

Nonparametric Profile Monitoring by Mixed Effects Modeling

Peihua QIU

School of Statistics
University of Minnesota
Minneapolis, MN 55455
(qiu@stat.umn.edu)

Changliang Zou and Zhaojun WANG

LPMC
and
Department of Statistics
Nankai University
Tianjing 300071
China

In some applications, the quality of a process is characterized by the functional relationship between a response variable and one or more explanatory variables. Profile monitoring is for checking the stability of this relationship over time. Control charts for monitoring nonparametric profiles are useful when the relationship is too complicated to be described parametrically. Most existing control charts in the literature are for monitoring parametric profiles. They require the assumption that within-profile measurements are independent of each other, which is often invalid in practice. This article focuses on nonparametric profile monitoring when within-profile data are correlated. A novel control chart is suggested, which incorporates local linear kernel smoothing into the exponentially weighted moving average (EWMA) control scheme. In this method, within-profile correlation is described by a nonparametric mixed-effects model. Our proposed control chart is fast to compute and convenient to use. Numerical examples show that it works well in various cases. Some technical details are provided in an Appendix available online as supplemental materials.

KEY WORDS: EWMA; Local linear kernel smoothing; Phase II; Statistical process control.

1. INTRODUCTION

In certain applications, the quality of a process is characterized by the functional relationship between a response variable and one or more explanatory variables. At each sampling stage one observes a set of data points of these variables that can be represented by a curve (or profile). Profile monitoring is mainly for checking the stability of this relationship over time based on observed profile data. In some applications (e.g., certain calibration applications), profiles can be described reasonably well by a linear regression model. But in some others, more flexible models are necessary for properly describing profiles. This article focuses on nonparametric profile monitoring when within-profile data are correlated.

As described by Woodall (2000), statistical process control (SPC) can generally be divided into two phases. In Phase I a set of process data are gathered and analyzed. Any unusual “patterns” in the data lead to adjustments and fine tuning of the process. Once all such assignable causes are accounted for, we are left with a clean set of data, gathered under stable operating conditions and illustrative of the actual process performance. This dataset, which is referred to as the in-control (IC) dataset hereafter, is then used for estimating certain IC parameters of the process. In Phase II SPC, the estimated IC process parameters are used, and the major goal of this phase is to detect any change in the profiles. Performance of a Phase II SPC procedure is often measured by the average run length (ARL), which is the average number of time points needed for the procedure to signal a change in profiles. The IC ARL value of the procedure is often controlled at a given level. Then the procedure performs better if its out-of-control (OC) ARL is shorter when detecting a specific profile change. In the literature, most SPC

control charts are for Phase II process monitoring, which is also the focus of the current article.

In recent years, Phase II profile monitoring has drawn much attention from statisticians. Early research on this topic focused on linear profile monitoring (see, e.g., Kang and Albin 2000; Kim, Mahmoud, and Woodall 2003; Mahmoud and Woodall 2004; Zou, Zhang, and Wang 2006; Mahmoud et al. 2007; among several others). Zou, Tsung, and Wang (2007) and Kazemzadeh, Noorossana, and Amiri (2008) considered cases when profiles can be described well by multiple and/or polynomial regression models. Some recent research concerns nonlinear profile monitoring. For instance, Williams et al. (2007) and Williams, Woodall, and Birch (2007) suggested three general approaches to nonlinear profile monitoring and used these approaches for monitoring nonlinear dose-response profiles. Colosimo and Pacella (2007) proposed methods for monitoring roundness profiles of some manufactured items. Lada, Lu, and Wilson (2002) and Ding, Zeng, and Zhou (2006) investigated a general class of nonlinear profiles, using techniques such as dimension-reduction, wavelet transformations, and independent component analysis. Zou, Tsung, and Wang (2008) discussed profile monitoring using nonparametric regression methods. A nice overview on profile monitoring can be found in Woodall et al. (2004).

In the literature, most existing profile monitoring control charts (e.g., Zou, Tsung, and Wang 2008) require a fundamental assumption that observations within a profile are independent

of each other, which is often invalid in applications. In practice, within-profile data are usually spatially or serially correlated. For instance, within-profile data of the vertical-density profiles (VDP's) considered by Walker and Wright (2002) and Williams, Woodall, and Birch (2007) are spatially correlated since the density measurements are taken in intervals that are close to each other along the vertical depth of a particle board.

As another example, within-profile data in the deep reactive ion etching (DRIE) example considered by Zou, Tsung, and Wang (2007) exhibit obvious serial correlation over time. As demonstrated in the following sections, when within-profile correlation is present, proper setup of the profile model becomes challenging, and estimation of certain IC process parameters becomes difficult as well. If it is ignored in profile monitoring, then the IC and OC properties of the related control charts will be adversely affected.

