

Monitoring Nonlinear Profiles Adaptively with a Wavelet-Based Distribution-Free CUSUM Chart

Huizhu Wang^a, Seong-Hee Kim^{a*}, Xiaoming Huo^a, Youngmi Hur^b, 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 Mathematics, Yonsei University, Seoul 120-749, Republic of Korea;* ^c*Edward P. Fitts Department of Industrial and Systems Engineering, North Carolina State University, Raleigh, NC 27695-7906, USA*

()

A wavelet-based distribution-free tabular CUSUM chart based on adaptive thresholding, $WDFTC_a$ is designed for rapidly detecting shifts in the mean of a high-dimensional profile whose noise components have a continuous non-singular multivariate distribution. First computing a discrete wavelet transform of the noise vectors for randomly sampled Phase I (in-control) profiles, $WDFTC_a$ uses a matrix-regularization method to estimate the covariance matrix of the wavelet-transformed noise vectors; then those vectors are aggregated (batched) so that the nonoverlapping batch means of the wavelet-transformed noise vectors have manageable covariances. Lower and upper in-control thresholds are computed for the resulting batch means of the wavelet-transformed noise vectors using the associated marginal Cornish-Fisher expansions that have been suitably adjusted for between-component correlations. From the thresholded batch means of the wavelet-transformed noise vectors, Hotelling's T^2 -type statistics are computed to set the parameters of a CUSUM procedure. To monitor shifts in the mean profile during Phase II (regular) operation, $WDFTC_a$ computes a similar Hotelling's T^2 -type statistic from successive thresholded batch means of the wavelet-transformed noise vectors using the in-control thresholds; then $WDFTC_a$ applies the CUSUM procedure to the resulting T^2 -type statistics. Experimentation with several normal and nonnormal test processes revealed that $WDFTC_a$ outperformed existing nonadaptive profile-monitoring schemes.

Keywords: Quality control; high dimension; profile; CUSUM chart; distribution-free; discrete wavelet transform

1. Introduction

With the rapid development of sensor technology, we can collect massive process and product data to monitor, control, and improve process performance and product quality. Massive data may lead to a data overabundance. For an example, Jin and Shi (1999) study a tonnage signal from an automotive stamping process, where a press can typically perform 200 strokes per minute with more than 6000 data points in each stroke, i.e., 1.2 million data points per minute. Staudhammer, Maness, and Kozak (2007) use laser range sensors (LRSs) to measure the thickness of sawed boards from a lumber manufacturing process, and each LRS provides more than 2000 data points for each sawed board.

These series of data points for each stroke or sawed board are called profiles. A profile describes the functional relationship between a response variable and one or more explanatory variables (Woodall 2007). In this paper, we restrict attention to one explanatory variable, the same setting used in Jeong, Lu, and Wang (2006), Chicken, Pignatiello Jr., and Simpson (2009) and Lee et al. (2012). The functional relationship sometimes has a linear or nonlinear representation, but it also can be too complicated to have a well-known functional form. Examples, including tonnage signals or LRS data, often have jumps, cusps, oscillations, and other types of non-smooth behavior.

*Corresponding author. Email: skim@isye.gatech.edu. Telephone: +1-404-894-4551

Monitoring linear profiles, such as some calibration processes, widely employ linear regression models. In order to detect a shift in the mean of successive observed profiles, control charts are applied to regression parameters, such as estimated intercepts and slopes. For simple linear profiles, Kang and Albin (2000) perform a single control chart on a multivariate T^2 -statistic constructed on estimated intercepts and slopes. For polynomial linear profiles and general linear profiles, Zou, Tsung, and Wang (2007) monitor all estimated regression parameters within a single chart using multivariate exponentially weighted moving average (EWMA) schemes.

Nonlinear profiles, such as product shapes and consecutive measurements of the same variable at different locations on individual products, are often modeled by nonlinear regression models; and a control chart is performed on estimated regression parameters from the observed profiles. Williams, Woodall, and Birch (2007) use a Hotelling's T^2 -chart to monitor parameters of a nonlinear regression model for a particle board manufacturing process. To detect several complex sawing defects in lumber manufacturing, Staudhammer, Maness, and Kozak (2007) model LRS profiles with a combination of two regression models, multiple linear and nonlinear, to describe roughness and waviness of surfaces, respectively; and then they monitor regression parameters from different models with separate multiple profile charts.

Even though Gupta, Montgomery, and Woodall (2006) point out that the use of control charts based on estimated regression parameters is effective in terms of average run length performance, there are some issues in using regression models for profile monitoring. First, regression methods tend to use smooth models to estimate in-control mean profiles; thus we lose some important local information, such as jumps and cusps. Secondly, the performance of control charts highly depends on the choice of models. If an inappropriate model is selected, then we may get misleading results and not detect shifts in the mean profile adequately. For example, Chicken, Pignatiello Jr., and Simpson (2009) show that when a change occurs in the mean profile, the selected regression model possibly still yields the same parameter estimates so that the shift is undetectable. In addition, fitting a sufficiently accurate regression model to observed profiles can be time-consuming. Lastly, some process profiles, such as tonnage signals and LRS data, can be too complex to be modeled adequately by parametric regression models.

As a consequence, some profile monitoring charts have been based on nonparametric regression techniques, such as smoothing splines, functional principal components analysis (FPCA), Fourier analysis, and wavelet analysis. Gardner et al. (1997) use a smoothing spline to model the thickness at selected locations of a wafer surface in a semiconductor manufacturing process. Ramsay and Silverman (1997) propose methods based on FPCA that can be used in nonparametric profile monitoring. Chen and Nembhard (2011) compute the Fast Fourier Transform (FFT) of profiles and construct an adaptive Neyman test statistic from the corresponding Fourier coefficients. However, splines, FPCA, and Fourier transforms cannot model locally sharp changes well—for example, see Ganesan et al. (2003), Jeong, Lu, and Wang (2006) and Chicken, Pignatiello Jr., and Simpson (2009). Instead, the discrete wavelet transform (DWT) has been suggested and shown to be effective for detecting and diagnosing process faults (Fan 1996; Jin and Shi 2001). Some advantages of using DWTs for profile monitoring are summarized as follows.

- Sparsity: the DWT can use substantially fewer wavelet basis functions to achieve a comparably accurate approximation for profiles, especially non-smooth profiles.
- Localization: the DWT is localized in both the frequency and time domains, that is, a frequency change causes changes only in certain time segments. By contrast, the Fourier transform is localized only in the frequency domain, which means a frequency change can cause changes everywhere in the time domain. Therefore the DWT can preserve locally sharp changes in profiles well and help diagnose process faults.
- Fast computation: For a profile of dimension n , the computational complexity of the profile's DWT is $O(n)$ (Mallat 1989); and this is smaller than the complexity $O(n \log_2 n)$ of the profile's FFT.

Therefore wavelet-based approaches have gained popularity, especially for monitoring high-dimensional profiles with non-smooth behaviors. The DWT is usually used as either a denoising method or a dimension-reduction method in nonparametric profile monitoring. Jin and Shi (2001) and Jeong, Lu, and Wang (2006) use a DWT to denoise observed profiles via universal thresholds and then monitor the preserved components with control charts. The number and locations of preserved components change from profile to profile;

therefore monitored wavelet components are selected adaptively. Their wavelet-based control charts with adaptive selection detect shifts effectively, but those control charts require assumptions of marginal normality and sometimes independence among components. Without the normality assumption, it is challenging to obtain distributions of monitored statistics and construct analytical control limits for control charts.

On the other hand, Jin and Shi (1999), Lada, Lu, and Wilson (2002), and Lee et al. (2012) employ the DWT as a dimension-reduction technique. Based on engineering knowledge of the automotive stamping operation that different process failures correspond to changes in different segments of the tonnage signal from the stamping process, Jin and Shi (1999) construct different thresholds for different profile segments. Then they monitor a fixed selection of wavelet components according to the determined thresholds, and detect process faults in the tonnage signals. Lada, Lu, and Wilson (2002) reduce profile dimensions by minimizing the weighted relative reconstruction error (WRRE), in order to balance model parsimony against data reconstruction error. Lee et al. (2012) apply WRRE to an in-control mean profile to determine locations of wavelet components whose values need to be monitored. A reduced-dimension DWT vector of a profile is formed by keeping wavelet components from the determined locations only; and then a distribution-free CUSUM chart is applied to Hotelling's T^2 -type statistics computed from the reduced-dimension DWT vectors. Since both control charts in Jin and Shi (1999) and Lee et al. (2012) select wavelet components from the same locations in each profile's DWT, the selection method for wavelet components is considered to be static or nonadaptive. Even though some wavelet-based control charts with static selection (specifically Lee et al. (2012)) are distribution-free, they are insensitive to some shifts or even fail to detect some shifts—for instance, a shift only affecting the unselected wavelet components. Extending static selection to dynamic (adaptive) selection is not trivial. Because the selected components change from profile to profile and the number of selected components is affected by the correlations between profile components, monitoring statistics tend to have a large variance; and the distribution of monitored statistics can be quite complicated.

In this article we formulate and evaluate $WDFTC_a$, a wavelet-based distribution-free CUSUM chart with adaptive selection that dynamically adjusts the number and locations of preserved components for each monitored DWT vector. In brief $WDFTC_a$ consists of the following steps. First computing a DWT of the noise vectors for randomly sampled Phase I (in-control) profiles, $WDFTC_a$ uses a matrix-regularization method to estimate the covariance matrix of the wavelet-transformed noise vectors; then those vectors are aggregated (batched) so that the nonoverlapping batch means of the wavelet-transformed noise vectors have manageable covariances. Lower and upper in-control thresholds are computed for the resulting batch means of the wavelet-transformed noise vectors using the associated marginal Cornish-Fisher expansions that have been suitably adjusted for correlations between the components of those vectors. From the thresholded batch means of the wavelet-transformed noise vectors, Hotelling's T^2 -type statistics are computed to set the parameters of a CUSUM procedure. To monitor shifts in the mean profile during Phase II (regular) operation, $WDFTC_a$ computes a similar Hotelling's T^2 -type statistic from successive thresholded batch means of the wavelet-transformed noise vectors using the upper and lower in-control thresholds; then $WDFTC_a$ applies the CUSUM procedure to the resulting T^2 -type statistics for rapid detection of an out-of-control condition.

Ideally, in Phase I (in-control) operation $WDFTC_a$ will retain relatively few detail coefficients; on the other hand, when the observed profiles are out-of-control, $WDFTC_a$'s estimators of both the scaling and detail coefficients will exhibit significant shifts from their in-control counterparts. Thus we can rapidly detect such shifts by monitoring changes in T^2 -type statistics computed from the selected components. In this paper, we extend the wavelet-based control chart with static selection of Lee et al. (2012) to formulate new control charts with adaptive selection. Experimental results show the new control charts can detect various shifts more effectively than the method of Lee et al. (2012) under a general noise distribution.

The paper is organized as follows. Section 2 briefly defines the problem and notation, provides motivating examples, and introduces the DWT. In Section 3, we describe the new profile-monitoring procedure $WDFTC_a$, and we discuss distribution-free thresholding methods and covariance matrix estimation. Section 4 presents experimental results for $WDFTC_a$ and its competitors, followed by conclusions in Section 5.

2. Background

In this section, we first define our problem and provide notation; then we discuss a motivating example and present a brief overview of DWT.

