Monitoring non-linear profiles using a wavelet-based distribution-free CUSUM chart

Joongsup (Jay) Lee[a], Youngmi Hur[b], Seong-Hee Kim[a] and James R. Wilson[c*]

[a]*H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, 30332-0205, USA*; [b]*Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, MD 21218-2682, USA*; [c]*Edward P. Fitts Department of Industrial and Systems Engineering, North Carolina State University, Raleigh, NC 27695-7906, USA*

()

WDFTC is a wavelet-based distribution-free CUSUM chart for detecting shifts in the mean of a profile with noisy components. Exploiting a discrete wavelet transform (DWT) of the mean in-control profile, WDFTC selects a reduced-dimension vector of the associated DWT components from which the mean in-control profile can be approximated with minimal weighted relative reconstruction error. Based on randomly sampled Phase I (in-control) profiles, the covariance matrix of the corresponding reduced-dimension DWT vectors is estimated using a matrix-regularisation method; then the DWT vectors are aggregated (batched) so that the non-overlapping batch means of the reduced-dimension DWT vectors have manageable covariances. To monitor shifts in the mean profile during Phase II operation, WDFTC computes a Hotelling's T^2–type statistic from successive non-overlapping batch means and applies a CUSUM procedure to those statistics, where the associated control limits are evaluated analytically from the Phase I data. Experimentation with several normal and non-normal test processes revealed that WDFTC was competitive with existing profile-monitoring schemes.

Keywords: average run length; CUSUM chart; profile; statistical process control; wavelets

1. Introduction

Rapid advancements in data-acquisition technology, such as the development of laser range sensors, have motivated researchers and practitioners to adapt conventional statistical process control (SPC) techniques for use with large data sets that are called profiles and that contain information about the relationship between the following: (1) a selected quality characteristic

*Corresponding author. Email: jwilson@ncsu.edu

(response); and (2) an input (design, decision) variable, where the input variable can be assigned values throughout the experimental region of interest. For such data, a single realization of an in-control process consists of n pairs $\{(x_i, y_i) : i = 1, \ldots, n\}$ of observations that can be described by the statistical model $y_i = f_0(x_i) + \varepsilon_i$, where $f_0(\cdot)$ is a given function that defines the in-control relationship between the input variable x_i and the corresponding mean response $\mathrm{E}[y_i] = f_0(x_i)$; and ε_i is a random noise term, which is typically assumed to be independent and identically distributed (i.i.d.) normal. This article details WDFTC, a wavelet-based distribution-free CUSUM chart that can detect shifts in the mean of a profile data set $\{(x_i, y_i) : i = 1, \ldots, n\}$, where the complexity of the functional relationship between the input variable x_i and the corresponding mean response $\mathrm{E}[y_i]$ may require a large number n of design points to yield a sufficiently accurate approximation of that relationship over the entire experimental region of interest. Moreover, WDFTC is designed to handle situations in which the noise components $\{\varepsilon_i : i = 1, \ldots, n\}$ associated with a complex profile may exhibit the following anomalous properties:

- heterogeneity of variance across all the design points in the experimental region of interest;
- marked departures from normality (for example, non-zero skewness that is frequently encountered in certain types of manufacturing operations — see Stanfield *et al.* (2004)); and
- substantial probabilistic dependencies (for example, non-zero correlations that arise because some of the corresponding points in the experimental region of interest are close to each other in space or time — see Lada *et al.* (2002) and Stanfield *et al.* (2004)).

Kang and Albin (2000) monitor a semiconductor manufacturing process that is characterized by a linear relationship between the following: (1) the expected value of the pressure y in the chamber where etching of the wafer occurs; and (2) the set point x for the mass flow controller that regulates the flow of gas into the etching chamber. Two quality characteristics (namely, the intercept a_0 and the slope a_1 in the linear statistical model $y = a_0 + a_1 x + \varepsilon$ for $x_{\mathrm{LO}} \leq x \leq x_{\mathrm{HI}}$) are monitored using Hotelling's T^2 chart. Kim *et al.* (2003) use two independent (univariate) exponentially weighted moving average (EWMA) charts to monitor the two regression parameters separately.

Although a linear form occurs frequently, many profile data sets (for example, radar signatures) exhibit non-linearities and other complicated features such as discontinuities, cusps, and other types of non-smooth, irregular behaviour (Chicken *et al.* 2009). Woodall *et al.* (2004) give an overview of using control charts to monitor both linear and non-linear profile data as an application of SPC. Ding *et al.* (2006) present a strategy for Phase I analysis of non-linear profile data, where the Phase I data may be contaminated by out-of-control realizations of the profile; and the objective is to identify and eliminate all out-of-control realizations so that the remaining Phase I data can be used to calibrate the profile-monitoring scheme that will be applied in Phase II operation. Williams *et al.* (2007) discuss an application of profile monitoring in the manufacture of particle board, and they extend Hotelling's T^2 chart to monitor the coefficients of a parametric non-linear regression model. Staudhammer *et al.* (2007) develop profile charts for monitoring the thickness of a sawn board at selected points along the length of the board as it leaves a sawing machine in a lumber mill. They also monitor regression parameters to detect complex sawing defects. However, as Chicken *et al.* (2009) point out, regression parameters may not adequately reflect the profile shifts; moreover, fitting a sufficiently accurate parametric model to a set of observed profiles can present substantial difficulties.

For most non-linear profile-monitoring charts, the power to detect shifts in the mean of a profile can drop significantly if the monitored profile consists of a large number of components (that is, if the profile is 'high-dimensional') (Fan 1996). Several dimension-reduction techniques have been proposed and incorporated into multivariate SPC charts for profile monitoring, including smoothing by regression (Kang and Albin 2000), functional principal component analysis

(Ramsay and Silverman 2006), and the use of the discrete wavelet transform (DWT) (Jin and Shi 1999, Lada *et al.* 2002, Jeong *et al.* 2006).

Among such dimension-reduction techniques, wavelet-based approaches have gained popularity, especially for monitoring profiles that have highly complex or non-smooth behaviour; and such methods have been shown to be effective (Ganesan *et al.* 2004). These profiles are usually multiscale in nature, exhibiting substantially different critical features at different times and frequencies; see Ganesan *et al.* (2004) and Kano *et al.* (2002). Jin and Shi (1999, 2001) use wavelets to monitor waveform signals (non-linear profiles) from an automotive steel-stamping operation. To detect shifts in antenna data, Jeong *et al.* (2006) apply a Hotelling's T^2–type chart to the wavelet coefficients of the observed non-linear profiles. To monitor shifts in the mean of a non-linear profile whose noise components are randomly sampled from a common normal distribution, Chicken *et al.* (2009) track shifts in the mean of the corresponding discrete wavelet transform (DWT) using a likelihood ratio test to detect the change point. Chicken *et al.* (2009) use trial-and-error simulations to estimate the upper control limit for the log-likelihood-ratio test statistic beyond which an associated series of sampled profiles is declared to be out of control.

Generally, a wavelet-based monitoring approach first uses wavelets to decompose a sampled profile into scaling and detail coefficients at various levels of resolution; then a noise-elimination method such as principal component analysis (Jolliffe 1986) or a thresholding method (Donoho and Johnstone 1994) is used to reduce in magnitude or eliminate (that is, set to zero) all the estimated wavelet coefficients that are considered to be 'unimportant' so that the surviving coefficients can be effectively monitored for possible shifts in the mean of the original sampled profiles. In this article, we exploit the capacity for parsimonious representation via wavelet coefficients in the formulation of WDFTC, a wavelet-based distribution-free tabular CUSUM chart for monitoring high-dimensional profiles; and the wavelet-based dimension reduction is achieved by minimizing the weighted relative reconstruction error (Lada *et al.* 2002).

Beyond the challenge of coping effectively with the 'curse of dimensionality', the assumption of i.i.d. normal errors is a severe constraint on the development of an effective wavelet-based control chart for monitoring profiles with deterministic and stochastic properties that may be irregular in some subregions of time or space. In our experience, we have found that SPC charts based on the assumption of i.i.d. normal noise components do not perform adequately when they are applied to processes whose responses (and hence the corresponding errors) exhibit substantial variance heterogeneity, pronounced non-normality, or significant correlations (Kim *et al.* 2007, Lee *et al.* 2009). Little has been done on the development and practical implementation of a monitoring scheme for high-dimensional profiles with non-normal, correlated responses. Qiu (2008) proposes a distribution-free multivariate CUSUM chart based on log-linear modelling, but the method is only applied to test processes with three quality characteristics; and as Qiu remarks, the performance of the proposed chart is unknown for high-dimensional profiles. Procedure WDFTC is intended to enable robust, efficient monitoring of high-dimensional profiles with anomalous distributional properties.

The rest of this article is organized as follows. In Section 2, we introduce the symbols and terminology used throughout the article. We also present two examples that motivate our development of WDFTC by illustrating the degradation in the performance of the profile-monitoring chart M^* of Chicken *et al.* (2009) when the profile responses exhibit substantial non-normality, variance heterogeneity, or correlation between the responses. We close Section 2 with a brief discussion of the wavelet transform as a tool for monitoring high-dimensional profiles. In Section 3 we present the WDFTC chart, which is specifically designed to monitor a profile whose components may have any non-singular joint probability density function — in particular, different profile components may have different continuous marginal distributions that may be non-normal; moreover, the covariance matrix of the profile's components is merely required to

be symmetric and positive definite. In Section 4 of this article, we summarize the results of a comprehensive performance evaluation of WDFTC in comparison with other existing profile-monitoring schemes. In Section 4, we also summarize the results of applying WDFTC to laser range sensor data that arise in lumber manufacturing. We conclude the article by recapitulating the main findings of this work in Section 5.

2. Background

To facilitate our discussion of the development of a distribution-free chart for monitoring high-dimensional profiles, we consider a vector-valued stochastic process of the form

$$\boldsymbol{Y}_j = \boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{\varepsilon}_j, \;\; j = 1,2,\ldots, \tag{1}$$

where: $\boldsymbol{x} = (x_1,\ldots,x_n)^{\mathrm{T}}$ is the $n \times 1$ vector consisting of n selected values of the input variable to be used in generating the jth observed profile (note that $\boldsymbol{x}$ is the same for all profiles; and throughout this article, we let $\mathbf{A}^{\mathrm{T}}$ denote the transpose of a vector or matrix $\mathbf{A}$); $\boldsymbol{Y}_j = (y_{1,j},\ldots,y_{n,j})^{\mathrm{T}}$ is the $n \times 1$ vector consisting of the n respective values of the response variable; $\boldsymbol{f}(\boldsymbol{x}) = [\,f(x_1),\ldots,f(x_n)\,]^{\mathrm{T}}$ is the $n \times 1$ vector consisting of the n respective expected values of the response variable; and $\boldsymbol{\varepsilon}_j = (\varepsilon_{1,j},\ldots,\varepsilon_{n,j})^{\mathrm{T}}$ is the associated $n \times 1$ noise (error) vector with mean $\mathrm{E}[\boldsymbol{\varepsilon}_j] = \mathbf{0}_n$ (the $n \times 1$ vector of zeros) and covariance matrix $\mathrm{Cov}[\boldsymbol{\varepsilon}_j] = \mathrm{E}[\boldsymbol{\varepsilon}_j \boldsymbol{\varepsilon}_j^{\mathrm{T}}] = \boldsymbol{\Sigma}_0$. The relevant univariate functional relationship holds for each point of the jth profile; thus we have $y_{i,j} = f(x_i) + \varepsilon_{i,j}$ for $i = 1,\ldots,n$, where $\varepsilon_{i_1,j}$ and $\varepsilon_{i_2,j}$ may be non-normal and correlated for $i_1 \neq i_2$. We distinguish two process states: (1) $\boldsymbol{Y}_j$ is in control when $\mathrm{E}[\boldsymbol{Y}_j] = \boldsymbol{f}_0 = [\,f_0(x_1),\ldots,f_0(x_n)\,]^{\mathrm{T}}$ for a given in-control function $f_0(\cdot)$ relating the input variable to the corresponding mean response; and (2) $\boldsymbol{Y}_j$ is out of control when $\mathrm{E}[\boldsymbol{Y}_j] = \boldsymbol{f}_1 = [\,f_1(x_1),\ldots,f_1(x_n)\,]^{\mathrm{T}} \neq \boldsymbol{f}_0$ for any other function $f_1(\cdot)$ relating the input variable to the corresponding mean response. Without loss of generality, throughout the rest of this article we assume the mean in-control profile $\boldsymbol{f}_0$ is centred so that $\mathbf{1}_n^{\mathrm{T}} \boldsymbol{f}_0 = \sum_{i=1}^{n} f_0(x_i) = 0$, where $\mathbf{1}_n$ is the $n \times 1$ vector of ones.

Whether it is in control or out of control, the jth observed profile $\boldsymbol{Y}_j$ (for $j = 1,2,\ldots$) is assumed to have the same covariance matrix $\mathrm{Cov}[\boldsymbol{Y}_j] = \boldsymbol{\Sigma}_0$. For the ith component $Y_{i,j}$ of the jth profile ($i = 1,\ldots,n$), we let $\sigma_i^2 = \left[\boldsymbol{\Sigma}_0\right]_{i,i}$ denote the component's marginal variance. Suppose that the profile length n has the form $n = 2^J$ for some positive integer J and that $\mathbf{W}$ denotes the corresponding DWT matrix defined by a given wavelet system with the coarsest level of resolution $L \in \{0,\ldots,J-1\}$ as elaborated in Section 2.2. Then $\boldsymbol{d}_j = \mathbf{W}\boldsymbol{Y}_j = (d_{1,j},\ldots,d_{n,j})^{\mathrm{T}}$ is the DWT of the jth profile, while $\boldsymbol{\theta}_0 = \mathbf{W}\boldsymbol{f}_0 = (\theta_{1,0},\ldots,\theta_{n,0})^{\mathrm{T}}$ is the DWT of the mean in-control profile $\boldsymbol{f}_0$ (Ogden 1997, Mallat 2009). For the leading components of $\boldsymbol{\theta}_0$ (or $\boldsymbol{d}_j$), we have 2^L scaling coefficients (or estimated scaling coefficients), representing the coarser features of the associated profile — i.e., the profile features that are prominent at the lower levels of resolution. Moreover for the remaining components of $\boldsymbol{\theta}_0$ (or $\boldsymbol{d}_j$), we have $n - 2^L$ detail coefficients (or estimated detail coefficients) representing the finer features of the associated profile — i.e., the profile features that are revealed only at the higher levels of resolution. The covariance matrix of $\boldsymbol{d}_j$ is given by $\mathrm{Cov}[\boldsymbol{d}_j] = \boldsymbol{\Lambda}_0 = \mathbf{W}\boldsymbol{\Sigma}_0\mathbf{W}^{\mathrm{T}}$.

