IISE Transactions

Taylor & Francis
Taylor & Francis Group

IISE Transactions

ISSN: 2472-5854 (Print) 2472-5862 (Online) Journal homepage: https://www.tandfonline.com/loi/uiie21

Statistical process control procedures for functional data with systematic local variations

Young-Seon Jeong, Myong K. Jeong, Jye-Chyi Lu, Ming Yuan & Jionghua (Judy) Jin

To cite this article: Young-Seon Jeong, Myong K. Jeong, Jye-Chyi Lu, Ming Yuan & Jionghua (Judy) Jin (2018) Statistical process control procedures for functional data with systematic local variations, IISE Transactions, 50:5, 448-462, DOI: 10.1080/24725854.2017.1419315

To link to this article: https://doi.org/10.1080/24725854.2017.1419315

Accepted author version posted online: 21 Dec 2017.
Published online: 13 Feb 2018.

IISE TRANSACTIONS
2018, VOL. 50, NO. 5, 448–462
https://doi.org/10.1080/24725854.2017.1419315

Taylor & Francis
Taylor & Francis Group

Statistical process control procedures for functional data with systematic local variations

Young-Seon Jeong[a], Myong K. Jeong[b], Jye-Chyi Lu[c], Ming Yuan[d] and Jionghua (Judy) Jin[e]

[a]Department of Industrial Engineering, Chonnam National University, Gwangju, South Korea; [b]Department of Industrial and Systems Engineering, Rutgers, the State University of New Jersey, Piscataway, NJ, USA; [c]School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, USA; [d]Department of Statistics, Columbia University, New York, NY, USA; [e]Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI, USA

ABSTRACT

Many engineering studies for manufacturing processes, such as for quality monitoring and fault detection, consist of complicated functional data with sharp changes. That is, the data curves in these studies exhibit large local variations. This article proposes a wavelet-based local random-effect model that characterizes the variations within multiple curves in certain local regions. An integrated mean and variance thresholding procedure is developed to address the large number of parameters in both the mean and variance models and keep the model simple and fit the data curves well. Guidelines are provided to select the regularization parameters in the penalized wavelet-likelihood method used for the parameter estimations. The proposed mean and variance thresholding procedure is used to develop new statistical procedures for process monitoring with complicated functional data. A real-life case study shows that the proposed procedure is much more effective in detecting local variations than existing techniques extended from methods based on a single data curve.

ARTICLE HISTORY
Received 15 December 2015
Accepted 4 December 2017

KEYWORDS
Data denoising; mixed effects; process monitoring; profile data; wavelet thresholding

1. Introduction

Advances in sensing technology have created opportunities to collect large amounts of functional data that can be used to solve problems that previously could not be explored. For example, Jin and Shi (1999) and Zhou *et al.* (2006) used functional data to detect the types of stamping/forging faults in metal forming; Figure 1 shows stamping tonnage signals that contain rich information about the features of stamping process failures (Jin and Shi, 1999, 2001). The individual die forces contribute to the total tonnage signals at different local segments, and different dies may generate different tonnage variations due to variations in the degradation rates, sensitivities to temperature fluctuations, material properties variations, etc. A local change in the tonnage signals can be justified since press machines and dies are designed to actively operate at certain crank angles. Thus, it is important to consider such inherent local variations in the development of monitoring charts. Similar examples can be seen in semiconductor manufacturing (Lada *et al.*, 2002) and nuclear power generation (Omitaou *et al.*, 2006). Functional data can be used in applications other than engineering studies, including on investigating a biomarker of early colon carcinogenesis (Morris *et al.*, 2003) and modeling functional sulfur dioxide samples for environmental monitoring (Castro *et al.*, 2005).

All datasets in the above studies show complicated, local nonstationary characteristics, especially when process faults occur. For example, a tonnage signal (Jin and Shi, 1999) has several regions of nonstationary patterns representing different process problems. The tonnage curves in Zhou and Jin (2005) show that a relatively small change around the center of the peak tonnage indicates a significant process fault. Figure 2 shows four strain gage sensors mounted on four pillars of a forging press machine. These sensors are used to measure the aggregated tonnage forces on the press created by all the dies, and the summation of these four tonnage sensor signals is referred to as the "total tonnage signal" in this article.

The mean profile of the total tonnage signal under normal operational conditions reflects the average properties of the incoming raw materials, workpiece geometry, process setups, and die working conditions, etc. Variations in these tonnage signals reflect natural process variations due to factors that inherently disturb the process, such as unevenly spraying lubricant over the surface of the workpiece, inevitable fluctuations in the temperature of the press and embedded dies, inherent variations in the incoming workpieces including their material properties, etc. For example, loose tie rods and worn bearings in the main crankshaft are reflected by changes in the tonnage profile around the peak. Also, cylinder cushion force stability can be reflected by variations in tonnage profiles within the local region when the upper die starts to separte from the low die (Jin and Shi, 1999). Since tonnage force signals can provide a wealth of information about those process variations and operating conditions, it is always desirable to effectively use these tonnage signals for online process monitoring and quality improvements.

Almost all of the discussed works have advocated using wavelets to handle complicated patterns in their functional data. Another reason for the popularity of wavelets in recent

CONTACT Jye-Chyi Lu jclu@isye.gatech.edu
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/uiie.

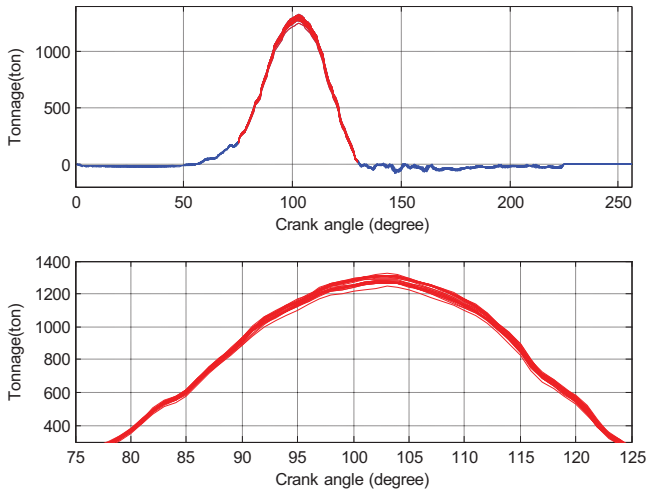


Figure 1. In-control tonnage curves with systematic local variations around the center.

engineering applications is the availability of a fast algorithm for Discrete Wavelet Transform (DWT) that is similar to the Fast Fourier Transform commonly used in signal processing. The computational efficiency of the DWT is better than that of other transforms.

In addition, there has been a considerable amount of literature that has elaborated on the importance of developing wavelet-based methods to analyze complicated high-dimensional data. For example, Jin and Shi (1999) used engineering knowledge to segment data to isolate fault types. Then, wavelet coefficients were selected using a procedure similar to the universal thresholding rule proposed by Donoho and Johnstone (1994). Both segmentation and thresholding reduce the amount of data, creating efficiency in further decision making. Jeong *et al.* (2006a) developed a procedure to balance the modeling accuracy and data reduction goals and treated selected wavelet coefficients as reduced-size data. These reduced-size data were then analyzed using statistical or data-mining procedures (e.g., classification). Soleymani *et al.* (2009) proposed a method combining a general linear model and wavelet transform technique to analyze fixed and random effects via functional magnetic resonance imaging. To identify the fixed and random coefficients, they applied the vertical energy thresholding (VET) method, which is a dimension reduction method used in the wavelet domain. Then, random effect coefficients are identified through the between-variations of different subjects. Morris and Carroll (2006) presented a Bayesian wavelet-based mixed-effect model to represent the broad range of the mean and between-curve correlation structure. The proposed functional mixed model is flexible, accommodating various types of nonstationary functional data. In addition, Guo *et al.* (2012) proposed a wavelet-based Exponentially Weighted Moving Average (EWMA) chart and Cumulative SUM (CUSUM) chart to monitor small mean shifts and variance changes, respectively. The proposed approach used multi-scale Haar wavelet coefficients selected by the optimal level of detail coefficients to detect process faults.

For multiple sets of complicated functional data, Jeong *et al.* (2006b) presented a Phase-II study that established Statistical Process Control (SPC) limits based on estimated model parameters from multiple sets of antenna data curves collected during in-control manufacturing. Lada *et al.* (2002) and Ganesan *et al.* (2003) collected several sets of semiconductor manufacturing data curves to model process behaviors and fault patterns. Castro *et al.* (2005) collected multiple sets of complicated functional data in different time periods for numerous environmental monitoring stations. Chicken *et al.* (2009) proposed a SPC model for a single profile based on the likelihood ratio test, for which all wavelet coefficients were considered. In all of these studies, the functional data curve serves as an experimental unit, and many *replicates* of these units are naturally needed in statistical modeling and data-mining analysis. Paynabar and Jin (2011) used a wavelet-based mixed-effect model to characterize within- and between-profile (nonlinear) variations. The profiles of pressing force signals obtained from a valve seat assembly procedure were used to illustrate the work. In cancer research, Morris and Carroll (2006) and Morris *et al.* (2006) used Bayesian wavelet-based functional mixed models to characterize complicated nonlinear profile data.

However, most existing random effect models, such as Bayesian-based procedures (Chipman *et al.*, 1997; Abramovich *et al.*, 1998; Crouse *et al.*, 1998; Vidakovic, 1998), have considered all coefficients to be random variables. Therefore, this article proposes a wavelet-based random effects model that *can capture between-curve variations in various sizes of local regions* depending on the support region covered by a few random wavelet coefficients. More important, *the proposed model makes no assumption regarding which wavelet coefficient is random and which is not*. In this article, we develop a formal variance-thresholding procedure to identify random wavelet coefficients. One of our goals is to develop a data reduction procedure to extract representative variance parameters. By using the selected wavelet coefficients, the proposed procedure can approximate a profile and can closely model its features, such as local variations or jumps. In addition, based on the proposed

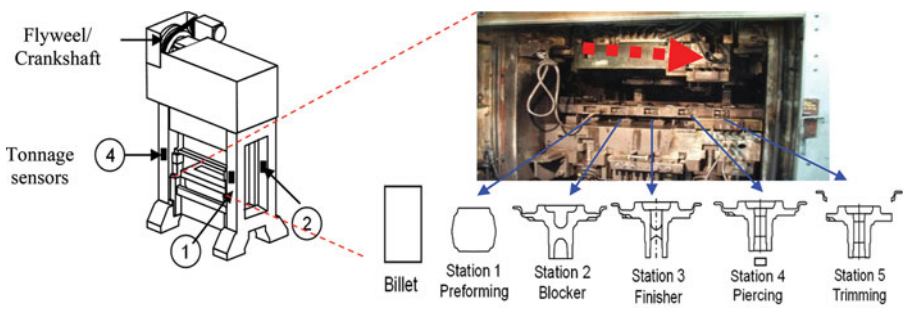


Figure 2. Forging a press with four strain sensors on the columns (adopted from Yang and Jin (2012)).

mean and variance thresholding procedure, we propose a profile monitoring procedure as an extension of process monitoring via wavelet-based mixed effects modeling. The proposed profile monitoring procedure is developed with a smaller number of selected components under a wavelet-based mixed model to detect changes in process mean and variance in functional data.

In summary, this article proposes an integrated Wavelet Mean and Variance Thresholding (WMVT) procedure to select representative coefficients to model nonstationary data patterns and local variations in multiple sensor signals together with SPC procedures for profile monitoring with the selected variables in the wavelet domain. By using the proposed WMVT, our model can detect both changes in the mean and variance, which have different root causes in diagnosis during process monitoring. The experimental results indicate that our procedures are much more effective in detecting local variations when handling complicated multiple functional data curves.

This article is organized as follows: after the introdution in Section 1, Section 2 briefly reviews the background of wavelets, and Section 3 proposes a wavelet-based local random-effect model. Section 4 develops the WMVT procedure and provides guidelines to select the regularization parameters. The profile monitoring procedures and the experimental results are presented in Section 5 and Section 6, respectively. A case study using a real-world example is presented in Section 7. Conclusions and suggestions for possible future works are offered in Section 8.