2.1 Notation and Problem

For $j = 1, 2, \dots$, we consider the j th observed profile $\mathbf{Y}_j = (Y_{j,1}, \dots, Y_{j,n})^T$ as an n -dimensional random vector; and for $i = 1, \dots, n$, the i th component $Y_{j,i}$ of that profile is the response associated with the i th prespecified level x_i of the designated explanatory variable so that we have the functional relationship

$$Y_{j,i} = f(x_i) + \varepsilon_{j,i} \text{ for } i = 1, \dots, n \text{ and } j = 1, 2, \dots, \quad (1)$$

where $\varepsilon_{j,i}$ is the error (noise) in the i th component of the j th profile. (Throughout this article, we let \mathbf{A}^T denote the transpose of a vector or matrix \mathbf{A} .) If we let $\mathbf{x} = (x_1, \dots, x_n)^T$ denote the $n \times 1$ vector of preselected levels of the explanatory variable that are used with each profile, and if we let $\boldsymbol{\varepsilon}_j = (\varepsilon_{j,1}, \dots, \varepsilon_{j,n})^T$ denote the corresponding $n \times 1$ vector of noise terms for the j th profile, then the functional relationship (1) may be compactly expressed as $\mathbf{Y}_j = \mathbf{f}(\mathbf{x}) + \boldsymbol{\varepsilon}_j$ for $j = 1, 2, \dots$, where $\mathbf{f}(\mathbf{x}) \equiv [f(x_1), \dots, f(x_n)]^T$. We assume that $E[\boldsymbol{\varepsilon}_j] = \mathbf{0}$ and that the covariance matrix $\text{Cov}[\boldsymbol{\varepsilon}_j] = E[\boldsymbol{\varepsilon}_j \boldsymbol{\varepsilon}_j^T] = \boldsymbol{\Sigma}_0$. In general the components of $\boldsymbol{\varepsilon}_j$ may be non-normal and correlated. With this setup, the j th profile \mathbf{Y}_j has mean vector $E[\mathbf{Y}_j] = \mathbf{f}(\mathbf{x})$ and covariance matrix $\boldsymbol{\Sigma}_0$. In addition, the marginal variance of $Y_{j,i}$ or $\varepsilon_{j,i}$ is denoted by $\sigma_i^2 = [\boldsymbol{\Sigma}_0]_{ii}$.

We assume that when $\mathbf{f}(\cdot)$ is the in-control function $\mathbf{f}_0(\cdot)$, i.e., $E[\mathbf{Y}_j] = \mathbf{f}_0(\mathbf{x})$, the monitored process is in control and \mathbf{Y}_j is an in-control profile. When $\mathbf{f}(\cdot)$ is some out-of-control function $\mathbf{f}_1(\cdot)$, i.e., $E[\mathbf{Y}_j] = \mathbf{f}_1(\mathbf{x}) \neq \mathbf{f}_0(\mathbf{x})$, the monitored process is out of control and \mathbf{Y}_j is an out-of-control profile. Our problem is to detect rapidly the onset of an out-of-control condition.

To simplify notation, we let $\mathbf{f}_0 \equiv \mathbf{f}_0(\mathbf{x})$ and $\mathbf{f}_1 \equiv \mathbf{f}_1(\mathbf{x}) \neq \mathbf{f}_0$, where $\mathbf{f}_0 = [f_{0,1}, \dots, f_{0,n}]^T$ and $\mathbf{f}_1 = [f_{1,1}, \dots, f_{1,n}]^T$. Without loss of generality, we assume \mathbf{f}_0 is centered, i.e., $\sum_{i=1}^n f_{0,i} = 0$. The profile dimension is set to $n = 2^J$ for some positive integer J . Let $L \in \{0, \dots, J-1\}$ denote the coarsest level of resolution in a given DWT system (see Section 2.3) and \mathbf{W} denote the associated DWT matrix for the given DWT system with the selected L . We apply the DWT to \mathbf{f}_0 , \mathbf{Y}_j and $\boldsymbol{\varepsilon}_j$ for $j = 1, 2, \dots$, obtaining the corresponding DWTs $\boldsymbol{\theta}_0 = \mathbf{W}\mathbf{f}_0 = (\theta_{0,1}, \dots, \theta_{0,n})^T$, $\mathbf{d}_j = \mathbf{W}\mathbf{Y}_j = (d_{j,1}, \dots, d_{j,n})^T$, and $\boldsymbol{\omega}_j = \mathbf{W}\boldsymbol{\varepsilon}_j = (\omega_{j,1}, \dots, \omega_{j,n})^T$, respectively. The first 2^L components in $\boldsymbol{\theta}_0$, \mathbf{d}_j and $\boldsymbol{\omega}_j$ are scaling coefficients and contain important features of the original vectors at the lower levels of resolution. The remaining $n - 2^L$ components are called detail coefficients. The covariance matrix of \mathbf{d}_j or $\boldsymbol{\omega}_j$ is $\boldsymbol{\Lambda}_0 = \mathbf{W}\boldsymbol{\Sigma}_0\mathbf{W}^T$. All notation is summarized below.

\mathbf{f}_0	The $n \times 1$ in-control mean profile;
\mathbf{Y}_j	The j th $n \times 1$ observed profile, for $j = 1, \dots$;
$\boldsymbol{\varepsilon}_j$	$= \mathbf{Y}_j - \mathbf{f}_0$, the $n \times 1$ noise vector of the j th observed profile \mathbf{Y}_j ;
$\boldsymbol{\theta}_0$	$= \mathbf{W}\mathbf{f}_0$, the $n \times 1$ DWT of the in-control mean profile;
\mathbf{d}_j	$= \mathbf{W}\mathbf{Y}_j$, the $n \times 1$ DWT of the j th observed profile \mathbf{Y}_j ;
$\boldsymbol{\omega}_j$	$= \mathbf{W}\boldsymbol{\varepsilon}_j = \mathbf{d}_j - \boldsymbol{\theta}_0$, the $n \times 1$ DWT of the noise vector of the j th observed profile \mathbf{Y}_j ;
$\boldsymbol{\omega}_k(r)$	$= r^{-1} \sum_{u=1}^r \boldsymbol{\omega}_{(k-1)r+u}$, the DWT of the k th non-overlapping batch mean of noise vectors with batch size r ;
$\boldsymbol{\lambda}^L$	$= \{\lambda_1^L, \dots, \lambda_n^L\}$, an $n \times 1$ lower threshold vector;
$\boldsymbol{\lambda}^U$	$= \{\lambda_1^U, \dots, \lambda_n^U\}$, an $n \times 1$ upper threshold vector;
$\boldsymbol{\omega}_k^*(r)$	$= \{\omega_{k,1}^*(r), \dots, \omega_{k,n}^*(r)\}$, the thresholded DWT of the k th non-overlapping batch mean of noise vectors, i.e., applying $\boldsymbol{\lambda}^L$ and $\boldsymbol{\lambda}^U$ on $\boldsymbol{\omega}_k(r)$, where $\omega_{k,i}^*(r) = \omega_{k,i}(r)(1 - \mathbf{1}_{\lambda_i^L < \omega_{k,i}(r) < \lambda_i^U})$ and $\mathbf{1}$ is an indicator function;
$\bar{\boldsymbol{\omega}}_N$	$= N^{-1} \sum_{j=1}^N \boldsymbol{\omega}_j$, the sample mean of $\boldsymbol{\omega}_1, \dots, \boldsymbol{\omega}_N$, where N is the size of a data set;
$\widehat{\boldsymbol{\Lambda}}_0$	$= (N-1)^{-1} \sum_{j=1}^N (\boldsymbol{\omega}_j - \bar{\boldsymbol{\omega}}_N)(\boldsymbol{\omega}_j - \bar{\boldsymbol{\omega}}_N)^T$, $n \times n$ sample covariance matrix of noise vectors;
$\widetilde{\boldsymbol{\Lambda}}_0$	The regularized version of $\widehat{\boldsymbol{\Lambda}}_0$, see discussion in Section 3.3;
$\widetilde{\boldsymbol{\Lambda}}_0(r)$	$= \widetilde{\boldsymbol{\Lambda}}_0/r$.

In this paper, we define the in-control average run length, ARL_0 , as the average number of in-control profiles that are observed before a (false) out-of-control alarm is raised; similarly the out-of-control average run length, ARL_1 , is the average number of out-of-control profiles that are observed before a (true) out-of-control alarm is raised. We compare and analyze different control charts in terms of ARL_0 and ARL_1 .

2.2 Motivation

In this subsection, we show that a wavelet-based monitoring chart with static selection can miss certain shifts, which in turn demonstrates the need for new charts with adaptive selection. More specifically, we examine the performance of $WDFTC_s$ from Lee et al. (2012), which is a distribution-free tabular CUSUM chart on T^2 -type statistics constructed from preselected locations of wavelet components. Recall that preselected locations are fixed for all profiles regardless of shift types. A brief description of Procedure $WDFTC_s$ is provided in the Appendix; also see Figure 2 in Lee et al. (2012) for more details.

In our preliminary experiments, we use Mallat’s piecewise smooth function as f_0 with $n = 512$ equispaced data points, as depicted in Figure 1. This function exhibits some problematic characteristics, such as cusps and jumps, which are often seen in manufacturing profiles. The noise components are assumed to follow independent and identically distributed (IID) standard normal distributions. We consider two types of shifts, namely the wavelet global (WG) shift and the wavelet local (WL) shift, with different values of the shift size η ; and thus the out-of-control mean profile is $f_1 = f_0 + \eta\zeta$, where ζ denotes the shift type and $\eta \in \{0.25, 0.5, 0.75, 1, 2\}$. These two types of shifts only affect unselected DWT components in $WDFTC_s$ and are shown in Figure 2. More details of WG and WL are in Section 4.1.1. The target ARL_0 is set to 200.

Table 1 shows ARL_1 of $WDFTC_s$ with different values of the shift size η . The performance of ARL_1 implies that $WDFTC_s$ fails to detect both types of shifts (WG and WL), even for a large shift size, such as $\eta = 2$. This is not surprising, because the WG and WL shifts do not affect the statistical properties of the preselected wavelet components; and thus constructed T^2 -type statistics in $WDFTC_s$ have the same statistical properties (i.e., the mean and variance of the relevant T^2 -type statistics do not change from in-control profiles to out-of-control profiles). Other existing wavelet-based control charts with a preselected subset of wavelet components also exhibit similar limitations. To overcome this problem, we consider DWTs of noise vectors of profiles and determine adaptively the locations of preserved wavelet components in the noise DWT vectors. More specifically, for each DWT of a noise vector, we preserve all scaling coefficients and a number of detail coefficients whose values turn out to be outliers. Therefore locations of preserved wavelet components can be different for each observed profile and can be affected by any type of shift.

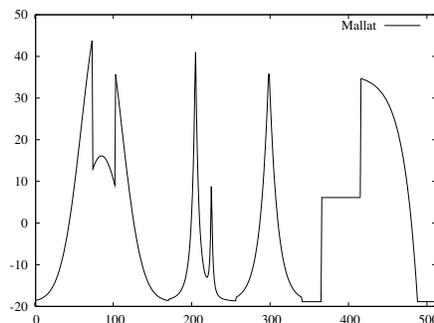


Figure 1. Mallat’s piecewise smooth function.

Table 1. ARL_1 delivered by $WDFTC_s$ for wavelet global shift and wavelet local shift.