Throughout the article, we compare and analyse different profile-monitoring charts based on the in-control average run length (ARL_0) and the out-of-control average run length (ARL_1) expressed in terms of the number of individual profiles $\{\boldsymbol{Y}_j : j = 1,2,\ldots\}$ that are observed before raising a false alarm (under the in-control condition) or a true alarm (under a specific out-of-control condition).

2.1. *Motivating examples*

In this subsection, we demonstrate the need for a distribution-free SPC chart that effectively monitors high-dimensional profiles exhibiting variance heterogeneity, non-normality, or stochastic dependencies among profile components. In particular, we examine the performance of the wavelet-based profile-monitoring chart M^* of Chicken *et al.* (2009), which is designed to monitor non-linear profiles described by equation (1), where for $j = 1, 2, \ldots$, the jth error vector $\boldsymbol{\varepsilon}_j = (\varepsilon_{1,j}, \ldots, \varepsilon_{n,j})^{\mathrm{T}}$ is assumed to consist of i.i.d. $N(0, \sigma_*^2)$ components; that is, the noise terms $\{\varepsilon_{i,j} : i = 1, \ldots, n\}$ are assumed to be independent normal random variables with mean 0 and standard deviation σ_*. Therefore each observed profile has the covariance matrix $\boldsymbol{\Sigma}_0 = \sigma_*^2 \mathbf{I}_n$, where $\mathbf{I}_n$ denotes the $n \times n$ identity matrix.

We consider two motivating examples in which the above assumptions on the error vectors $\{\boldsymbol{\varepsilon}_j\}$ are violated. In the first motivating example (ME_1), we explore the effect on the performance of the chart M^* arising from correlated normal noise components with heterogeneous variances as detailed below.

- The autocorrelation function for the noise components of each profile in ME_1 is taken from von Sachs and MacGibbon (2000, p. 484) — namely, the damped sinusoidal form

$$\rho(\ell) = \mathrm{Corr}[y_{i,j}, y_{i+\ell,j}] = (-\alpha_2)^{|\ell/2|} \left[\frac{\sin(|\ell|\omega + \xi)}{\sin(\xi)} \right] \text{ for } \ell = 0, \pm 1, \ldots, \pm(n-1), \quad (2)$$

  where we take $\alpha_1 = 4/3$, $\alpha_2 = -8/9$, the angular frequency $\omega = \cos^{-1}\left[\alpha_1/2\sqrt{-\alpha_2}\right] \cong 0.785$, and the phase constant $\xi = \tan^{-1}[\tan(\omega)(1-\alpha_2)/(1+\alpha_2)] \cong 1.51$. This gives, for example, $\rho(1) = 0.71$ and $\rho(2) = 0.052$.
- The marginal variances for the components of each profile in ME_1 are similar to those used in Example 2 of Gao (1997),

$$\sigma_i^2 = \mathrm{Var}[y_{i,j}] = \sigma_0^2 \left(1 + \left\{ 0.5 - 2.5\left[(i-1)/n - 0.515 \right]^2 \right\}^2 \right)^2 \quad (3)$$

  for $i = 1, \ldots, n$, where $\sigma_0^2 = 9.50$. The resulting marginal variances σ_i^2 (for $i = 1, \ldots, n$) take values between 9.5 and 14.8; and the componentwise correlations take values between -0.71 and 0.71.

The covariances between pairs of profile components in ME_1 are then given by $\mathrm{Cov}[y_{i_1,j}, y_{i_2,j}] = \left[\boldsymbol{\Sigma}_0\right]_{i_1,i_2} = \sigma_{i_1}\sigma_{i_2}\rho(i_1 - i_2)$ for $i_1, i_2 = 1, \ldots, n$.

In the second motivating example (ME_2), we explore the effect on the performance of the chart M^* arising from non-normal marginal distributions for the components of the profiles that are randomly sampled during Phase II operation. By contrast with ME_1, test process ME_2 has noise components that are mutually independent shifted exponential random variables with mean 0 and variance 1 for $j = 1, 2, \ldots$ and $i = 1, \ldots, n$.

In both test processes ME_1 and ME_2, we add out-of-control shifts and noise terms to the mean in-control profile $\boldsymbol{f}_0$ defined by $n = 512$ equally spaced points on the piecewise smooth function of Mallat (2009, p. 458) as depicted in Figure 1; and we monitor the observed profiles $\{\boldsymbol{Y}_j : j = 1, 2, \ldots\}$ in Phase II operation using procedure M^*.

When monitoring non-normal profiles, we consider two different simulation-based methods to calibrate (estimate) the control limits for an SPC chart that was originally developed under the assumption of normally distributed profile components, possibly with non-zero componentwise correlations.

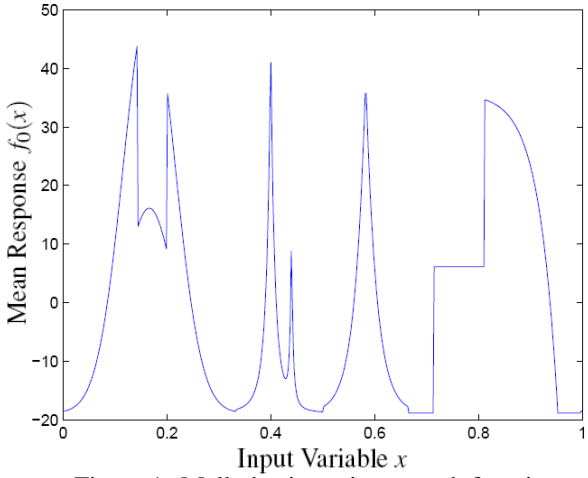

Figure 1. Mallat's piecewise smooth function.

Calibration Method $\mathrm{CM_A}$: Generate a preliminary (Phase I) data set consisting of normally distributed profiles that have the same in-control mean vector and the same covariance matrix as the non-normal profiles to be monitored. Obtain the required control limit(s) for the normally distributed profiles via trial-and-error simulations designed to yield the prespecified target value of ARL_0. Use the resulting control limit(s) to monitor the non-normal profiles in regular Phase II operation.

Calibration Method $\mathrm{CM_B}$: Obtain the required control limit(s) via trial-and-error simulations using the same type of in-control, non-normal profiles that are to be monitored in Phase II; then use the resulting control limit(s) to detect out-of-control conditions in Phase II operation.

Exploiting the $\mathrm{CM_A}$-based control limit(s), we can illustrate the risk of monitoring non-normal profiles with existing SPC charts that were originally designed for normal profiles. A similar approach is taken by Qiu (2008), wherein he demonstrates how excessively large rates of occurrence for false alarms (or equivalently, values for ARL_0 that are substantially below the user-specified nominal level) can occur when SPC charts based on the normality assumption are applied to non-normal profiles. On the other hand, $\mathrm{CM_B}$ enables us to compare the performance of different SPC charts in terms of the resulting values of ARL_1 (or equivalently, the rates of occurrence of true alarms) for a specific out-of-control condition, because each chart's control limits have been calibrated to yield the target value of ARL_0 when the monitored non-normal profiles are in control.

The profile-monitoring chart M^* was applied as follows:

Chart M^*: Calculate $\boldsymbol{\theta}_0 = \mathbf{W}\boldsymbol{f}_0$. Let j_0 denote the unknown change point (profile index) after which out-of-control profiles occur in Phase II operation, where $j_0 \geq 0$. Calculate $\boldsymbol{d}_j = \mathbf{W}\boldsymbol{Y}_j$ for $j = 1,\ldots,G$, where G is assumed to be large enough so that $G \geq j_0$. For $j = 1,\ldots,G$, calculate the basic statistic $w_j = (n/\sigma_*^2)\sum_{i=1}^{n}(d_{i,j} - \theta_{i,0})^2$ measuring the standardized discrepancy between $\boldsymbol{d}_j$ and $\boldsymbol{\theta}_0$ as well as its 'thresholded' version $\widetilde{w}_j = (n/\sigma_*^2)\sum_{i=1}^{n}[\mathrm{thr}(d_{i,j} - \theta_{i,0})]^2$, where the VisuShrink thresholding operator $\mathrm{thr}(\cdot)$ of Donoho and Johnstone (1994) is applied to each component of the difference $\boldsymbol{d}_j - \boldsymbol{\theta}_0$. Given a candidate value $u \in \{0,1,\ldots,G-1\}$ for the unknown change point j_0, calculate the associated likelihood-ratio parameter estimator $\widehat{\gamma} = (G-u)^{-1}\left(\sum_{j=u+1}^{G}\widetilde{w}_j\right) - u^{-1}\left(\sum_{j=1}^{u}\widetilde{w}_j\right)$ for $u \neq 0$ and $\widehat{\gamma} = G^{-1}\sum_{j=1}^{G}\widetilde{w}_j$ for $u = 0$; and finally evaluate the log-likelihood-ratio statistic $h(u) = (\widehat{\gamma}/2)\sum_{j=u+1}^{G}[(w_j/n) - 1]$ so as to find

$\widehat{j_0} = \arg\max\{h(u) : u = 0, 1, \ldots, G-1\}$, the estimated change point. For an upper control limit (UCL) obtained via trial-and-error simulations designed to yield the target value $\mathrm{ARL}_0 = 200$, raise an out-of-control alarm at time (profile) index G if $h(\widehat{j_0}) > \mathrm{UCL}$.

Following the approach of Chicken *et al.* (2009), we express the overall size of a shift $\boldsymbol{f}_1 - \boldsymbol{f}_0$ in the mean profile in terms of the squared Euclidean distance between $\boldsymbol{f}_0$ and $\boldsymbol{f}_1$,

$$a = \|\boldsymbol{f}_1 - \boldsymbol{f}_0\|_2^2 = \sum_{i=1}^{n} [\, f_1(x_i) - f_0(x_i)\,]^2 .$$

Recall that $n = 512$; and in Phase II operation of M^*, we add uniform local shifts to $\boldsymbol{f}_0$ for the component indices $i \in \{89, 90, \ldots, 96\}$ (i.e., 8 shifted components) and for $i \in \{241, 242, \ldots, 256\}$ (i.e., 16 shifted components) so as to yield a selected value of the overall shift size $\sqrt{a}$. This local shift was also used by Jeong *et al.* (2006) and Chicken *et al.* (2009).

Table 1 contains the estimated ARLs and the associated standard errors delivered by M^* based on 1,000 independent replications of the test process ME_1 when a uniform local shift of overall size $\sqrt{a}$ was added to the in-control mean profile $\boldsymbol{f}_0$ to yield $\boldsymbol{f}_1$, with the same values of $\sqrt{a}$ used by Chicken *et al.* (2009). To apply M^*, we estimated σ_* using the average of the median absolute deviations of the $n/2$ highest-level detail coefficients from each observed profile as proposed by Chicken *et al.* (2009). Note that in test problem ME_1, the calibration methods CM_A and CM_B coincide. Comparing the values of ARL_1 in Table 1 with the corresponding values of ARL_1 in Table 1 of Chicken *et al.* (2009), we concluded that the performance of M^* was unacceptable for all the specified out-of-control conditions.

Table 1. ARLs delivered by M^* for test process ME_1.

Shift Type	$\sqrt{a}$	Est. ARL	Std. Err.
In-Control	0	199.	0.83
Local Shift	0.1	198.	0.71
	0.2	198.	0.71
	0.3	196.	0.71
	0.4	194.	0.69
	0.5	189.	0.68

Table 2 contains the estimated ARLs and the associated standard errors delivered by M^* based on 1,000 independent replications of the non-normal test process ME_2 when a uniform local shift of overall size $\sqrt{a}$ was added to the in-control mean profile $\boldsymbol{f}_0$ to yield $\boldsymbol{f}_1$. The performance of M^* was evaluated using both calibration methods CM_A and CM_B. The control limit obtained from CM_A resulted in an extremely small value of ARL_0 for M^*, which translated into an unacceptably large rate of occurrence of false alarms; and in this highly irregular situation, we omitted applying M^* to ME_2 with the specified out-of-control conditions. Comparing the values of ARL_1 in Table 2 with the corresponding values of ARL_1 in Table 1 of Chicken *et al.* (2009), we concluded that when procedure M^* was calibrated using method CM_B, the performance of M^* was unacceptable for all the specified out-of-control conditions.

It was clear from the results for both test processes ME_1 and ME_2 that the performance of M^* became problematic in the presence of stochastic dependence, heterogeneous variances, and non-normality of the sampled profiles. Such characteristics are common in high-dimensional profile data, but most existing profile-monitoring charts, including M^*, require the monitored profile to have i.i.d. normal noise components for successful application of the chart. This con-

Table 2. ARLs delivered by M^* for test process ME_2.

Shift Type	$\sqrt{a}$	Calibration Method			
		$\mathrm{CM_A}$		$\mathrm{CM_B}$	
		Est. ARL	Std. Err.	Est. ARL	Std. Err.
In-Control	0	3.32	0.09	200.	0.81
Local Shift	0.1	–	–	191.	0.98
	0.2	–	–	165.	0.86
	0.3	–	–	110.	0.62
	0.4	–	–	59.	0.37
	0.5	–	–	30.	0.20

clusion will be placed into a more complete perspective in Section 4 of this article, where we summarize the results of a comprehensive experimental performance evaluation of WDFTC versus M^* and some other commonly used profile-monitoring schemes.

2.2. *Wavelet transform overview*

In this subsection, we briefly review the wavelet transform. Let $\mathscr{L}^2[0,1]$ denote the space of real-valued square-integrable functions defined on the unit interval $[0,1]$. The wavelet transform of a function $g \in \mathscr{L}^2[0,1]$ is used to obtain a representation of g as an infinite series involving orthonormal basis functions. A scaling function $\phi \in \mathscr{L}^2[0,1]$ has several key properties that give rise to the associated wavelet function $\psi \in \mathscr{L}^2[0,1]$; and from ψ, we can derive an orthonormal set of basis functions for $\mathscr{L}^2[0,1]$ analogous to the trigonometric functions used in the Fourier series representation. For simplicity in the following discussion, we assume that ϕ and ψ are the Haar scaling and wavelet functions, respectively; see Ogden (1997, pp. 7–23) or Mallat (2009, p. 291).