There is no existing research on Phase II nonparametric profile monitoring in cases when within-profile data are correlated. Recent articles by Jensen, Birch, and Woodall (2008) and Jensen and Birch (2009) discussed linear and nonlinear profile monitoring in Phase I analysis, using linear and nonlinear mixed-effects modeling (cf., e.g., Laird and Ware 1982). Their methods can accommodate certain within-profile correlation. However, besides the fundamental difference between the Phase I linear/nonlinear profile monitoring considered in their articles and the Phase II nonparametric profile monitoring considered here, their approaches assume that both the fixed and random effects terms in their models follow certain parametric models, and that the covariance matrix of the random errors also follows a parametric form, such as the autoregressive or compound symmetry form. While parametric methods are useful in certain applications, questions will always arise about the adequacy of these parametric model assumptions and about the potential impact of model misspecifications on profile monitoring performance (see Hart 1997 for related discussion). In addition, design points are assumed to be deterministic (i.e., non-random) in Jensen, Birch, and Woodall (2008) and Jensen and Birch (2009), and unchanged from profile to profile in Jensen, Birch, and Woodall (2008). In practice, however, different profiles often have different design points (i.e., the so-called unbalanced design cases). In some cases they might even be random (i.e., the random design cases). Phase II profile monitoring in such cases is particularly challenging, which is also discussed in this article.

To properly describe within-profile correlation, we propose to use a nonparametric mixed-effects model (cf., e.g., Shi, Weiss, and Taylor 1996; Rice and Wu 2001; Wu and Zhang 2002), which allows a flexible variance-covariance structure. Based on estimated variance structure from an IC dataset, we propose a novel Phase II control chart for monitoring nonparametric profiles, which can accommodate within-profile correlation and arbitrary design. Our proposed control chart is based on local linear kernel smoothing of profile data and on the EWMA weighting scheme as well. It incorporates properly both the exponential weights used in the EWMA scheme at different time points and the heteroscedasticity of observations within each profile into the local linear kernel smoother. Numerical results show that this approach performs well in various cases.

Our proposed control chart is described in detail in Section 2. Its numerical performance is investigated in Section 3. In Section 4, we apply this method to a dataset from a manufacturing process of aluminium electrolytic capacitors. Several remarks conclude the article in Section 5. Some technical details are provided in an Appendix, which is available online as supplementary materials.

2. METHODOLOGY

This section is organized in five parts. In Section 2.1, nonparametric mixed-effects modeling of an IC dataset is introduced. Its model estimation is discussed in Section 2.2. In Section 2.3, a new Phase II nonparametric profile control chart is proposed, which can accommodate within-profile correlation and arbitrary design. Certain computational issues are addressed in Section 2.4. Some practical guidelines regarding the design and implementation of the proposed control chart are given in Section 2.5.

2.1 Nonparametric Mixed-Effects Modeling

The Phase II nonparametric profile control chart proposed in this article does not require IC process parameters to be known. Instead, we estimate the related IC process parameters from an IC dataset using nonparametric mixed-effects modeling. In the literature, mixed-effects modeling is often used in longitudinal data analysis (cf., e.g., Laird and Ware 1982; Diggle, Liang, and Zeger 1994). It has become a major tool for accommodating the possible correlation among observed data. Nonparametric mixed-effects (NME) modeling for analyzing longitudinal data was discussed by several authors, including Shi, Weiss, and Taylor (1996) and Rice and Wu (2001). Here we follow this framework for modeling the within-profile correlation of an IC dataset. In what follows, we use the term "profile" throughout; but it should be noted that, in the literature on mixed-effects modeling, it is often referred to as a "cluster" or "subject."

To simplify the presentation, we choose to discuss cases with a single covariate here; this discussion can be easily generalized to cases with multiple covariates. In the IC dataset, assume that there are m profiles and the i th profile has n_i observations, for $i = 1, 2, \dots, m$. Then, the NME model can be written as

$$y_{ij} = g(x_{ij}) + f_i(x_{ij}) + \varepsilon_{ij} \quad \text{for } j = 1, 2, \dots, n_i, i = 1, 2, \dots, m, \quad (1)$$

where g is the population profile function (i.e., the fixed-effects term), f_i is the random-effects term describing the variation of the i th individual profile from g , $\{x_{ij}, y_{ij}\}_{j=1}^{n_i}$ is the sample collected for the i th profile, and ε_{ij} 's are iid random errors with mean 0 and variance σ^2 . In Equation (1), it is routinely assumed that the random-effects term f_i and the errors ε_{ij} are independent of each other, and f_i is a realization of a mean 0 process with a common covariance function

$$\gamma(x_1, x_2) = E[f_i(x_1)f_i(x_2)].$$

Without loss of generality, we further assume that $x_{ij} \in [0, 1]$, for all i and j .

Equation (1) is fairly flexible. It includes many common correlation structures as special cases. For instance, if $f_i(x_{ij}) = \alpha_i$

and α_i is a mean 0 random variable, then within-profile correlation will have the compound symmetry form. If $\text{Corr}(f_i(x_1), f_i(x_2)) = \rho(|x_1 - x_2|; \alpha)$ for some correlation function ρ and a coefficient α , then the correlation structure includes the nonhomogeneous Ornstein–Uhlenbeck process and the Gaussian correlation model (cf. Zhang et al. 1998). When the design points are equally spaced and unchanged among different profiles, this model can also be used for describing the autoregressive correlation structure. Because of its flexibility, Equation (1) requires a relatively large set of IC profiles for model estimation and calibration compared to its parametric counterparts. Thanks to fast progress in sensor and information technology, automatic data acquisition has become increasingly common in the industry. Consequently, a large amount of IC data are often available, and Equation (1) allows us to make use of such data without imposing a parametric model form.