η	Wavelet global shift (WG)					Wavelet local shift (WL)				
	0.25	0.5	0.75	1	2	0.25	0.5	0.75	1	2
ARL_1	202.72	195.54	198.19	203.19	189.25	212.47	206.64	202.42	201.03	202.22

To emphasize the need for a new chart, extreme shifts are considered in this section. Comparisons between $WDFTC_s$ and the proposed new chart on more realistic shifts are discussed in Section 4.

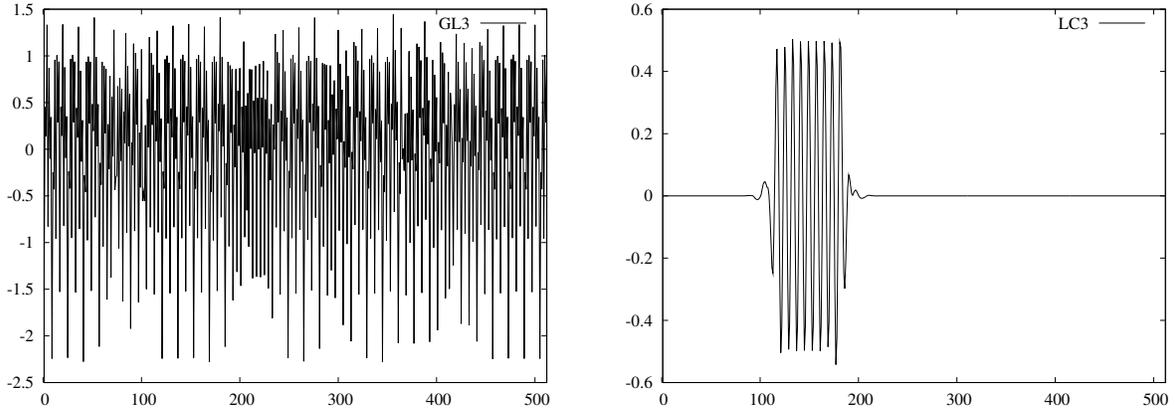


Figure 2. Wavelet global shift function (left) and wavelet local shift function (right).

2.3 Discrete Wavelet Transform

The wavelet transform is analogous to the Fourier transform, but it employs scaling and wavelet basis functions instead of the sine and cosine basis functions used in the Fourier transform. Let $\mathcal{L}^2[0, 1]$ denote the space of real-valued square-integrable functions defined on the unit interval $[0, 1]$. The wavelet transform of a function $g \in \mathcal{L}^2[0, 1]$ represents g as an infinite series involving orthonormal scaling or wavelet basis functions. A scaling function $\phi \in \mathcal{L}^2[0, 1]$ has an associated wavelet function $\psi \in \mathcal{L}^2[0, 1]$; and from ϕ or ψ , we can derive an orthonormal set of basis functions for $\mathcal{L}^2[0, 1]$ similar to the trigonometric functions used in the Fourier series representation. Assuming that ϕ and ψ are based on the Haar wavelet for simplicity, we explain the mechanism of the wavelet transform.

The wavelet representation of a function $g \in \mathcal{L}^2[0, 1]$ is given by

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

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 $\langle g_1, g_2 \rangle = \int_0^1 g_1(z) g_2(z) dz$ is defined as the inner product operator for $g_1, g_2 \in \mathcal{L}^2[0, 1]$.

Let $P_B(g)$ represent the B th partial sum on the far right-hand side of Equation (2). It is clear that the quantity $P_B(g)$ becomes more accurate as B increases, and it provides an approximation to g for a finite B . The scaling coefficients of g are defined as the quantities $\{C_{\ell,m} = \langle g, \phi_{\ell,m} \rangle\}$, which represent the low-frequency components or the smooth parts of $g(z)$. The detail coefficients of g are the quantities $\{D_{\ell,m} = \langle g, \psi_{\ell,m} \rangle\}$; and they represent the high-frequency components or local behaviors of $g(z)$. Due to the discrete nature of measurements from a physical device, suppose that we take $g \approx P_J(g)$ for some finest (highest) level of resolution J ; and we stop the successive function-approximation operations at some coarsest (lowest) level of resolution L , where $L < J$. Then we obtain an approximate representation of g based on its DWT,

$$g(z) \approx \sum_{m=0}^{2^J - 1} C_{J,m} \phi_{J,m}(z) = \sum_{m=0}^{2^L - 1} C_{L,m} \phi_{L,m}(z) + \sum_{\ell=L}^{J-1} \sum_{m=0}^{2^\ell - 1} D_{\ell,m} \psi_{\ell,m}(z) \quad (3)$$

for almost all $z \in [0, 1]$.

The DWT can also be presented in terms of matrices, and we use the matrix representation throughout the paper. Let W denote an $n \times n$ orthogonal matrix associated with the selected functions $\phi(t)$ and $\psi(t)$, where n has the form $n = 2^J$ for some positive integer J . For any $n \times 1$ vector Y , the matrix-vector multiplication $d = WY$ yields the corresponding DWT of Y . Let $L \in \{0, 1, \dots, J-1\}$ denote the coarsest level of resolution in the given wavelet system. Thus the first 2^L components of d are scaling coefficients, and all remaining components are detail coefficients.

3. New Procedure

In this section, we introduce a new chart that monitors statistics constructed from adaptively thresholded DWT vectors. We first present our proposed procedure $WDFTC_a$, and then discuss different adaptive thresholding methods and covariance matrix estimation.

3.1 Procedure $WDFTC_a$

$WDFTC_a$ first applies DWT to each observed profile, thresholds components of DWT vectors adaptively, and then constructs T^2 -type statistics from thresholded DWT vectors. In order to maintain the power of control charts (Fan (1996)), a desirable thresholding method selects a small number of components (e.g., mainly scaling coefficients) when in-control, so that monitored statistics have manageable variability and can be approximated as a Brownian motion process properly. On the other hand, the thresholding method should select more components (e.g., including some extreme outliers of detail coefficients) when out-of-control, so that monitored statistics have significant changes and can quickly detect shifts in the mean profile. Additionally, scaling coefficients contain prominent information of a profile due to the nature of wavelet transform, while extreme outliers in DWT vectors vary for different observed profiles and thus should be selected adaptively.

Combining with an adaptive thresholding method, $WDFTC_a$ becomes more sensitive to detect a wide range of shift types than existing wavelet-based control charts with static selection. The detailed steps in $WDFTC_a$ are described as below.

Procedure $WDFTC_a$

Phase I — Obtain vectors $\mathbf{d}_j = \mathbf{Y}_j - \mathbf{f}_0$, from observed profiles $\{\mathbf{Y}_j : j = 1, \dots, N\}$ where N is the size of Phase I data set. Calculate DWTs $\boldsymbol{\omega}_j = \mathbf{W}\mathbf{d}_j$, and then perform the following steps:

1. Compute sample covariance matrix $\widehat{\boldsymbol{\Lambda}}_0$ from $\{\boldsymbol{\omega}_j : j = 1, \dots, N\}$, see Notation. Apply Algorithm CMR as follow to regularize $\widehat{\boldsymbol{\Lambda}}_0$ and obtain $\widetilde{\boldsymbol{\Lambda}}_0$. Next determine the batch size r by Algorithm BSD as follow and let $\widetilde{\boldsymbol{\Lambda}}_0(r) = \widetilde{\boldsymbol{\Lambda}}_0/r$. For $k = 1, 2, \dots, \lfloor N/r \rfloor$, compute the k th non-overlapping batch mean vector $\boldsymbol{\omega}_k(r) = r^{-1} \sum_{u=1}^r \boldsymbol{\omega}_{(k-1)r+u}$.
2. Determine the lower and upper threshold vectors $\boldsymbol{\lambda}^L$ and $\boldsymbol{\lambda}^U$ as Section 3.2.
3. Threshold $\boldsymbol{\omega}_k(r)$ with $\boldsymbol{\lambda}^L$ and $\boldsymbol{\lambda}^U$; and obtain $\boldsymbol{\omega}_k^*(r)$, see Notation. Compute the Hotelling's T^2 -type statistics

$$T_k^2(r) = \left(\boldsymbol{\omega}_k^*(r) \right)^T \left(\widetilde{\boldsymbol{\Lambda}}_0(r) \right)^{-1} \left(\boldsymbol{\omega}_k^*(r) \right) \quad (4)$$

4. Calculate sample mean $\widehat{\boldsymbol{\mu}}_{T^2(r)}$ and sample variance $\widehat{\boldsymbol{\sigma}}_{T^2(r)}^2$ from $\{T_k^2(r) : k = 1, 2, \dots, \lfloor N/r \rfloor\}$. Solve a root H from Equation (5) with $K = 0.1\widehat{\boldsymbol{\sigma}}_{T^2(r)}$ and prespecified ARL_0 .

$$\frac{\widehat{\boldsymbol{\sigma}}_{T^2(r)}^2}{2K^2} \left(\exp \left\{ \frac{2K[H + 1.166\widehat{\boldsymbol{\sigma}}_{T^2(r)}]}{\widehat{\boldsymbol{\sigma}}_{T^2(r)}^2} \right\} - 1 - \left\{ \frac{2K[H + 1.166\widehat{\boldsymbol{\sigma}}_{T^2(r)}]}{\widehat{\boldsymbol{\sigma}}_{T^2(r)}^2} \right\} \right) = \frac{2ARL_0}{r}, \quad (5)$$

Phase II — Obtain DWTs $\boldsymbol{\omega}_j = \mathbf{W}(\mathbf{Y}_j - \mathbf{f}_0)$ from new observed profiles $\mathbf{Y}_j, j = 1, 2, \dots$, and non-overlapping batch mean vector $\boldsymbol{\omega}_k(r) = r^{-1} \sum_{u=1}^r \boldsymbol{\omega}_{(k-1)r+u}, k = 1, 2, \dots$

5. Threshold $\boldsymbol{\omega}_k(r)$ with $\boldsymbol{\lambda}^L$ and $\boldsymbol{\lambda}^U$; and obtain thresholded DWTs $\boldsymbol{\omega}_k^*(r)$. Calculate the associated statistic $T_k^2(r)$ from (4).
6. Raise an alarm for $\boldsymbol{\omega}_k(r)$ if $S^+(k) \geq H$ or $S^-(k) \geq H$, where

$$S^\pm(k) = \begin{cases} 0, & \text{for } k = 0, \\ \max \left\{ 0, S^\pm(k-1) \pm (T_k^2(r) - \widehat{\boldsymbol{\mu}}_{T^2(r)} - K) \right\}, & \text{for } k = 1, 2, \dots \end{cases} \quad (6)$$

Remark 1: In some applications, the in-control mean profile \mathbf{f}_0 may not be known. As an alternative, we use sample mean of Phase I data $\widehat{\mathbf{f}}_0 = \sum_{j=1}^N \mathbf{Y}_j / N$ as an estimator of \mathbf{f}_0 . Correspondingly, $\boldsymbol{\theta}_0$ is replaced with $\widehat{\boldsymbol{\theta}}_0 = \mathbf{W}\widehat{\mathbf{f}}_0$.

Remark 2: Lee et al. (2012) show batching reduces covariances and improves the ARL_1 performance of control charts.