For a function $g \in \mathscr{L}^2[0,1]$, the representation of g in terms of the Haar scaling and wavelet functions is given by

$$g(z) = \lim_{B\to\infty} \sum_{\ell=-\infty}^{B-1} \sum_{m=0}^{\lceil 2^\ell \rceil - 1} \langle g, \psi_{\ell,m}\rangle \psi_{\ell,m}(z) = \lim_{B\to\infty} \sum_{m=0}^{2^B-1} \langle g, \phi_{B,m}\rangle \phi_{B,m}(z) \tag{4}$$

for almost all $z \in [0,1]$, where: $h_{\ell,m}(z) = 2^{\ell/2} h(2^\ell z - m)$ for $h = \psi, \phi$; and for $g_1, g_2 \in \mathscr{L}^2[0,1]$ we let $\langle g_1, g_2\rangle = \int_0^1 g_1(z) g_2(z)\,\mathrm{d}z$ denote the inner product operator (Ogden 1997). The Bth partial sum $P_B(g)$ on the far right-hand side of equation (4) can be viewed as an approximation to g that becomes progressively more accurate as B increases. In equation (4), the quantities $\{\mathsf{C}_{\ell,m} = \langle g, \phi_{\ell,m}\rangle\}$ are called the scaling coefficients of g and the quantities $\{\mathsf{D}_{\ell,m} = \langle g, \psi_{\ell,m}\rangle\}$ are called the detail coefficients of g. In practice, a physical measuring device can only measure a signal (function) g to a finite level of resolution; thus we take $g \approx P_J(g)$ for some finest (highest) level of resolution J; furthermore, the successive function-approximation operations must stop at some coarsest (lowest) level of resolution L, where $L < J$. As a result, one obtains an approximate representation of g based on its DWT,

$$g(z) \approx \sum_{m=0}^{2^J-1} \mathsf{C}_{J,m} \phi_{J,m}(z) = \sum_{m=0}^{2^L-1} \mathsf{C}_{L,m} \phi_{L,m}(z) + \sum_{\ell=L}^{J-1} \sum_{m=0}^{2^\ell-1} \mathsf{D}_{\ell,m} \psi_{\ell,m}(z) \tag{5}$$

for almost all $z \in [0,1]$, where: (1) the scaling functions $\{\phi_{L,m}(z)\}$ represent the low-frequency components of $g(z)$ — that is, the smooth parts of $g(z)$; and (2) the wavelet functions $\{\psi_{\ell,m}(z)\}$ represent the high-frequency components of $g(z)$ — that is, the local behaviour of $g(z)$.

To monitor deviations from an in-control profile defined by the function $f_0(x)$ for $x \in [x_{\mathrm{LO}}, x_{\mathrm{HI}}]$, we exploit the wavelet transform by taking $g(z) = f_0[x_{\mathrm{LO}} + z(x_{\mathrm{HI}} - x_{\mathrm{LO}})]$ for $z \in [0,1]$ in equations (4) and (5). Because the in-control function $f_0(x) = g[(x - x_{\mathrm{LO}})/(x_{\mathrm{HI}} - x_{\mathrm{LO}})]$ for $x \in [x_{\mathrm{LO}}, x_{\mathrm{HI}}]$ as approximated via equation (5) is originally represented using $n = 2^J$ scaling coefficients $\{\mathsf{C}_{J,m} : m = 0, 1, \ldots, 2^J - 1\}$ of $f_0(\cdot)$ at the finest level of resolution, we see that $f_0(\cdot)$ can also be represented using the 2^L coarsest-level scaling coefficients $\{\mathsf{C}_{L,m} : m = 0, 1, \ldots, 2^L - 1\}$ of $f_0(\cdot)$ together with the $n - 2^L$ detail coefficients $\{\mathsf{D}_{\ell,m} : \ell = L, L+1, \ldots, J-1; m = 0, 1, \ldots, 2^\ell - 1\}$ of $f_0(\cdot)$. Therefore monitoring deviations from an $n \times 1$ in-control mean profile vector $\boldsymbol{f}_0$ defined by the function $f_0(\cdot)$ is equivalent to monitoring deviations from the $n \times 1$ vector consisting of the 2^L coarsest-level scaling coefficients and the $n - 2^L$ detail coefficients that together constitute the DWT of $f_0(\cdot)$ for the Haar wavelet system with a given value of L. Let $\mathbf{W}$ denote the $n \times n$ orthogonal matrix associated with the DWT of $n \times 1$ vectors based on the Haar wavelet system with coarsest level of resolution L. Given a randomly sampled $n \times 1$ in-control profile $\boldsymbol{Y}_j$, the linear transformation $\boldsymbol{d}_j = \mathbf{W}\boldsymbol{Y}_j$ yields estimates of the scaling and detail coefficients of $f_0(\cdot)$, where if necessary the rows of $\mathbf{W}$ have been suitably interchanged to ensure that the first 2^L components of $\boldsymbol{d}_j$ are the estimated scaling coefficients of $f_0(\cdot)$, and the last $n - 2^L$ components of $\boldsymbol{d}_j$ are the estimated detail coefficients of $f_0(\cdot)$.

Because of its simplicity, the Haar wavelet is frequently used in existing wavelet-based SPC schemes (Ganesan *et al.* 2004, Jeong *et al.* 2006), especially when the in-control function $f_0(\cdot)$ is piecewise constant. For smoother functions, other wavelet systems such as the Daubechies or symmlet wavelets are often used (Lada *et al.* 2002, Ganesan *et al.* 2004). In this article we use the symmlet wavelet with the number of vanishing moments equal to eight because the symmlet 8 wavelet yields a smoother approximation to $f_0(\cdot)$ than the Haar wavelet does.

3. Procedure WDFTC: A wavelet-based distribution-free tabular CUSUM chart for profile monitoring

Procedure WDFTC combines the DWT with the distribution-free tabular CUSUM chart of Kim *et al.* (2007) and Lee *et al.* (2009) and focuses on monitoring key components of the DWT determined by a wavelet-based dimension-reduction technique that will be explained in Section 3.1. Table 3 provides a list of all key notation needed in the formulation of WDFTC.

WDFTC begins by computing the wavelet coefficient vector $\boldsymbol{\theta}_0 = \mathbf{W}\boldsymbol{f}_0$ for the in-control mean profile $\boldsymbol{f}_0$. As described in the next subsection, we seek an 'optimal' set of p wavelet coefficients selected from the components of $\boldsymbol{\theta}_0$ to constitute the respective non-zero components of the $n \times 1$ vector $\boldsymbol{\theta}_0^{\#}$ so that the following conditions hold: (1) we take $2^L \le p \le n$, selecting all 2^L scaling coefficients and the $p - 2^L$ largest-magnitude detail coefficients of $\boldsymbol{\theta}_0$ to form the non-zero components of $\boldsymbol{\theta}_0^{\#}$; and (2) as an approximation to $\boldsymbol{f}_0$, the inverse transform $\mathbf{W}^{-1}\boldsymbol{\theta}_0^{\#}$ minimizes the weighted relative reconstruction error (WRRE) evaluated over all $p \in \{2^L, \ldots, n\}$. Let $\boldsymbol{\vartheta}_0^{\#}$ denote the $p \times 1$ reduced-dimension version of $\boldsymbol{\theta}^{\#}$ in which all the non-selected (zero-valued) components have been deleted; and let $\boldsymbol{d}_j^{\#}$ denote the corresponding $p \times 1$ reduced-dimension version of the DWT of the jth profile $\boldsymbol{Y}_j$ for $j = 1, 2, \ldots$. Let $\boldsymbol{\Lambda}_0^{\#}$ denote the $p \times p$ covariance matrix of $\boldsymbol{d}_j^{\#}$, and let $\widetilde{\boldsymbol{\Lambda}}_0^{\#}$ denote the regularised (thresholded) estimator of $\boldsymbol{\Lambda}_0^{\#}$ computed from the Phase I data. WDFTC computes the batch-means vectors $\boldsymbol{d}_k^{\#}(r) = r^{-1}\sum_{u=1}^{r} \boldsymbol{d}_{(k-1)r+u}^{\#}$ based on non-overlapping batches of size r observed in Phase I for $k = 1, \ldots, \lfloor N/r \rfloor$.

Within the kth batch of r profiles observed in Phase I of Procedure WDFTC, all the

Table 3. Notation summary.

$\boldsymbol{f}_0$	=	the $n \times 1$ in-control mean profile, which is assumed to satisfy the centring condition $\mathbf{1}_n^{\mathrm{T}} \boldsymbol{f}_0 = 0$;
$\boldsymbol{\theta}_0$	=	$\mathbf{W}\boldsymbol{f}_0$, the $n \times 1$ DWT of the in-control mean profile $\boldsymbol{f}_0$, where the first 2^L components of $\boldsymbol{\theta}_0$ are the scaling coefficients and the last $n - 2^L$ components of $\boldsymbol{\theta}_0$ are the detail coefficients;
$\boldsymbol{\theta}_0^{\#}$	=	the $n \times 1$ version of $\boldsymbol{\theta}_0$ in which p elements are selected for retention and $n-p$ elements are set to zero and so as to minimize the weighted relative reconstruction error that is defined by equation (7) and that is incurred by using $\mathbf{W}^{-1}\boldsymbol{\theta}_0^{\#}$ as an approximation to $\boldsymbol{f}_0$;
$\boldsymbol{\vartheta}_0^{\#}$	=	the $p \times 1$ version of $\boldsymbol{\theta}_0^{\#}$ in which the $n-p$ non-selected elements of $\boldsymbol{\theta}_0^{\#}$ have been deleted;
$\boldsymbol{f}_0^{\#}$	=	$\mathbf{W}^{-1}\boldsymbol{\theta}_0^{\#}$, the approximate in-control mean profile reconstructed from $\boldsymbol{\theta}_0^{\#}$;
$\boldsymbol{Y}_j$	=	the jth $n \times 1$ observed profile for $j = 1, \ldots, N$ in Phase I and for $j = 1, 2, \ldots$ in Phase II;
$\boldsymbol{Y}_k(r)$	=	$r^{-1}\sum_{u=1}^{r} \boldsymbol{Y}_{(k-1)r+u}$, the kth $n \times 1$ batch-means vector based on non-overlapping batches of size r for $k = 1, \ldots, \lfloor N/r \rfloor$ in Phase I and for $k = 1, 2, \ldots$ in Phase II;
$\boldsymbol{d}_j$	=	$\mathbf{W}\boldsymbol{Y}_j$, the $n \times 1$ DWT of the jth observed profile $\boldsymbol{Y}_j$;
$\boldsymbol{d}_k(r)$	=	$r^{-1}\sum_{u=1}^{r} \boldsymbol{d}_{(k-1)r+u}$, the kth $n \times 1$ batch-means DWT vector computed from non-overlapping batches of size r;
$\boldsymbol{d}_j^{\#}$	=	the $p \times 1$ reduced-dimension version of $\boldsymbol{d}_j$ in which the $n-p$ elements of $\boldsymbol{d}_j$ corresponding to the non-selected (zero-valued) elements of $\boldsymbol{\theta}_0^{\#}$ have been deleted to yield $\boldsymbol{d}_j^{\#}$;
$\boldsymbol{d}_k^{\#}(r)$	=	$r^{-1}\sum_{u=1}^{r} \boldsymbol{d}_{(k-1)r+u}^{\#}$, the kth $p \times 1$ batch-means vector of reduced-dimension DWTs based on non-overlapping batches of size r;
$\boldsymbol{\Lambda}_0$	=	$\mathrm{E}[(\boldsymbol{d}_j - \mathrm{E}[\boldsymbol{d}_j])(\boldsymbol{d}_j - \mathrm{E}[\boldsymbol{d}_j])^{\mathrm{T}}]$, the $n \times n$ covariance matrix of $\boldsymbol{d}_j$, assumed to be the same for both in-control and out-of-control conditions;
$\boldsymbol{\Lambda}_0(r)$	=	$\boldsymbol{\Lambda}_0/r$, the $n \times n$ covariance matrix of $\boldsymbol{d}_k(r)$;
$\boldsymbol{\Lambda}_0^{\#}$	=	$\mathrm{E}[(\boldsymbol{d}_j^{\#} - \mathrm{E}[\boldsymbol{d}_j^{\#}])(\boldsymbol{d}_j^{\#} - \mathrm{E}[\boldsymbol{d}_j^{\#}])^{\mathrm{T}}]$, the $p \times p$ covariance matrix of the reduced-dimension DWT $\boldsymbol{d}_j^{\#}$;
$\boldsymbol{\Lambda}_0^{\#}(r)$	=	$\boldsymbol{\Lambda}_0^{\#}/r$, the $p \times p$ covariance matrix of the reduced-dimension batch-means DWT $\boldsymbol{d}_k^{\#}(r)$;
$\overline{\boldsymbol{d}}_N^{\#}$	=	$N^{-1}\sum_{j=1}^{N} \boldsymbol{d}_j^{\#}$, the $p \times 1$ sample mean of the reduced-dimension DWTs $\{\boldsymbol{d}_j^{\#} : j = 1, \ldots, N\}$ computed from the profiles observed in Phase I;
$\widehat{\boldsymbol{\Lambda}}_0^{\#}$	=	$(N-1)^{-1}\sum_{j=1}^{N} (\boldsymbol{d}_j^{\#} - \overline{\boldsymbol{d}}_N^{\#})(\boldsymbol{d}_j^{\#} - \overline{\boldsymbol{d}}_N^{\#})^{\mathrm{T}}$, the $p \times p$ sample covariance matrix of the reduced-dimension DWTs $\{\boldsymbol{d}_j^{\#} : j = 1, \ldots N\}$ computed from the profiles observed in Phase I;
$\widetilde{\boldsymbol{\Lambda}}_0^{\#}$	=	version of $\widehat{\boldsymbol{\Lambda}}_0^{\#}$ that has been regularised (thresholded) according to Algorithm CMR below;
$\widetilde{\boldsymbol{\Lambda}}_0^{\#}(r)$	=	$\widetilde{\boldsymbol{\Lambda}}_0^{\#}/r$, the $p \times p$ estimated covariance matrix of the reduced-dimension DWTs $\{\boldsymbol{d}_k^{\#}(r) : k = 1, \ldots, \lfloor N/r \rfloor\}$ based on the regularised sample covariance matrix $\widetilde{\boldsymbol{\Lambda}}_0^{\#}$.

sample information about in-control deviations from $\boldsymbol{\theta}_0$ is combined in Hotelling's statistic $T_k^2(r) = \left[\boldsymbol{d}_k^{\#}(r) - \boldsymbol{\vartheta}_0^{\#}\right]^{\mathrm{T}} \left[\widetilde{\boldsymbol{\Lambda}}_0^{\#}/r\right]^{-1} \left[\boldsymbol{d}_k^{\#}(r) - \boldsymbol{\vartheta}_0^{\#}\right]$ for $k = 1, \ldots, \lfloor N/r \rfloor$. Procedure WDFTC determines its control limit analytically for a given target value of ARL_0 using an approach adapted from Kim *et al.* (2007) based on the sample mean and variance of the statistics $\{T_k^2(r) : k = 1, \ldots, \lfloor N/r \rfloor\}$ observed in Phase I. Then in Phase II (regular) operation, the CUSUM procedure of Lee *et al.* (2009) is applied to the associated statistics $\{T_k^2(r) : k = 1, 2, \ldots\}$ to detect out-of-control conditions. A formal algorithmic statement of WDFTC is given in Figure 2.