2. Introduction to wavelets

Let $\mathbf{y}_i = [y_{i1}, y_{i2}, \ldots, y_{iN}]^T$ be a vector of N equally spaced data points (or a curve) at time i obtained from a manufacturing process, where $N = 2^J$ with some positive integer J and $i = 1, 2, \ldots, M$ for independently replicated curves. The superscript T represents the transpose operator. Let $\mathbf{Y} = [\mathbf{y}_1^T, \mathbf{y}_2^T, \ldots, \mathbf{y}_M^T]^T$. When a DWT $\mathbf{W}$ is applied to the data $\mathbf{Y}$, the vector of wavelet coefficients obtained from this transformation is $\mathbf{D} = \mathbf{YW}$, where $\mathbf{D} = [\mathbf{d}_1^T, \mathbf{d}_2^T, \ldots, \mathbf{d}_M^T]^T$, $\mathbf{d}_i = [d_{i1}, d_{i2}, \ldots, d_{iN}]^T$, d_{im} is the wavelet coefficient at the mth wavelet position for the ith data curve, and $\mathbf{W} = [h_{ij}]$, for $i, j = 1, 2, \ldots, N$, is the orthonormal $N \times N$ wavelet-transform matrix. In addition, by using the inverse DTW $\mathbf{W}^{-1}$, the original observations $\mathbf{Y}$ can be reconstructed through $\mathbf{Y} = \mathbf{DW}^{-1}$. See Mallat (1998, Chapter 4) for details of the DWT.

The statistical literature has focused on single-curve data (Jeong *et al.*, 2006a, 2006b). A popular underlying model with a certain constant variance random-error structure—e.g., $\mathbf{y} = \mathbf{f} + \boldsymbol{\varepsilon}$, or $\mathbf{d} = \boldsymbol{\theta} + \boldsymbol{\varepsilon}_d$—is assumed to generate N data points. Then, a few wavelet coefficients can be selected based on a thresholding procedure to estimate the true model (Donoho and Johnstone, 1994). In the signal processing literature, a few of the largest coefficients are selected, and other coefficients are set to zero, to use the inverse DWT to approximate the original data curve (Mallat, 1998, Section 9.2). The selected wavelet coefficients in both the statistical and signal processing literature can be used as "reduced size data" in the follow-up decision analysis (e.g., Jeong *et al.* (2006a)).

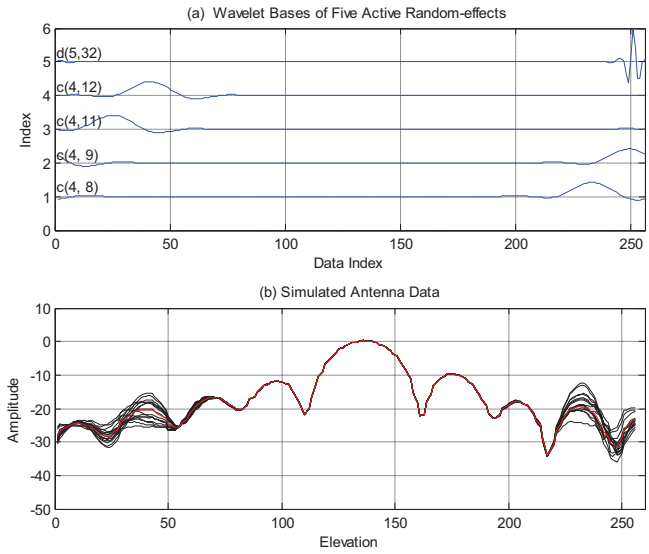


Figure 3. Support areas of active random wavelet coefficients.

Figure 3 yields a better understanding of the relationship between $\mathbf{f}$ in the time domain and its DWT $\boldsymbol{\theta}$ in the wavelet domain. Based on the Symmlet-8 wavelet, Figure 3(a) shows that each wavelet coefficient will only affect the original data curve in its support area. Using all of these coefficients together with the local random-effect model proposed in Section 3, Figure 3(b) illustrates that the original data curves and their local variations can be generated. In Figure 3(b), the red line indicates a baseline, and the 16 black lines represent curves with local variations at both ends, characterized by several wavelet coefficients. Note that the center areas of the curves (black color) overlap with a baseline curve (red color), and the data noise was not added into Figure 3(b).

Considering only the case where a local segment of data has shifted, Jeong *et al.* (2006b) presented an analytical mapping between the time and wavelet domain data. By analyzing the support of the changed wavelet coefficients, the second result of Theorem 1 by Jeong *et al.* (2006b) can be used to identify the locations of the changes in the original time domain. This is important in process fault analysis.

3. Wavelet-based local random-effect model

To elaborate the proposed model, denote θ_{ij} by the jth true wavelet coefficient for the ith curve and d_{ij} the sample version of θ_{ij}. The d_{ij}s are assumed to be independent and to follow a distribution of $N(\theta_{ij}, \sigma^2)$, where the θ_{ij}s and σ^2 are unknown parameters to be estimated. In some support regions of the wavelet coefficients, the data from different replicates have similar behaviors and thus we assume that $\theta_{1j} = \cdots = \theta_{Mj} \equiv \theta_{\cdot j}$ to keep a simple model. In other regions, the data curves differ significantly (see Fig. 3(b)). Then, the θ_{ij}s are modeled as independent random effects with $\theta_{ij} \sim N(\theta_{\cdot j}, \tau_j^2)$, where $\theta_{\cdot j}$ measures the average value of the wavelet coefficients in the jth position and τ_j^2 is the wavelet position–dependent variance. To simplify the expressions, we assume that all $\theta_{ij} \sim N(\theta_{\cdot j}, \tau_j^2)$ with the convention that $\tau_j^2 = 0$ implies a fixed-effect model of $\theta_{\cdot j}$. By using the inverse DWT to approximate the original data curve, replicated curves from the wavelet-based mixed model in

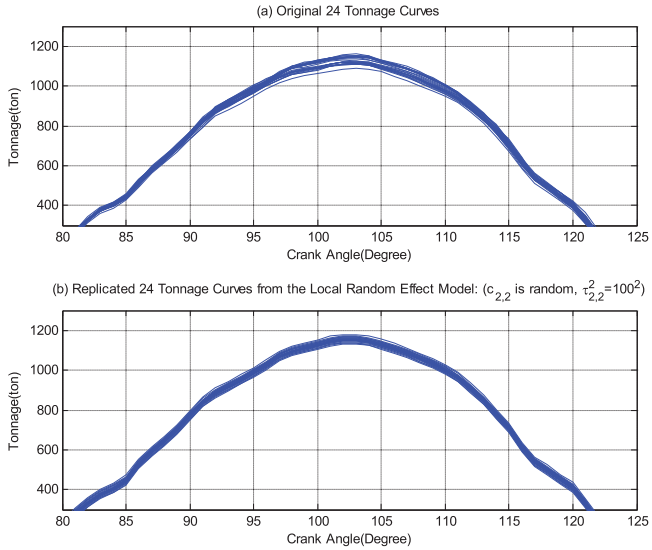


Figure 4. Replicated tonnage curves from the random-effect model.

the time domain will have systematic variations over the region covered by the wavelet coefficients $c_{i,j}$ or $d_{i,j}$, which indicates an approximate coefficient vector and a detail coefficicent vector, respectively.

Example 1 (*Local variations around center—tonnage data*): Focusing only on the center portion of the tonnage signals, Figure 4(a) shows the original data curves in the normal-condition stamping process. Figure 4(b) shows 24 replicated curves from a random-effect model with a variance $\tau^2_{2,2} = 100^2$ for the coefficient $c_{2,2}$, which supports the area $[t_{65}, t_{128}]$ as studied in Figure 2(a). By transforming the coefficient $c_{2,2}$ into the original domain, the local variations around the center similar to the original data are restored, as shown in Figure 4(b). Thus, a wavelet random-effect model can capture the local variations by handling the corresponding wavelet coefficients.

Example 2 (*local variations at side-regions—antenna data*): Suppose that only the following five wavelet coefficients $c_{4,8}$, $c_{4,9}$, $c_{4,11}$, $c_{4,12}$, and $d_{4,16}$ are random. See Figure 3(a) for the support areas of these wavelet bases in the case of the Symmlet-8 wavelet family. Note that all of the support areas from these random effects are only on two sides of the antenna data. Figure 3(b) shows simulated curves with σ^2 set to zero in order to display the impact of these random effects. These examples show that in addition to the typical mean modeling with thresholded wavelet methods (e.g., Jeong *et al.* (2006a)), it is important to decide which wavelet coefficients should be random and to estimate the variances in the random effects. The next section proposes a thresholding method that simultaneously captures both the mean pattern and the local variation for multiple curves.

4. The WMVT procedure

The local random-effect model proposed in Section 3 can be summarized as follows:

$$d_{ij} = \theta_{ij} + z_{ij}, \tag{1}$$

where $\theta_{ij} \sim N(\theta_{\cdot j}, \tau_j^2)$ and $z_{ij} \overset{iid}{\sim} N(0, \sigma^2)$. *Note that one does not know which true wavelet coefficients are random effects, and this will be determined based on the procedure we propose below.*

Let us start with the situation where all coefficients are random and also assume that random-effect coefficients are independent (Guo, 2002; Morris and Carroll, 2006). Estimating the mean and variance parameters can be achieved by minimizing the following negative log-likelihood:

$$M\sum_{j=1}^{N} \ln(\sigma^2 + \tau_j^2) + \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}(d_{ij} - \theta_{\cdot j})^2}{\sigma^2 + \tau_j^2}. \tag{2}$$

To encourage sparsity among $\theta_{\cdot j}$s and $\tau_{\cdot j}$s so that the number of coefficients remains small and data reduction can be achieved, we impose two penalties at the end of the log-likelihood function:

$$M\sum_{j=1}^{N} \ln(\sigma^2 + \tau_j^2) + \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}(d_{ij} - \theta_{\cdot j})^2}{\sigma^2 + \tau_j^2} + \lambda_1 \sum_{j=1}^{N} |\theta_{\cdot j}| + \lambda_2 \sum_{j=1}^{N} \tau_j^2. \tag{3}$$

The minimization of Equation (3) follows the spirit of soft-thresholding. See Remark 1 below for details.

The first penalty term with a regularization parameter λ_1 encourages sparsity among mean parameters $\theta_{\cdot j}$s. The second term with λ_2 encourages some of the τ_js to be zero, which implies that the jth position wavelet coefficient is a fixed effect. See the Remarks after the parameter estimation algorithm for further insight regarding the thresholding effects. Tuning parameters λ_1 and λ_2 control the trade-off between the modeling accuracy (in terms of maximizing the likelihood function) and sparsity. Note that tuning parameters λ_1 and λ_2 should be different, as the number of random-effect coefficients influences the smoothness of the curves, whereas the heterogeneity of multiple curves affects the selection of λ_2. Thus, a larger λ_2 can model multiple homogeneous curves (see Section 4.1 for further details). By sharing information across all multiple curves, the proposed approach achieves both mean and variance thresholding.

Algorithm for Parameter Estimation:

Given λ_1 and λ_2,

1. Initialize an estimate of σ^2.
   Based on our experiments, wavelet coefficients at the finest level are less likely to be random effects. Thus, an initial estimate of σ^2 can be obtained from the following pooled variance idea. For each curve, obtain an estimate of σ^2 based on Donoho and Johnston's (1994) robust estimate. Then, the common variance σ^2 for M curves can be estimated by averaging these robust estimates: $\hat{\sigma} = M^{-1}\sum_{i=1}^{M} 0.6745^{-1}\, median\,(|d_{im}| : N/2 + 1 \le m \le N)$, where the index m indicates wavelet coefficients at the finest level.
2. Initialize an estimate of τ_js.
   (i) If the sample variance of $d_{\cdot j}$ is larger than the current estimate of σ^2, estimate τ_j^2 by the difference between the two. That is, this position of the wavelet coefficients has a random effect.

(ii) Otherwise, estimate τ_j^2 by zero.