WDFTC_a and WDFTC_s have quite different mechanisms. WDFTC_s analyzes the DWT of the in-control mean profile and only monitors a subset of DWT components, while WDFTC_a watches nonzero components from thresholded DWTs, whose locations vary for different observed profiles. More specifically, WDFTC_s constructs T^2 -type statistics by $p \times 1$ reduced-dimension DWTs of profiles and the corresponding $p \times p$ covariance matrix. On the other hand, WDFTC_a constructs T^2 -type statistics by $n \times 1$ thresholded DWTs of profiles and the $n \times n$ covariance matrix, but the number of nonzero components of thresholded DWTs is far smaller than n when in-control, so matrix multiplications to calculate monitored statistics can be accelerated by exploring sparsity of thresholded DWTs.

Algorithm CMR

1. Divide the Phase I data set into two disjoint subsets \mathcal{S}_1 and \mathcal{S}_2 with size $N_1 = \lfloor N(1 - \log N) / \log N \rfloor$ and $N_2 = N - N_1$, respectively.
2. Calculate the sample covariance matrices $\widehat{\Lambda}_{\mathcal{S}_1}^\#$ and $\widehat{\Lambda}_{\mathcal{S}_2}^\#$.
3. Compute the estimated threshold,

$$\widehat{\tau} = \arg \min_{\widehat{\tau} \geq 0} \sum_{\substack{(\mu \neq \nu) \text{ and} \\ (2^L < \mu \text{ or } 2^L < \nu)}} \left\{ [\widehat{\Lambda}_{\mathcal{S}_1}^\#]_{\mu, \nu} \mathbf{1}(|[\widehat{\Lambda}_{\mathcal{S}_1}^\#]_{\mu, \nu}| \geq \widehat{\tau}) - [\widehat{\Lambda}_{\mathcal{S}_2}^\#]_{\mu, \nu} \right\}^2. \quad (7)$$

4. Calculate the sample covariance matrix $\widehat{\Lambda}_0^\#$ using the entire Phase I data set of size N and apply the threshold $\widehat{\tau}$ from (7) to $\widehat{\Lambda}_0^\#$ and obtain the regularized sample covariance matrix $\widetilde{\Lambda}_0^\#$ as below.

$$\widetilde{\Lambda}_0^\# = \begin{cases} [\widehat{\Lambda}_0^\#]_{\mu, \nu}, & \text{if (i) } \mu \neq \nu \text{ or (ii) } \mu \leq 2^L \text{ and } \nu \leq 2^L, \\ [\widehat{\Lambda}_0^\#]_{\mu, \nu} \mathbf{1}(|[\widehat{\Lambda}_0^\#]_{\mu, \nu}| \geq \widehat{\tau}), & \text{otherwise,} \end{cases}$$

where $\mathbf{1}(\cdot)$ is the indicator function.

Algorithm BSD

1. Obtain $\widehat{\tau}$ and the regularized sample covariance matrix $\widetilde{\Lambda}_0^\#$ from Algorithm CMR. Set a subset $\mathcal{O} = \{(\mu, \nu) : \mu \neq \nu, 2^L < \mu \leq n, \text{ and } 2^L < \nu \leq n\}$ containing all off-diagonal elements of $\widetilde{\Lambda}_0^\#$ but excluding the estimated covariances between pairs of scaling coefficients.
2. Let \mathcal{Q} denote the number of nonzero elements in \mathcal{O} .

$$\mathcal{Q} = \sum_{(\mu, \nu) \in \mathcal{O}} \mathbf{1}(|[\widetilde{\Lambda}_0^\#]_{\mu, \nu}| > 0).$$

- 2a. If $\mathcal{Q} = 0$, then set $r = 1$ and stop. Otherwise, go to step [2b].
- 2b. Calculate the average magnitude ζ of nonzero elements in \mathcal{O} . Then set the batch size $r = \lceil \sqrt{2}\zeta/\widehat{\tau} \rceil$ and stop.

$$\zeta = \sum_{(\mu, \nu) \in \mathcal{O}} |[\widetilde{\Lambda}_0^\#]_{\mu, \nu}| / \mathcal{Q}.$$

3.2 Thresholding Methods

Many researchers have developed wavelet thresholding methods with Gaussian noise as a canonical model. For example, minimax thresholding method and universal thresholding method from Donoho and Johnstone (1994) are used for Gaussian white noise, while a level-dependent thresholding method from Johnstone (1999) works for cross-correlated Gaussian noise. Since WDFTC_a is designed to be distribution-free, we need a thresholding method that is both distribution-free and effective in the presence of cross correlations among noise components. In this subsection, we propose a distribution-free thresholding method to incorporate with WDFTC_a.

When the monitored process is in-control, the difference between an observed profile and the in-control mean profile has zero expected value, $E[\mathbf{d}_j] = E[\mathbf{Y}_j - \mathbf{f}_0] = \mathbf{0}$, so does its DWT vector $E[\boldsymbol{\omega}_j] = E[\mathbf{W}\mathbf{d}_j] = \mathbf{0}$. Thus it would be best to suppress all components of $\boldsymbol{\omega}_j$. Some existing thresholding methods can achieve this goal at least in the asymptotic sense for Gaussian white noise model, for example, universal thresholds that are mentioned above. On the other hand, when a shift occurs, the expected value of the difference between an out-of-control profile \mathbf{Y}_j and an in-control mean profile \mathbf{f}_0 becomes $\mathbf{f}_1 - \mathbf{f}_0 \neq \mathbf{0}$, so it is desirable to keep more components after thresholding so that shifts can be detected faster. This implies that a desired thresholding method in WDFTC_a should satisfy the following three critical properties: (a) it is distribution-free, (b) it keeps scaling coefficients but filters out most detail coefficients when in-control, and (c) it keeps some detail coefficients in addition to scaling coefficients when out-of-control.

Property (b) helps monitored statistics satisfy the underlying assumption of Equation (5) (i.e., cumulative sum statistics $S^\pm(k)$ behave like reflected Brownian motion processes when in-control). If we threshold both scaling coefficients and detail coefficients, it is possible that there is no component survived after thresholding for some profiles. Thus we may often obtain statistics with zero values, and behaviors of such statistics are hardly approximated well by a Brownian motion process. Additionally, scaling coefficients contain important information of profiles due to the nature of wavelet transform, such as the magnitude of noise. Therefore we prefer to keep all scaling coefficients without thresholding. On the other hand, if the number of preserved detail coefficients changes significantly from one profile to another when in-control, which is often caused by small thresholds or high correlations between components, the variability of T^2 -type statistics becomes extremely large and it takes many observations until $S^\pm(k)$ exhibits appropriate asymptotic behaviors. As a result, actual ARL_0 of the control chart may be larger than the target value for finite target ARL_0 . To ensure well-behaved T^2 -type statistics when in-control, Property (b) is required.

Property (c) is critical to the effectiveness of control charts. Thresholds cannot be too large to filter out all of the important coefficients when out-of-control, since shift information can also be filtered out. This would cause a control chart inefficient in shift detection. It is recommended that one chooses relatively large but not extremely large thresholds and keep more detail coefficients when out-of-control, so that T^2 -type statistics become larger and $S^\pm(k)$ exit control limits faster when a shift occurs. To ensure Properties (b) and (c), we design our thresholding method in a way that it keeps all scaling coefficients and relatively large outliers of detail coefficients only.

Next, we review the universal thresholding method and then propose our new thresholding method developed from the fourth moment Cornish-Fisher expansion from Bekki et al. (2009).

3.2.1 Universal Thresholds

Universal thresholds, developed by Donoho and Johnstone (1994), are one of the most well-known wavelet thresholds. They provide some desirable asymptotic characteristics, which are stated below.

Lemma 1. *If z_1, z_2, \dots, z_n are independent and identically distributed as a standard normal distribution, $P\{\max_{1 \leq i \leq n} |z_i| > \sqrt{2 \log n}\} \rightarrow 0$ as $n \rightarrow \infty$. Therefore,*

$$\lim_{n \rightarrow \infty} \frac{\max_{1 \leq i \leq n} |z_i|}{\sqrt{2 \log n}} = 1, \text{ almost surely.}$$

The universal threshold for the i th noise component is defined as $\lambda_i = \sigma_i \sqrt{2 \log n}$, where σ_i is the marginal standard deviation of the i th component. Because the term $\sqrt{2 \log n}$ is independent from distribution parameters, this thresholding method is called universal and known to be robust to different signals or profiles.

Supported by Lemma 1, universal thresholds are designed to filter out all detail coefficients asymptotically, so that zero components in the underlying mean profile are estimated as zero with high probability. This implies that universal thresholds satisfy Property (b) and (c). However, those properties do not hold any longer when noises are non-normal and/or have cross correlations. Averkamp and Houdré (2003) and Averkamp and Houdré (2005) present thresholding methods for non-normal noises. But their thresholds vary from distributions and have no closed forms, making it hard to implement. Besides, their thresholds also tend to select more components than universal thresholds, violating Property (b). Instead, we present a distribution-free thresholding method that has similar properties as the universal thresholding method.

3.2.2 Cornish-Fisher Expansion Thresholds

We interpret the universal threshold by connecting it to a hypothesis testing:

$$H_0 : E[\omega_{j,i}] = 0 \text{ versus } H_1 : E[\omega_{j,i}] \neq 0, \text{ for } 2^L < i \leq n \text{ and } j = 1, 2, \dots$$

When $|\omega_{j,i}| > \widehat{\sigma}_i \sqrt{2 \log n}$, we reject H_0 and keep $\omega_{j,i}$. Otherwise, we filter out $\omega_{j,i}$ and set $\omega_{j,i} = 0$. Let $q = \Phi(\sqrt{2 \log n})$, where Φ represents the standard normal cumulative distribution function. If noises are assumed to follow IID normal distributions, universal thresholds can be considered as q -quantile estimates. We extend this idea to a general noise distribution as follows.

- If $\omega_{j,i} < z_i^{1-q}$ or $\omega_{j,i} > z_i^q$, reject H_0 and keep $\omega_{j,i}$, where z_i^{1-q} and z_i^q are the $(1-q)$ - and q -quantiles of the underlying distribution of $\omega_{j,i}$.
- Otherwise, filter out $\omega_{j,i}$ and set $\omega_{j,i} = 0$.

In other words, critical quantile values z_i^{1-q} and z_i^q serve as the lower and upper thresholds for $\omega_{j,i}$. By using quantiles, we can determine whether a wavelet component is significant with an approximately similar confidence level as universal thresholds, no matter what underlying distributions are. Moreover, as $n \rightarrow 0$, $q \rightarrow 1$, Lemma 1 still holds.

For profiles with a moderate dimension size $n = 512$, the probability from universal thresholds is $q = \Phi(\sqrt{2 \log 512}) = 0.9998$. The 0.9998-quantile is considered as an extreme quantile. It is known that accurate estimation of extreme quantiles is difficult. However, for our purpose, we only need relatively good thresholds that are large enough to guarantee Properties (a)–(c), instead of accurate quantile estimates. We consider indirect quantile estimation with the Cornish-Fisher (CF) expansion from Bekki et al. (2009), because it can provide good quantile estimates with a moderate data set and low data storage requirements. Additionally, experimental results in Bekki et al. (2009) show applaudable performance with the fourth moment expansion. A q -quantile estimator from the CF expansion is defined as follows.

$$z_i^q = \widehat{\mu}_i + \widehat{\sigma}_i x_i^q, \text{ where } x_i^q = z_q + \frac{1}{6}(z_q^2 - 1)\widehat{\gamma}_{1i} + \frac{1}{24}(z_q^3 - 3z_q)\widehat{\gamma}_{2i} - \frac{1}{36}(2z_q^3 - 5z_q)\widehat{\gamma}_{3i}, \quad (8)$$

where z_q is the q -quantile of the standard normal distribution, and $\widehat{\mu}_i$, $\widehat{\gamma}_{1i}$ and $\widehat{\gamma}_{2i}$ are sample mean, sample central standardized skewness, and sample central standardized excess kurtosis for the i th component from Phase I data, respectively.