Procedure WDFTC

Phase I — Using the in-control mean profile $\boldsymbol{f}_0$ and the randomly sampled in-control profiles $\{\boldsymbol{Y}_j : j = 1,\ldots,N\}$, perform the following steps:

[1] Choose the optimal set of p non-zero components for the $n \times 1$ vector $\boldsymbol{\theta}_0^{\#}$ by selecting from the associated components of $\boldsymbol{\theta}_0$ so as to minimize the weighted relative reconstruction error that is defined by equation (7) and that is incurred by using $\mathbf{W}^{-1}\boldsymbol{\theta}_0^{\#}$ to approximate $\boldsymbol{f}_0$, where $2^L \le p \le n$. Assign $\boldsymbol{\vartheta}_0^{\#}$ as the $p \times 1$ version of $\boldsymbol{\theta}_0^{\#}$ from which all non-selected (zero-valued) components have been deleted.

[2] Apply the covariance-matrix regularisation scheme of Algorithm CMR (Figure 3 in Section 3.2) to the sample covariance matrix $\widehat{\boldsymbol{\Lambda}}_0^{\#}$ of the reduced-dimension DWTs $\{\boldsymbol{d}_j^{\#} : j = 1,\ldots,N\}$, thereby obtaining the regularised sample covariance matrix $\widetilde{\boldsymbol{\Lambda}}_0^{\#}$; then let $\widetilde{\boldsymbol{\Lambda}}_0^{\#}(r) = \widetilde{\boldsymbol{\Lambda}}_0^{\#}/r$.

[3] Compute the batch size r using Algorithm BSD (Figure 4 in Section 3.3).

[3a] For $k = 1,2,\ldots,\lfloor N/r \rfloor$, compute the kth batch-means vector $\boldsymbol{Y}_k(r) = r^{-1}\sum_{u=1}^{r}\boldsymbol{Y}_{(k-1)r+u}$, the associated batch-means DWT $\boldsymbol{d}_k(r) = \mathbf{W}\boldsymbol{Y}_k(r)$, and the reduced-dimension batch-means DWT $\boldsymbol{d}_k^{\#}(r)$ to obtain the Hotelling's statistic

$$T_k^2(r) = [\boldsymbol{d}_k^{\#}(r) - \boldsymbol{\vartheta}_0^{\#}]^{\mathrm{T}}[\widetilde{\boldsymbol{\Lambda}}_0^{\#}(r)]^{-1}[\boldsymbol{d}_k^{\#}(r) - \boldsymbol{\vartheta}_0^{\#}]. \tag{6}$$

[3b] From the Phase I statistics $\{T_k^2(r) : k = 1,\ldots,\lfloor N/r \rfloor\}$, compute the usual sample mean $\widehat{\mu}_{T^2(r)}$ and the sample variance $\widehat{\sigma}^2_{T^2(r)}$.

[4] Calculate the root H of the equation

$$\frac{\widehat{\sigma}^2_{T^2(r)}}{2K^2}\left(\exp\left\{\frac{2K[H + 1.166\widehat{\sigma}_{T^2(r)}]}{\widehat{\sigma}^2_{T^2(r)}}\right\} - 1 - \left\{\frac{2K[H + 1.166\widehat{\sigma}_{T^2(r)}]}{\widehat{\sigma}^2_{T^2(r)}}\right\}\right) = 2\mathrm{ARL}_0,$$

where $K = 0.1\widehat{\sigma}_{T^2(r)}$.

Phase II — For $k = 1,2,\ldots$, compute the kth batch-means vector $\boldsymbol{Y}_k(r)$ from the latest non-overlapping batch of r profiles $\{\boldsymbol{Y}_j : j = (k-1)r+1,\ldots,kr\}$ observed in Phase II, and perform the following steps:

[5] Compute $\boldsymbol{d}_k(r)$ and its reduced-dimension counterpart $\boldsymbol{d}_k^{\#}(r)$ to obtain the associated Hotelling's statistic $T_k^2(r)$ as in equation (6).

[6] Raise an alarm after observing the kth batch-means vector $\boldsymbol{Y}_k(r)$ if $S^+(k) \ge H$ or $S^-(k) \ge H$, where

$$S^{\pm}(k) = \begin{cases} 0, & \text{if } k = 0, \\ \max\left\{0,\ S^{\pm}(k-1) \pm [T_k^2(r) - \widehat{\mu}_{T^2(r)}] - K\right\}, & \text{if } k = 1,2,\ldots. \end{cases}$$

Figure 2. Algorithmic description of WDFTC.

3.1. *Dimension reduction*

In this subsection, we discuss WDFTC's dimension-reduction technique. Jin and Shi (1999) use a universal thresholding scheme for wavelet shrinkage, but such a scheme assumes uncorrelated normal components and thus does not always work for non-normal components. Instead, we propose an extension of the method of Lada *et al.* (2002) that exploits the concept of weighted relative reconstruction error. We seek to select a (relatively) small number p of the components of $\boldsymbol{\theta}_0 = (\theta_{1,0},\ldots,\theta_{n,0})^{\mathrm{T}} = \mathbf{W}\boldsymbol{f}_0$, including all 2^L scaling coefficients and the $p - 2^L$ largest-magnitude detail coefficients (provided $p > 2^L$); and the modified vector $\boldsymbol{\theta}_0^{\#} = (\theta_{1,0}^{\#},\ldots,\theta_{n,0}^{\#})^{\mathrm{T}}$

is obtained from $\boldsymbol{\theta}_0$ by setting to zero the $n-p$ non-selected components of $\boldsymbol{\theta}_0$ so that the reconstructed vector $\boldsymbol{f}_0^{\#} = \mathbf{W}^{-1}\boldsymbol{\theta}_0^{\#}$ is a sufficiently accurate approximation to $\boldsymbol{f}_0$. In the following discussion, we will write $\boldsymbol{\theta}_0^{\#}$ and $\boldsymbol{f}_0^{\#}$ as $\boldsymbol{\theta}_0^{\#}(p)$ and $\boldsymbol{f}_0^{\#}(p)$, respectively, to emphasize the dependence of these vectors on p. When we use $\boldsymbol{f}_0^{\#}(p)$ as an approximation to $\boldsymbol{f}_0$, the relative reconstruction error is $\|\boldsymbol{f}_0^{\#}(p)-\boldsymbol{f}_0\|_2/\|\boldsymbol{f}_0\|_2$ (Lada *et al.* 2002); and the corresponding data-compression ratio is p/n. For a given value of $p \in \{2^L,\ldots,n\}$ and weight $q \in [0,1]$ assigned to the data-compression ratio, we define the weighted relative reconstruction error (WRRE) as follows:

$$\begin{aligned}\mathrm{WRRE}(p;\boldsymbol{f}_0,q) &= (1-q)\left[\frac{\|\mathbf{W}^{-1}\boldsymbol{\theta}_0^{\#}(p)-\boldsymbol{f}_0\|_2}{\|\boldsymbol{f}_0\|_2}\right]+q\left(\frac{p}{n}\right)\\ &= (1-q)\left[\frac{\|\boldsymbol{f}_0^{\#}(p)-\boldsymbol{f}_0\|_2}{\|\boldsymbol{f}_0\|_2}\right]+q\left(\frac{p}{n}\right); \qquad (7)\end{aligned}$$

and we choose p (and implicitly, $\boldsymbol{\theta}_0^{\#}(p)$) to minimize $\mathrm{WRRE}(p;\boldsymbol{f}_0,q)$,

$$p = \underset{u=2^L,\ldots,n}{\arg\min}\ \mathrm{WRRE}(u;\boldsymbol{f}_0,q)\,. \qquad (8)$$

Remark 3.1 : There is a potential problem in using the dimension-reduction scheme of equations (7) and (8) if $\mathbf{1}_n^{\mathrm{T}}\boldsymbol{f}_0 \neq 0$ and all the components of $\boldsymbol{f}_0$ have large magnitudes. In this situation, the relative reconstruction error can be negligibly small in comparison with the data-compression ratio for all feasible values of p so that equation (8) yields $p = 2^L$; and then the only non-zero components of $\boldsymbol{\theta}_0^{\#}(p)$ are the scaling coefficients in $\boldsymbol{\theta}_0$, which can yield a low-resolution approximation to $\boldsymbol{f}_0$. The centring condition $\mathbf{1}_n^{\mathrm{T}}\boldsymbol{f}_0 = 0$ avoids this problem.

In the formulation of $\mathrm{WRRE}(p;\boldsymbol{f}_0,q)$ given by equation (7), the weight q can be adjusted to achieve an effective trade-off between the relative reconstruction error and the data-compression ratio. In many applications of profile monitoring, the reduced dimension p must be sufficiently small to ensure that the Hotelling's statistics $\{T_k^2(r) : k = 1,2,\ldots\}$ computed in Phase II have adequate power to detect shifts in the mean profile. On the other hand, p must be sufficiently large so that the selected scaling and detail coefficients in the DWT of an out-of-control profile can accurately represent deviations from the in-control mean profile. Setting the weight $q = 0.5$ yields the same value of p as for the method of Lada *et al.* (2002). For profiles of moderate dimension (that is, $n \leq 1{,}000$), we found that $q = 0.5$ generally yielded satisfactory results. On the other hand, for profiles of dimension $n > 1{,}000$, we found that $q > 0.5$ was required to obtain acceptable results. In this article, we use $q = 0.7$ to handle profiles of dimension $n = 2{,}048$.

The effectiveness of the dimension-reduction scheme in WDFTC also depends on the coarsest level of resolution, L, based on the application at hand. For the choice of L to be used with WDFTC, we adapt the approach of Lada and Wilson (2006) and use the default value $L = \lceil J/2 \rceil$, where $J = \log_2(n)$. In some cases we also use slightly smaller values of L than the default value (for example, $\lceil J/2 \rceil - 1$ or $\lceil J/2 \rceil - 2$), but only if such values of L yield a meaningful dimension reduction compared with that of the default value.

In some applications, the mean in-control profile $\boldsymbol{f}_0$ and its DWT $\boldsymbol{\theta}_0$ may not be known exactly. To estimate $\boldsymbol{f}_0$ in such cases, we use the centred sample mean

$$\widehat{\boldsymbol{f}_0} = \left(\mathbf{I}_n - n^{-1}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}\right)\left(N^{-1}\sum_{j=1}^{N}\boldsymbol{Y}_j\right)$$

of the profiles observed in Phase I. Moreover from the DWT $\widehat{\boldsymbol{\theta}}_0 = \mathbf{W}\widehat{\boldsymbol{f}}_0$, we obtain the associated estimators $\widehat{\boldsymbol{\theta}}_0^{\#}(p)$ and $\widehat{\boldsymbol{f}}_0^{\#}(p) = \mathbf{W}^{-1}\widehat{\boldsymbol{\theta}}_0^{\#}(p)$ to be used in equations (7) and (8) of the dimension-reduction scheme as well as the estimator $\widehat{\boldsymbol{\vartheta}}_0^{\#}$ to be used in computing Hotelling's statistic $T_k^2(r) = \left[\boldsymbol{d}_k^{\#}(r) - \widehat{\boldsymbol{\vartheta}}_0^{\#}\right]^{\mathrm{T}} \left[\widetilde{\boldsymbol{\Lambda}}_0^{\#}(r)\right]^{-1} \left[\boldsymbol{d}_k^{\#}(r) - \widehat{\boldsymbol{\vartheta}}_0^{\#}\right]$ for $k = 1, 2, \ldots$ in both Phases I and II of WDFTC.

3.2. *Covariance-matrix regularisation*

In this section, we explain the covariance-matrix regularisation step **[2]** of WDFTC that is applied to the sample covariance matrix $\widehat{\boldsymbol{\Lambda}}_0^{\#} = (N-1)^{-1} \sum_{j=1}^{N} \left(\boldsymbol{d}_j^{\#} - \overline{\boldsymbol{d}}_N^{\#}\right)\left(\boldsymbol{d}_j^{\#} - \overline{\boldsymbol{d}}_N^{\#}\right)^{\mathrm{T}}$ of the reduced-dimension DWTs $\{\boldsymbol{d}_j^{\#} : j = 1, \ldots, N\}$ computed from the profiles observed in Phase I, where $\overline{\boldsymbol{d}}_N^{\#} = N^{-1} \sum_{j=1}^{N} \boldsymbol{d}_j^{\#}$. Commenting on the wavelet-based method of Jin and Shi (2001) for diagnosis of process faults, Woodall *et al.* (2004) state that the use of Hotelling's T^2 statistic may not be efficient because high correlations between the components of each profile $\boldsymbol{Y}_j$ may lead to over-parameterisation — that is, an excessive value for the dimension p of the $\{\boldsymbol{d}_j^{\#}\}$. Moreover if $p > 200$, then estimating the $p \times p$ covariance matrix $\boldsymbol{\Lambda}_0^{\#}$ can also be difficult, especially if there is a limited amount of Phase I (training) data (see, for example, Hoffbeck and Landgrebe (1996), Daniels and Kass (2001), and Ledoit and Wolf (2002)). In particular if the size N of the Phase I data set is insufficient or the joint distribution of each in-control random vector $\boldsymbol{d}_j^{\#}$ is singular, then $\widehat{\boldsymbol{\Lambda}}_0^{\#}$ is not guaranteed to be positive definite so that the associated Hotelling's T^2 statistic is not guaranteed to exist.