3. Update $\hat{\theta}_{\cdot j}$s by minimizing Equation (4) with respect to $\theta_{\cdot j}$s.
   By minimizing the penalized log-likelihood function with respect to $\theta_{\cdot j}$s, we obtain the following closed-form solution for the estimate of $\hat{\theta}_j$s (see Appendix for its detailed derivation).

$$\hat{\theta}_{\cdot j} = \left(|\bar{d}_{\cdot j}| - \lambda_1(\hat{\sigma}^2 + \hat{\tau}_j^2)/(2M)\right)_+ sign(\bar{d}_{\cdot j}), \qquad (4)$$

   where $(x)_+ = \max(x, 0)$ and $\bar{d}_{\cdot j} = (d_{1j} + \cdots + d_{Mj})/M$.
4. Update $\hat{\tau}_j^2$ by minimizing Equation (3) with respect to τ_js.
   Similarly, by minimizing the penalized log-likelihood function with respect to τ_js and by defining $s_j^2 = \sum_{i=1}^{M}(d_{ij} - \hat{\theta}_{\cdot j})^2/M$, we can also obtain a closed-form solution for the updated estimate of $\hat{\tau}_j^2$ as follows (see Appendix for its derivation):

$$\hat{\tau}_j^2 = \left(\frac{-1 + \sqrt{1 + 4s_j^2\lambda_2/M}}{2\lambda_2/M} - \hat{\sigma}^2\right)_+. \qquad (5)$$

5. Update $\hat{\sigma}^2$ by minimizing Equation (3) with respect to σ^2.
   We can solve the following equation to obtain the updated estimate of σ^2 (see Appendix for the detailed derivation):

$$\sum_{j=1}^{N} \frac{\hat{\sigma}^2 + \hat{\tau}_j^2 - s_j^2}{(\hat{\sigma}^2 + \hat{\tau}_j^2)^2} = 0. \qquad (6)$$

6. Repeat Steps (3) to (5) until the differences between the updated estimates for $\theta_{\cdot j}$'s, τ_j^2, and σ^2 and their previous values are smaller than the prespecified precision threshold values.

Remark 1: The algorithm presented above reveals some operating characteristics of the proposed approach. Step 3 is similar to soft-thresholding. Although a hard-thresholding procedure (set smaller $\hat{\theta}_{ij}$ to zero if it is less than the threshold) will retain fewer coefficients and thus achieve better data reduction, soft-thresholding offers various advantages, including continuity of the shrinkage rule (Bruce and Gao, 1996). Hard-thresholding also leads to a larger variance in the estimates, and it is also sensitive to small changes in the data. Interestingly, the minimization of Equation (3) leads to the use of *varying* threshold values for means at different wavelet positions when different *variability* at different positions is considered. Thus, our estimate is expected to outperform soft-thresholding with a *fixed* threshold value developed under a constant variance model used in most of the wavelet-thresholding literature (e.g., Donoho and Johnstone (1994) and Jung *et al.* (2006)).

Remark 2: Step 4 discloses the mechanism behind variance thresholding. By going through some algebric simplification, one can see that Equation (5) implies that $\hat{\tau}_j^2 = 0$ if

$$s_j^2 < \hat{\sigma}^2 + \lambda_2\hat{\sigma}^4/M. \qquad (7)$$

Therefore, the positions whose coefficients display limited variation will be set as fixed effects by shrinking the $\hat{\tau}_j^2$s to zero. Then, these zero coefficients will not be used as reduced-size data in later decision analysis.

Remark 3: Step 5 updates the estimate of σ^2. Although this is an easy one-dimensional optimization problem, it is the most time-consuming step of the algorithm, due to the lack of a closed-form solution. Also, our procedure finds the estimates based on an iterative algorithm. Given λ_1 and λ_2, Equation (3) is a convex function of each of the $\theta_{\cdot j}$s, τ_js, and σ^2 by fixing the others (Yuan and Wahba, 2004). Thus, we can minimize Equation (3) with respect to $\theta_{\cdot j}$s, τ_js, and σ^2 using an iterative method such as the Newton iteration, which is known to be globally convergent for functions convex on R (Stoer and Bulirsch, 2002).

Remark 4: With the estimates of the model parameters, M multiple curves can be reconstructed in the following way. First, obtain the estimate of the θ_{ij}s as follows: (i) for a random-effect position, obtain $\hat{\theta}_{ij}$ from simulated normal random variates with mean $\hat{\theta}_{\cdot j}$ and variance $\hat{\sigma}^2 + \hat{\tau}_j^2$; (ii) for a fixed-effect position, simulate $\hat{\theta}_{ij}$ from a normal distribution with mean $\hat{\theta}_{\cdot j}$ and variance $\hat{\sigma}^2$. Then, apply the inverse DWT with these estimates to reconstruct multiple curves.

Remark 5: We assumed that the wavelet coefficients within a given curve are independent. When this assumption is not satisfied, we need to extend our model by incorporating the covariance regularization into the model in Equations (2) and (3) (Bickel and Levina, 2008), where many more samples are needed due to the large dimensions of the covariance matrix.

4.1. Guideline for the selection of tuning parameters

The effectiveness of the proposed WMVT procedure depends on the tuning parameters λ_1 and λ_2. For the tuning parameters, we apply the leave-one-out cross-validation technique, as the sample size is small. Let the $\hat{\theta}_{\cdot j}^{[k]}$s, $\hat{\tau}_j^{2[k]}$s, and $\hat{\sigma}^{2[k]}$ be the estimates obtained by minimizing the penalized log-likelihood function in Equation (3) based on all data curves except the kth one. The measure of the quality of these estimates is based on the log-likelihood for data d_k (see below). Then, the cross-validation estimate of λ_1 and λ_2 is defined to be the minimizer of the following log-likelihood function for all M curves being left out one at a time in the cross-validation process:

$$V_0^*(\lambda_1, \lambda_2) = \sum_{k=1}^{M}\left[\sum_{j=1}^{N} \ln\left(\hat{\sigma}^{2[k]} + \hat{\tau}_j^{2[k]}\right) + \frac{\sum_{j=1}^{N}(d_{kj} - \hat{\theta}_{\cdot j}^{[k]})^2}{\hat{\sigma}^{2[k]} + \hat{\tau}_j^{2[k]}}\right].$$

If we let K_1 be number of nonzero mean wavelet coefficients and K_2 the number of wavelet random-effect positions, from our experimental experience (not reporpted here) based on various types of curves, the following general guidelines are provided to select λ_1 and λ_2:

1. λ_1: The smoothness of the curves affects λ_1. In general, smoother curves require a smaller K_1. Since a larger λ_1 penalizes the likelihood function more (see

Equation (3)), K_1 becomes smaller. It is then recommended to use a larger λ_1 to model smoother curves, such as tonnage curves.

2. λ_2: The heterogeneity of multiple curves affects the selection of λ_2. For more homogeneous curves, such as tonnage curves, a smaller K_2 is sufficient to capture the local variation around the center peak (see Fig. 3). Thus, a larger λ_2 is recommended to model multiple homogeneous curves. On the other hand, a larger K_2 is needed to model more heterogeneous curves, such as antenna data, and thus a smaller λ_2 is recommended.

5. Profile monitoring via the wavelet-based local random-effect model

This section shows how to use selected wavelets to monitor possible systematic changes of curves at certain local regions. Suppose that an automobile sheet metal stamping process has a process change at time i and thus tonnage curves collected after time i have larger systematic "local variations" compared with the original curves. Figure 5 shows an example of certain local changes around the center. Multiple curves from the normal condition are presented in blue. The curves from the abnormal condition are plotted in red lines that have greater variation around the center than the blue lines. Note that the curves from both conditions match well in regions other than the center area; e.g., data below 60 units and beyond 125 units in the x-axis. There are very few publications on variance monitoring, especially for monitoring abnormal variations in local regions. Therefore, unlike the traditional time domain–based SPC procedures in profile monitoring (Reynolds and Cho, 2006; Huwang *et al.*, 2007; Zou *et al.*, 2007), this section presents *a SPC procedure to monitor a few selected wavelet coefficients capturing systematic local variations*. Comparison studies show a much smaller Average Run Length (ARL) for our proposed SPC procedure. In this study, the following two scenarios are examined: (i) only one process variance changes and (ii) both the mean and variance change.

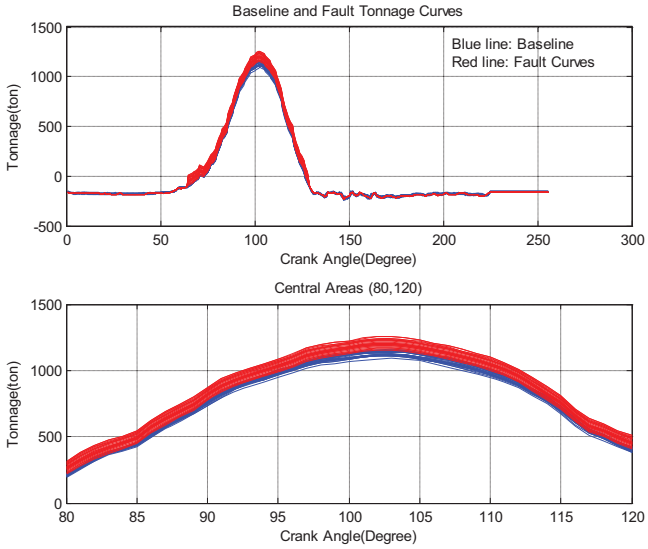


Figure 5. Local variations around the center area in tonnage curves.

5.1. Only the process variance changes

Let $\mathbf{d}_i = (d_{i,1}, d_{i,2}, \ldots, d_{i,N})$ be wavelet coefficients of the ith observation $\mathbf{y}_i$. When the process is in control, the $d_{i,j}$s are independent and are $N(\theta_{\cdot j}, \sigma^2 + \tau_j^2)$ distributed where $\theta_{\cdot j}$s, σ^2 and τ_j^2 were estimated using the WMVT procedure described in Section 4. After identifying the fixed- and random-effect wavelet coefficients through the WMVT procedure, we rearranged the positions of the wavelet coefficients so that the first p_1 variables are random-effect coefficients, the next p_2 variables are fixed-effect coefficients, and the remaining are shrunken coefficients. For these selected coefficients to be monitored, the covariance matrix of wavelet coefficients $\mathbf{\Sigma}_0$ can be expressed as

$$\mathbf{\Sigma}_0 = \begin{bmatrix} \mathbf{\Sigma}_0^{r(p_1\times p_1)} & \mathbf{0}^{p_1\times p_2} & \mathbf{0}^{p_1\times(N-p_1-p_2)} \\ \mathbf{0}^{p_2\times p_1} & \mathbf{\Sigma}_0^{f(p_2\times p_2)} & \mathbf{0}^{p_2\times(N-p_1-p_2)} \\ & \mathbf{0}^{(N-p_1-p_2)\times N} & \end{bmatrix}_{N\times N},$$

where $\Sigma_0^r = (\sigma^2 + \tau_j^2)\mathbf{I}_{p_1}$ and $\Sigma_0^f = \sigma^2\mathbf{I}_{p_2}$ are the covariance matrices of random- and fixed-effect variables, respectively, and $\mathbf{I}_{p_i}$ is the $p_i \times p_i$ identity matrix. Let $\tilde{\boldsymbol{d}}_i^{rf} = (d_{i,1}, \ldots, d_{i,p_1}, d_{i,p_1+1}, \ldots, d_{i,n_1})$ be the vector of only random- and fixed-effect wavelet coefficients and $n_1 = p_1 + p_2$. Then, the mean and covariance matrices of $\tilde{\boldsymbol{d}}_i^{rf}$ are given by

$$\boldsymbol{\theta}_0^{rf} = [\boldsymbol{\theta}_0^r \vdots \boldsymbol{\theta}_0^f]'; \quad \mathbf{\Sigma}_0^{rf} = \begin{bmatrix} \mathbf{\Sigma}_0^{r(p_1\times p_1)} & \mathbf{0}^{p_1\times p_2} \\ \mathbf{0}^{p_2\times p_1} & \mathbf{\Sigma}_0^{f(p_2\times p_2)} \end{bmatrix}_{n_1\times n_1},$$

where $\boldsymbol{\theta}_0^r$ and $\boldsymbol{\theta}_0^f$ are the mean vectors of random- and fixed-effect variables, respectively. Both $\boldsymbol{\theta}_0^{rf}$ and $\mathbf{\Sigma}_0^{rf}$ are known quantities from baseline profiles.

Under the assumption that only the process variance is changed, the hypothesis-testing formulation for a process-monitoring procedure in the wavelet domain is given as follows:

$$\mathrm{H}_0: \theta_{ij} \sim \begin{cases} N(\theta_{\cdot j}, \tau_j^2 + \sigma^2), & j \in D \\ N(\theta_{\cdot j}, \sigma^2), & j \notin D, \end{cases} \quad \text{versus}$$

$$\mathrm{H}_1: \theta_{ij} \sim \begin{cases} N(\theta_{\cdot j}, \tau_j^2 + \eta_j + \sigma^2), & j \in D \\ N(\theta_{\cdot j}, \sigma^2), & j \notin D, \end{cases}$$

where D is the set of random coefficients, $\theta_{\cdot j}$ measures the average value of wavelet coefficients in the jth position, and η_j is the change in process variance for the jth random-effect coefficient.