2.2 Estimation of the NME Model

In this section we discuss estimation of the IC g , γ , and σ^2 [cf. Equation (1)] from an IC dataset. These quantities will be used in constructing a Phase II nonparametric profile control chart when within-profile correlation is present and can be described by the NME Equation (1) (see related discussion in Section 2.3). In the literature there are some existing discussions about statistical analysis of correlated data under various settings and assumptions, including those in Altman (1990), Hart (1991), Hoover et al. (1998), Wang (1998), Zhang et al. (1998), Fan and Zhang (2000), Lin and Carroll (2000), and many others. Wu and Zhang (2002) proposed a method for estimating Equation (1) by combining linear mixed-effects (LME) modeling and local linear kernel smoothing (cf. Fan and Gijbels 1996). They demonstrate that their estimator of g , which is referred to as LLME, is often more efficient than certain alternative estimators in terms of the mean squared errors. Furthermore, by their approach, it is fairly easy to obtain consistent estimators of γ and σ^2 , which is important for constructing a Phase II control chart in the current study. For these reasons, we adopt Wu and Zhang’s method here, which is briefly described in the following.

For a given point $s \in [0, 1]$, LLME’s of $g(s)$ and $f_i(s)$ are obtained by minimizing the following penalized, negative log, local linear kernel likelihood function

$$\sum_{i=1}^m \left\{ \frac{1}{\sigma^2} \sum_{j=1}^{n_i} [y_{ij} - \mathbf{z}_{ij}^T(\boldsymbol{\beta} + \boldsymbol{\alpha}_i)]^2 K_h(x_{ij} - s) + \boldsymbol{\alpha}_i^T \mathbf{D}^{-1} \boldsymbol{\alpha}_i + \ln |\mathbf{D}| + n_i \ln(\sigma^2) \right\}, \quad (2)$$

where $K_h(\cdot) = K(\cdot/h)/h$, K is a symmetric density kernel function, h is a bandwidth, $\mathbf{z}_{ij}^T = (1, x_{ij} - s)$, $\boldsymbol{\beta}$ is a deterministic two-dimensional coefficient vector, and $\boldsymbol{\alpha}_i \sim (\mathbf{0}, \mathbf{D})$ is a two-dimensional vector of the random effects. Minimization of Equation (2) can be accomplished by the following iterative procedure:

Step 1. Set the initial values for \mathbf{D} and σ^2 , denoted as $\mathbf{D}_{(0)}$ and $\sigma_{(0)}^2$.

Step 2. At the k th iteration, for $k \geq 0$, compute estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}_i$ by solving the so-called mixed-model equation (cf. Davidian and Giltinan 1995; Wu and Zhang 2002), and the resulting estimates are

$$\widehat{\boldsymbol{\beta}}^{(k)} = \left\{ \sum_{i=1}^m \mathbf{Z}_i^T \boldsymbol{\Sigma}_i \mathbf{Z}_i \right\}^{-1} \left\{ \sum_{i=1}^m \mathbf{Z}_i^T \boldsymbol{\Sigma}_i \mathbf{y}_i \right\}, \quad (3)$$

$$\widehat{\boldsymbol{\alpha}}_i^{(k)} = \left\{ \mathbf{Z}_i^T \mathbf{K}_i \mathbf{Z}_i + \sigma_{(k)}^2 [\mathbf{D}_{(k)}]^{-1} \right\}^{-1} \mathbf{Z}_i^T \mathbf{K}_i (\mathbf{y}_i - \mathbf{Z}_i \widehat{\boldsymbol{\beta}}^{(k)}), \quad (4)$$

where $\mathbf{Z}_i = (\mathbf{z}_{i1}, \dots, \mathbf{z}_{in_i})^T$, $\mathbf{y}_i = (y_{i1}, \dots, y_{in_i})^T$, $\boldsymbol{\Sigma}_i = (\mathbf{Z}_i \mathbf{D}_{(k)} \mathbf{Z}_i^T + \sigma_{(k)}^2 \mathbf{K}_i^{-1})^{-1}$, and $\mathbf{K}_i = \text{diag}\{K_h(x_{i1} - s), \dots, K_h(x_{in_i} - s)\}$.