In this article we make the following (mild) assumptions: (1) the $n \times 1$ profile vector $\boldsymbol{Y}_j$ has a non-singular joint probability density function that depends on the current in-control or out-of-control condition; and (2) the covariance matrix $\mathrm{Cov}[\boldsymbol{Y}_j]$ is the same for both in-control and out-of-control conditions. Under assumptions (1) and (2), different profile components may have different continuous marginal distributions that may be non-normal. In this broadly applicable setting if $N \geq p + 1$, then $\widehat{\boldsymbol{\Lambda}}_0^{\#}$ is positive definite with probability one (see, for example, Proposition 2 of Porta Nova and Wilson (1989)).

To avoid problems with Hotelling's T^2 statistic in situations for which $p > 200$, we adapt the covariance-regularisation method of Bickel and Levina (2008) and use $\widetilde{\boldsymbol{\Lambda}}_0^{\#}$, the resulting thresholded version of $\widehat{\boldsymbol{\Lambda}}_0^{\#}$ in WDFTC. Although the main asymptotic results of Bickel and Levina (2008) are based on the assumption that the profiles $\{\boldsymbol{Y}_j\}$ are randomly sampled from a Gaussian (normal) or sub-Gaussian distribution, we have found the authors' approach to be useful in formulating a covariance-matrix regularisation procedure for WDFTC that is reasonably robust against violations of the normality assumption. As we shall see in Section 3.3, the batch-size determination Algorithm BSD is also designed to avoid large departures from normality in the basic random vectors from which the relevant Hotelling's T^2 statistic is computed.

In the context of profile monitoring with WDFTC, the basic idea of the covariance-matrix regularisation method of Bickel and Levina (2008) is that if p and N are sufficiently large and $\log(p)/N$ is sufficiently small, then the $p \times p$ sample covariance matrix $\widehat{\boldsymbol{\Lambda}}_0^{\#}$ can be (hard) thresholded at a positive level τ depending on N and p such that with high probability, the thresholded covariance matrix $\widetilde{\boldsymbol{\Lambda}}_0^{\#}$ is positive definite and close to the theoretical covariance matrix $\boldsymbol{\Lambda}_0^{\#} = \mathrm{E}\left[\left(\boldsymbol{d}_j^{\#} - \mathrm{E}[\boldsymbol{d}_j^{\#}]\right)\left(\boldsymbol{d}_j^{\#} - \mathrm{E}[\boldsymbol{d}_j^{\#}]\right)^{\mathrm{T}}\right]$ in a certain sense. We adapt the thresholding scheme of Bickel and Levina (2008) to WDFTC so that when it is applied to $\widehat{\boldsymbol{\Lambda}}_0^{\#}$, the following elements remain intact (i.e., are not subject to the thresholding operation): (1) the $2^L \times 2^L$ submatrix of sample covariances of the estimated scaling coefficients (i.e., $\left[\widehat{\boldsymbol{\Lambda}}_0^{\#}\right]_{u,v}$ for $u, v = 1, \ldots, 2^L$);

and (2) the diagonal elements (i.e., $\big[\widehat{\mathbf{\Lambda}}_0^{\#}\big]_{u,u}$ for $u = 1,\ldots,p$). With the threshold τ, WDFTC's covariance-regularisation scheme maps $\widehat{\mathbf{\Lambda}}_0^{\#}$ into the matrix $\mathbf{R}\big(\widehat{\mathbf{\Lambda}}_0^{\#};L,\tau\big)$ whose (u,v) element is

$$\big[\mathbf{R}\big(\widehat{\mathbf{\Lambda}}_0^{\#};L,\tau\big)\big]_{u,v} = \begin{cases} \big[\widehat{\mathbf{\Lambda}}_0^{\#}\big]_{u,v}, & \text{if } (u \le 2^L \text{ and } v \le 2^L) \text{ or } (u = v) \\ \big[\widehat{\mathbf{\Lambda}}_0^{\#}\big]_{u,v}\mathbb{I}\big(\big|\big[\widehat{\mathbf{\Lambda}}_0^{\#}\big]_{u,v}\big| \ge \tau\big), & \text{otherwise,} \end{cases} \tag{9}$$

where $\mathbb{I}(\cdot)$ is the indicator function. Algorithm CMR in Figure 3 determines the estimated threshold $\widehat{\tau}$ and the 'regularised' version $\widetilde{\mathbf{\Lambda}}_0^{\#}$ of the sample covariance matrix $\widehat{\mathbf{\Lambda}}_0^{\#}$ based on that threshold.

Algorithm CMR

[1] Divide the Phase I data set into two disjoint subsets of size $N_1 = \lfloor 0.4N \rfloor$ and $N_2 = N - N_1$, respectively.

[2] For $\ell = 1,2$, calculate the sample covariance matrix $\widehat{\mathbf{\Lambda}}_0^{\#}(\ell,N_\ell)$ from the Phase I data subset of size N_ℓ.

[3] Compute the estimated threshold,

$$\widehat{\tau} = \operatorname*{arg\,min}_{\tau \ge 0} \sum_{\substack{(u \ne v) \text{ and} \\ (2^L < u \text{ or } 2^L < v)}} \Big\{ [\widehat{\mathbf{\Lambda}}_0^{\#}(1,N_1)]_{u,v}\mathbb{I}[|[\widehat{\mathbf{\Lambda}}_0^{\#}(1,N_1)]_{u,v}| \ge \tau] - [\widehat{\mathbf{\Lambda}}_0^{\#}(2,N_2)]_{u,v} \Big\}^2.$$

[4] Calculate the sample covariance matrix $\widehat{\mathbf{\Lambda}}_0^{\#}$ using the entire Phase I data set of size N and apply the covariance-regularisation map of equation (9) to $\widehat{\mathbf{\Lambda}}_0^{\#}$ using the threshold $\widehat{\tau}$:

$$\widetilde{\mathbf{\Lambda}}_0^{\#} = \mathbf{R}(\widehat{\mathbf{\Lambda}}_0^{\#};L,\widehat{\tau}).$$

Figure 3. Algorithmic description of CMR.

Remark 3.2 : The estimated threshold $\widehat{\tau}$ can also be interpreted as the minimal magnitude for the sample covariances in $\widehat{\mathbf{\Lambda}}_0^{\#}$ to be considered 'significant'; and this interpretation will play an important role in Algorithm BSD for determining the batch size r as detailed in the next subsection.

Remark 3.3 : After a suitable batch size r is obtained from Algorithm BSD, we use $\widetilde{\mathbf{\Lambda}}_0^{\#}(r) = \widetilde{\mathbf{\Lambda}}_0^{\#}/r$ as our estimator of the covariance matrix of the reduced-dimension batch-means DWTs $\{\boldsymbol{d}_k^{\#}(r) : k = 1,\ldots,\lfloor N/r \rfloor\}$ computed in Phase I; and then in both Phases I and II of WDFTC, we use $\widetilde{\mathbf{\Lambda}}_0^{\#}(r)$ to calculate the Hotelling's statistic $T_k^2(r) = \big[\boldsymbol{d}_k^{\#}(r) - \widehat{\boldsymbol{\vartheta}}_0^{\#}\big]^{\mathrm{T}} \big[\widetilde{\mathbf{\Lambda}}_0^{\#}(r)\big]^{-1} \big[\boldsymbol{d}_k^{\#}(r) - \widehat{\boldsymbol{\vartheta}}_0^{\#}\big]$ for $k = 1,2,\ldots$.

3.3. *Batch size determination*

In this subsection, we explain the method used in WDFTC to determine the batch size r. In our experience, we have found that excessive covariances between the components of the dimension-reduced DWTs $\{\boldsymbol{d}_j^{\#}\}$ can seriously distort the performance of a profile-monitoring chart based on a Hotelling's T^2–type statistic computed from the $\{\boldsymbol{d}_j^{\#}\}$ obtained in Phase I of the chart's operation. In this situation we have obtained substantial improvements in the performance of WDFTC

by reducing the magnitudes of the covariances between pairs of components of the dimension-reduced DWTs to manageable levels. The desired covariance reductions are achieved indirectly by aggregating the observed profiles $\{\boldsymbol{Y}_j : j = 1,2,\ldots\}$ into non-overlapping batches of size r so that the associated non-overlapping batch means $\left\{\boldsymbol{Y}_k(r) = r^{-1}\sum_{u=1}^{r}\boldsymbol{Y}_{(k-1)r+u} : k = 1,2,\ldots\right\}$ yield batch-means DWT vectors $\left\{\boldsymbol{d}_k(r) = \mathbf{W}\boldsymbol{Y}_k(r) : k = 1,2,\ldots\right\}$ for which $\mathrm{Cov}[\boldsymbol{d}_k(r)] = \mathrm{Cov}[\boldsymbol{d}_j]/r = \boldsymbol{\Lambda}_0/r$ and $\mathrm{Cov}[\boldsymbol{d}_k^{\#}(r)] = \mathrm{Cov}[\boldsymbol{d}_j^{\#}]/r = \boldsymbol{\Lambda}_0^{\#}/r$, where r is taken to be just large enough to achieve effective covariance reductions. The formal statement of Algorithm BSD is given in Figure 4.

Algorithm BSD

[1] Apply Algorithm CMR to obtain $\widehat{\tau}$, the estimated threshold, and $\widetilde{\boldsymbol{\Lambda}}_0^{\#}$, the regularised (thresholded) version of $\widehat{\boldsymbol{\Lambda}}_0^{\#}$.

[2] Let Q denote the number of non-zero off-diagonal elements of $\widetilde{\boldsymbol{\Lambda}}_0^{\#}$, excluding the estimated covariances between pairs of scaling coefficients,

$$Q = \sum_{\substack{(u\neq v) \text{ and} \\ (2^L<u \text{ or } 2^L<v)}} \mathbb{I}([\widetilde{\boldsymbol{\Lambda}}_0^{\#}]_{u,v} \neq 0)\,.$$

[2a] If $Q = 0$, then return $r \leftarrow 1$ and stop; otherwise, go to step **[2b]**.

[2b] Calculate the average magnitude ζ of the non-zero off-diagonal elements of $\widetilde{\boldsymbol{\Lambda}}_0^{\#}$, excluding the estimated covariances between pairs of scaling coefficients,

$$\zeta = \frac{1}{Q}\sum_{\substack{(u\neq v) \text{ and} \\ (2^L<u \text{ or } 2^L<v)}} |[\widetilde{\boldsymbol{\Lambda}}_0^{\#}]_{u,v}|\,.$$

[3] Set the batch size $r \leftarrow \lceil \sqrt{2}\zeta/\widehat{\tau} \rceil$ and stop.

Figure 4. Algorithmic description of BSD.

The basic idea of Algorithm BSD is first to compute the average magnitude of the elements of the regularised sample covariance matrix $\widetilde{\boldsymbol{\Lambda}}_0^{\#}$ as delivered by Algorithm CMR, where the average is taken only over the elements that were subjected to the thresholding operation and survived in Algorithm CMR; then the ratio of this average to the estimated threshold $\widehat{\tau}$ is an estimate of the batch size r necessary to reduce the magnitudes of all relevant covariances between pairs of components of the reduced-dimension batch-means vector $\boldsymbol{d}_j^{\#}(r)$ to 'non-significant' levels.

Remark 3.4 : Algorithm BSD is designed to yield a batch size r sufficiently large so that all the off-diagonal elements of the regularised sample covariance matrix $\widetilde{\boldsymbol{\Lambda}}_0^{\#}(r) = \widetilde{\boldsymbol{\Lambda}}_0^{\#}/r$ have sufficiently small magnitudes to avoid aberrant behaviour of the profile-monitoring statistic $T_k^2(r)$. In particular, the inflation factor $\sqrt{2}$ in step **[3]** of Algorithm BSD yields a batch size $r > 1$ for most processes, provided that Algorithm CMR delivers at least one non-zero off-diagonal element in the regularised sample covariance matrix $\widetilde{\boldsymbol{\Lambda}}_0^{\#}$, excluding the estimated covariances between pairs of scaling coefficients.

Remark 3.5 : When the true covariance matrix $\boldsymbol{\Lambda}_0^{\#}(r)$ is used to calculate the profile-monitoring statistic $T_k^2(r)$, then we have the in-control mean $\mathrm{E}[T_k^2(r)] = p$ regardless of the distribution of the profiles $\{\boldsymbol{Y}_j\}$, provided that the latter distribution is non-singular. Thus one can check if the regularised matrix $\widetilde{\boldsymbol{\Lambda}}_0^{\#}(r)$ is a good estimate of $\boldsymbol{\Lambda}_0^{\#}(r)$ by comparing the sample average of the in-control statistics $\{T_k^2(r) : k = 1,\ldots,\lfloor N/r \rfloor\}$ with the corresponding theoretical mean value p.

4. Experiments

In this section, we present experimental results for WDFTC in comparison with other existing profile-monitoring charts. The following three charts are considered: (1) HTW_n, the classical Hotelling's T^2 chart based on the full $n \times 1$ vector of wavelet coefficients for each observed profile; (2) HTW_p, a reduced-dimension variant of HTW_n that is based on p preselected wavelet coefficients for each observed profile as detailed below; and (3) the M^* chart of Chicken *et al.* (2009) as described in the second section of this article. Concise summaries of the steps of procedures HTW_n and HTW_p are given below.

Chart HTW_n: Compute the exact covariance matrix $\mathbf{\Lambda}_0 = \mathbf{W}\mathbf{\Sigma}_0\mathbf{W}^{\text{T}}$ for the DWTs $\{\boldsymbol{d}_j : j = 1,\ldots,N\}$ of the profiles observed in Phase I, where $\mathbf{\Sigma}_0$ is assumed to be known. In terms of the prespecified false-alarm rate $\text{FAR} = 1/\text{ARL}_0$, calculate the upper control limit UCL_1 for the 'ideal' profile-monitoring statistic $T_j^2 = (\boldsymbol{d}_j - \boldsymbol{\theta}_0)^{\text{T}}\ \mathbf{\Lambda}_0^{-1}\ (\boldsymbol{d}_j - \boldsymbol{\theta}_0)$ as the $1 - \text{FAR}$ quantile of the chi-squared distribution with n degrees of freedom. Therefore UCL_1 is the solution of the equation $\Pr\{\chi_n^2 \le \text{UCL}_1\} = 1 - \text{FAR}$, where χ_n^2 denotes a chi-squared random variable with n degrees of freedom. After the jth profile is observed in Phase II, an out-of-control alarm is raised if $T_j^2 > \text{UCL}_1$.