The process changes under the above situation can be detected by solely monitoring the random-effect variables, $\tilde{\mathbf{d}}_i^r = (d_{i,1}, \ldots, d_{i,p_1})$, in the wavelet domain. In addition, the standardized version of $\tilde{\mathbf{d}}_i^r$, $\tilde{\mathbf{u}}_i = \mathbf{\Sigma}_0^{r-1/2}(\tilde{\mathbf{d}}_i^r - \boldsymbol{\theta}_0^r)$, follows the normal distribution with mean $\boldsymbol{\theta}_i^s = \mathbf{\Sigma}_0^{r-1/2}(\boldsymbol{\theta}_i^r - \boldsymbol{\theta}_0^r)$ and covariance $\mathbf{\Sigma}_i^s = \mathbf{\Sigma}_0^{r-1/2}\mathbf{\Sigma}_i^r\mathbf{\Sigma}_0^{r-1/2}$ where $\boldsymbol{\theta}_i^r$ and $\mathbf{\Sigma}_i^r$ are the mean and covariance vectors of $\tilde{\mathbf{d}}_i^r$, respectively. Thus, when the process is in control, $\tilde{\mathbf{u}}_i$ is distributed as $N(\mathbf{0}, \mathbf{I}_{p_1})$. The SPC procedure presented below will monitor the standardized coefficients $\tilde{\mathbf{u}}_i$.

In the case of Phase II process monitoring with an individual observation, $\mathbf{A}_i = \tilde{\mathbf{u}}_i\tilde{\mathbf{u}}_i'$ can be used as an estimator for $\mathbf{\Sigma}_i^s$ when the process mean does not change. The threshold estimate, $\mathbf{A}_i = \tilde{\mathbf{u}}_i\tilde{\mathbf{u}}_i'$, is not unbiased, but it is consistent in the operator

norm, uniformly over the class of matrices as long as $(\log p)/N \to 0$ (Bickel and Levina, 2008). The estimator does not need to be unbiased, as the monitoring statistics based on a consistent estimator $\mathbf{A}_i = \tilde{\mathbf{u}}_i\tilde{\mathbf{u}}_i'$ can capture the changes in $\mathbf{\Sigma}_i^s$ when the process mean does not change. One way to improve the performance is to combine as much information contained in $\mathbf{A}_i$ as possible from data collected over time to utilize a EWMA chart (Macgregor and Harris, 1993).

Define the EWMA of $\mathbf{A}_i$ at the ith profile observation as

$$\mathbf{Z}_i = \delta\mathbf{A}_i + (1-\delta)\mathbf{Z}_{i-1}, \quad 1 \le i \le M, \tag{8}$$

where $0 < \delta < 1$ is a smoothing constant and $\mathbf{Z}_0 = \tilde{\mathbf{u}}_1\tilde{\mathbf{u}}_1'$ is an initial estimate of the covariance (Macgregor and Harris, 1993). After some algebraic manipulations, Equation (8) can be expressed alternatively as

$$\mathbf{Z}_i = \sum_{k=1}^{i}\delta(1-\delta)^{i-k}\mathbf{A}_k \text{ with } \sum_{k=1}^{i}\delta(1-\delta)^{i-k} = 1.$$

When the process mean does not change, $\mathrm{E}(\mathbf{Z}_i) = \sum_{k=1}^{i}\delta(1-\delta)^{i-k}\mathrm{E}(\mathbf{A}_i) = \mathbf{\Sigma}_i^s$. Thus, $\mathbf{Z}_i$ can be used to estimate $\mathbf{\Sigma}_i^s$. Since the trace, which is the sum of the diagonal of the covariance matrix, measures the overall variability in a covariance matrix, we propose the following monitoring statistic:

$$\mathrm{W}_i = tr(\mathbf{Z}_i) = \sum_{k=1}^{i}\delta(1-\delta)^{i-k}\left(\sum_{j=1}^{p_1}u_{k,j}^2\right). \tag{9}$$

By using the large-sample normal approximation theory, control limits for W_i are given by

$$CL_{\mathrm{W}_i} = p_1 \pm \Phi^{-1}(1-\alpha)\sqrt{\sum_{k=1}^{i}\left(\delta(1-\delta)^{i-k}\right)^2 2p_1},$$

where α is the significance level and Φ is the standard normal distribution function. This confidence interval is utilized to detect local variations for SPC monitoring and will also be used in the later comparisons of simulated results (see Appendix for the detailed derivations of CL_{W_i}).

5.2. Both the process mean and the variance change

Under the assumption that both the process mean and variance may change during the monitoring period, we can detect the process change by monitoring both random- and fixed-effect variables. When the process is in control, we can assume that $\boldsymbol{\theta}_i^{rf} = \boldsymbol{\theta}_0^{rf}$ and $\mathbf{\Sigma}_i^{rf} = \mathbf{\Sigma}_0^{rf}$ where both $\boldsymbol{\theta}_0^{rf}$ and $\mathbf{\Sigma}_0^{rf}$ are the mean and covariance vectors of $\tilde{\mathbf{d}}_i^{rf}$. The standardized version of $\tilde{\mathbf{d}}_i^{rf}$, $\tilde{\mathbf{v}}_i = \mathbf{\Sigma}_0^{rf-1/2}(\tilde{\mathbf{d}}_i^{rf} - \boldsymbol{\theta}_i^{rf})$, follows the normal distribution with mean $\boldsymbol{\theta}_i^{v} = \mathbf{\Sigma}_0^{rf-1/2}(\boldsymbol{\theta}_i^{rf} - \boldsymbol{\theta}_0^{rf})$ and covariance $\mathbf{\Sigma}_i^{v} = \mathbf{\Sigma}_0^{rf-1/2}\mathbf{\Sigma}_i^{rf}\mathbf{\Sigma}_0^{rf-1/2}$. Thus, when the process is in control, $\tilde{\mathbf{v}}_i$ is normally distributed as $N(\mathbf{0}, \mathbf{I}_{n_1})$ where $n_1 (= p_1 + p_2)$ is the number of fixed- and random-effect variables.

Assuming that the process mean and variance change, the hypothesis-testing formulation for a process-monitoring procedure in the wavelet domain is given as

$$\mathrm{H}_0: \ \theta_{ij} \sim \begin{cases} N(\theta_{.j}, \tau_j^2 + \sigma^2), & j \in A \\ N(\theta_{.j}, \sigma^2), & j \notin A, \end{cases} \quad \text{versus}$$

$$\mathrm{H}_1: \ \theta_{ij} \sim \begin{cases} N(\theta_{.j} + \upsilon_j, \tau_j^2 + \eta_j + \sigma^2), & j \in A \\ N(\theta_{.j}, \sigma^2), & j \notin A, \end{cases}$$

where A is the set of shifted wavelet coefficients, υ_j is the change in the process mean for the jth coefficient, and η_j is the change in the process variance for the jth random-effect coefficient.

When the process mean changes during the monitoring period, $\mathbf{Z}_i$ is modified by

$$\mathbf{C}_i = \sum_{k=1}^{i}\delta(1-\delta)^{i-k}(\tilde{\mathbf{v}}_k - \boldsymbol{\gamma}_k)(\tilde{\mathbf{v}}_k - \boldsymbol{\gamma}_k)', \quad 1 \le i \le M, \tag{10}$$

where $\boldsymbol{\gamma}_k$ is the estimate of the process mean. The optimal estimate $\boldsymbol{\gamma}_i$ for process mean at time i is $\varphi\tilde{\mathbf{v}}_i + (1-\varphi)\boldsymbol{\gamma}_{i-1}$ with smoothing weight $0 < \phi < 1$ (Macgregor and Harris, 1993). As

$$\mathrm{E}(\mathbf{C}_i) \to \frac{2(1-\phi)}{(2-\phi)}\mathbf{\Sigma}_i^{v} \text{ as } i \to \infty,$$

then $((2-\phi)/2(1-\phi))\mathbf{C}_i$ can be used as the estimator of $\mathbf{\Sigma}_i^{v}$ (see the Appendix for a derivation of the estimator $\mathbf{\Sigma}_i^{v}$). Since $(2-\phi)/2(1-\phi)$ is a constant, we can propose the following monitoring statistic:

$$\mathrm{Q}_i = tr(\mathbf{C}_i) = \sum_{k=1}^{i}\delta(1-\delta)^{k-i}tr((\tilde{\mathbf{v}}_k - \boldsymbol{\gamma}_k)(\tilde{\mathbf{v}}_k - \boldsymbol{\gamma}_k)'),$$

By using large-sample normal approximation theory, the control limit of Q_i is given by

$$CL_{\mathrm{Q}_i} = n_1\sum_{k=1}^{i}\pi_{kk} \pm \Phi^{-1}(1-\alpha)\sqrt{2n_1\sum_{k=1}^{i}\sum_{j=1}^{i}\pi_{kj}^2}.$$

(See Appendix for the derivation of CL_{Q_i}.)

6. Performance evaluation for process monitoring

This section presents the results of a simulation using tonnage stamping signals to compare the Average Run Length (ARL_1) values from the proposed statistic W_i and Q_i and six Hotelling T^2 procedures extracted from the literature: (i) all wavelet coefficients (denoted as T_A^2) and selected wavelet coefficients based on (ii) VisuShrink method (denoted as T_{Is}^2); (iii) VET method (denoted as T_{Vet}^2); (iv) VertiShrink method (denoted as T_{Vs}^2); (v) Principal Component Analysis (PCA) method (denoted as T_{PCA}^2); and (vi) Paynabar and Jin's method (2011; denoted as T_{PJ}^2). See Equation (11) for a detailed definition. Most of the existing wavelet-based monitoring approaches (e.g., Jin and Shi (2003), Jeong *et al.* (2006b), Zhou *et al.* (2006), and Jin and Li (2009)) consider only within-profile variation, so this article selects the wavelet-based mixed-effect model studied in Paynabar and Jin (2011) to conduct a benchmark performance comparison. Their mixed-effect model captures both within-profile and between-profile variations and is thus the most relevant to our work, which characterizes the local between-profile variations.

The control limits are set to make in-control $ARL_0 = 200$ (Kang and Albin, 2000). The comparison study connects our wavelet selection procedure described in Section 4 to the SPC procedure. That is, we investigate the impact of different selections of wavelet coefficients to the performance of the profile monitoring. This comparison study uses 24 simulated tonnage signals with $N = 256$ baseline observations, as shown in Figure 1. Since the types of wavelet transform are not critical for the simulated performance, in this simulation, the Symmlet-8 wavelet is used with the lowest resolution level (L) set to a value of four. Random noise from the normal distribution $N(0, \sigma^2)$ with $\sigma^2 = 1$ is added to generate 1000 replications for each study.

In order to improve the efficiency of a Hotelling T^2 statistic, several authors have proposed testing a subset of the coefficients. For example, Jin and Shi (2001) applied the VisuShrink data reduction procedure (Donoho and Johnstone, 1994) to select smaller important wavelet coefficients for process monitoring. Jung *et al.* (2006) proposed a VET data reduction procedure with multiple curves. They constructed the Hotelling T^2 with screened coefficients by using the data reduction procedure

$$T_S^2 = (\mathbf{d}_{i,S} - \boldsymbol{\theta}_{0,S})^T \boldsymbol{\Sigma}_{0,S}^{-1} (\mathbf{d}_{i,S} - \boldsymbol{\theta}_{0,S}), \tag{11}$$

where $\boldsymbol{\Sigma}_{0,S} = (\sigma_j^2)\mathbf{I}_q, j \in S, q = |S|, \sigma_j^2$ is the variance of the jth wavelet position for the baseline profiles, and S is a set of the selected wavelet coefficients obtained by using data reduction methods, such as VisuShrink, VET, and VertiShrink (Donoho and Johnstone 1994; Chang and Vidakovic, 2002; Jung *et al.* 2006). $\mathbf{d}_{i,S}$ is the vector of wavelet coefficients included in the set S for the ith data curve, and $\boldsymbol{\theta}_{0,S}$ is the vector of the average value of wavelet coefficients in the set S for the baseline profiles. The control limit then becomes

$$UCL = \chi_{\alpha,q}^2,$$

where $\chi_{\alpha,q}^2$ is the upper αth percentile point of the chi-squared distribution with q degrees of freedom.