Step 3. Based on $\widehat{\boldsymbol{\beta}}^{(k)}$ and $\widehat{\boldsymbol{\alpha}}_i^{(k)}$, update the estimates of \mathbf{D} and σ^2 by

$$\mathbf{D}_{(k+1)} = \frac{1}{m} \sum_{i=1}^m \widehat{\boldsymbol{\alpha}}_i^{(k)} [\widehat{\boldsymbol{\alpha}}_i^{(k)}]^T, \quad (5)$$

$$\sigma_{(k+1)}^2 = \frac{1}{m} \sum_{i=1}^m \frac{1}{n_i} [\mathbf{y}_i - \mathbf{Z}_i (\widehat{\boldsymbol{\beta}}^{(k)} + \widehat{\boldsymbol{\alpha}}_i^{(k)})]^T \mathbf{K}_i \times [\mathbf{y}_i - \mathbf{Z}_i (\widehat{\boldsymbol{\beta}}^{(k)} + \widehat{\boldsymbol{\alpha}}_i^{(k)})]. \quad (6)$$

Step 4. Repeat Steps 2 and 3 until the following condition is satisfied:

$$\|\mathbf{D}_{(l)} - \mathbf{D}_{(l-1)}\|_1 / \|\mathbf{D}_{(l-1)}\|_1 \leq \epsilon,$$

where ϵ is a prespecified small positive number (e.g., $\epsilon = 10^{-4}$), and $\|\mathbf{A}\|_1$ denotes the sum of absolute values of all elements of \mathbf{A} . Then the algorithm stops at the l th iteration.

Note that in Step 4 we use the relative error of the successive estimates of \mathbf{D} in the convergence criterion. In fact, other estimates can also be used for this purpose. We use \mathbf{D} here because our simulation shows that it gives good results in various cases. As a side note, similar to the estimation of LME models, nonconvergence of the above iterative procedure can occasionally happen, although we found that the frequency of nonconvergence is negligible in all our simulation studies, except certain extreme cases such as the ones when m or n_i ’s are too small. To reduce the frequency of nonconvergence, it is suggested in the literature to use good initial values for \mathbf{D} and σ^2 . A simple but effective method is to set $\mathbf{D}_{(0)}$ to be the identity matrix and

$$\sigma_{(0)}^2 = \frac{1}{m} \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} [y_{ij} - \widehat{g}^{(P)}(x_{ij})]^2,$$

where $\widehat{g}^{(P)}(x_{ij})$ is the standard local linear kernel estimator constructed from the pooled data (cf. Hoover et al. 1998).

After obtaining estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}_i$ using the above algorithm, we can define

$$\widehat{g}(s) = \mathbf{e}_1^T \widehat{\boldsymbol{\beta}}(s), \quad \widehat{f}_i(s) = \mathbf{e}_1^T \widehat{\boldsymbol{\alpha}}_i(s), \quad (7)$$

$$\widehat{\gamma}(s_1, s_2) = \frac{1}{m} \sum_{i=1}^m \widehat{f}_i(s_1) \widehat{f}_i(s_2) \quad \text{for any } s_1, s_2 \in [0, 1],$$

where $\mathbf{e}_1 = (1, 0)^T$. Note that the variance estimator from the previous iterative procedure depends on s . Since σ^2 is a population parameter that does not depend on s , we suggest estimating it by

$$\hat{\sigma}^2 = \frac{1}{m} \sum_{i=1}^m \frac{1}{n_i} \sum_{j=1}^{n_i} [y_{ij} - \hat{g}(x_{ij}) - \hat{f}_i(x_{ij})]^2, \quad (8)$$

which is similar to the nonparametric estimator proposed by Hall and Marron (1990).

The following proposition investigates the asymptotic properties of the one-step estimators of g , γ , and σ^2 . For given initial values $\mathbf{D}_{(0)}$ and $\sigma_{(0)}^2$, the one-step estimators are those calculated by Equations (3), (7), and (8) when $k = 1$.

Proposition 1. Under the conditions in Equations (C1)–(C6), (C8)–I, and (C9) given in Appendix A, for any points $s_1, s_2 \in [0, 1]$, we have (i) $\hat{g}(s_1) = g(s_1)\{1 + O_p[m^{-1/2} + O(h^2)]\}$; (ii) $\hat{\gamma}(s_1, s_2) = \gamma(s_1, s_2)\{1 + O_p[h^2 + (nh)^{-1/2} + m^{-1/2} + (mnh^3)^{-1/2}]\}$; and (iii) $\hat{\sigma}^2 = \sigma^2\{1 + O_p[h^2 + (nh)^{-1/2} + m^{-1/2} + (mnh^3)^{-1/2}]\}$, where Equation (C6) assumes that $n_i \sim n$, for all i .

Result (i) of Proposition 1 is a direct conclusion of theorem 1 in Wu and Zhang (2002), except that certain conditions were modified. The other two results establish the consistency of the estimators of γ and σ^2 , which is important for the Phase II profile monitoring problem discussed in the following sections.

2.3 Phase II Nonparametric Profile Monitoring

In this section we present a Phase II nonparametric profile monitoring scheme in the general case when within-profile data might be correlated and the design points within and between profiles are arbitrary. This is a challenging task due to the following two major reasons. First, because the within-profile data might be correlated, estimation of the profile function g involves a considerable amount of computation if the NME modeling is also used in Phase II SPC, as described in Section 2.2. However, a good online control chart should maintain a reasonable efficiency while being effective in detecting profile shifts. Second, in cases when the design points $\mathbf{X}_i = \{x_{i1}, x_{i2}, \dots, x_{in_i}\}$ are unchanged from profile to profile, one method that comes to mind is to first average observed responses y_{ij} 's across individual profiles and then detect potential profile shifts using a generalized likelihood ratio test statistic (cf. Fan et al. 2001). This idea cannot be applied to the current problem directly because the response is observed at different design points in different profiles. One immediate alternative is to estimate g from individual profile data at a given set of points in $[0, 1]$. But the resulting estimates will be inefficient since they are constructed from individual profile data instead of from all observed data.