Chart HTW_p: Compute the exact covariance matrix $\mathbf{\Lambda}_0 = \mathbf{W}\mathbf{\Sigma}_0\mathbf{W}^{\text{T}}$ for the DWTs $\{\boldsymbol{d}_j : j = 1,\ldots,N\}$ of the profiles observed in Phase I, where $\mathbf{\Sigma}_0$ is assumed to be known. Select the p largest-magnitude components of the DWT $\boldsymbol{\theta}_0 = \mathbf{W}\boldsymbol{f}_0$ of the mean in-control profile; and for the corresponding $p \times 1$ subvectors $\{\boldsymbol{d}_j^{\#} : j = 1,\ldots,N\}$ extracted from the DWTs of the profiles observed in Phase I, let $\mathbf{\Lambda}_0^{\#}$ denote the associated covariance matrix (a submatrix of $\mathbf{\Lambda}_0$). In terms of the prespecified false-alarm rate $\text{FAR} = 1/\text{ARL}_0$, calculate the upper control limit UCL_2 for the 'ideal' reduced-dimension profile-monitoring statistic $T_j^2 = \left(\boldsymbol{d}_j^{\#} - \boldsymbol{\theta}_0^{\#}\right)^{\text{T}}\left(\mathbf{\Lambda}_0^{\#}\right)^{-1}\left(\boldsymbol{d}_j^{\#} - \boldsymbol{\theta}_0^{\#}\right)$ as the $1 - \text{FAR}$ quantile of the chi-squared distribution with p degrees of freedom. Therefore UCL_2 is the solution of the equation $\Pr\{\chi_p^2 \le \text{UCL}_2\} = 1 - \text{FAR}$. An out-of-control alarm is raised after the jth profile if $T_j^2 > \text{UCL}_2$.

Remark 4.1 : The p components of $\boldsymbol{\theta}_0$ that are selected for use in HTW_p may be different from the p components of $\boldsymbol{\theta}_0$ that minimize the weighted relative reconstruction error defined by equation (7).

In all the experiments reported below, we used the exact values of the covariance matrices $\mathbf{\Sigma}_0$, $\mathbf{\Lambda}_0$, and $\mathbf{\Lambda}_0^{\#}$ as required for procedures HTW_n and HTW_p. Recall that procedure M^* estimates σ_* from the average of the median absolute deviations of the $n/2$ coefficients at the highest levels of resolution for each of the profiles observed so far in Phase II operation. Moreover, WDFTC uses the regularised sample covariance matrix, $\widetilde{\mathbf{\Lambda}}_0^{\#}$, computed from the Phase I data set of size N. In this respect the procedures HTW_n and HTW_p have some advantage over procedures M^* and WDFTC in the experimental performance evaluation that may not carry over to practical applications in which $\mathbf{\Sigma}_0$, $\mathbf{\Lambda}_0$, and $\mathbf{\Lambda}_0^{\#}$ are unknown and must be estimated from a Phase I data set. For profiles of dimension $n = 512$, we applied WDFTC with a Phase I data set of size $N = 3{,}000$; and for profiles of dimension $n = 2{,}048$, we applied WDFTC with a Phase I data set of size $N = 5{,}000$.

In the first part of the experimental performance evaluation of WDFTC and its competitors HTW_n, HTW_p, and M^*, we applied those procedures to both normal and non-normal profiles having both independent and correlated components such that the mean in-control profile $\boldsymbol{f}_0$ is defined by $n = 512$ equally spaced points on Mallat's piecewise smooth function as depicted in Figure 1. In the second part of the experimental performance evaluation, we applied WDFTC to a lumber manufacturing process (Staudhammer 2004) in which the mean in-control profile had $n = 2{,}048$ points. In both applications, we estimated the relevant in-control and out-of-

control ARLs based on 1,000 independent replications of each test process. Corresponding to each table of estimated ARLs given in this section, there is a matching table of standard errors for those estimated ARLs that is given in the Online Supplement to this article. The data set for Mallat's piecewise smooth function and the MATLAB codes (Hanselman and Littlefield 2001) needed to reproduce the results presented in Section 4.1 below are available online via http://www.ise.ncsu.edu/jwilson/files/wdftc-codes.zip.

4.1. *Profiles based on Mallat's piecewise smooth function*

In the experiments with the mean in-control profile $\boldsymbol{f}_0$ based on Mallat's piecewise smooth function, we set the target value of $\text{ARL}_0 = 200$. The mean out-of-control profile has the form $\boldsymbol{f}_1 = \boldsymbol{f}_0 + \eta\boldsymbol{\Delta}\boldsymbol{\sigma}$, where: (1) the shift-size parameter $\eta \in \{0.25, 0.5, 0.75, 1, 2\}$; (2) the $n \times n$ shift-sign matrix $\boldsymbol{\Delta} = \text{diag}(\delta_1, \ldots, \delta_n)$ is a diagonal matrix with $\delta_i \in \{-1, 0, 1\}$ for $i = 1, \ldots, n$; and (3) $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_n)^{\text{T}}$ is the vector of marginal standard deviations of the respective components of $\boldsymbol{\varepsilon}_j$. Whereas procedure M^* is based on the Haar wavelet system, we used the symmlet 8 wavelet system in procedures WDFTC, HTW_n, and HTW_p. Because $n = 512$, the highest level of resolution $J = \log_2(n) = 9$; and selecting the coarsest level of resolution $L = \lceil J/2 \rceil = 5$ and the weight $q = 0.5$ in equation (7) for the weighted relative reconstruction error, we obtain the 'optimal' reduced dimension $p = 62$ from equation (8). To make a fair comparison of WDFTC with HTW_p, we also set $p = 62$ in the latter chart.

In the following tables, Global Shift 1 refers to the situation in which $\delta_i = 1$ for $i = 1, \ldots, n$ so that there is a positive shift of size $\eta\sigma_i$ in the ith component of the mean profile for $i = 1, \ldots, n$. By contrast, Global Shift 2 refers to the situation in which $\delta_i = 1$ for $i = 1, \ldots, n/2$ and $\delta_i = -1$ for $i = (n/2)+1, \ldots, n$; therefore in the first half of the components of the mean profile, there are positive shifts of the respective amounts $\eta\sigma_1, \ldots, \eta\sigma_{n/2}$, and in the last half of the components of the mean profile there are negative shifts of the respective amounts $-\eta\sigma_{(n/2)+1}, \ldots, -\eta\sigma_n$. Local Shift 1 is specified as follows: $\delta_i = 1$ for $i \in \mathscr{A}_1 = \{73, 74, \ldots, 76\} \cup \{288, 289, \ldots, 296\}$, and $\delta_i = 0$ for $i \notin \mathscr{A}_1$. Therefore with Local Shift 1, the 13 selected components of the mean profile are increased by the respective amounts $\eta\sigma_i$ for $i \in \mathscr{A}_1$, while all other components of the mean profile remain unchanged. Local Shift 2 is specified as follows: $\delta_i = 1$ for $i \in \mathscr{A}_2 = \{3, 4, \ldots, 15\} \cup \{344, 345, \ldots, 347\}$, and $\delta_i = 0$ for $i \notin \mathscr{A}_2$. Therefore with Local Shift 2, the 17 selected components of the mean profile are increased by the respective amounts $\eta\sigma_i$ for $i \in \mathscr{A}_2$, while other components of the mean profile remain unchanged.

4.1.1. Multivariate normal errors

Most existing profile-monitoring charts assume that the observed profiles $\{\boldsymbol{Y}_j : j = 1, 2, \ldots\}$ are i.i.d. multivariate normal vectors with a common marginal variance and zero correlations between each pair of components. With $\boldsymbol{f}_0$ based on Mallat's piecewise smooth function, we first consider the following three cases in which the error vector $\boldsymbol{\varepsilon}_j$ is multivariate normal with mean $\boldsymbol{0}_n$ and covariance matrix $\boldsymbol{\Sigma}_0$: (1) the components of $\boldsymbol{\varepsilon}_j$ are independent standard normal random variables so that $\boldsymbol{\Sigma}_0 = \mathbf{I}_n$; (2) the components of $\boldsymbol{\varepsilon}_j$ are correlated standard normal random variables with common correlation 0.5 so that $\boldsymbol{\Sigma}_0$ has all its diagonal elements equal to 1.0 and all its off-diagonal elements equal to 0.5; and (3) the components of $\boldsymbol{\varepsilon}_j$ are correlated normal random variables with mean zero, marginal variances given by equation (3), and pairwise correlations given by equation (2) so that $\left[\boldsymbol{\Sigma}_0\right]_{u,v} = \sigma_u\sigma_v\rho(u-v)$ for $u, v = 1, \ldots, n$ as for the test processes ME_1 and ME_2.

Case (1): Error vector has independent standard normal components. Table 4 shows the values of ARL_0 and ARL_1 delivered by WDFTC and its competitors for Case (1). WDFTC required the average batch size $\bar{r} = 3$ observed profiles. (Henceforth, the term 'observation' will be

used to mean a single observed profile.) All four charts yielded values for ARL_0 close to the target value of 200 observations. To detect Global Shifts 1 and 2 in Phase II (regular) operation with $\eta > 0.25$, WDFTC required one vector of batch means (that is, 3 observations), whereas each of the other charts required 1 observation. To detect Global Shifts 1 and 2 in Phase II operation with $\eta = 0.25$, WDFTC sometimes required two batch means so that on average WDFTC required about 4 observations; by contrast, HTW_n required about 16 observations, while HTW_p and M^* each required about 2 observations. For the Local Shifts 1 and 2 with $0.25 \leq \eta \leq 1$, WDFTC significantly outperformed all the other charts, and HTW_n usually delivered the worst performance. The latter conclusion is not surprising, because high dimensionality degrades the performance of Hotelling's T^2–type charts (Fan 1996). For Local Shifts 1 and 2 with $0.25 \leq \eta \leq 1$, the performance of M^* was often similar to that of HTW_n and was always much worse than that of WDFTC. For example, to detect Local Shift 1 with $\eta = 0.5$, charts M^* and HTW_n each required about 145 observations, while WDFTC required about 35 observations.

Table 4. ARLs for error vector with independent standard normal components.

Shift Type	Shift Size	WDFTC $\bar{r} = 3$	HTW_n	HTW_p	M^*
In-Control	0	189.97	210.56	197.81	196.07
Global Shift 1	0.25	3.80	16.13	2.22	2.12
	0.5	3.00	1.18	1.00	1.04
	0.75	3.00	1.00	1.00	1.00
	1	3.00	1.00	1.00	1.00
	2	3.00	1.00	1.00	1.00
Global Shift 2	0.25	3.86	16.32	2.34	1.42
	0.5	3.00	1.15	1.00	1.05
	0.75	3.00	1.00	1.00	1.00
	1	3.00	1.00	1.00	1.00
	2	3.00	1.00	1.00	1.00
Local Shift 1	0.25	114.41	191.78	164.74	183.95
	0.5	35.04	145.24	95.74	145.35
	0.75	16.04	101.19	44.07	88.77
	1	9.47	65.16	18.58	43.98
	2	3.06	6.35	1.39	2.65
Local Shift 2	0.25	112.09	179.01	166.38	197.08
	0.5	33.06	135.86	89.34	131.01
	0.75	15.14	84.56	40.69	73.69
	1	8.82	49.35	15.57	28.19
	2	3.04	3.16	1.31	2.05

Case (2): Error vector has correlated standard normal components. Table 5 shows the values of ARL_0 and ARL_1 delivered by WDFTC and its competitors for Case (2). As we saw in Case (1), WDFTC required the average batch size $\bar{r} = 3$ observations, and all four charts yielded values for ARL_0 close to the target value of 200 observations. However for Global Shift 1 and all levels of η, the introduction of a common correlation of 0.5 significantly increased the value of ARL_1 for all four charts compared with the results for Case (1). For example, in Case (1) to detect Global Shift 1 with $\eta = 0.5$, WDFTC required about 3 observations while HTW_n, HTW_p, and M^* each required about 1 observation; by contrast in Case (2) the corresponding values of ARL_1 for WDFTC, HTW_n, HTW_p, and M^* were about 134, 190, 180, and 86 observations, respectively. Overall in Case (2) for Global Shift 1, M^* significantly outperformed WDFTC, which in turn significantly outperformed HTW_n and HTW_p. To detect Global Shift 2 at all levels of η, WDFTC required about 3 observations while HTW_p required about 1 observation; on

the other hand, the values of ARL_1 for HTW_n ranged from approximately 4 observations (for $\eta = 0.25$) to approximately 1 observation (for $\eta > 0.25$), and the values of ARL_1 for M^* ranged from approximately 9 observations (for $\eta = 0.25$) to approximately 2 observations (for $\eta = 2$). For Local Shifts 1 and 2 and all levels of η, WDFTC substantially outperformed M^*; and for Local Shifts 1 and 2 with $0.25 \leq \eta \leq 0.75$, WDFTC substantially outperformed HTW_p, which in turn outperformed HTW_n. For example, to detect Local Shift 2 with $\eta = 0.5$, WDFTC, HTW_n, HTW_p, and M^* required approximately 18, 93, 48, and 199 observations, respectively.

Table 5. ARLs for error vector with correlated standard normal components.

Shift Type	Shift Size	WDFTC $\bar{r} = 3$	HTW_n	HTW_p	M^*
In-Control	0	188.73	210.65	198.30	200.48
Global Shift 1	0.25	174.96	210.21	197.38	153.84
	0.5	134.04	189.50	180.23	85.59
	0.75	78.60	171.14	149.24	39.77
	1	47.08	163.73	110.20	20.75
	2	12.46	94.86	29.09	3.02
Global Shift 2	0.25	3.01	3.79	1.07	8.84
	0.5	3.01	1.15	1.00	5.11
	0.75	3.01	1.00	1.00	3.89
	1	3.01	1.00	1.00	2.97
	2	3.01	1.00	1.00	1.97
Local Shift 1	0.25	71.47	177.53	140.86	200.55
	0.5	18.54	112.10	50.69	203.18
	0.75	8.75	58.10	16.30	198.04
	1	5.56	25.88	4.95	196.10
	2	3.01	1.54	1.00	168.72
Local Shift 2	0.25	65.51	163.70	135.48	201.31
	0.5	17.57	92.67	47.87	198.74
	0.75	8.38	43.98	13.69	197.23
	1	5.41	15.86	4.44	196.54
	2	3.01	1.13	1.00	164.76

Case (3): Error vector has a general normal distribution. Table 6 shows the values of ARL_0 and ARL_1 delivered by WDFTC and its competitors for Case (3). For this test process, WDFTC required the average batch size $\bar{r} = 8$ observations. As we saw in Cases (1) and (2), all four charts yielded values for ARL_0 close to the target value of 200 observations. Because of the batching operation, WDFTC usually required at least 8 observations to detect shifts of any type. For Global Shift 1 with all levels of η, HTW_n and HTW_p outperformed WDFTC, and WDFTC substantially outperformed M^*. For example, to detect Global Shift 1 with $\eta = 0.5$, WDFTC, HTW_n, HTW_p, and M^* delivered ARL_1 values of approximately 8, 1, 1, and 175 observations, respectively. To detect Global Shift 2 at all levels of η, WDFTC required about 8 observations, while HTW_p required about 1 observation; on the other hand the values of ARL_1 for HTW_n ranged from approximately 3 observations (for $\eta = 0.25$) to approximately 1 observation (for $\eta > 0.25$), and the values of ARL_1 for M^* ranged from 5 observations (for $\eta = 0.25$) to 1 observation (for $\eta = 2$). For Local Shifts 1 and 2 and all levels of η, WDFTC substantially outperformed M^*; and for Local Shifts 1 and 2 with $0.25 \leq \eta \leq 0.75$, WDFTC substantially outperformed HTW_p, which in turn outperformed HTW_n. For example, to detect Local Shift 2 with $\eta = 0.5$, WDFTC, HTW_n, HTW_p, and M^* required approximately 12, 73, 39, and 201 observations, respectively.