To evaluate the capability of the proposed methods to detect different types of process changes, two scenarios with different changes are investigated.

Case *I: Only the process variance has changed*

The simulated curve at time i is generated under a random shift $\gamma_{t_j} \sim N(0, \kappa\sigma_\gamma^2)$ in the wavelet domain as follows:

$$y_i(t_j) = \begin{cases} f_0(t_j) + \gamma_{t_j} + \varepsilon(t_j), & t_j \in A \\ f_0(t_j) + \varepsilon(t_j), & \text{elsewhere,} \end{cases}$$

where $\sigma_\gamma^2 = \sum_{k \in D} h_{kj}^2 \tau_k^2$, D is a set of the selected random wavelet coefficients for tonnage signals, A is the shift area [91, 110] units in the time domain, and $\kappa\,(>1)$ is the level of shift for the variance $\kappa\sigma_\gamma^2$. $\kappa = 1$ implies that there is no change in variance. Specifically, based on the proposed WMVT, in this experiment, three wavelet coeffeicnts, $(c_{4,6}, c_{4,7}, c_{4,8})$, are selected as random-effect coefficients. The support areas in the time domain of these coefficients is covered from t_{81} to t_{128}.

Table 1 reports the ARL_1 values for the Hotelling T^2 charts with different thresholding results under different amounts of variance change. The results indicate that all other procedures show the improvement when detecting process changes compared with the T_A^2 chart. Due to the high dimenisonality of the data, the T_A^2 chart does not work well. The proposed W_i chart produces the smallest ARL_1 to detect process changes than the methods extended from ideas given in the literature. For example, compared with the T_{PCA}^2 chart, our proposed W_i chart has improved ARLs on average of 20%. In addition, the W_i chart performs much better than the other charts in detecting smaller levels of change in variance.

Table 1. Comparison of ARLs for Hotelling T^2 charts with different thresholding results.

Level of variance shift (κ)	T_A^2	T_{Js}^2	T_{Vs}^2	T_{Vet}^2	T_{PCA}^2	T_{PJ}^2	W_i	Q_i
1	201.29	199.69	199.64	201.11	201.33	201.56	201.26	199.47
1.5	196.32	171.84	190.51	172.28	180.72	176.84	143.95	155.28
2	161.13	151.79	152.1	124.8	134.06	129.77	106.54	120.53
2.5	117.64	105.63	98.22	88.28	83.10	86.03	69.63	69.63
3	123.46	81.24	75.23	56.48	50.78	53.97	33.33	53.27
3.5	83.14	56.22	60.66	33.26	33.99	33.97	25.54	30.24
4	56.12	46.74	41.75	18.42	20.54	19.82	20.67	21.12
4.5	38.76	25.09	22.97	11.84	13.45	12.99	10.76	10.45
5	34.32	18.94	18.95	9.03	8.75	9.24	7.38	8.12

In the case of a PCA-based approach, we used the first six Principal Components (PCs) since six PCs are enough to explain most of the total variations. However, as shown in Figure 6, the selected PCs could not capture the local variations in the original data, leading to a higher ARL.

In addition, we compare the proposed control chart with the chart designed to detect changes in the variance, such as the |S| chart, which is designed to detect changes in the variance when both the mean and variance are changed (Aparisi *et al.*, 1999). Since our problem is process monitoring for an individual observation (curve), it is not possible to obtain the sample variance-covariance matrix in the |S| chart. Instead, we estimate the variance of each curve based on Donoho and Johnstone's (1994) robust estimate method, such that $\hat{\sigma} = M^{-1} \sum_{i=1}^{M} 0.6745^{-1}\, median(|d_{im}| : N/2 + 1 \le m \le N)$, where the index m indicates wavelet coefficients at the finest level (in this case, $M = 1$). In this experiment, an in-control ARL of 200 is chosen as the critical limit L obtained from 1000 simulation runs.

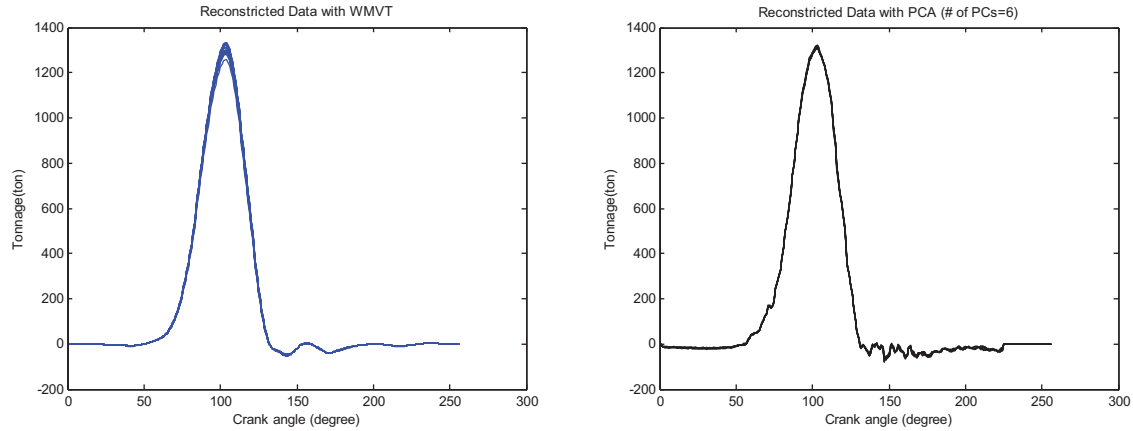
Table 2 compares the ARLs of the proposed W_i chart to that of the |S| chart. In this experiment, we can see that the performance of the W_i chart is much better than that of the |S| chart on all change levels since the |S| chart has a much larger value of critical limit L to meet the given false alarm rate.

Case *II: Both the process mean and variance change*

In this case, both the process mean and variance shifted at different local segments. We compare the proposed method to others in terms of the ARL_1 performance. The monitoring curves

Table 2. Comparison of ARLs for the proposed W_i charts with the |S| chart.

Level of variance shift (κ)	|S| *chart*	W_i *chart*
1	201.23	201.26
1.5	180.38	143.95
2	177.98	106.54
2.5	168.55	69.63
3	158.58	33.33
3.5	152.34	25.54
4	145.32	20.67
4.5	132.45	10.76
5	118.37	7.38



(a) Reconstructed curves using WMVT (b) Reconstructed curves using PCA

Figure 6. Reconstructed curves using (a) WMVT and (b) PCA.

are generated under a random mean shift $\psi_{t_l} \sim N(0, \upsilon\sigma_\psi^2)$ and variance shift $\gamma_{t_j} \sim N(0, \kappa\sigma_\gamma^2)$ as follows:

$$y_i = \begin{cases} f_0(t_j) + \gamma_{t_j} + \varepsilon(t_j), & t_j \in A_j \\ f_0(t_l) + \psi_{t_l} + \varepsilon(t_l), & t_l \in A_l \\ f_0(t_j) + \varepsilon(t_j), & \text{elsewhere,} \end{cases}$$

where $\sigma_\psi^2 = \sum_{k*\in D} h_{k*l}\xi_{k*}$ and $\sigma_\gamma^2 = \sum_{k\in D} h_{kj}^2\tau_k^2$. Note that ξ_{k*} is the change indicator for the mean shift. The $\upsilon(>1)$ and $\kappa(>1)$ are the level of shift for the mean and variance, respectively. In addition, D is a set of selected wavelet coefficients for tonnage signals, A_l and A_j are the shift area [65, 80], [91, 110] units in the time domain. Specifically, based on the proposed WMVT, in this experiment, three random-effect coefficients ($c_{4,6}$, $c_{4,7}$, $c_{4,8}$) and eight fixed-effect coefficients are selected ($c_{4,5}$, $d_{4,5}$, $d_{4,6}$, $d_{4,8}$, $d_{3,11}$, $d_{3,12}$, $d_{3,15}$, $d_{3,16}$).

Table 3 gives the ARL_1 values for the T^2 chart for eight thresholding methods. This table indicates that the Q_i chart shows the best performance to detect the mean and variance changes, as the Q_i chart accounts for the regions for both mean and variance changes with only selected coefficients. In addition, the W_i chart has similar ARL_1 results within the same variance shift, due to the random effect variables not covering the areas of the mean shift in this experiment. In addition, Table 4 compares the proposed control charts (W_i and Q_i charts) with the |S| chart. Table 4 demonstrates that the performance of the Q_i chart is better than

Table 3. Comparison of ARLs for Hotelling T^2 charts with different thresholding results.

Variance level (κ)	*Mean level* (υ)	T_A^2	T_{Js}^2	T_{Vs}^2	T_{Vet}^2	T_{PCA}^2	T_{PJ}^2	W_i	Q_i
1	1	199.19	201.1	200.9	199.5	199.66	199.40	199.23	199.2
	1.5	180.23	155.23	165.23	135.3	157.37	158.81	197.34	145.23
	2	170.45	135.47	150.45	100.9	138.83	139.65	198.54	120.34
	2.5	135.64	120.23	130.87	75.55	113.71	114.03	197.23	69.64
	3	119.23	111.87	112.43	59.89	100.16	103.66	199.35	55.23
1.5	1	190.59	183.46	193.71	178.5	186.17	165.25	145.23	151.11
	1.5	177.58	164.89	159.21	135.4	158.77	136.68	142.49	110.06
	2	142.58	117.38	112.86	80.33	112.06	103.35	149.34	71.65
	2.5	105.75	96.44	84.23	54.43	84.27	84.42	145.45	53.54
	3	77.68	60.32	50.73	40.02	56.48	66.96	147.85	23.5
2	1	160.45	150.4	148.23	130.8	145.81	122.44	105.23	107.93
	1.5	143.3	134.64	132.03	124.4	132.42	107.64	103.29	70.45
	2	103.56	124.69	125.84	90.33	110.01	88.95	107.34	48.12
	2.5	92.65	89.39	82.91	50.98	77.15	66.71	106.66	32.06
	3	89.18	64.73	51.68	37.46	60.19	55.68	101.56	23.52
2.5	1	114.76	105.16	100.8	91.5	101.54	84.02	70.32	72.72
	1.5	101.5	95.55	90.29	89.3	92.65	74.92	66.78	50.94
	2	99.68	86.81	79.64	77.32	85.10	69.86	71.56	45.44
	2.5	80.95	71.31	78.9	50.46	69.27	55.95	68.76	35.32
	3	66.34	54.32	44.22	34.49	49.69	43.90	67.67	23.74
3	1	120.49	89.03	75.12	72.2	89.10	58.55	35.32	37.58
	1.5	105.46	80.44	70.53	70.6	80.07	52.15	34.56	23.35
	2	86.03	66.72	60.07	64.97	67.89	47.50	34.34	22.79
	2.5	80.47	60.89	54	56.05	60.89	43.34	33.57	22.84
	3	71.9	59.94	40.05	31.3	50.54	34.71	36.56	20.43

Table 4. Comparison of ARLs for the proposed charts with the |*S*|-chart

Variance level (κ)	*Mean level* (υ)	*|S| chart*	*W_i chart*	*Q_i chart*
1	1	200.34	199.23	199.2
	1.5	189.34	197.34	145.23
	2	180.23	198.54	120.34
	2.5	155.19	197.23	69.64
	3	139.52	199.35	55.23
1.5	1	190.34	145.23	151.11
	1.5	180.24	142.49	110.06
	2	162.34	149.34	71.65
	2.5	135.25	145.45	53.54
	3	99.23	147.85	23.5
2	1	165.23	105.23	107.93
	1.5	158.23	103.29	70.45
	2	123.78	107.34	48.12
	2.5	89.32	106.66	32.06
	3	79.34	101.56	23.52
2.5	1	155.28	70.32	72.72
	1.5	141.36	66.78	50.94
	2	121.45	71.56	45.44
	2.5	115.65	68.76	35.32
	3	90.25	67.67	23.74
3	1	140.45	35.32	37.58
	1.5	105.34	34.56	23.35
	2	96.34	34.34	22.79
	2.5	78.67	33.57	22.84
	3	65.96	36.56	20.43

that of the W_i and |S| charts when both the mean and variance change. In particular, similar to a previous experiment (Table 2), the |S| chart has a much larger value of the critical limit L to meet the given false alarm rate.