To overcome the above difficulties, at any point $s \in [0, 1]$, we consider the following local weighted negative log likelihood:

$$\begin{aligned} \text{WL}(a, b; s, \lambda, t) &= \sum_{i=1}^t \sum_{j=1}^{n_i} [y_{ij} - a - b(x_{ij} - s)]^2 \\ &\quad \times K_h(x_{ij} - s)(1 - \lambda)^{t-i} / v^2(x_{ij}), \end{aligned}$$

where λ is a weighting parameter and $v^2(x) = \gamma(x, x) + \sigma^2$ is the variance function of the response. Note that $\text{WL}(a, b; s, \lambda, t)$ combines the exponential weighting scheme used in EWMA at different time points through the term $(1 - \lambda)^{t-i}$ and the local linear kernel smoothing procedure (cf. Fan and Gijbels 1996). At the same time, it takes into account the heteroscedasticity of observations by using $v^2(x_{ij})$. Then the local linear kernel estimator of $g(s)$, defined as the solution to a of the minimization problem $\min_{a,b} \text{WL}(a, b; s, \lambda, t)$, has the expression

$$\hat{g}_{t,h,\lambda}(s) = \frac{\sum_{i=1}^t \sum_{j=1}^{n_i} U_{ij}^{(t,h,\lambda)}(s) y_{ij}}{\sum_{i=1}^t \sum_{j=1}^{n_i} U_{ij}^{(t,h,\lambda)}(s)}, \quad (9)$$

where

$$\begin{aligned} U_{ij}^{(t,h,\lambda)}(s) &= \frac{(1 - \lambda)^{t-i} K_h(x_{ij} - s)}{v^2(x_{ij})} \\ &\quad \times [m_2^{(t,h,\lambda)}(s) - (x_{ij} - s)m_1^{(t,h,\lambda)}(s)], \\ m_l^{(t,h,\lambda)}(s) &= \sum_{i=1}^t (1 - \lambda)^{t-i} \sum_{j=1}^{n_i} (x_{ij} - s)^l \\ &\quad \times K_h(x_{ij} - s) / v^2(x_{ij}), \quad l = 0, 1, 2. \end{aligned} \quad (10)$$

Note that $m_0^{(t,h,\lambda)}(s)$ is not used in Equation (9), but it will be used in Section 2.4.

From Equations (9) and (10) we can see that $\hat{g}_{t,h,\lambda}(s)$ makes use of all the available observations up to the current time point t , and different profiles are weighted as in a conventional EWMA chart (i.e., more recent profiles get more weight and the weight changes exponentially over time). When $\lambda = 0$ (i.e., all profiles receive equal weight), the resulting estimator is similar to the local linear generalized estimating equations (GEE) estimator considered in Lin and Carroll (2000). The GEE estimator can accommodate within-profile correlation without specifying the correlation structure (it uses the so-called independent working correlation matrix). Under certain mild conditions, Lin and Carroll showed that it is asymptotically the best estimator. Although Wu and Zhang (2002) demonstrated that their LLME estimator performs better in certain cases, especially when within-profile correlation is strong, this latter estimator involves a considerable amount of computation, and may not be feasible for Phase II profile monitoring, which is an online sequential procedure. As a comparison, the estimator in Equation (9) has an explicit formula, and the related computation is relatively fast.

Following the convention in Phase II analysis, we assume that the IC regression function, denoted as g_0 , and the variance function $v^2(\cdot)$ are both known. In practice, they need to be estimated from an IC dataset, as described in Section 2.2. Let $\xi_{ij} = [y_{ij} - g_0(x_{ij})]$, for all i and j , and $\hat{\xi}_{t,h,\lambda}(s)$ be the estimator defined in Equation (9) after y_{ij} are replaced by ξ_{ij} . Then the IC distribution of $\hat{\xi}_{t,h,\lambda}(s)$ does not depend on g_0 , and the original testing problem with $H_0 : g = g_0$ versus $H_1 : g \neq g_0$, which is associated with the profile monitoring problem, is changed to the one with $H_0 : g = 0$ versus $H_1 : g \neq 0$. Consequently, the IC distribution of the proposed control chart defined in the following and all quantities related to this distribution (e.g., the control limit L) do not depend on g_0 either, which will simplify the design and implementation of our proposed control chart.