Bekki et al. (2009) also point out that quantile estimates based on the CF expansion may be inaccurate for extreme quantiles for skewed distributions, but perform reasonably well for symmetric distributions. Since wavelet coefficients are weighted averages and tend to be more symmetric than the raw data, so the CF expansion provides relatively good thresholds for WDFTC_a.

Consider profiles with IID exponential noise distributions. We set the profile dimension $n = 512$ and each noise component to follow centralized exponential distributions with parameter 1 (i.e., $\varepsilon_{ij} \sim \exp(1) - 1$, $i = 1, 2, \dots, n$, and $j = 1, 2, \dots$). If we apply Haar DWT, the highest level of wavelet components is difference of two exponential random variables, which follows the distribution, Laplace(0, 1). We calculate theoretical values of the 0.0002- and 0.9998-quantiles from Laplace(0, 1) and also estimate them with the CF expansion based on 500 randomly sampled noise vectors. Since wavelet components at the highest level ($n/2 < i \leq n$) follow the same distribution, we can obtain average and standard error (SE) of the 0.0002- and 0.9998-quantile estimates based on results of $n/2$ components. Table 2 provides average and SE for the CF expansion (CF), and also provides theoretical values for the Laplace distribution (Laplace) and universal thresholds (Universal). For IID exponential noises, the CF expansion provides more accurate quantile estimates than universal thresholds and the CF estimates are also close to the theoretical values. On the other hand, universal thresholds ± 3.53 are clearly off for such a non-normal distribution.

3.2.3 Inflation Factor

With the probability $q = \Phi(\sqrt{2 \log n})$ from universal thresholds, the CF expansion provides good thresholds for independent normal and non-normal profiles. However, when positive cross correlations exist, more components tend to survive after thresholding and associated T^2 -type statistics change significantly from one profile to another even when in-control. It becomes more difficult to achieve a good approximation of

Table 2. Average and SE of the 0.0002- and 0.9998-quantile estimates for wavelet components at the highest level.

	The 0.0002-quantile			The 0.9998-quantile		
	CF	Universal	Laplace	CF	Universal	Laplace
Average	7.324	3.53	7.824	-7.537	-3.53	-7.824
SE	0.091	N/A	N/A	0.105	N/A	N/A

a Brownian motion process, and thus Equation (5) does not work well. We want to adjust q value when there are strong correlations. To this end, we employ an inflation factor γ and redefine $q = \Phi(\gamma\sqrt{2\log n})$ for estimating quantiles z_{1-q} and z_q .

Recall that Λ denotes the profile covariance matrix. When all components in a profile's DWT are independent, $\Lambda_{ij} = 0$ for $i \neq j$. When there exist cross correlations among components in a profile's DWT, Λ_{ij} becomes nonzero for some $i \neq j$. Since positive correlations among detail coefficients or among detail coefficients and scaling coefficients, cause more survived components after thresholding, we determine an inflation factor by comparing the actual covariance matrix Λ and its "ideal" covariance matrix S , where S has the same marginal variances for all coefficients as Λ (i.e., $S_{ii} = \Lambda_{ii}$), but zero correlation (i.e., $S_{ij} = 0$ for $i \neq j$). Parameter t , set in Equation (9), can be interpreted as a measure of similarity between the actual covariance matrix Λ and its "ideal" covariance matrix S . If $\Lambda = S$, i.e., all profile variables are independent, t is equal to 1; while t becomes small if Λ and S are different.

When Λ and S are quite different, $1/\sqrt{t}$ turns to be very large. However, if an inflation factor is too large, less components survive and Property (c) is violated. Therefore we set an upper bound for an inflation factor in Equation (10). We recommend to take $\gamma_{max} = 1.5$ to maintain Property (c). Note that without an inflation factor, estimated quantiles by CF expansion are already quite large, therefore an inflation factor should be relatively small and 1.5 is a good empirical choice. Algorithm CFI as follow states procedures for computing an inflation factor γ .

Algorithm CFI

The inflation factor γ is determined by the following steps.

1. Let $\Lambda = \tilde{\Lambda}_0(r)$, where $\tilde{\Lambda}_0(r)$ is a covariance matrix estimate in (4).
2. Set an $n \times n$ matrix S with $S_{ii} = \Lambda_{ii}$ and $S_{ij} = 0$ for $i \neq j$ and $1 \leq i, j \leq n$.
3. Calculate a correlation-type statistic t as follows.

$$t = \frac{\sum_{1 \leq i, j \leq n} (\Lambda_{ij} - \bar{\Lambda})(S_{ij} - \bar{S})}{\sqrt{(\sum_{1 \leq i, j \leq n} (\Lambda_{ij} - \bar{\Lambda})^2) (\sum_{1 \leq i, j \leq n} (S_{ij} - \bar{S})^2)}}, \quad (9)$$

where $\bar{A} = \sum_{1 \leq i, j \leq n} A_{ij}/n^2$ for any matrix A .

4. Set an inflation factor

$$\gamma = \min\left(\frac{1}{\sqrt{t}}, \gamma_{max}\right) \quad (10)$$

and let $q = \Phi(\gamma\sqrt{2\log n})$.

We prove $0 < t \leq 1$ and $1 \leq \gamma \leq 1.5$ in the Appendix. If profiles have independent noises, t equals to 1 and inflation factor γ also equals to 1, so the probability q is as same as the probability from universal thresholds. Additionally, a small t implies that a big discrepancy between the actual covariance matrix and the "ideal" case, and thus both γ and thresholds take larger values. On the other hand, when cross correlation is low, t takes large values and γ is small. Additionally, when negative correlations exist in Λ , we get a larger t (or γ close to 1), compared to the case where all correlations are positive with the same level of correlation. WDFTC_a with CF thresholds is given as below.

Step 2 of Procedure WDFTC_a

By employing CF thresholds, Step 2 of WDFTC_a states as follows.

2. For $i = 1, \dots, n$, calculate sample mean $\hat{\mu}_i$, sample standard deviation $\hat{\sigma}_i$, sample central standardized skewness $\hat{\gamma}_{1i}$, and sample central standardized excess kurtosis $\hat{\gamma}_{2i}$, from $\{\omega_{k,i}(r) : k = 1, \dots, \lfloor N/r \rfloor\}$, where $\omega_{k,i}(r)$ is the i th component of the k th batch means vector $\omega_k(r)$. Then determine the inflation factor γ as Equation (10) from Algorithm CFI. Obtain $q = \Phi(\gamma\sqrt{2\log n})$, and calculate corresponding estimated quantiles z_i^{1-q} and z_i^q from Equation (8). The lower and upper thresholds λ^L and λ^U are determined by

$$\lambda^L = \{z_1^{1-q}, \dots, z_n^{1-q}\} \text{ and } \lambda^U = \{z_1^q, \dots, z_n^q\}.$$

3.3 Covariance Matrix Estimation

When WDFTC_a constructs T^2 -type statistics, a covariance matrix estimate of profile DWTs is required, if the true covariance matrix is unknown. Because the locations of survived DWT components vary from one profile to another, we need the information of the entire $n \times n$ covariance matrix.

Several researchers developed some covariance matrix estimation methods for high-dimensional vectors, such as Bickel and Levina (2008) and Cai and Yuan (2012). We employ the regularization method from Bickel and Levina (2008), because their method is robust to different forms of high-dimensional covariance matrices and easy to implement as shown in Lee et al. (2012). Bickel and Levina (2008) divide the data set into two groups with the ratio $(\log N - 1) : 1$, where N is the size of the data set, and obtain sample covariance matrices for each group, $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$, and search for a threshold s that minimizes the difference between the thresholded Σ_1 by s and Σ_2 in the Frobenius metric. Then the covariance matrix estimate is calculated by thresholding the sample covariance of the entire data set with s .

Algorithm CMR is the regularization method from Bickel and Levina (2008) and also used in Lee et al. (2012), but Lee et al. (2012) use a splitting ratio $2 : 3$ that does not perform well for higher dimensional profiles. Note that we work with $n \times n$ matrix, while Lee et al. (2012) work with $p \times p$ matrix, where $p \ll n$. We find that the original splitting ratio $(\log N - 1) : 1$ in Bickel and Levina (2008) performs better. Thus Algorithm CMR employs the splitting ratio $(\log N - 1) : 1$ for covariance matrix estimation.

4. Experiments

In this section, we test two different in-control mean profiles: Mallat's piecewise smooth function and the LRS data from a lumber manufacturing process. We present comprehensive experimental results of WDFTC_a and compare with its competitor WDFTC_s mentioned in Section 2.2. Recall that WDFTC_s is an effective wavelet-based control chart with static selection, i.e., only monitoring a subset of wavelet components. On the other hand, WDFTC_a monitors adaptively selected wavelet components that vary for different sample profiles.

4.1 Mallat's Piecewise Function

To investigate the effectiveness of WDFTC_a, we take Mallat's piecewise smooth function as the in-control mean profile and test processes on various noise distributions and shifts. We assume that the true covariance matrix of profiles is known for Mallat's piecewise smooth function, but this assumption is relaxed for the LRS data in the next subsection.

We take $n = 512$ equally spaced data points from each observed profile and apply Symmlet 8 wavelets with the coarsest level of resolution $L = 5$. The size of Phase I data set is $N = 20,000$. Set the target $ARL_0 = 200$ and actual ARL_0 and ARL_1 are calculated based on 1000 independent replications. Since we use the true covariance matrix, the batch size cannot be determined by a sample covariance matrix, we use the average batch size reported in Lee et al. (2012) as our batch size for both WDFTC_s and WDFTC_a. For example, the average batch size of WDFTC_s for IID standard normal noises is 3 from Table 4 in Lee et al. (2012), and thus we use 3 as batch size in both WDFTC_s and WDFTC_a. The inflation factor is determined by comparing the true covariance matrix with its "ideal" case.

Additionally, the weight in WRRE for $WDFTC_s$ is set to 0.5 for Mallat's piecewise smooth function, and thus 62 fixed components are selected to construct their statistics.

4.1.1 Shift Configurations

The out-of-control mean profile is set to be $f_1 = f_0 + \eta \zeta \sigma$, where shift size $\eta \in \{0.25, 0.5, 0.75, 1, 2\}$ and $\sigma = (\sigma_1, \dots, \sigma_n)^T$ is the vector of marginal standard deviations of profiles. We define distinctive shifts by using different ζ . First, ζ has a form of a diagonal matrix, i.e., $\zeta = \Delta$, where $\Delta = \text{diag}(\delta_1, \dots, \delta_n)$. We define four shifts as follows.