In summary, based on these experiments, our studies indicate that the proposed process monitoring methods offer better performance in shifts in the variance only as well as shifts in both the mean and variance parameters.

7. Case study

A complex, real-world forging process consisting of multiple embedded operation dies is used as an example to show the effectiveness of our proposed methodology. In this forging process, five embedded dies are designed to work together to produce a complete product at the end of each operation stroke. As shown in Figure 2, the final product is produced by passing a raw billet through five embedded dies that perform five different operations in the following sequence: (i) pre-forming; (ii) blocking; (iii) finishing; (iv) piercing; and (v) trimming. The active working range of each die operation has been shown to be only a limited portion of the complete operating cycle (Jin, 2004; Jin and Shi, 2005). Based on the method presented by Jin and Shi (2005), a set of offline station-by-station tests can be conducted to obtain the tonnage signals generated at individual die stations under normal operating conditions, as shown in Figure 7. The active working range for each die is marked as $\mathbf{L}_i (i = 1, \ldots, 5)$, which indicates that the contribution of the individual die force to the monitored total tonnage signals occurs at different local segments of the total tonnage signals. Also, different dies may generate different tonnage variations, due to their different degradation rates and/or different sensitivities to temperature fluctuations, variations in material properties, etc. Therefore, it

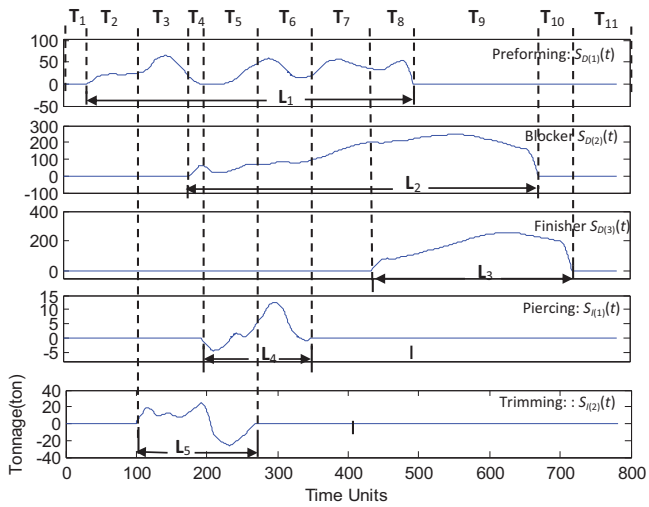


Figure 7. Tonnage signals generated at individual die stations (adopted from Yang and Jin (2012)).

is important to consider the inherent local signal variations in the development of a monitoring chart.

Among these five die operations, blocking (station 2) and finishing (station 3) operations produce a significant shape change on the workpieces and thus generate large tonnage forces at the center peak area around [500–660], as shown in Figure 7. In contrast, the other three operations only make small contributions to this peak tonnage area. Therefore, localized monitoring of the signal profile around this center peak area can expedite the detection and diagnosis of potential faults at stations 2 and 3. Jin and Li (2009) also showed that a few of the wavelet coefficients can be selected to adequately represent the tonnage signals of individual die tonnage forces. Furthermore, most of those selected wavelet coefficients can be mapped to a single or a few operational stations, as the wavelets offer an advantage in efficiently representing signals in both localized time and frequency domains. Therefore, monitoring of those wavelet coefficients that capture local variations, instead of the original signals, can reduce the dimension of the monitoring features. This in turn can enhance the diagnosis of a faulty station after an alarm in the control chart.

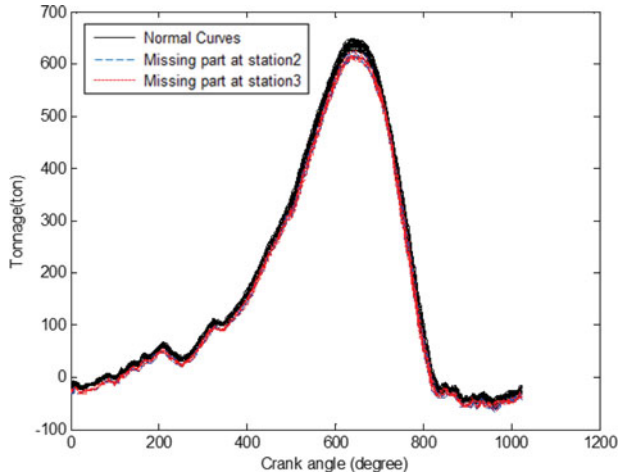


Figure 8. In-control and out-of-control profiles.

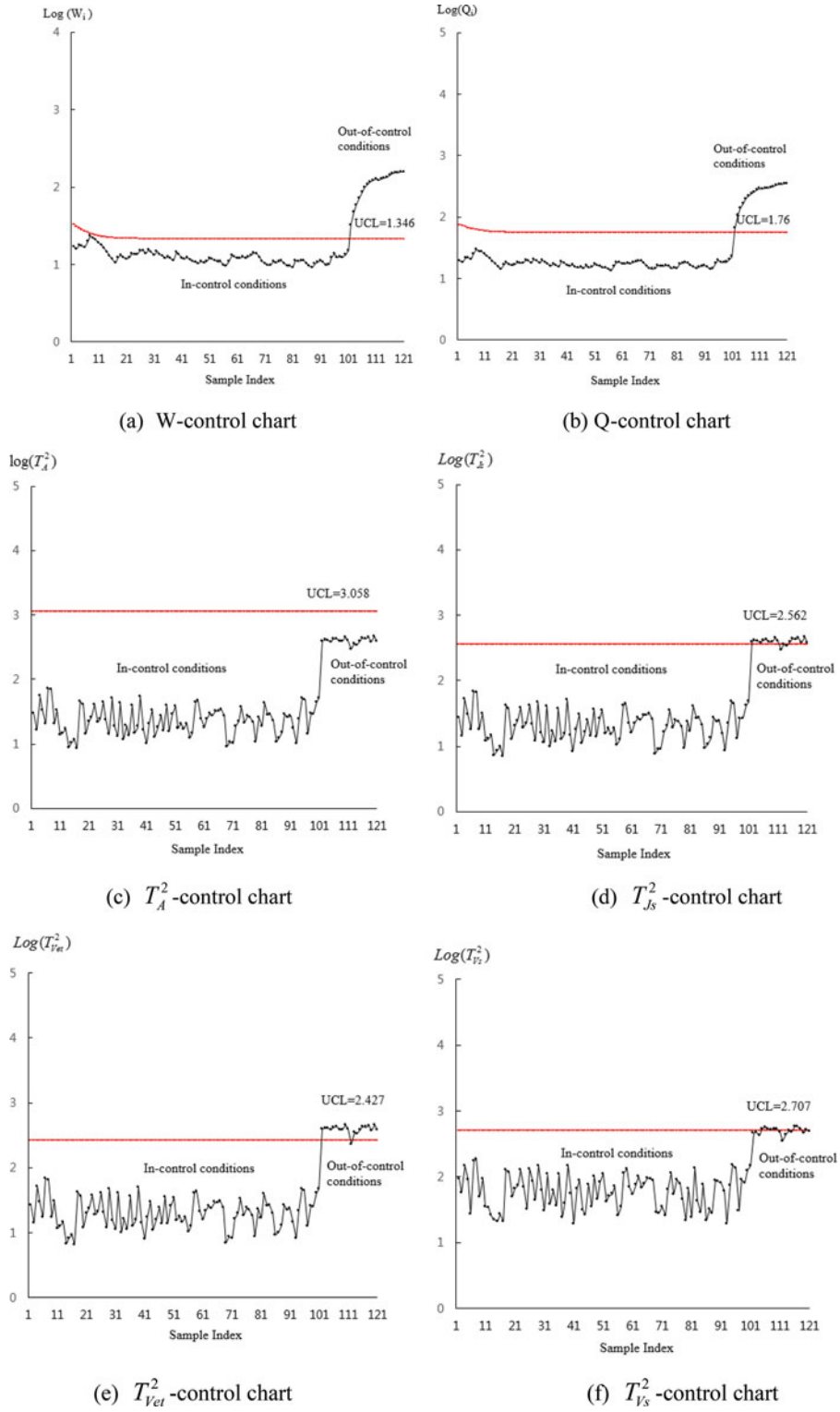


Figure 9. Control charts to detect the out-of-control condition.

To establish process monitoring charts, 100 samples of total tonnage signals were collected under normal operating conditions. Another set of 110 samples was collected, and out of 110 samples, the first 100 samples are in-control profiles (see black profiles in Fig. 8 for in-control profiles) and the last 10 samples were collected under the out-of-control condition corresponding to the missing workpiece at station 2 (blue profiles) and station 3 (red profiles), respectively, in Figure 8. These two faulty conditions result in a significant shift in the mean and variance of the measured tonnage profiles. According to Anderson–Darling's test (Stephens, 1974) and Ljung–Box's test (Ljung and Box, 1978), the normality and independence assumptions defined in Equation (1) are true in real-life studies. Figure 9 shows the plotted values of statistics W_i and Q_i, and all of the first 100 in-control samples are within the control limit. Figure 9 shows similar plots for four existing SPC charts.

As shown in Figures 8 and 9, the proposed methods, W- and Q-control charts, have better detection power than existing ones, especially for the T_A^2-control chart and $T_{V_s}^2$-control chart. For example, the T_A^2-control chart and $T_{V_s}^2$-control chart did not detect many of the out-of-control conditions, and the $T_{J_s}^2$ chart and $T_{V_{et}}^2$ chart missed a few out-of-control samples.

8. Conclusions and future research

This article proposed a wavelet-based local random-effects model to characterize between-curve local variations. The proposed WMVT model with penalized likelihood functions is easy to understand and implement. Closed-form expressions were provided to estimate the mean and variance thresholding parameters. Based on the proposed WMVT model, this article developed a SPC procedure for profile monitoring via wavelet-based local mixed-effect models. Analyses of real-life data and simulations indicate that the proposed SPC procedures based on the WMVT model adequately detect local variations and uses fewer coefficients.

Although the penalized likelihood method limits the number of coefficients in the model, this research did not delve deeper into data reduction, as it was considered beyond its scope. Although procedures such as those used by Jeong *et al.* (2006a) could be used to formulate data reduction metrics, these procedures may not be sufficient in these circumstances, as the penalized likelihood procedure usually uses cross-validations to determine the tuning parameters. Consequently, the problem is more complicated than the single-curve studies of data reduction contained in Jeong *et al.* (2006a), and further work is needed in this direction. Also, the proposed procedure needs to be extended to multiple curves with local variations when the wavelet coefficients are not independent. Furthermore, the extension of the proposed procedure to root cause diagnosis would be an interesting topic for future research.

Funding

Part of this work is supported by the National Science Foundation under awards #0644830, #0853894, and #1233800 and NRF-2015R1C1A1A01051487 from the National Research Foundation of Korea.

Notes on contributors

Young-Seon Jeong is an assistant professor in the Department of Industrial Engineering at Chonnam National University, Gwangju, South Korea. He received his Ph.D. degree from Rutgers University, New Jersey, in 2011. His recent research focuses on statistical decision-making models for monitoring of process control systems in semiconductor manufacturing systems, development of a statistical data mining methodology with diverse applications such as manufacturing systems, and intelligent transportation systems.

Myong K. Jeong is a professor in the Department of Industrial and Systems Engineering at Rutgers University. His research interests are focused on developing data-mining techniques, process monitoring and control procedures, and optimization techniques for machine learning. His research has been supported by the National Science Foundation, National Transportation Research Center, United States Department of Agriculture, Qatar National Research Fund, Electronics and Telecommunications Research Institute, and various industries. He has been a consultant for Samsung Electronics, ETRI, KISTI, and other companies.

Jye-Chyi Lu received a Ph.D. degree in statistics from the University of Wisconsin, Madison, in 1998. He is currently a professor in the School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta. His current research interests include decision and data analytics, supply chain management, data mining, industrial statistics, and reliability.

Ming Yuan is currently a professor in the Department of Statistics, Columbia University.

