beta_0^7 \bar{\beta}_3} \quad \text{(A9)}$$

In general, the equation for the ARL for MSSPC-dyadic at any depth of decomposition can be written from the wavelet decomposition diagram for that depth. Figure 27 shows the probability of not detecting the shift for each wavelet and last scaled signal coefficient for $L = 1$ and $L = 3$. The first term in ARL_L in Eqs. A6, A8, and A9 and is always 1. Each of the next 2^{L-1} terms is obtained by traversing the grid in time until the first $\bar{\beta}_L$ is reached. Each term is two times the product of all the β's up to the time point. For example, the fourth term in Eq. A9 corresponds to x_6 in Figure 27b and is two times the product of β_1, β_2, and the two β_0's at scale $m = 1$. The last term in the ARL equation can be written as

$$\frac{2\beta_0^{2^L - L - 1} \bar{\beta}_L \Pi_{i=1}^{L} \beta_i}{1 - \beta_0^{2^L - 1} \bar{\beta}_L} \left(1 + \sum_{k=1}^{2^{L-1}-1} \beta_0^{\sum_{j=1}^{k} \zeta(j)}\right) \quad \text{(A10)}$$

where

$$\zeta(j) = \begin{cases} 1 & \text{if } j \text{ is odd} \\ 1 + \log_2(j) & \text{if } j \text{ is dyadic} \\ 1 + \log_2\left(j - 2^{\lfloor \log_2(j) \rfloor}\right) & \text{if } j \text{ is even but nondyadic} \end{cases} \quad \text{(A11)}$$

ARL results based on this theoretical approach are compared with those from the Monte Carlo simulation in Figures 10 and 15 for uncorrelated and autocorrelated measurements. These results show a good match between theoretical and simulated results for uncorrelated measurements. For autocorrelated measurements, the assumption of uncorrelated wavelet coefficients is less accurate, and the match between the theoretical and simulated results deteriorates slightly, as shown in Figure 15. The theoretical ARL tends to be smaller than the simulated ARL. The reason for this underestimation is discussed in Appendix A2.

A2. ARL with integer discretization

When integer discretization is used, wavelet coefficients are not downsampled. As in the previous subsection, consider a zero-mean Gaussian random process $X[k]$ with a mean shift of size δ introduced starting from time step k. Thus, $X[0]$ is a random variable with mean 0, whereas $X[k]$, $k > 0$ are IID random variables with mean δ.

Let the event that the shift is missed at time $k > 0$ be denoted as E_k. The runlength (RL) is the time step at which the fault is first detected. Thus, we have

$$P(RL = k) = P(\overline{E_k}, E_{k-1}, E_{k-2}, \ldots, E_2, E_1) \quad \text{(A12)}$$

The *a priori* probability of the event E_k is $\beta[k]$. Note that the notation β_m used in the previous subsection denotes the probability of the shift being missed at scale m, whereas the notation $\beta[k]$ used here is the *overall* probability of the shift being missed at time k. Given the number of scales chosen for wavelet decomposition, L, this probability is a function of the distributions of the variables $X[k-2^L+1]$, $X[k-2^L+2]$, …, $X[k]$. This implies that $\beta[k]$ is constant for $k \geq 2^L$. Let us choose $L=1$ for the sake of this derivation. In that case, let p_1 denote $\beta[1]$ and let p_2 denote $\beta[k]$, $k>1$.

Since the wavelet decomposition without downsampling is not orthogonal, the events E_k are, in general, not independent unlike the previous section. Let us first derive an expression for the ARL assuming independence of these events.

For $k>1$

$$\begin{aligned} P(RL=k) &= P\left(\overline{E_k}, E_{k-1}, E_{k-2}, \ldots, E_2, E_1\right) \\ &= P\left(\overline{E_k}\right) \times P(E_{k-1}) \times P(E_{k-2}) \cdots P(E_2) \\ &\quad \times P(E_1) \\ &= (1-p_2)p_2^{k-2}p_1 \end{aligned} \tag{A13}$$

For $k=1$, $P(RL=k)=(1-p_1)$.

The ARL, which is the expected value of the random variable RL, can be derived in a similar way to Eq. A5 as

$$\begin{aligned} ARL_1 &= \Sigma_{k=1}^{\infty} k \times P(RL=k) \\ &= (1-p_1) + \Sigma_{k=2}^{\infty} k \times (1-p_2)p_2^{k-2}p_1 \\ &= (1-p_1) + p_1\frac{2-p_2}{1-p_2} \end{aligned} \tag{A14}$$

Let $\boldsymbol{E}$ represent the MSSPC discrete decision function. Since $L=1$, $\boldsymbol{E}$ is only a function only of $\{X[k-1], X[k]\}$ at time k. If the values of the measurements were a and b at times $k-1$ and k, respectively, and the shift was not detected at time k, then we have $\boldsymbol{E}(X[k-1]=a, X[k]=b)=1$. The value for $\boldsymbol{E}(X[k-1]=a, X[k]=b)=0$ if the shift was detected at time k. Given the values of the detection thresholds for the wavelet coefficients at each scale, as well as the thresholds for the reconstructed signal, $\boldsymbol{E}$ is a deterministic function in two dimensions. We have

$$p_1 = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \boldsymbol{E}(x_0, x_1) f(x_0, \sigma) f(x_1-\delta, \sigma)\, dx_0\, dx_1 \tag{A15}$$

Similarly

$$p_2 = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \boldsymbol{E}(x_1, x_2) f(x_1-\delta, \sigma)\, f(x_2-\delta, \sigma)\, dx_1\, dx_2 \tag{A16}$$

The terms p_1 and p_2 are computed by numerical integration using the preceding formulation. Our ongoing work has focused on deriving closed-form expressions for the probabilities just given, although it is beyond the scope of this work.