When the process is IC, $|\widehat{\xi}_{t,h,\lambda}(s)|$ should be small. So, a natural statistic that can be used for SPC will be

$$T_{t,h,\lambda} = c_{0,t,\lambda} \int \frac{[\widehat{\xi}_{t,h,\lambda}(s)]^2}{v^2(s)} \Gamma_1(s) ds,$$

where

$$\begin{aligned} c_{t_0,t_1,\lambda} &= a_{t_0,t_1,\lambda}^2 / b_{t_0,t_1,\lambda}, \\ a_{t_0,t_1,\lambda} &= \sum_{i=t_0+1}^{t_1} (1-\lambda)^{t_1-i} n_i, \\ b_{t_0,t_1,\lambda} &= \sum_{i=t_0+1}^{t_1} (1-\lambda)^{2(t_1-i)} n_i, \end{aligned}$$

and Γ_1 is some prespecified density function. In the expression of $T_{t,h,\lambda}$, quantities $c_{0,t,\lambda}$ and $v(\cdot)$ are used for unifying its asymptotic variance. See Theorem 1 and its proof in the Appendix for details. In practice, we suggest using the following discretized version:

$$T_{t,h,\lambda} \approx \frac{c_{0,t,\lambda}}{n_0} \sum_{k=1}^{n_0} \frac{[\widehat{\xi}_{t,h,\lambda}(s_k)]^2}{v^2(s_k)}, \tag{11}$$

where $\{s_k, k = 1, \dots, n_0\}$ are some iid random numbers generated from Γ_1 . Then, the chart triggers a signal if

$$T_{t,h,\lambda} > L,$$

where $L > 0$ is a control limit chosen to achieve a specific IC ARL, denoted as ARL_0 . Hereafter, this chart is referred to as the mixed-effects nonparametric profile control (MENPC) chart.

In Phase II SPC, it is a convention that the IC distribution of the process measurements y_{ij} 's is assumed known. Then the control limit L can be searched for by simulation based on this distribution. In practice, the IC distribution is often unknown. Instead, we usually have a quite large IC dataset. In such cases, L can be searched for by a resampling algorithm, briefly described in the following. In each simulation run, we resample the IC dataset by randomly choosing a sequence of profiles with replacement. The sequence of profiles is sequentially chosen until a signal of shift is triggered by chart MENPC. Then an estimated ARL_0 value is computed based on B simulation runs, and L is searched for by matching the estimated ARL_0 value to the nominal value. In our numerical examples in Section 3, $B = 10,000$.

It should be pointed out that it is computationally faster to use points s_k rather than the original design points x_{ij} in Equation (11). As shown in Section 2.4, $T_{t,h,\lambda}$ can be calculated in a recursive manner when s_k are used in Equation (11), and this recursive feature will be lost if x_{ij} are used. Further, from the theoretical properties of $T_{t,h,\lambda}$ given in Theorem 2 and from our numerical results, the selection of $\{s_k, k = 1, 2, \dots, n_0\}$ has little effect on the performance of the MENPC chart, as long as n_0 is not too small. (See the related discussion in Section 2.5 about practical guidelines on the selection of certain procedure parameters.) In the special case when design points \mathbf{X}_i are unchanged for different profiles, we can use \mathbf{X}_i directly (instead of $\{s_k, k = 1, 2, \dots, n_0\}$) when computing the charting statistic. Next, we give some asymptotic properties of the charting statistic $T_{t,h,\lambda}$, which can justify the performance of the MENPC

chart to a certain degree and shed some light on practical design of the chart as well. Theorem 1 gives the asymptotic null distribution of $T_{t,h,\lambda}$, where design points x_{ij} 's in each IC profile are assumed to be iid with a density Γ_2 .

Theorem 1. Assume that the process is IC and that Equations (C1)–(C7) given in Appendix A all hold. Then we have the following results

- (i) If $n_i h$ is bounded for each i and Equation (C8)-II in Appendix A holds, then

$$(T_{t,h,\lambda} - \tilde{\mu}_h) / \tilde{\sigma}_h \xrightarrow{\mathcal{L}} N(0, 1),$$

where

$$\begin{aligned} \tilde{\mu}_h &= \frac{\int [K(u)]^2 du}{h} \int \frac{\Gamma_1(x)}{\Gamma_2(x)} dx, \\ \tilde{\sigma}_h^2 &= \frac{2 \int [K * K(u)]^2 du}{h} \int \frac{\Gamma_1^2(x)}{\Gamma_2^2(x)} dx. \end{aligned}$$

- (ii) If $n_i h \rightarrow \infty$ for each i and Equations (C8)-III and (C10) in Appendix A hold, then

$$\frac{1}{d_{0,t,\lambda}} T_{t,h,\lambda} \overset{D}{\sim} \frac{1}{n_0} \boldsymbol{\zeta}^T \boldsymbol{\zeta},$$

where $\overset{D}{\sim}$ denotes asymptotic equivalence in the distribution $d_{0,t,\lambda} = \sum_{i=t_0+1}^{t_1} (1-\lambda)^{2(t_1-i)} n_i^2 / b_{t_0,t_1,\lambda}$, and $\boldsymbol{\zeta}$ is an n_0 -dimensional multivariate normal random vector with mean $\mathbf{0}$ and covariance matrix

$$\boldsymbol{\Omega} = \begin{pmatrix} \frac{\gamma(s_1, s_1)}{v^2(s_1)} & \dots & \frac{\gamma(s_1, s_{n_0})}{v(s_1)v(s_{n_0})} \\ \vdots & \ddots & \vdots \\ \frac{\gamma(s_{n_0}, s_1)}{v(s_{n_0})v(s_1)} & \dots & \frac{\gamma(s_{n_0}, s_{n_0})}{v^2(s_{n_0})} \end{pmatrix}.$$