- Global straight line shift (G1): $\delta_i = 1$ for $i = 1, \dots, n$.
- Global stepwise shift (G2): $\delta_i = 1$ for $i = 1, \dots, n/2$ and $\delta_i = -1$ for $i = n/2 + 1, \dots, n$.
- Local straight line shift (L1): $\delta_i = 1$ for $i \in \mathcal{A}_1 = \{3, 4, \dots, 15\} \cup \{344, 345, \dots, 347\}$ and $\delta_i = 0$ for $i \notin \mathcal{A}_1$.
- Local flare shift (L2): $\delta_i = (i - 480)/32$ for $i \in \mathcal{A}_2 = \{481, \dots, 512\}$ and $\delta_i = 0$ for $i \notin \mathcal{A}_2$.

Next, ζ changes its form to $\zeta = W^{-1}\Theta$, where $\Theta = \text{diag}(\theta_1, \dots, \theta_n)$, a diagonal matrix defined as the same as Section 2.2, only affecting non-selected components from the static selection in $WDFTC_s$.

- Wavelet global shift (WG): $\theta_i = 1$ in ζ_1 , for $i = 1, \dots, n$, see Figure 2 (Left).
- Wavelet local shift (WL): $\theta_i = 1$ in ζ_2 , for $i \in \mathcal{B}_1 = \{80, 81, \dots, 88\}$ and $\theta_i = 0$ for $i \notin \mathcal{B}_1$, see Figure 2 (Right).

4.1.2 Noise Distributions

We consider three multivariate normal distributions, SMN, CMN and GMN, with mean zero and covariance matrices Σ_S , Σ_{CN} , and Σ_G , respectively, and two multivariate shifted exponential distributions, EXP and CEXP, with mean zero and covariance matrices Σ_E and Σ_{CE} , respectively. All five noise distributions SMN, CMN GMN, EXP and CEXP, with their covariance matrices are defined as below.

- SMN: $\Sigma_S = I_n$, where I_n is an $n \times n$ identity matrix.
- CMN: All diagonal elements in Σ_{CN} are equal to 1 and all off-diagonal elements in Σ_{CN} have value 0.5.
- GMN: Diagonal elements $[\Sigma_G]_{ii}$ are taken from Example 2 of Gao (1997), defined as Equation (11) with values from 9.5 to 14.8, and off-diagonal elements are $[\Sigma_G]_{i_1, i_2} = [\Sigma_G]_{i_1, i_1}^{1/2} [\Sigma_G]_{i_2, i_2}^{1/2} \rho(i_1 - i_2)$ for $i_1 \neq i_2$, where $\rho(\cdot)$ are determined by the following Equation (12), originally taken from von Sachs and MacGibbon (2000).

$$[\Sigma_G]_{ii} = \sigma_0^2 \left(1 + \left\{ 0.5 - 2.5 \left[(i-1)/n - 0.515 \right]^2 \right\}^2 \right)^2, \text{ where } \sigma_0^2 = 9.50. \quad (11)$$

$$\rho(\ell) = \text{Corr}[\varepsilon_{i,j}, \varepsilon_{i+\ell,j}] = (-\alpha_2)^{|\ell/2|} \left[\frac{\sin(|\ell|\omega + \xi)}{\sin(\xi)} \right], \text{ where } \ell = 0, \pm 1, \dots, \pm(n-1), \alpha_1 = \frac{4}{3},$$

$$\alpha_2 = -\frac{8}{9}, \omega = \cos^{-1} \left[\frac{\alpha_1}{2\sqrt{-\alpha_2}} \right] \cong 0.785, \xi = \tan^{-1} \left[\frac{1 - \alpha_2}{1 + \alpha_2} \tan(\omega) \right] \cong 1.51. \quad (12)$$

- EXP: $\Sigma_E = I_n$, the same as Σ_S .
- CEXP: We employ the NORTA method in Cario and Nelson (1996) to generate CEXP, by transforming a CMN vector into a shifted exponential (CEXP) vector. Thus the covariance matrix Σ_{CE} has diagonal elements $[\Sigma_{CE}]_{ii} = 1$ and off-diagonal elements $[\Sigma_{CE}]_{i_1, i_2}$ close to 0.5 but slightly less than 0.5 on the average.

4.1.3 Results

We first present the actual values of ARL_0 delivered by $WDFTC_s$ and $WDFTC_a$ for various distributions. Table 3 shows that $WDFTC_s$ delivers an actual value of ARL_0 close to the target, while $WDFTC_a$ delivers an actual value of ARL_0 that slightly deviates from the target. This is mainly due to the larger variance of

the T^2 -type statistics used in $WDFTC_a$ compared with $WDFTC_s$. Thus we get a less accurate estimate of the relevant covariance matrix with 20,000 Phase I data. However, we confirm that this problem goes away when we increase Phase I data. Nevertheless, the deviation from target 200 in Table 3 is acceptable.

Table 3. Actual ARL_0 from $WDFTC_s$ and $WDFTC_a$ for Mallat’s piecewise smooth function under various noise distributions.

Noise type	\bar{r}	γ	$WDFTC_s$	$WDFTC_a$
SMN	3	1.00	202.61	190.62
CMN	3	1.50	204.22	199.58
GMN	8	1.29	196.42	197.26
EXP	3	1.00	196.57	195.91
CEXP	3	1.50	203.99	214.33

Tables 4–5 compare ARL_1 between $WDFTC_s$ and $WDFTC_a$. For all tested noise distributions, both control charts are able to detect the first four shifts (G1, G2, L1, and L2), but $WDFTC_a$ outperforms $WDFTC_s$. Moreover, $WDFTC_a$ performs significantly better than $WDFTC_s$ on small shift size. For example, when shift size $\eta = 0.25$, $WDFTC_a$ detects local shifts L1 and L2 by 9.4% to 32.2% faster than $WDFTC_s$, where local shifts are more difficult to detect than global ones. Furthermore, $WDFTC_a$ can effectively detect wavelet shifts ζ_1 and ζ_2 , where $WDFTC_s$ fails. Even though correlations in the CEXP noise distribution lead to a large inflation factor and make it difficult for $WDFTC_a$ to detect wavelet shifts until the shift size η reaches 2, $WDFTC_a$ still performs better than $WDFTC_s$ because $WDFTC_s$ cannot detect wavelet shifts of any size. Results with the theoretical covariance matrix for Mallat’s piecewise smooth function show that $WDFTC_a$ performs better than $WDFTC_s$ in most scenarios and also detects shifts missed by $WDFTC_s$.

4.2 Lumber Manufacturing Process

We compare experimental results between $WDFTC_s$ and $WDFTC_a$ for LRS data from a lumber manufacturing process. LRS data are generated by a statistical model developed in Staudhammer et al. (2005) and implemented in Staudhammer, Kozak, and Maness (2006), Staudhammer, Maness, and Kozak (2007) and Lee et al. (2012). An LRS profile is the thickness of a large number of tested points on a sawed board, measured along a vertical line from an edge of the board, by an LRS at a certain location of the sawed board.

Given a saw configuration, a board type and a LRS location, a sampled LRS profile is $Y_j = \{Y_{j,1}, \dots, Y_{j,n}\}$, where $Y_{j,i} = f_{0,i} + \varepsilon_{j,i}$ is the measured thickness of the j th board at the i th horizontal distance x_i cm along the length of the board, $i = 1, \dots, n$, and $f_{0,i}$ denotes the true mean of thickness at the same tested point. $f_0 = \{f_{0,1}, \dots, f_{0,n}\}$ is the in-control mean of LRS data, taken over the population. However, f_0 shown in Figure 3, cannot be presented by an explicit form.

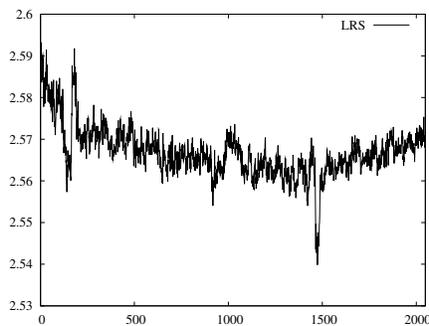


Figure 3. LRS data from a lumber manufacturing process (cm).

We take $n = 2048$ tested points on each sawed board and again use Symmlet 8 wavelets with the coarsest level of resolution $L = 5$. Set the target ARL_0 to 370 and the size of Phase I data to $N = 30,000$. Delivered

Table 4. ARL_1 for Mallat's piecewise smooth function with SMN, CMN and GMN noise vectors.

Shift type	η	SMN		CMN		GMN	
		WDFTC _s	WDFTC _a	WDFTC _s	WDFTC _a	WDFTC _s	WDFTC _a
G1	0.25	3.81	3.44	195.45	191.12	8.00	8.00
	0.5	3.00	3.00	150.01	123.36	8.00	8.00
	0.75	3.00	3.00	86.27	62.03	8.00	8.00
	1	3.00	3.00	49.86	34.14	8.00	8.00
	2	3.00	3.00	12.69	9.38	8.00	8.00
G2	0.25	3.88	3.51	3.00	3.00	8.00	8.00
	0.5	3.00	3.00	3.00	3.00	8.00	8.00
	0.75	3.00	3.00	3.00	3.00	8.00	8.00
	1	3.00	3.00	3.00	3.00	8.00	8.00
	2	3.00	3.00	3.00	3.00	8.00	8.00
L1	0.25	123.58	103.38	73.12	51.56	40.42	28.34
	0.5	33.63	30.83	18.15	13.16	11.81	9.08
	0.75	15.58	13.78	8.65	6.69	8.02	8.00
	1	9.09	8.08	5.55	4.12	8.00	8.00
	2	3.03	3.00	3.00	3.00	8.00	8.00
L2	0.25	141.21	116.22	84.31	60.26	44.88	30.43
	0.5	40.26	35.06	21.31	15.20	13.58	9.76
	0.75	17.84	15.89	9.80	7.38	8.07	8.00
	1	10.47	9.47	6.18	4.77	8.00	8.00
	2	3.16	3.07	3.00	3.00	8.00	8.00
WG	0.25	200.64	58.88	191.77	189.33	203.66	8.00
	0.5	197.93	11.96	198.71	139.10	203.87	8.00
	0.75	198.33	4.19	213.91	31.00	201.18	8.00
	1	195.84	3.01	202.34	6.14	202.12	8.00
	2	201.35	3.00	198.42	3.00	190.07	8.00
WL	0.25	205.53	161.06	199.45	196.02	203.02	190.21
	0.5	207.34	49.86	210.75	125.44	188.08	145.39
	0.75	202.78	11.03	202.32	13.24	204.72	58.94
	1	200.93	4.39	200.85	3.43	200.82	18.97
	2	200.46	3.00	193.36	3.00	202.68	8.00

ARL 's are calculated based on 1000 independent replications. For LRS data, we estimate covariance matrix by using Algorithm CMR with ratio $(\log N - 1) : 1$.

Additionally, the weight in WRRE for WDFTC_s is set to 0.7 for LRS data, and thus 92 fixed components are selected to construct their statistics.