Table 6. ARLs for error vector with a general normal distribution.

Shift Type	Shift Size	WDFTC $\bar{r}=8$	HTW_n	HTW_p	M^*
In-Control	0	198.99	210.65	201.02	201.44
Global Shift 1	0.25	8.01	3.00	1.04	197.46
	0.5	8.01	1.00	1.00	175.29
	0.75	8.01	1.00	1.00	123.74
	1	8.01	1.00	1.00	77.63
	2	8.01	1.00	1.00	14.97
Global Shift 2	0.25	8.01	2.98	1.03	5.29
	0.5	8.01	1.00	1.00	3.03
	0.75	8.01	1.00	1.00	2.11
	1	8.01	1.00	1.00	1.98
	2	8.01	1.00	1.00	1.00
Local Shift 1	0.25	57.99	179.06	144.43	201.04
	0.5	17.22	113.26	63.98	200.99
	0.75	9.19	56.98	21.46	199.10
	1	8.08	25.91	7.74	198.38
	2	8.01	1.50	1.03	194.82
Local Shift 2	0.25	39.30	145.23	119.47	201.03
	0.5	11.96	73.24	38.99	200.64
	0.75	8.17	26.50	10.18	199.89
	1	8.01	9.13	2.84	199.82
	2	8.01	1.02	1.00	198.89

4.1.2. Multivariate shifted exponential errors

To demonstrate the distribution-free aspect of WDFTC, in this subsection we consider two cases in which the error vector $\boldsymbol{\varepsilon}_j$ has a multivariate exponential distribution, but $\boldsymbol{f}_0$ is still based on Mallat's piecewise smooth function: (1) the components of $\boldsymbol{\varepsilon}_j$ are independent shifted standard exponential random variables with mean zero and standard deviation one; and (2) the components of $\boldsymbol{\varepsilon}_j$ are shifted standard exponential random variables generated via the NORTA method (Cario and Nelson 1996) so that a standard normal vector with common correlation 0.5 between each pair of components is transformed into $\boldsymbol{\varepsilon}_j$, yielding a pairwise correlations between components of $\boldsymbol{\varepsilon}_j$ that are slightly less than 0.5 on the average.

In the following tables, we only report the ARLs delivered by HTW_n and HTW_p using the control limits based on calibration method CM_A. In Section 2.1, we concluded that the performance of M^* was not acceptable when the noise components have exponential marginals. Therefore in the following tables, we only report the ARLs delivered by M^* using the control limits based on calibration method CM_B.

Case (1): Error vector has independent shifted standard exponential components. Table 7 shows the values of ARL_0 and ARL_1 delivered by WDFTC and its competitors for Case (1). WDFTC required the average batch size $\bar{r}=3$ observations. The small values of ARL_0 for HTW_n and HTW_p (approximately 11 and 36 observations, respectively) led us to conclude that those charts were not robust against departures from normality. On the other hand by exploiting its readily computed, distribution-free control limits, WDFTC delivered $\text{ARL}_0 \approx 194$ observations, which did not deviate significantly from the target value of 200 observations; moreover, WDFTC substantially outperformed M^* for Global Shift 1 and for Local Shifts 1 and 2 at all levels of η. For Global Shift 2, WDFTC delivered values of ARL_1 ranging from approximately 4 observations (for $\eta = 0.25$) to 3 observations (for $0.5 \leq \eta \leq 2$), while M^* delivered values of ARL_1 ranging from approximately 5 observations (for $\eta = 0.25$) to 1 observation (for $\eta = 2$). All in all, the performance of WDFTC in the case of shifted standard exponential errors provided

good evidence of the chart's effectiveness and robustness.

Table 7. ARLs for error vector with independent shifted standard exponential components.

Shift Type	Shift Size	WDFTC $\bar{r}=3$	HTW_n CM_A	HTW_p CM_A	M^* CM_B
In-Control	0	193.85	11.42	35.90	200.24
Global Shift 1	0.25	4.24	–	–	82.77
	0.5	3.00	–	–	14.43
	0.75	3.00	–	–	4.40
	1	3.00	–	–	2.01
	2	3.00	–	–	1.00
Global Shift 2	0.25	4.22	–	–	4.74
	0.5	3.00	–	–	2.90
	0.75	3.00	–	–	2.00
	1	3.00	–	–	1.78
	2	3.00	–	–	1.00
Local Shift 1	0.25	123.55	–	–	200.26
	0.5	39.20	–	–	197.22
	0.75	18.07	–	–	194.17
	1	10.55	–	–	186.59
	2	3.17	–	–	110.55
Local Shift 2	0.25	122.26	–	–	200.87
	0.5	35.62	–	–	198.08
	0.75	16.54	–	–	197.37
	1	9.87	–	–	189.51
	2	3.07	–	–	96.01

Case (2): Error vector has correlated shifted standard exponential components. Table 8 shows the values of ARL_0 and ARL_1 delivered by WDFTC and its competitors for Case (2). WDFTC required the average batch size $\bar{r}=3$ observations. The extremely small values of ARL_0 for HTW_n and HTW_p (approximately 3 and 5 observations, respectively) reinforced our conclusion that those charts were not robust against departures from normality. Both WDFTC and M^* delivered values of ARL_0 close to the target value of 200 observations; but whereas the control limits for WDFTC are easily evaluated, the control limits for M^* must be estimated by cumbersome, compute-intensive simulation experiments. For Global Shift 1 with $\eta = 0.25$, WDFTC and M^* performed about the same, delivering ARL_1 values of approximately 170 observations and 176 observations, respectively; but M^* significantly outperformed WDFTC for $0.5 \leq \eta \leq 2$. For Global Shift 2, WDFTC delivered values of ARL_1 ranging from about 6 observations (for $\eta = 0.25$) to about 3 observations (for $0.5 \leq \eta \leq 2$), while M^* delivered values of ARL_1 ranging from about 10 observations (for $\eta = 0.25$) to about 2 observations (for $\eta = 2$). For Local Shifts 1 and 2 with all values of η, WDFTC substantially outperformed M^*. For example, to detect Local Shift 2 with $\eta = 0.5$, WDFTC required approximately 60 observations, while M^* required approximately 224 observations. The results for Case (2) provided further evidence of WDFTC's robustness and effectiveness.

4.2. *Laser range sensor data*

In this subsection, we summarize the experimental results for an application of WDFTC to laser range sensor (LRS) data observed in a lumber-manufacturing process. LRS equipment can measure the thickness of a sawed board with a high degree of accuracy, and the development

Table 8. ARLs for error vector with correlated shifted standard exponential components.

Shift Type	Shift Size	WDFTC $\bar{r}=3$	HTW$_n$ CM$_A$	HTW$_p$ CM$_A$	M^* CM$_B$
In-Control	0	197.06	2.83	4.68	200.85
Global Shift 1	0.25	170.16	–	–	176.49
	0.5	163.11	–	–	115.93
	0.75	141.20	–	–	65.55
	1	110.05	–	–	39.91
	2	34.84	–	–	5.90
Global Shift 2	0.25	5.90	–	–	9.69
	0.5	3.01	–	–	5.61
	0.75	3.01	–	–	4.01
	1	3.01	–	–	3.17
	2	3.01	–	–	1.93
Local Shift 1	0.25	171.70	–	–	230.75
	0.5	61.63	–	–	233.84
	0.75	28.96	–	–	223.00
	1	15.94	–	–	217.93
	2	5.13	–	–	197.92
Local Shift 2	0.25	165.25	–	–	226.96
	0.5	59.68	–	–	224.39
	0.75	26.22	–	–	227.72
	1	14.37	–	–	228.05
	2	4.89	–	–	198.83

of such equipment has provided ample opportunities for quality engineers in the industry to improve and maintain the quality of the manufactured boards (Staudhammer 2004).

Figure 5 is a plot of a sample stream of board-thickness measurements taken along the length of a certain type of board from a particular sensor location as detailed in Staudhammer *et al.* (2006). Four laser sensors are set up to measure the thickness of sawed boards of various types at two different locations on both sides of the board. In this subsection, we use the thickness measurements from one laser location only, but the measurements from all four laser locations can easily be incorporated to monitor various kinds of board defects as detailed below.

For each sawed board, over 2,000 thickness measurements are taken from each laser location; and the physical proximity of the locations on the board for successive thickness measurements naturally induces correlation between those measurements. On the other hand, Staudhammer (2004) finds that there is no significant correlation between measurements taken on different boards, and she formulates statistical models to describe the variation in board thickness along the length of each individual board. Staudhammer proposes new profile-monitoring charts to detect various types of board defects, and she evaluates the performance of those charts using a comprehensive simulation study based on the proposed statistical models.

From equation (1) of Staudhammer *et al.* (2006), we see that for a particular saw configuration, type of board, and side of the board, the statistical model for the thickness of the uth board (expressed in cm) as measured from the vth laser location at the ith horizontal distance x_i cm along the length of the board has the form

$$y_{uvi} = \mu_0 + \mathcal{B}_u + \mathcal{L}_v + \mathcal{BL}_{uv} + \varepsilon_{uvi} \text{ for } i = 1, \ldots, n, \tag{10}$$

where: (1) μ_0 is the true mean in-control board thickness taken over the population of sawed boards defined by the given saw configuration, type of board, and side of the board; (2) $\mathcal{B}_u$ is the random board effect for the uth sample board, which is assumed to be i.i.d. $N(0, \sigma^2_{\mathcal{B}})$; (3)

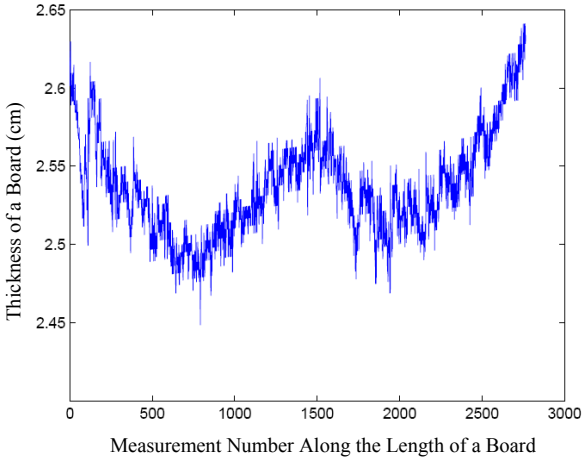

Figure 5. Sample stream of board-thickness measurements.

$\mathcal{L}_v$ is the random effect of the vth laser location, which is assumed to be i.i.d. $N(0,\sigma^2_{\mathcal{L}})$; (4) $\mathcal{BL}_{uv}$ is the random effect arising from the interaction of the board and laser-location effects, which is assumed to be i.i.d. $N(0,\sigma^2_{\mathcal{BL}})$; and (5) ε_{uvi} is the residual error associated with the thickness measurement taken on the uth board from the vth laser location at the ith distance x_i along the board so that the error process $\{\varepsilon_{uvi} : i = 1,\ldots n\}$ is assumed to be stationary and correlated with marginal distribution $N(0,\sigma^2_{\varepsilon})$. Staudhammer *et al.* (2006) obtain the following parameter estimates for the model defined by equation (10): $\widehat{\sigma}_{\mathcal{B}} = 0.0204$ cm, $\widehat{\sigma}_{\mathcal{L}} = 0.0052$ cm, and $\widehat{\sigma}_{\mathcal{BL}} = 0.0238$ cm.

The authors find that for the board type BB considered in this article and for each value of u and v, the error process $\{\varepsilon_{uvi} : i = 1,\ldots n\}$ can be adequately represented by an ARIMA(1,1,1) time series model,

$$(1-\alpha B)(\varepsilon_i - \varepsilon_{i-1}) = (1-\beta B)\zeta_i \quad \text{for} \quad i = 1,2,\ldots, \tag{11}$$

where: (1) B is the backward shift operator so that $(1-\alpha B)\varepsilon_i = \varepsilon_i - \alpha\varepsilon_{i-1}$; and (2) the white-noise process $\{\zeta_i : i = 1,2,\ldots\}$ consists of i.i.d. $N(0,\sigma^2_{\zeta})$ random variables. The authors obtain the following parameter estimates for the error model defined by equation (11): the autoregressive parameter $\widehat{\alpha} = 0.00053$ cm, the moving-average parameter $\widehat{\beta} = 0.00178$ cm, and the white-noise standard deviation $\widehat{\sigma}_{\zeta} = 0.00967$ cm. We apply WDFTC to board-thickness data generated according to the statistical model specified by equations (10) and (11); and we compare the performance of WDFTC with that of the profile-monitoring charts proposed by Staudhammer *et al.* (2006, 2007) for detecting various types of defects in the lumber-manufacturing process.

Rasmussen *et al.* (2004) identify common defects that can arise in lumber manufacturing. In the experimental performance evaluation, we consider four such defects: the machine positioning problem, taper, flare, and snake. Taken from Staudhammer (2004) with permission from the author, Figure 6 illustrates all four types of defects.