Jionghua (Judy) Jin is a professor in the Department of Industrial and Operations Engineering at the University of Michigan. She received her Ph.D. degree from the University of Michigan in 1999. Her recent research focuses on data fusion and analytics for system monitoring, diagnosis, quality control, and decision making. Her research emphasizes a multidisciplinary approach by integrating applied statistics, machine learning, signal processing, reliability engineering, system control, and decision-making theory. She has received a number of awards, including NSF CAREER and PECASE Awards, and 12 Best Paper Awards since 2000. She is an elected Fellow of ASME and IIE; a senior member of ASQ and ISI; and a member of IEEE, INFORMS, and SME.

References

Abramovich, F., Sapatinas, T. and Silverman, B. (1998) Wavelet thresholding via a Bayesian approach. *Journal of Royal Statistical Society: Series B*, **60**(4), 725–749.

Aparisi, F., Jabaloyes, J. and Carrion, A. (1999) Statistical properties of the |S| multivariate control chart. *Communications in Statistics—Theory and Methods*, **28**(11), 2671–2686.

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

Bruce, A.G. and Gao, H.-Y. (1996) Understanding wave shrink: Variance and bias estimation. *Biometrika*, **83**(4), 727–745.

Castro, B.F., Guillas, S. and Manteiga, W.G. (2005) Functional samples and bootstrap for predicting sulfur dioxide levels. *Technometrics*, **47**(2), 212–222.

Chang, W. and Vidakovic, B. (2002) Wavelet estimation of a baseline signal from repeated noisy measurements by vertical block shrinkage. *Computational Statistics and Data Analysis*, **40**(2), 317–328.

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

Chipman, H., Kolaczyk, E. and McCulloch, R. (1997) Adaptive Bayesian wavelet shrinkage. *Journal of the American Statistical Association*, **92**(440), 1413–1421.

Crouse, M., Nowak, R. and Baraniuk, R. (1998) Wavelet-based statistical signal processing using hidden Markov models. *IEEE Transactions on Signal Processing*, **46**(4), 886–902.

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

Ganesan, R., Das, T.K., Sikder, A.K. and Kumar, A. (2003) Wavelet based identification of delamination emission signal. *IEEE Transactions on Semiconductor Manufacturing*, **16**(4), 677–685.

Guo, H., Paynabar, K. and Jin, J. (2012) Multiscale monitoring of autocorrelated processes using wavelets analysis. *IIE Transactions*, **44**(4), 312–326.

Guo, W. (2002) Functional mixed effects models. *Biometrics*, **58**(1), 121–128.

Huwang, L., Yeh, A.B. and Wu, C.W. (2007) Monitoring multivariate process variability for individual observations. *Journal of Quality Technology*, **39**(3), 258–278.

Jeong, M.K., Lu, J.C., Huo, X., Vidakovic, B. and Chen, D. (2006a) Wavelet-based data reduction techniques for fault detection. *Technometrics*, **48**(1), 26–40.

Jeong, M.K., Lu, J.C. and Wang, N. (2006b) Wavelet-based SPC procedure for complicated functional data. *International Journal of Production Research*, **44**(4), 1–16.

Jin, J. (2004) Individual station monitoring using press tonnage sensors for multiple operation stamping processes. *ASME Transactions, Journal of Manufacturing Science and Engineering*, **126**(1), 83–90.
Jin, J. and Li, J. (2009) Multiscale wavelet analysis for mapping aggregated signal features to embedded operations. *IIE Transactions on Quality and Reliability Engineering*, **41**(7), 615–625.
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**(3), 257–268.
Jin, J. and Shi, J. (2003) Press tonnage signal decomposition and validation analysis for transfer or progressive die processes. *Journal of Manufacturing Science and Engineering*, **127**(1), 231–235.
Jin, J. and Shi, J. (2005) Press tonnage signal decomposition and validation for transfer or progressive die processes. *ASME Transactions, Journal of Manufacturing Science and Engineering*, **127**(1), 231–235.
Jung, U., Jeong, M.K. and Lu, J.C. (2006) A vertical energy thresholding procedure for data reduction with multiple complex curves. *IEEE Transactions on Systems, Man, Cybernetics, Part B*, **36**(5), 1128–1138.
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.
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.
Ljung, G.M. and Box, G.E.P. (1978) On a measure of lack of fit in time series models. *Biometrika*, **65**(2), 297–303.
Macgregor, J.F. and Harris, T.J. (1993) The exponentially weighted moving variance. *Journal of Quality Technology*, **25**(2), 106–118.
Mallat, S.G. (1998) *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, CA.
Morris, J.S., Arroyo, C., Coull, B.A., Ryan. L.M., Herrick, R. and Gortmaker, S.L. (2006) Using wavelet-based functional mixed models to characterize population heterogeneity in accelerometer profiles. *Journal of the American Statistical Association*, **101**(476), 1352–1364.
Morris, J.S. and Carroll, R.J. (2006) Wavelet-based functional mixed models. *Journal of the Royal Statistical Society: Series B*, **68**(2), 179–199.
Morris, J.S., Vannucci, M., Brown, P.J. and Carroll, R.J. (2003) Wavelet-based nonparametric modeling of hierarchical functions in colon carcinogenesis. *Journal of the American Statistical Association*, **98**(463), 573–597.
Omitaomu, F., Jeong, M.K., Badiru, A. and Hines, J.W. (2006) On-line prediction of motor shaft misalignment using Fourier-transformed acceleration data and support vector regression. *ASME Transactions, Journal of Manufacturing Science and Engineering*, **128**(4), 1019–1024.
Paynabar, K. and Jin, J. (2011) Characterization of non-linear profiles variations using mixed-effect models and wavelets. *IIE Transactions*, **43**(4), 275–290.
Reynolds, M.R. and Cho, G.-Y. (2006) Multivariate control charts for monitoring the mean vector and covariance matrix. *Journal of Quality Technology*, **38**(3), 230–253.
Soleymani, M., Hossein-Zadeh, G.A. and Soltanian-Zadeh, H. (2009) Fixed and random effect analysis of multi-subject fMRI data using wavelet transform. *Journal of Neuroscience Methods*, **176**(2), 237–245.
Stephens, M.A. (1974) EDF statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association*, **69**(347), 730–737.
Stoer, J. and Bulirsch, R. (2002) *Introduction to Numerical Analysis*, Springer-Verlag, New York.
Vidakovic, B. (1998) Nonlinear wavelet shrinkage with Bayes rules and Bayes factors. *Journal of the American Statistical Association*, **93**(441), 173–179.
Yang, Q. and Jin, J. (2012) Separation of individual operation signals from mixed sensor measurements. *IIE Transactions on Quality and Reliability Engineering*, **44**(9), 780–792.
Yuan, M. and Wahba, G. (2004) Doubly penalized likelihood estimator in heteroscedastic regression. *Statistics and Probability Letters*, **69**(1), 11–20.
Zhou, S. and Jin, J. (2005) An unsupervised clustering method for cycle-based waveform signals in manufacturing processes. *IIE Transactions*, **37**(6), 569–584.
Zhou, S., Sun, B. and Shi, J. (2006) An SPC monitoring systems for cycle-based waveform signals using Haar transform. *IEEE Transactions on Automation Science and Engineering*, **3**(1), 60–72.
Zou, C., Tsung, F. and Wang, Z. (2007) Monitoring general linear profiles using multivariate EWMA schemes. *Technometrics*, **49**(4), 395–408.

Appendix

1. Mapping property of the wavelet transforms

To better understand the relationship between the original profiles in the time domain and its random coefficients in the wavelet domain in the case where a local segment can be used to identify the locations of systematic variations in the time domain, the mapping property of the wavelet transforms is very important to search for process faults and their causes.

Assume that there is a set (denoted by D) of random coefficients that are independent in the wavelet domain; i.e., $\tau_j^2 \neq 0$ for $j \in D$ and zero elsewhere. Then, replicated curves from the wavelet-based random-effect model in the time domain will have the following *systematic* variations over the region A, where A is the support area covered by the wavelet coefficients in the set D:

$$y_i(t_j) \sim \begin{cases} N(f_{.j}, \sum_{k\in D} h_{kj}^2 \tau_k^2 + \sigma^2), & t_j \in A \\ N(f_{.j}, \sigma^2), & \text{elsewhere,} \end{cases}$$

where $y_i(t_j)$ is the original time domain data for the ith curve at time point t_j and $f_{.j}$ is the mean curve f evaluated at t_j.

Justification: The replicated curves can be reconstructed from the inverse DWT; that is,

$$\begin{aligned} y_i(t_j) &= \sum_{k=1}^{N} h_{kj} d_{i,k}, \quad (j = 1, 2, \ldots, N) \\ &= \sum_{k\in D} h_{kj} d_{ik} + \sum_{l\in S/D} h_{lj} d_{il}, \end{aligned}$$

where S is the set of all wavelet positions. Then, the variability across the curves is given by

$$\begin{aligned} Var(y_i(t_j)) &= \sum_{k\in D} h_{kj}^2 Var(d_{ik}) + \sum_{l\in S/D} h_{lj}^2 Var(d_{il}), \\ &= \sum_{k\in D} h_{kj}^2 \tau_k^2 + \sigma^2. \end{aligned}$$

2. Derivation of parameter estimates for $\theta_{.j}$, τ_j^2, and σ^2

The penalized log-likelihood function is given by

$$\begin{aligned} h(\theta_{.j}, \tau_j^2, \sigma^2) = M \sum_{j=1}^{N} \ln(\sigma^2 + \tau_j^2) + \frac{\sum_{i=1}^{M} \sum_{j=1}^{N} (d_{ij} - \theta_{.j})^2}{\sigma^2 + \tau_j^2} \\ + \lambda_1 \sum_{j=1}^{N} |\theta_{.j}| + \lambda_2 \sum_{j=1}^{N} \tau_j^2. \end{aligned}$$

Taking the partial derivative of h with respect to $\theta_{.j}$, we obtain

$$\frac{\partial h}{\partial \theta_{.j}} = -2\sum_{i=1}^{M}\frac{d_{ij}-\theta_{.j}}{\sigma^2+\tau_j^2} + \lambda_1 \operatorname{sign}(\theta_{.j}) = \frac{-2M}{\sigma^2+\tau_j^2}(\bar{d}_{.j}-\theta_{.j}) + \lambda_1 \operatorname{sign}(\theta_{.j}) = 0,$$

where

$$\bar{d}_{.j} = \frac{1}{M}\sum_{i=1}^{M} d_{ij}.$$

Therefore,

$$\hat{\theta}_{.j} = \left(|\bar{d}_{.j}| - \lambda_1(\hat{\sigma}^2+\hat{\tau}_j^2)/2M\right)_+ sign(\bar{d}_{.j}), \text{ where } (y)_+ = \max(y,0).$$

In a similar way, by taking the partial derivative of h with respect to τ_j^2, we obtain

$$\frac{\partial h}{\partial \tau_j^2} = \frac{M}{\sigma^2+\tau_j^2} - \sum_{i=1}^{M}\frac{(d_{ij}-\theta_{.j})^2}{(\sigma^2+\tau_j^2)^2} + \lambda_2 = \frac{M}{(\sigma^2+\tau_j^2)^2}(\sigma^2+\tau_j^2 - s_j^2 + \lambda_2(\sigma^2+\tau_j^2)^2/M) = 0,$$

where $s_j^2 = \sum_{i=1}^{M}(d_{ij}-\theta_{.j})^2/M$. Letting $a = \sigma^2+\tau_j^2$, we obtain $\lambda_2 a^2/M + a - s_j^2 = 0$ and

$$a = \sigma^2+\tau_j^2 = -1 + \sqrt{1+4s_j^2\lambda_2/M}2\lambda_2/M$$

because $a > 0$. Therefore,

$$\hat{\tau}_j^2 = \left(\frac{-1+\sqrt{1+4s_j^2\lambda_2/M}}{2\lambda_2/M} - \hat{\sigma}^2\right)_+.$$

Finally, we can obtain the estimate of σ^2 by solving the following equation:

$$\frac{\partial h}{\partial \sigma^2} = M\sum_{j=1}^{N}\frac{1}{\sigma^2+\tau_j^2} - \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}(d_{ij}-\theta_{.j})^2}{(\sigma^2+\tau_j^2)^2} = M\sum_{j=1}^{N}\frac{\sigma^2+\tau_j^2-s_j^2}{(\sigma^2+\tau_j^2)^2} = 0,$$

which is equivalent to Equation (7).