4.2.1 Noise Distributions

As mentioned early, Staudhammer, Kozak, and Maness (2006) points that the thickness of a sawed board is affected by a saw configuration, a board type and an LRS location, and thus noise components have the form

$$\varepsilon_{j,i} = f(x_i) + \mathcal{B}_j + \mathcal{L}_v + \mathcal{B}\mathcal{L}_{jv} + \ell_{jvi} \text{ for } i = 1, \dots, n, \quad (13)$$

where: (a) $\mathcal{B}_j \sim N(0, \sigma_{\mathcal{B}}^2)$ with $\sigma_{\mathcal{B}} = 0.0204$ cm, is the random effect of the j th sample board; (b) $\mathcal{L}_v \sim N(0, \sigma_{\mathcal{L}}^2)$ with $\sigma_{\mathcal{L}} = 0.0052$ cm, is the random effect of the v th laser location; (c) $\mathcal{B}\mathcal{L}_{jv} \sim N(0, \sigma_{\mathcal{B}\mathcal{L}}^2)$ with $\sigma_{\mathcal{B}\mathcal{L}} = 0.0238$ cm, is the random effect caused by the interaction of the board and laser-location effects; (d) ℓ_{jvi} is the random affect arising from the interaction of the board, laser-location and the distance x_i along the board, so that the process $\{\ell_{jvi} : i = 1, \dots, n\}$ is assumed to be stationary and represented by an ARIMA(1,1,1) time series model, where

$$(1 - \alpha\mathcal{B})(\ell_{jvi} - \ell_{jv(i-1)}) = (1 - \beta\mathcal{B})\varepsilon_i \text{ for } i = 1, 2, \dots, \quad (14)$$

Table 5. ARL_1 for Mallat's piecewise smooth function with EXP and CEXP noise vectors.

Shift type	η	EXP		CEXP	
		WDFTC _s	WDFTC _a	WDFTC _s	WDFTC _a
G1	0.25	4.23	4.32	199.33	172.44
	0.5	3.00	3.00	171.77	153.04
	0.75	3.00	3.00	157.02	111.57
	1	3.00	3.00	131.06	75.63
	2	3.00	3.00	43.50	19.82
G2	0.25	4.35	4.39	6.05	3.34
	0.5	3.00	3.00	3.00	3.00
	0.75	3.00	3.00	3.00	3.00
	1	3.00	3.00	3.00	3.00
	2	3.00	3.00	3.00	3.00
L1	0.25	139.06	124.92	185.56	138.96
	0.5	38.49	36.97	66.38	34.26
	0.75	17.20	16.54	27.54	15.12
	1	9.93	9.67	15.37	8.84
	2	3.09	3.02	5.06	3.00
L2	0.25	145.67	131.99	191.12	148.23
	0.5	44.50	43.58	76.19	41.38
	0.75	19.94	19.45	33.38	17.61
	1	11.22	11.08	17.87	10.23
	2	3.43	3.33	5.57	3.01
WG	0.25	206.36	117.92	202.43	218.82
	0.5	199.38	38.63	209.50	215.50
	0.75	201.11	13.64	208.58	211.14
	1	197.65	6.39	210.53	207.17
	2	208.70	3.00	203.63	160.85
WL	0.25	203.22	178.99	207.62	211.78
	0.5	203.96	99.86	210.83	206.97
	0.75	202.04	22.81	201.89	211.72
	1	211.41	6.96	209.04	203.71
	2	197.41	3.00	211.53	5.31

where: (a) \mathcal{B} is the backshift operator so that $(1 - \alpha\mathcal{B})\ell_{jvi} = \ell_{jvi} - \alpha\ell_{jv(i-1)}$; and (b) $\{\varepsilon_i : i = 1, 2, \dots\}$ is a white noise process, where $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ with $\widehat{\sigma}_\varepsilon = 0.00967$ cm. Staudhammer, Kozak, and Maness (2006) use the autoregressive parameter $\widehat{\alpha} = 0.00053$ cm and the moving-average parameter $\widehat{\beta} = 0.00178$ cm.

4.2.2 Shift Configurations

We take the same four types of process faults as Lee et al. (2012): machine positioning problem (MPP), taper, flare and snake, see Figure 4, and briefly explain their causes and associated shifts as follows. Set $\delta = \{\delta_1, \dots, \delta_n\}$ as an associated shift vector, and thus the out-of-control mean profile has the form $f_1 = f_0 + \delta$. More details can be found in Staudhammer, Kozak, and Maness (2006) and Staudhammer, Maness, and Kozak (2007).

- MPP is caused by incorrect positions of saw guides and results in a uniform change in thickness along the length of the board, with $\delta_i = \tau$ for $i = 1, \dots, n$
- Taper is caused by machine misalignment and results in a gradual increment or decrement in the board thickness along the length of the board, with $\delta_i = x_i\tau/x_n$ for $i = 1, \dots, n$.
- Flare is caused when feed roll does not engage at a proper time, and results in an unexpected increment in the board thickness at the end of the board, with $\delta_i = 0$ for $i < i_0$ and $\delta_i = (x_i - x_{i_0})\tau/(x_n - x_{i_0})$ for $i \geq i_0$, where $i_0 = \max\{i : x_i < x_n - 15\}$.
- Snake is usually caused by several saw problems and results in a waveform with the period along the

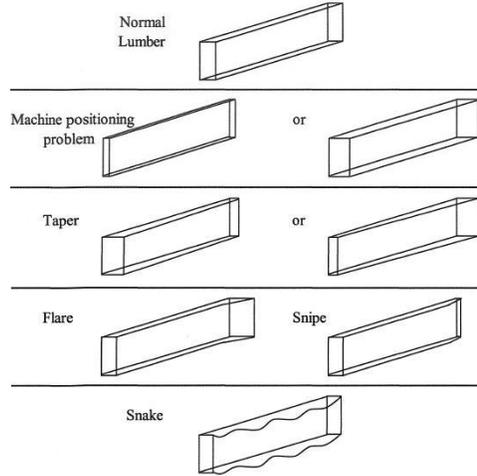


Figure 4. Four major types of process faults.

length of the board, with $\delta_i = \tau \sin(2\pi x_i/P)$ for $i = 1, \dots, n$, where τ is the amplitude and the period $P = 182.88$ cm.

Parameter τ takes values from $\{0.0254, 0.0508, 0.0762, 0.1016\}$ cm for MPP, while $\tau \in \{0.0508, 0.1016, 0.1524, 0.2032\}$ cm for others.

4.2.3 Results

From Algorithm CFI, we obtain an inflation factor for $WDFTC_a$ from each replication and the average of 1000 replications is $\bar{\gamma} = 1.35$. Table 6 show ARL_0 and ARL_1 for both $WDFTC_s$ and $WDFTC_a$. Both charts deliver actual ARL_0 close to the target value, but $WDFTC_a$ performs significantly better than $WDFTC_s$ for all shifts and detects some shifts with almost half observed profiles compared to $WDFTC_s$. Similar to Mallat case, $WDFTC_a$ achieves more ARL_1 saving on small shifts. For some shifts that are difficult to detect, such as flare, $WDFTC_a$ uses less than 28.9% to 42.9% observed profiles than $WDFTC_s$, and this saving is larger than some global shifts, such as MPP.

$WDFTC_a$ keeps only 32.000022 components on the average when in-control, including all $2^L = 32$ scaling coefficients. On the other hand, $WDFTC_s$ selects 92 components for each observed profile. $WDFTC_a$ does not only select significantly fewer wavelet components than $WDFTC_s$, but also chooses components adaptively with as much information as possible. Moreover, fewer monitored components speed up shift detecting and especially benefits small shifts, see Jeong, Lu, and Wang (2006). Hence $WDFTC_a$ outperforms $WDFTC_s$ on LRS data and achieves practical improvement.

5. Conclusion

We develop a wavelet-based control chart with adaptive thresholds for high-dimensional profiles with general noise distributions and some correlations. The main idea of $WDFTC_a$ is to employ an adaptive thresholding method to select a small number of wavelet components when in-control and more components when out-of-control. CF thresholds are the adaptive thresholds implemented in $WDFTC_a$ and can be considered as the lower and upper quantiles associated with the probability of universal thresholds in the standard normal distribution. $WDFTC_a$ performs well for high-dimensional profiles and is especially sensitive to small shifts and/or local shifts. It can also detect shifts that are missed by existing wavelet-based control charts with static selection. Experimental results show $WDFTC_a$ outperforms $WDFTC_s$ for both Mallat's piecewise smooth function and LRS data. $WDFTC_a$ achieves significant improvement with LRS data and reveals its potential in real applications.

Table 6. ARL's for LRS data.

	τ	WDFTC _{τ}	WDFTC _{α}
ARL ₀	0	361.72	367.23
MPP	0.0254	156.48	138.74
	0.0508	43.18	38.87
	0.0762	20.94	19.86
	0.1016	13.65	12.75
Taper	0.0508	168.92	89.06
	0.1016	46.47	27.11
	0.1524	22.24	13.40
	0.2032	14.07	9.71
Flare	0.0508	269.26	153.87
	0.1016	91.75	65.27
	0.1524	42.30	27.81
	0.2032	24.45	14.08
Snake	0.0508	88.29	57.07
	0.1016	23.85	16.73
	0.1524	12.54	9.49
	0.2032	10.17	8.19

Acknowledgements

The first two authors were supported by the NSF under Grant CMMI-1131047. The third author is supported by (while serving at) the National Science Foundation. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

- Averkamp, R., and C. Houdré. 2003. "Wavelet Thresholding for non necessarily Gaussian Noise: Idealism." *The Annals of Statistics* 31 (1): 110–151.
- Averkamp, R., and C. Houdré. 2005. "Wavelet Thresholding for non necessarily Gaussian Noise: Functionality." *The Annals of Statistics* 33 (5): 2164–2193.
- Bekki, J. M., J. W. Fowler, G. T. Mackulak, and B. L. Nelson. 2009. "Indirect cycle time quantile estimation using the Cornish–Fisher expansion." *IIE Transactions* 42 (1): 31–44.
- Bickel, Peter J., and Elizaveta Levina. 2008. "Covariance regularization by thresholding." *The Annals of Statistics* 36 (6): 2577–2604.
- Cai, T. Tony, and Ming Yuan. 2012. "Adaptive Covariance Matrix Estimation Through Block Thresholding." *The Annals of Statistics* 40 (4): 2014–2042.
- Cario, M. C., and B. L. Nelson. 1996. "Autoregressive to Anything: Time-Series Input Processes for Simulation." *Operations Research Letters* 19: 51–58.
- Chen, Shuohui, and Harriet Black Nembhard. 2011. "A High-Dimensional Control Chart for Profile Monitoring." *Quality and Reliability Engineering International* 27 (4): 451–464.
- Chicken, E., J. J. Pignatiello Jr., and J. R. Simpson. 2009. "Statistical Process Monitoring of Nonlinear Profiles Using Wavelets." *Journal of Quality Technology* 41 (2): 198–212.
- Donoho, D. L., and I. M. Johnstone. 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., T. K. Das, A. K. Sikder, and A. Kumar. 2003. "Wavelet based identification of delamination emission signal." *IEEE Transactions on Semiconductor Manufacturing* 16 (4): 677–685.
- Gao, H.-Y. 1997. *Wavelet shrinkage estimates for heteroscedastic regression models*. Tech. rep.. Seattle: MathSoft, Inc.
- Gardner, M. M., Jye-Chyi Lu, R. S. Gyurcsik, J. J. Wortman, B. E. Hornung, H. H. Heinisch, E. A. Rying, S. Rao,