The machine positioning problem (MPP) is one of the simplest defects, resulting in a uniform change in thickness along the length of the board. The taper defect results in a gradual thickening or thinning along the length of the board. The flare is one of the more complex defects, which results in progressive board thickening only at the end of the board. The snake is another complex

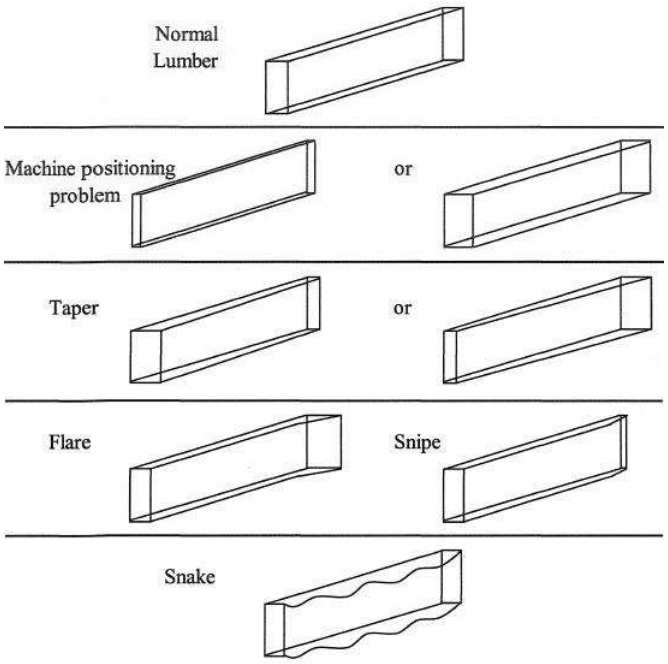

Figure 6. Common defects in lumber manufacturing.

defect that causes high within-board variation of the board's thickness along the length of the board. For more-detailed descriptions of the these defects, see Staudhammer (2004). We use the following synthetic out-of-control conditions with various levels of severity to simulate all four types of defects as follows:

- For the MPP defect, we used the out-of-control mean $\mu_1 = \mu_0 + \delta$, where the shift $\delta \in \{0.0254, 0.0508, 0.0762, 0.1016\}$ (expressed in cm).
- For the taper defect, we took $\mathrm{E}[y_{uvi}] = \mu_0 + x_i\delta/x_n$ for $i = 1,\ldots,n$ and $\delta \in \{0.0508, 0.1016, 0.1524, 0.2032\}$ (expressed in cm) so that the mean deviation from the in-control board thickness μ_0 increased in proportion to the horizontal distance x_i along the length of the board (where the board length $x_n = 244$ cm).
- For the flare defect, we took

$$\mathrm{E}[y_{uvi}] = \begin{cases} \mu_0\,, & \text{if } x_i < x_n - 15 \text{ cm}\,, \\ \mu_0 + (x_i - x_{i_0})\delta/(x_n - x_{i_0})\,, & \text{if } x_i \geq x_n - 15 \text{ cm}\,, \end{cases}$$

  for $i_0 = \max\{i : x_i < x_n - 15\}$ so that tapering occurs only along the last 15 cm of the board's length.
- For the snake defect, we took $\mathrm{E}[y_{uvi}] = \mu_0 + \mathsf{A}\sin(2\pi x_i/\mathsf{P})$, adding a waveform with the period $\mathsf{P} = 182.88$ cm and with the amplitude $\mathsf{A} \in \{0.0508, 0.1016, 0.1524, 0.2032\}$ (all in cm) for $i = 1,\ldots,n = 2{,}048$.

Table 9 summarizes the ARLs delivered by WDFTC when it is applied to the LRS data for the target false alarm rate FAR $= 0.0027$ alarms/profile (sampled board), which is equivalent to setting the target value $\mathrm{ARL}_0 = 370$ profiles (boards). For ease of comparison, the last column of Table 9 shows the results reported by Staudhammer *et al.* (2006, 2007) for their four profile-monitoring charts that are specifically designed to detect sawing defects of type MPP, taper, flare, and snake, respectively. Staudhammer *et al.* (2006, 2007) report their results using graphs

of the corresponding rates of occurrence for true and false alarms; in Table 9 we convert those rates into the associated values of ARL_0 and ARL_1.

Table 9. ARLs for WDFTC and the profile charts (PCs) of Staudhammer (2004)

Shift Type	Shift Size δ or A (cm)	WDFTC $\bar{r} = 6$	PC
MPP	0	358.15	333.33
	0.0254	208.35	20.00
	0.0508	60.60	3.33
	0.0762	30.05	1.25
	0.1016	17.85	1.00
Taper	0	358.15	200.00
	0.0508	190.50	
	0.1016	57.16	
	0.1524	29.07	
	0.2032	17.23	
Flare	0	358.15	50.00
	0.0508	282.76	
	0.1016	109.78	
	0.1524	50.34	
	0.2032	28.57	
Snake	0	358.15	76.92
	0.0508	112.80	
	0.1016	31.84	
	0.1524	16.62	
	0.2032	9.38	

Because $n = 2{,}048$, we considered this problem to exemplify high-dimensional profile monitoring; and therefore we set $q = 0.7$ to obtain more effective dimension reduction when minimizing the WRRE. With this choice of q, we solved equation (8) for various values of L. Ultimately we decided to set $L = 5$ because that choice resulted in a good data-compression ratio, and further meaningful dimension-reduction was not achieved by using smaller values of L. With $q = 0.7$ and $L = 5$, equation (8) yielded $p = 92$, achieving a data-compression ratio of approximately 4.5%. WDFTC delivered the average batch size $\bar{r} = 6$ observations.

In Staudhammer *et al.* (2006, 2007), various Shewhart-type profile-monitoring charts are tailored respectively to the detection of a specific type of defect; and the development of such highly specialized charts can require an extensive modelling-and-analysis effort. See, for example, the authors' approach to detecting the snake defect. Such modelling efforts are not required to apply WDFTC. It is also noteworthy that WDFTC can detect all the different types of defects without the need for frequent recalibration, although some defects are harder to detect than others (for example, the flare defect).

From Table 9 we see the profile chart of Staudhammer *et al.* (2006, 2007) that is specifically designed for the MPP defect delivered substantially smaller values of ARL_1 than WDFTC delivered for this particular defect. For other kinds of defects, however, the profile charts of Staudhammer *et al.* (2006, 2007) delivered values of ARL_0 that were far below the target value of 370 observations; for example, the values of ARL_0 for the charts designed to detect taper, flare, and snake defects were 200, 50, and 77 observations, respectively. Staudhammer *et al.* (2006, 2007) acknowledge the difficulty of adjusting their profile charts to obtain the target ARL_0, because it will require estimating the tails of the run length distribution, which is a challenging problem. Overall, we concluded that WDFTC outperformed the profile-monitoring charts of Staudhammer *et al.* (2006, 2007) in this application.

5. Conclusion

In this article we described WDFTC, a wavelet-based distribution-free chart for monitoring high-dimensional profiles whose components can have non-normal distributions, variance heterogeneity, or substantial intercomponent correlations. We also formulated the following: (1) an effective dimension-reduction technique based on the discrete wavelet transform and the concept of weighted relative reconstruction error; and (2) a covariance-matrix regularisation scheme and a batch-size determination procedure that significantly improve the effectiveness of the associated Hotelling's T^2–type statistics. When tested on normal or non-normal profiles with dimension $n = 512$ and with independent or correlated components, WDFTC was competitive with other commonly used charts, including the chart M^* of Chicken *et al.* (2009); moreover, WDFTC substantially outperformed all those charts for small- to medium-size local shifts. We found another advantage of WDFTC is that its control limits are rapidly evaluated numerically instead of requiring calibration via cumbersome, time-consuming trial-and-error simulations.

When WDFTC was applied to lumber-manufacturing profiles of dimension $n = 2{,}048$, we found that WDFTC was sufficiently versatile to detect a wide variety of defect types with reasonable sensitivity while maintaining the user-specified overall rate of generating false alarms. By contrast each of the profile-monitoring charts of Staudhammer *et al.* (2006, 2007) was specifically designed to detect a single defect type; and although we found that each such chart often outperformed WDFTC in detecting its relevant defect, we encountered substantial difficulties in trying to calibrate those specialized charts so as to deliver the target false-alarm rate when those charts are operated separately or jointly. Overall we concluded that WDFTC also outperformed the profile-monitoring charts of Staudhammer *et al.* (2006, 2007).

Acknowledgements

The authors thank Dr Christina Staudhammer (University of Florida) and Dr Eric Chicken (Florida State University) for numerous enlightening discussions on this work. The authors also thank the editors and referees for suggestions that improved the clarity and accessibility of this article.

References

Bickel, P.J. and Levina, E. 2008. Covariance regularization by thresholding. *Annals of Statistics*, 36 (6), 2577–2604.

Cario, M.C. and Nelson, B.L. 1996. Autoregressive to anything: Time-series input processes for simulation. *Operations Research Letters*, 19, 51-58.

Chicken, E., Pignatiello, Jr., J.J., and Simpson, J.R. 2009. Statistical process monitoring of nonlinear profiles using wavelets. *Journal of Quality Technology*, 41 (2), 198–212.

Daniels, M.J. and Kass, R.E. 2001. Shrinkage Estimators for covariance matrices. *Biometrics*, 57 (4), 1173–1184.

Ding, Y. Zeng, L., and Zhou, S. 2006. Phase I analysis for monitoring nonlinear profiles in manufacturing processes. *Journal of Quality Technology*, 38 (3), 199–216.

Donoho, D.L. and Johnstone, I.M. 1994. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81 (3), 425–455.

Fan, J. 1996. Test of significance based on wavelet thresholding and Neyman's truncation. *Journal of American Statistical Association*, 91 (434), 674–688.

Ganesan, R., Das, T.K., and Venkataraman, V. 2004. Wavelet-based multiscale statistical process monitoring: A literature review. *IIE Transactions*, 36 (9), 787–806.

Gao, H.-Y. 1997. Wavelet shrinkage estimates for heteroscedastic regression models. Technical Report. Seattle: MathSoft, Inc.

Hanselman, D., and Littlefield, B. 2001. *Mastering* MATLAB *6: A comprehensive tutorial and reference.* Upper Saddle River, NJ: Prentice Hall.

Hoffbeck, J.P. and Landgrebe, D.A. 1996. Covariance matrix estimation and classification with limited training data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18 (7), 763–767.

Jeong, M.K., Lu, J.-C., and Wang, N. 2006. Wavelet-based SPC procedure for complicated functional data. *International Journal of Production Research*, 44 (4), 729–744.

Jin, J. and Shi, J. 1999. Feature-preserving data compression of stamping tonnage information using wavelets. *Technometrics*, 41 (4), 327–339.

Jin, J. and Shi, J. 2001. Automatic feature extraction of waveform signals for in-process diagnostic performance improvement. *Journal of Intelligent Manufacturing*, 12, 257–268.

Jolliffe, I.T. 1986. *Principal component analysis*. New York: Springer-Verlag.

Kang, L. and Albin, S.L. 2000. On-line monitoring when the process yields a linear profile. *Journal of Quality Technology*, 32 (4), 418–426.

Kano, M., Nagao, K., Hasebe, S., Hashimoto, I., Ohno, H., Strauss, R., and Bakshi, B.R. 2002. Comparison of statistical process monitoring methods with applications to the Eastman challenge problem. *Computers & Chemical Engineering*, 26, 161–174.

Kim, K., Mahmoud, M.A., and Woodall, W.H. 2003. On the monitoring of linear profiles. *Journal of Quality Technology*, 35 (3), 317–328.

Kim, S.-H., Alexopoulos, C., Tsui, K.-L., and Wilson, J.R. 2007. A distribution-free tabular CUSUM chart for autocorrelated data. *IIE Transactions*, 39, 317–330.

Lada, E.K., Lu, J.-C., and Wilson, J.R. 2002. A wavelet-based procedure for process fault detection. *IEEE Transactions on Semiconductor Manufacturing*, 15 (1), 79–90.

Lada, E.K. and Wilson, J.R. 2006. A wavelet-based spectral procedure for steady-state simulation analysis. *European Journal of Operations Research*, 174, 1769–1801.

Ledoit, O. and Wolf, M. 2002. Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *The Annals of Statistics*, 30 (4), 1081–1102.

Lee, J., Alexopoulos, C., Goldsman, D., Kim, S.-H., Tsui, K.-L., and Wilson, J.R. 2009. Monitoring autocorrelated processes using a distribution-free tabular CUSUM chart with automated variance estimation. *IIE Transactions*, 41, 979–994.

Mallat, S.G. 2009. *A wavelet tour of signal processing: The sparse way*. 3rd ed. Boston: Elsevier/Academic Press.

Montgomery, D.C. 2005. *Introduction to statistical quality control*. 5th ed. New York: John Wiley & Sons.

Ogden, R.T. 1997. *Essential wavelets for statistical application and data analysis*. Boston: Birkhäuser.

Porta Nova, A.M.O., and Wilson, J.R. 1989. Estimation of multiresponse simulation metamodels using control variates. *Management Science*, 35 (11), 1316–1333.

Qiu, P. 2008. Distribution-free multivariate process control based on log-linear modeling. *IIE Transactions* 40 (7), 664–677.

Ramsay, J.O. and Silverman, B.W. 2006. *Functional data analysis*. 2nd ed. New York: Springer.

Rasmussen, H.K., Kozak, R.A., and Maness, T.C. 2004. An analysis of machine-caused lumber shape defects in British Columbia sawmills. *Forest Products Journal*, 54 (6), 47–56.

Stanfield, P.M., Wilson, J.R., and King, R.E. 2004. Flexible modelling of correlated operation times with application in product-reuse facilities. *International Journal of Production Re-*

search, 42 (11), 2179–2196.

Staudhammer, C. 2004. Statistical procedures for development of real-time statistical process control (SPC) in lumber manufacturing. Thesis (PhD). The University of British Columbia, Canada.

Staudhammer, C., Kozak, R.A., and Maness, T.C. 2006. SPC methods for detecting simple sawing defects using real-time laser range sensor data. *Wood and Fiber Science*, 38 (4), 696–716.

Staudhammer, C., Maness, T.C., and Kozak, R.A. 2007. Profile charts for monitoring lumber manufacturing using laser range sensor data. *Journal of Quality Technology*, 39 (3), 224–240.

von Sachs, R. and MacGibbon, B. 2000. Non-parametric curve estimation by wavelet thresholding with locally stationary errors. *Scandinavian Journal of Statistics*, 27, 475–499.

Williams, J.D., Woodall, W.H., and Birch, J.B. 2007. Statistical monitoring of nonlinear product and process quality profiles. *Quality and Reliability Engineering International*, 23, 925–941.

Woodall, W.H., Spitzner, D.J., Montgomery, D.C., and Gupta, S. 2004. Using control charts to monitor process and product quality profiles. *Journal of Quality Technology*, 36 (3), 309–320.