3. Derivation of the control limits of W_i

The monitoring statistic

$$W_i = tr(\mathbf{Z}_i) = \sum_{k=1}^{i}\delta(1-\delta)^{i-k}\left(\sum_{j=1}^{p_1}u_{k,j}^2\right)$$

is a modified version of Huwang's process variability monitoring statistic for an individual observation (Huwang *et al.*, 2007). A larger W_i value indicates that the variance of a process has increased. When the process is in control, $\sum_{j=1}^{p_1}u_{k,j}^2$ follows a χ^2-distribution with degrees of freedom p_1. Then, the mean and variance of W_i are given as follows:

$$E(W_i) = \sum_{k=1}^{i}\delta(1-\delta)^{i-k}E\left(\sum_{j=1_1}^{p_1}u_{k,j}^2\right) = p_1,$$

$$Var(W_i) = \sum_{k=1}^{i}\left(\delta(1-\delta)^{i-k}\right)^2 Var\left(\sum_{j=1}^{p_1}u_{k,j}^2\right) = \sum_{k=1}^{i}\left(\delta(1-\delta)^{i-k}\right)^2 2p_1.$$

Thus, by using the large-sample normal approximation theory, control limits of W_i are given by

$$CL_{W_i} = p_1 \pm \Phi^{-1}(1-\alpha)\sqrt{\sum_{k=1}^{i}\left(\delta(1-\delta)^{i-k}\right)^2 2p_1}.$$

4. Derivation of the estimator Σ_i^v

By definition of EWMA($\tilde{\mathbf{v}}_i$) with smoothing weight $0<\alpha<1$, $\mathbf{r}_i = \sum_{k=1}^{i}\alpha(1-\alpha)\tilde{\mathbf{v}}_k$.

Thus, $\tilde{\mathbf{v}}_i - \boldsymbol{\gamma}_i = \tilde{\mathbf{v}}_i - \sum_{k=1}^{i}\alpha(1-\alpha)\tilde{\mathbf{v}}_k$ and

$$E\left(\tilde{\mathbf{v}}_i - \boldsymbol{\gamma}_i\right) = E\left(\tilde{\mathbf{v}}_i - \sum_{k=1}^{i}\alpha(1-\alpha)\tilde{\mathbf{v}}_k\right) = E(\tilde{\mathbf{v}}_i) - \sum_{k=1}^{i}\alpha(1-\alpha)E(\tilde{\mathbf{v}}_k) = \boldsymbol{\mu}_i - [1-(1-\alpha)^i]\boldsymbol{\mu}_i = (1-\alpha)^i\boldsymbol{\mu}_i.$$

In addition,

$$\begin{aligned}
&E[(\tilde{\mathbf{v}}_i-\boldsymbol{\gamma}_i)(\tilde{\mathbf{v}}_i-\boldsymbol{\gamma}_i)'] \\
&= Cov(\tilde{\mathbf{v}}_i-\boldsymbol{\gamma}_i) + E(\tilde{\mathbf{v}}_i-\boldsymbol{\gamma}_i)E(\tilde{\mathbf{v}}_i-\boldsymbol{\gamma}_i)' \\
&= Cov\left[\tilde{\mathbf{v}}_i - \sum_{k=1}^{i}\alpha(1-\alpha)^{i-k}\tilde{\mathbf{v}}_k\right] + (1-\alpha)^{2i}\boldsymbol{\mu}_i^v \\
&= Cov[\tilde{\mathbf{v}}_i] - Cov\left[\sum_{t=1}^{n}\alpha(1-\alpha)^{i-k}\tilde{\mathbf{v}}_k\right] + (1-\alpha)^{2i}\boldsymbol{\mu}_i^v \\
&= \left[1+\frac{\alpha}{2-\alpha}\left((1-\alpha)^{2i}-1\right)\right]\boldsymbol{\Sigma}_0^{rf} + (1-\alpha)^{2i}\boldsymbol{\mu}_i^v \\
&= \frac{1}{2-\alpha}\left[2(1-\alpha)+\alpha(1-\alpha)^{2i}\right]\boldsymbol{\Sigma}_0^{rf} + (1-\alpha)^{2i}\boldsymbol{\mu}_i^v.
\end{aligned}$$

Based on these results, we can obtain

$$\begin{aligned}
E(\mathbf{C}_i) &= \sum_{k=1}^{i}\delta(1-\delta)^{i-k}E((\tilde{\mathbf{v}}_i-\boldsymbol{\gamma}_k)(\tilde{\mathbf{v}}_i-\boldsymbol{\gamma}_k)') \\
&= \sum_{k=1}^{i}\delta(1-\delta)^{i-k}\left[\frac{1}{2-\alpha}\left(2(1-\alpha)+\alpha(1-\alpha)^{2i}\right)\right]\boldsymbol{\Sigma}_i^v \\
&\quad + (1-\alpha)^{2i}\boldsymbol{\mu}_i^v \\
&= \frac{2(1-\alpha)}{2-\alpha}\boldsymbol{\Sigma}_i^v + \sum_{k=1}^{i}\delta(1-\delta)^{i-k}\frac{\alpha}{2-\alpha}(1-\alpha)^{2k}\boldsymbol{\Sigma}_i^v
\end{aligned}$$

$$+(1-\alpha)^{2i}\boldsymbol{\mu}_i^v$$
$$=\frac{2(1-\alpha)}{2-\alpha}\boldsymbol{\Sigma}_i^v \quad \text{as } i\to\infty.$$

5. Matrix form of Q_i

By using matrix forms, the monitoring statistic Q_i, where

$$Q_i=\sum_{k=1}^{i}\delta(1-\delta)^{k-i}tr((\tilde{\mathbf{v}}_k-\boldsymbol{\gamma}_k)(\tilde{\mathbf{v}}_k-\boldsymbol{\gamma}_k)'),$$

can be simplified as follows: $\mathbf{Q}_i=tr((\tilde{\mathbf{V}}_i-\mathbf{H}_i)'\boldsymbol{\Delta}(\tilde{\mathbf{V}}_i-\mathbf{H}_i)$, where

$$\tilde{\mathbf{V}}_i=[\tilde{\mathbf{v}}_1,\tilde{\mathbf{v}}_2,\ldots,\tilde{\mathbf{v}}_i]'$$
$$\mathbf{H}_i=[\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2,\ldots,\boldsymbol{\gamma}_i]'.$$
$$\boldsymbol{\Delta}=diag((1-\delta)^{i-1},\delta(1-\delta)^{i-2},\ldots,\delta(1-\delta),\delta).$$

In addition,

$$(\tilde{\mathbf{V}}_i-\tilde{\mathbf{H}}_i)=\begin{pmatrix}(\tilde{\mathbf{v}}_1-\boldsymbol{\gamma}_1)'\\(\tilde{\mathbf{v}}_2-\boldsymbol{\gamma}_2)'\\\vdots\\(\tilde{\mathbf{v}}_i-\boldsymbol{\gamma}_i)'\end{pmatrix}$$
$$=\begin{pmatrix}(1-\varphi)\tilde{\mathbf{v}}_1\\(1-\varphi)\tilde{\mathbf{v}}_2-\varphi(1-\varphi)\tilde{\mathbf{v}}_1\\\vdots\\(1-\varphi)\tilde{\mathbf{v}}_i-\varphi(1-\varphi)\tilde{\mathbf{v}}_{i-1}-\cdots-\varphi(1-\varphi)^{i-1}\tilde{\mathbf{v}}_1\end{pmatrix}$$
$$=\mathbf{B}\tilde{\mathbf{V}}_i,$$

where

$$\mathbf{B}=\begin{pmatrix}1-\varphi & 0 & \cdots & 0\\-\varphi(1-\varphi) & 1-\varphi & \cdots & 0\\\vdots & \vdots & \ddots & \vdots\\-\varphi(1-\varphi)^{i-1} & \cdots & -\varphi(1-\varphi) & (1-\varphi)\end{pmatrix}.$$

Thus,

$$\mathbf{Q}_i=tr((\tilde{\mathbf{V}}_i-\mathbf{H}_i)'\boldsymbol{\Delta}(\tilde{\mathbf{V}}_i-\mathbf{H}_i))=tr(\tilde{\mathbf{V}}_i'\mathbf{B}'\boldsymbol{\Delta}\mathbf{B}\tilde{\mathbf{V}}_i)$$
$$=\sum_{k=1}^{i}\sum_{j=1}^{i}\pi_{kj}\left(\sum_{l=1}^{n_1}\tilde{v}_{k,l}\tilde{v}_{j,l}\right).$$

The matrix $\mathbf{B}'\boldsymbol{\Delta}\mathbf{B}$ indicates that recent profiles are heavily weighted such as $\pi_{ii}=\delta(1-\varphi)^2$, $\pi_{(i-1)(i-1)}=\delta(1-\delta)(1-\varphi)^2+\delta[-\varphi(1-\varphi)]^2$, and $\pi_{(i-2)(i-2)}=\delta(1-\delta)^2(1-\varphi)^2+\delta[-\varphi(1-\varphi)^2]^2+\delta(1-\delta)[-\varphi(1-\varphi)]^2$, and so on.

6. Derivation of mean, variance, and control limit of Q_i

$$E(Q_i)=E\left(\sum_{k=1}^{i}\sum_{j=1}^{i}\pi_{kj}\left(\sum_{l=1}^{n_1}\tilde{v}_{kl}\tilde{v}_{jl}\right)\right)$$
$$=\sum_{k=1}^{i}\pi_{kk}E\left(\sum_{l=1}^{n_1}\tilde{v}_{kl}^2\right)+\sum_{k=1}^{i}\sum_{k\neq j}^{i}\pi_{kj}E\left(\sum_{l=1}^{n_1}\tilde{v}_{kl}\tilde{v}_{jl}\right)$$
$$=\sum_{k=1}^{i}\pi_{kk}E\left(\sum_{l=1}^{n_1}\tilde{v}_{kl}^2\right)$$
$$+\sum_{k=1}^{i}\sum_{k\neq j}^{i}\pi_{kj}E(\tilde{v}_{11}\tilde{v}_{21}+\tilde{v}_{11}\tilde{v}_{22}+\cdots+\tilde{v}_{in_1}\tilde{v}_{i-1n_1})$$
$$=n_1\sum_{k=1}^{i}\pi_{kk}.$$

$$Var(Q_i)=Var\left(\sum_{k=1}^{i}\sum_{j=1}^{i}\pi_{kj}\left(\sum_{l=1}^{n_1}\tilde{v}_{kl}\tilde{v}_{jl}\right)\right)$$
$$=Var\left[\sum_{k=1}^{i}\pi_{kk}\sum_{l=1}^{n_1}\tilde{v}_{kl}^2+2\sum_{k=1}^{i}\sum_{j<k}^{i}\pi_{kj}\sum_{l=1}^{n_1}\tilde{v}_{kl}\tilde{v}_{jl}\right]$$
$$=\sum_{k=1}^{i}\pi_{kk}^2Var\left(\sum_{l=1}^{n_1}\tilde{v}_{kl}^2\right)+4\sum_{k=1}^{i}\sum_{j<k}^{i}\pi_{kj}^2Var\left(\sum_{l=1}^{n_1}\tilde{v}_{kl}\tilde{v}_{jl}\right)$$
$$=\sum_{k=1}^{i}\pi_{kk}^2Var\left(\sum_{l=1}^{n_1}\tilde{v}_{kl}^2\right)+4\sum_{k=1}^{i}\sum_{j<k}^{i}\pi_{kj}^2$$
$$\times\left[Var(\tilde{v}_{11})+Var(\tilde{v}_{12})+\cdots+Var((\tilde{v}_{in_1})Var(\tilde{v}_{i-1n_1})\right]$$
$$=2n_1\sum_{k=1}^{i}\pi_{kk}^2+4n_1\sum_{k=1}^{i}\sum_{j<k}^{i}\pi_{kj}^2$$
$$=2n_1\left(\sum_{k=1}^{i}\pi_{kk}^2+2\sum_{k=1}^{i}\sum_{j<k}^{i}\pi_{kj}^2\right)$$
$$=2n_1\sum_{k=1}^{i}\sum_{j=1}^{i}\pi_{kj}^2.$$

By using large-sample normal approximation theory, the control limit of Q_i is given by

$$CL_{Q_i}=n_1\sum_{k=1}^{i}\pi_{kk}\pm\Phi^{-1}(1-\alpha)\sqrt{2n_1\sum_{k=1}^{i}\sum_{j=1}^{i}\pi_{kj}^2}.$$