From Theorem 1(i), we can see that $T_{t,h,\lambda}$ is asymptotically independent of the nuisance parameters $\gamma(\cdot, \cdot)$ and σ^2 . The condition that $n_i h$ is bounded for each i is satisfied when n_i is finite and bounded for each i , which is often the case in practice. When $n_i h \rightarrow \infty$ for each i , the within-profile correlation will play an important role in the expansion of the variance of $T_{t,h,\lambda}$, which leads to a different asymptotic distribution, as described in Theorem 1(ii). In such situations it seems desirable to incorporate the covariance matrix $\boldsymbol{\Omega}$ into the test statistic. However, $\boldsymbol{\Omega}$ may not be positive-definite in certain cases (e.g., the case of compound symmetry correlation). Therefore, it is not obvious how to do so, which is left to our future research.

The next theorem investigates the asymptotic behavior of $T_{t,h,\lambda}$ under the OC model

$$y_{ij} = \begin{cases} g_0(x_{ij}) + f_i(x_{ij}) + \varepsilon_{ij}, & \text{if } 1 \leq i \leq \tau \\ g_1(x_{ij}) + f_i(x_{ij}) + \varepsilon_{ij}, & \text{if } i > \tau, \end{cases} \tag{12}$$

where τ is an unknown change point, and $g_1(x) = g_0(x) + \delta(x)$ is the unknown OC profile function. In the theorem we use the following notations:

$$\begin{aligned} \zeta_\delta &= \int \left[\delta(u) + \frac{h^2 \eta_1}{2} \delta''(u) \right]^2 \frac{\Gamma_1(u)}{v^2(u)} du, & \eta_1 &= \int K(t) t^2 dt, \\ \zeta_1 &= \int \delta^2(u) \frac{\Gamma_1(u) \gamma(u, u)}{v^2(u)} du, & \zeta_2 &= \int [\delta''(u)]^2 \Gamma_1(u) du. \end{aligned}$$

Theorem 2. Under Equations (C1)–(C7) given in Appendix A and the extra condition that $\zeta_2 < M$ for some constant $M > 0$, we have

- (i) If $n_i h$ is bounded for each i , $c_{0,t,\lambda} n h \zeta_1 \rightarrow 0$, and Equation (C8)–IV in Appendix A holds, then $(T_{t,h,\lambda} - \tilde{\mu}_h - c_{0,t,\lambda} \zeta_\delta) / \tilde{\sigma}_h \xrightarrow{\mathcal{L}} N(0, 1)$.
- (ii) If $n_i h$ is bounded for each i , $\zeta_2 \rightarrow 0$, and Equation (C8)–IV in Appendix A holds, then $T_{t,h,\lambda}$ has nontrivial power (i.e., greater than the nominal level) when $\delta \propto c_{0,t,\lambda}^{-4/9}$ and $h = O(c_{0,t,\lambda}^{-2/9})$.
- (iii) If $n_i h \rightarrow \infty$ for each i , and Equations (C8)–III and (C10) in Appendix A hold, then $\frac{1}{d_{0,t,\lambda}} T_{t,h,\lambda} \stackrel{D}{\sim} \frac{1}{n_0} \boldsymbol{\zeta}^T \boldsymbol{\zeta}$, where $\boldsymbol{\zeta}$ is an n_0 -dimensional multivariate normal random vector with mean $\boldsymbol{\delta} = [\delta(s_1), \dots, \delta(s_{n_0})]^T$ and covariance matrix $\boldsymbol{\Omega}$.

2.4 Some Computational Issues

Although computing power has improved dramatically and it is computationally trivial to do nonparametric function estimation for individual profiles, for online process monitoring, which generally handles a large number of profiles, fast implementation is still important and some computational issues deserve our careful examination. For the proposed chart, computing the test statistic $T_{t,h,\lambda}$ by Equations (9) through (11) requires a considerable amount of computing time and a substantial amount of storage space as well to save all past profile observations. In this section we provide updating formulas for computing $T_{t,h,\lambda}$, which can greatly simplify the computation and reduce the storage requirement. Let

$$\begin{aligned} \tilde{m}_l^{(t,h)}(s) &= \sum_{j=1}^{n_k} (x_{ij} - s)^l K_h(x_{ij} - s) / v^2(x_{ij}), \quad l = 0, 1, 2, \\ \tilde{q}_l^{(t,h)}(s) &= \sum_{j=1}^{n_k} (x_{ij} - s)^l K_h(x_{ij} - s) y_{ij} / v^2(x_{ij}), \quad l = 0, 1. \end{aligned}$$