- J. C. Davis, and P. K. Mozumder. 1997. "Equipment Fault Detection Using Spatial Signatures." *IEEE Transactions on Components, Packaging, and Manufacturing Technology, Part C* 20 (4): 295–304.
- Gupta, S., D. C. Montgomery, and W. H. Woodall. 2006. "Performance Evaluation of Two Methods for Online Monitoring of Linear Calibration Profiles." *International Journal of Production Research* 44 (10): 1927–1942.
- Jeong, M. K., J. C. Lu, and N. Wang. 2006. "Wavelet-Based SPC Procedure for Complicated Functional Data." *International Journal of Production Research* 44 (4): 729–744.
- Jin, J., and J. Shi. 1999. "Feature-Preserving Data Compression of Stamping Tonnage Information Using Wavelets." *Technometrics* 41 (4): 327–339.
- Jin, J., and J. Shi. 2001. "Automatic Feature Extraction of Waveform Signals for In-Process Diagnostic Performance Improvement." *Journal of Intelligent Manufacturing* 12 (3): 257–268.
- Johnstone, I. M. 1999. "Wavelet shrinkage for correlated data and inverse problems adaptivity results." *Statistica Sinica* 9 (1): 51–83.
- Kang, L., and S. L. Albin. 2000. "On-Line Monitoring When the Process Yields a Linear Profile." *Journal of Quality Technology* 32: 418–426.
- Lada, E. K., J. C. Lu, and J. R. Wilson. 2002. "A Wavelet Based Procedure for Process Fault Detection." *IEEE Transactions on Semiconductor Manufacturing* 15: 79–90.
- Lee, J., Y. Hur, S.-H. Kim, and J. R. Wilson. 2012. "Monitoring nonlinear profiles using a wavelet-based distribution-free CUSUM chart." *International Journal of Production Research* .
- Mallat, S. G. 1989. "Multifrequency channel decompositions of images and wavelet models." *IEEE Transactions on Acoustics, Speech and Signal Processing* 37 (12): 2091–2110.
- Ramsay, J. O., and B. W. Silverman. 1997. *Functional Data Analysis*. New York, NY: Springer.
- Staudhammer, C. L., R. A. Kozak, and T. C. Maness. 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. L., V. M. LeMay, T. C. Maness, and R. A. Kozak. 2005. "Mixed-Model Development of Real-time Statistical Process Control Data in Wood Products Manufacturing." *Forestry Biometry, Modelling and Information Sciences* 1: 19–35.
- Staudhammer, C. L., T. C. Maness, and R. A. Kozak. 2007. "Profile Charts for Monitoring Lumber Manufacturing Using Laser Range Sensor Data." *Journal of Quality Technology* 39 (3): 224–240.
- von Sachs, Rainer, and Brenda MacGibbon. 2000. "Non-parametric curve estimation by wavelet thresholding with locally stationary errors." *Scandinavian Journal of Statistics* 27 (3): 475–499.
- Williams, J. D., W. H. Woodall, and J. B. Birch. 2007. "Statistical Monitoring of Nonlinear Product and Process Quality Profiles." *Quality and Reliability Engineering International* 23: 925–941.
- Woodall, William H. 2007. "Current research on profile monitoring." *Produção* 17 (3): 420–425.
- Zou, Changliang, Fugee Tsung, and Zhaojun Wang. 2007. "Monitoring general linear profiles using multivariate exponentially weighted moving average schemes." *Technometrics* 49 (4): 395–408.

Appendix

Procedure WDFTC_s

Procedure WDFTC_s

Phase I—Obtain the $n \times 1$ DWT of the in-control mean profile θ_0 and assign weights q and $1 - q$ ($0 \leq q \leq 1$), to the compression ratio and reconstructed error of the in-control mean profile f_0 in WRRE, which is defined in Equation (6) from Lada, Lu, and Wilson (2002). Then determine p ($p < n$) wavelet locations of θ_0 that minimize WRRE. WDFTC_s monitors p components from selected locations instead of all n components in each DWT. For observed in-control profiles $\{Y_j : j = 1, \dots, N\}$, follow the following steps.

1. For $j = 1, \dots, N$, obtain the reduced-dimension DWT $\omega_j^\#$, by keeping components from the selected wavelet locations in the DWT $\omega_j = W(Y_j - f_0)$. Calculate sample covariance matrix, denoted as $\Lambda_0^\#$, from $\{\omega_j^\# : j = 1, \dots, N\}$.
2. Apply Algorithm CMR (see Figure 3 in Lee et al. (2012)) to regularize $\Lambda_0^\#$ and obtain $\tilde{\Lambda}_0^\#$.
3. Determine the batch size r by Algorithm BSD (see Figure 4 in Lee et al. (2012)) and calculate non-overlapping batch means, $\omega_1^\#(r), \dots, \omega_{\lfloor N/r \rfloor}^\#(r)$, where $\omega_k^\#(r) = r^{-1} \sum_{u=1}^r \omega_{(k-1)r+u}^\#$. Let $\tilde{\Lambda}_0^\#(r) = \tilde{\Lambda}_0^\# / r$. The Hotelling's T^2 -type statistic is constructed as

$$T_k^2(r) = (\omega_k^\#(r))^T \left(\tilde{\Lambda}_0^\#(r) \right)^{-1} \omega_k^\#(r), \text{ for } k = 1, \dots, \lfloor N/r \rfloor. \quad (15)$$

4. Calculate sample mean $\hat{\mu}_{T^2(r)}$ and sample variance $\hat{\sigma}_{T^2(r)}^2$ of $T_k^2(r)$. With $K = 0.1 \hat{\sigma}_{T^2(r)}$ and a prespecified in-control average run length ARL_0 , solve the root H from Equation (5).

Phase II—Obtain new observed profiles $\{Y_j : j = 1, 2, \dots\}$ and compute non-overlapping batch means $Y_k(r) = r^{-1} \sum_{u=1}^r Y_{(k-1)r+u}$, for $k = 1, 2, \dots$

5. Compute the DWT of the k th batch mean $\omega_k(r) = WY_k(r)$ and its reduced-dimension version $\omega_k^\#(r)$, for $k = 1, 2, \dots$
6. Raise an alarm when $S^+(k) \geq H$ or $S^-(k) \geq H$, where $S^\pm(k)$ is defined in (6).

Proof of $0 < t \leq 1$

We prove $0 < t \leq 1$ for Equation (9) and thus $\gamma = 1/\sqrt{t} \geq 1$. First set the difference between two matrix averages to be $\nabla = \bar{\Lambda} - \bar{S}$. Next we show

$$\bar{\Lambda} < n\bar{S}. \quad (16)$$

Note that $1 \leq i, j \leq n$, and

$$\begin{aligned} n^2 \bar{\Lambda} &= \sum_i \sum_j \Lambda_{ij} = \sum_i \Lambda_{ii} + \sum_i \sum_{j \neq i} \Lambda_{ij} \\ &\leq \sum_i \Lambda_{ii} + \sum_i \sum_{j \neq i} \sqrt{\Lambda_{ii} \Lambda_{jj}} \\ &\leq \sum_i \Lambda_{ii} + \sum_i \sum_{j \neq i} \frac{\Lambda_{ii} + \Lambda_{jj}}{2} \\ &= n \sum_i \Lambda_{ii} = n \sum_i S_{ii} = n^3 \bar{S}, \end{aligned} \quad (17)$$

where the first inequality holds due to the property of correlation; the second inequality holds due to the inequality of arithmetic and geometric means and $\Lambda_{ii} \geq 0$, $1 \leq i \leq n$. On the other side, the equality holds only when all elements in Λ are the same, i.e., $\Lambda_{i,j} = c$, $1 \leq i, j \leq n$ and c is a fixed constant. Such Λ implies that all variables are perfectly positive correlated, which is impossible in reality. Therefore strict inequality (16) holds. Since the denominator of t is always positive, we first check the sign of the numerator.

$$\begin{aligned}
\sum_i \sum_j (\Lambda_{ij} - \bar{\Lambda})(S_{ij} - \bar{S}) &= \sum_i \sum_{j \neq i} (\Lambda_{ij} - \bar{\Lambda})(0 - \bar{S}) + \sum_i (\Lambda_{ii} - \bar{\Lambda})(S_{ii} - \bar{S}) \\
&= \sum_i \sum_{j \neq i} (\Lambda_{ij} - \bar{\Lambda})(-\bar{S}) + \sum_i (S_{ii} - \bar{S})^2 - \nabla \sum_i (S_{ii} - \bar{S}). \tag{18}
\end{aligned}$$

Equation (18) holds due to the definition of \mathbf{S} , $\Lambda_{ii} = S_{ii}$ for $1 \leq i \leq n$, and $\bar{\Lambda} = \bar{S} + \nabla$. Further,

$$\begin{aligned}
\sum_i \sum_{j \neq i} (\Lambda_{ij} - \bar{\Lambda}) &= \sum_i \sum_{j \neq i} \Lambda_{ij} - n(n-1)\bar{\Lambda} = \sum_i \sum_j \Lambda_{ij} - \sum_i \Lambda_{ii} - n(n-1)\bar{\Lambda} \\
&= n^2\bar{\Lambda} - n^2\bar{S} - n(n-1)\bar{\Lambda} = n\bar{\Lambda} - n^2\bar{S} = n(\bar{S} + \nabla) - n^2\bar{S} \\
&= n(1-n)\bar{S} + n\nabla \tag{19}
\end{aligned}$$

$$\nabla \sum_i (S_{ii} - \bar{S}) = \nabla \left(\sum_i S_{ii} - n\bar{S} \right) = \nabla (n^2\bar{S} - n\bar{S}) \tag{20}$$

where (19) holds due to $\sum_i \sum_j \Lambda_{ij} = n^2\bar{\Lambda}$ and (20) holds due to (17). Then

$$\begin{aligned}
(18) &= n(n-1)\bar{S}^2 - n\nabla\bar{S} + \sum_i (S_{ii} - \bar{S})^2 - n^2\nabla\bar{S} + n\nabla\bar{S} \\
&= n^2\bar{S}^2 - n\bar{S}^2 + \sum_i S_{ii}^2 - 2\sum_i S_{ii}\bar{S} + n\bar{S}^2 - n^2\bar{S}\nabla \\
&= \sum_i S_{ii}^2 - n^2\bar{S}^2 - n^2\bar{S}\nabla \tag{21} \\
&= \sum_i S_{ii}^2 - n^2\bar{S}\bar{\Lambda}
\end{aligned}$$

$$> \sum_i S_{ii}^2 - n^2\bar{S}n\bar{S} = \sum_i S_{ii}^2 - \frac{(\sum_i S_{ii})^2}{n} \tag{22}$$

$$\geq \sum_i S_{ii}^2 - \frac{n\sum_i S_{ii}^2}{n} = 0 \tag{23}$$

Inequality (22) holds due to Inequality (16). The last inequality (23) employs Jensen's inequality. Therefore $t > 0$ holds for any $\mathbf{\Lambda}$. According to Cauchy-Schwarz inequality,

$$\left[\sum_i \sum_j (\Lambda_{ij} - \bar{\Lambda})(S_{ij} - \bar{S}) \right]^2 \leq \left(\sum_i \sum_j (\Lambda_{ij} - \bar{\Lambda})^2 \right) \left(\sum_i \sum_j (S_{ij} - \bar{S})^2 \right),$$

thus it is guaranteed that $t \leq 1$ and thus $\gamma \geq 1$. By definition of γ in (10), $1 \leq \gamma \leq 1.5$.

Additionally, when negative correlations exist in $\mathbf{\Lambda}$, ∇ , the difference between $\bar{\Lambda}$ and \bar{S} gets smaller and the numerator of t gets larger due to (21), compared to the covariance matrix with the same level of correlation but all positively correlated, i.e., a matrix with the absolutely value of $\mathbf{\Lambda}$.