The ARL values tabulated in Table A1 show that the assumption of the events E_k being independent is not valid except for large shifts. The ARL estimates from the preceding analysis are compared to those generated from Monte Carlo simulations. For each experiment, the run lengths for simulated data were averaged over 100,000 instances. As the shift magnitude increases from small to medium, the signal-to-noise ratio increases. As a result, the detection probabilities are more correlated. This is reflected in the increased relative error in ARL estimates. For large shifts, the probability that the shift is detected in the very first step itself grows larger, and as a result, the dependence of detection probabilities at successive time steps is less important. Hence, the relative error in ARL estimates decreases for large shifts. It is worth noting that the ARL estimates derived from Eq. A13 are always lower than the actual values. This is due to the fact that an unusually high or unusually low signal value at time $k-1$ is likely to trigger detection at time-step $k-1$ as well as time step k. Thus, there exists a positive correlation between detection probabilities at consecutive time steps that Eq. A13 fails to model, and, hence, an underestimation of the ARL values results. In other words, Eq. A12 assumes $P(E_k|E_{k-1}) = P(E_k)$, whereas in reality, $P(E_k|E_{k-1}) > P(E_k)$.

To obtain more accurate ARL estimates, let us now model the detection event as a one-step Markov chain. This assumption allows us to model the joint probability distribution as

$$\begin{aligned} P(RL=k, k>2) &= P\left(\overline{E_k}, E_{k-1}, E_{k-2}, \ldots, E_2, E_1\right) \\ &= P\left(\overline{E_k}|E_{k-1}\right) \times P\left(E_{k-1}|E_{k-2}\right) \\ &\quad \times P\left(E_{k-2}|E_{k-3}\right) \cdots P\left(E_2|E_1\right) \times P(E_1) \\ &= (1-p_3)p_3^{k-3}p_4p_1 \end{aligned} \tag{A17}$$

where $p_3 = P(E_k|E_{k-1})$, $k>2$ and $p_4 = P(E_2|E_1)$. The values of p_3 and p_4 are computed by numerical integration

$$p_4 = \frac{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \boldsymbol{E}(x_0, x_1)\boldsymbol{E}(x_1, x_2) f(x_0, \sigma) f(x_1-\delta, \sigma) f(x_2-\delta, \sigma)\, dx_0\, dx_1\, dx_2}{p_1} \tag{A18}$$

Similarly

$$p_3 = \frac{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \boldsymbol{E}(x_1, x_2)\boldsymbol{E}(x_2, x_3) N(x_1-\delta, \sigma) N(x_2-\delta, \sigma) N(x_3-\delta, \sigma)\, dx_1\, dx_2\, dx_3}{p_2} \tag{A19}$$

Table A1. Theoretical ARL Estimation with Assumed Independence of Detection Probabilities (Eq. A13)

Shift Size (δ)	Exp. ARL	Theor. ARL	% Error
0	2,937.1	2,840.9	3.2756
1	115.05	101.85	11.483
2	8.1691	6.5034	20.390
3	2.5130	2.3066	8.2124
4	1.7087	1.6954	0.7836

Table A2. Theoretical ARL Estimation with 1-Step Markov Model for Detection Probabilities (Eq. A19)

Shift Size (δ)	Exp. ARL	Theor. ARL	% Error
0	2,937.1	2,975.1	−1.2930
1	115.05	115.95	−0.7775
2	8.1691	8.4056	−2.8961
3	2.5130	2.5553	−1.6826
4	1.7087	1.7101	−0.0819

Also, $P(RL=2)=(1-p_4)p_1$ and $P(RL=1)=1-p_1$. We now have the expression

$$\begin{aligned} ARL &= \Sigma_{k=1}^{\infty} k \times P(RL=k) \\ &= (1-p_1)+2(1-p_4)p_1+\Sigma_{k=3}^{\infty} k \times (1-p_3)p_3^{k-3}p_4p_1 \\ &= (1-p_1)+2(1-p_4)p_1+p_1p_4\frac{3-2p_3}{1-p_3} \end{aligned} \tag{A20}$$

Table A2 shows the relative error of this ARL estimate. A family of theoretical and experimental ARL curves for different in-control runlengths is shown in Figure 12. Tables A1 and A2 indicate that by taking into account a 1-step Markov chain dependence, the ARL estimates improve in accuracy. Furthermore, the ARL values are always overestimated due to the negative correlation between shift detections at time steps k and $k-2$, given the detection at time step $k-1$. Equation A16 assumes $P(E_k|E_{k-1},\ E_{k-2})=P(E_k|E_{k-1})$, whereas in reality, $P(E_k|E_{k-1},\ E_{k-2})<P(E_k|E_{k-1})$. This observation indicates that the accuracy of the theoretical ARL estimation can be improved further by modeling with a 2-step Markov chain dependence. Inaccuracies due to a finite simulation sample size and numerical integration step-size are other sources of error.

Manuscript received Apr. 16, 2002, and revision received Sept. 30, 2002.