Then, $m_l^{(t,h,\lambda)}(s)$ in Equation (10) can be recursively updated by

$$m_l^{(t,h,\lambda)}(s) = (1 - \lambda) m_l^{(t-1,h,\lambda)}(s) + \tilde{m}_l^{(t,h)}(s), \quad l = 0, 1, 2,$$

where $m_l^{(0,h,\lambda)}(s) = 0$, for $l = 0, 1, 2$. Let $q_l^{(t,h,\lambda)}(s)$, for $l = 0, 1$, be two working functions defined by the recursive formula

$$q_l^{(t,h,\lambda)}(s) = (1 - \lambda) q_l^{(t-1,h,\lambda)}(s) + \tilde{q}_l^{(t,h)}(s), \quad l = 0, 1,$$

where $q_l^{(0,h,\lambda)}(s) = 0$, for $l = 0, 1$. Then, we have

$$\begin{aligned} \widehat{g}_{t,h,\lambda}(s) &= [M^{(t,h,\lambda)}]^{-1} \{ (1 - \lambda)^2 M^{(t-1,h,\lambda)} \widehat{g}_{t-1,h,\lambda} \\ &\quad + [\tilde{q}_0^{(t,h)} m_2^{(t,h,\lambda)} - \tilde{q}_1^{(t,h)} m_1^{(t,h,\lambda)}] \\ &\quad + (1 - \lambda) [q_0^{(t-1,h,\lambda)} \tilde{m}_2^{(t,h)} - q_1^{(t-1,h,\lambda)} \tilde{m}_1^{(t,h)}] \}, \end{aligned} \tag{13}$$

where $M^{(t,h,\lambda)}(s) = m_2^{(t,h,\lambda)}(s) m_0^{(t,h,\lambda)}(s) - [m_1^{(t,h,\lambda)}(s)]^2$. On the right-hand side of the above equation the dependence on s in each function is not made explicit in the notation for simplicity, which should not cause any confusion.

Using the previous updating formulas, implementation of the MENPC chart can be briefly described as follows. At time point t , we first compute quantities $\tilde{m}_l^{(t,h)}(s)$, for $l = 0, 1, 2$, and $\tilde{q}_l^{(t,h)}(s)$, for $l = 0, 1$, at n_0 predetermined s locations (see related discussion in Sections 2.3 and 2.5 about the selection of $\{s_k, k = 1, \dots, n_0\}$). Then, $m_l^{(0,h,\lambda)}(s_k)$, for $l = 0, 1, 2$, and $q_l^{(0,h,\lambda)}(s_k)$, for $l = 0, 1$, are updated by the previous formulas. Finally, $\widehat{g}_{t,h,\lambda}(s)$ is computed from Equation (13), and the test statistic $T_{t,h,\lambda}$ is computed by $\widehat{g}_{t,h,\lambda}(s)$ after y_{ij} is replaced by ξ_{ij} . This algorithm only requires $O(n_0 n_i h)$ operations for monitoring the i th profile, which is the same order as the computation involved in conventional local linear kernel smoothing. If n_i and n_0 are both large, we can further decrease the computation to the order of $O(n_i h)$ by using the updating algorithm proposed by Seifert et al. (1994) (see Fan and Marron 1994 for a similar algorithm). Obviously, using the proposed updating formulas, computer storage does not grow sequentially with time t .

2.5 Certain Practical Guidelines

On the Sizes of m and n_i . The number of IC profiles should be large enough to generate accurate estimates of IC g , γ , and σ^2 . This has become a less significant issue nowadays because a large amount of IC data are often available due to advances in data acquisition techniques. Empirically speaking, to attain desirable IC distributional properties, we recommend using IC data with $n_i \geq 20$ and $m \geq 500$, although more systematic future research is required to determine the size of a necessary IC dataset.

On Choosing the Bandwidth. In estimation of the NME model in Equation (1) by the iterative procedure described in Section 2.2 for Phase I analysis, people often use data-driven bandwidth selection techniques, such as the least-squares cross-validation (CV) and the generalized cross-validation (GCV) procedures. Wu and Zhang (2002) proposed a CV method by combining leave-one-subject-out and leave-one-point-out CV schemes. Their study showed that this method can effectively track estimates of both g and f_i . We adopt it in our numerical analysis of the IC data. With respect to Phase II online profile monitoring, like many other smoothing-based tests, the performance of the MENPC chart depends on the selection of the bandwidth parameter h used in Equation (9). The optimal selection of h remains an open problem in this area, and it is widely recognized that optimal h for nonparametric curve estimation is generally not optimal for testing (cf., e.g., Hart 1997). A uniformly most powerful test usually does not exist due to the fact that nonparametric regression functions have infinite dimensions. We suggest using the following empirical bandwidth formula

$$h_E = \begin{cases} c_1 n^{-1/5} \left(\sum_{i=1}^n (x_i - \bar{x})^2 / n \right)^{1/2} & \text{for balanced design} \\ c_2 [\tilde{n}(2 - \lambda) / \lambda]^{-1/5} \sqrt{\text{Var}(x)} & \text{for random design,} \end{cases} \tag{14}$$

where $\bar{x} = \sum_{i=1}^n x_i$ is the mean of the n design points in the balanced design case, \tilde{n} and $\text{Var}(x)$ are the average number of

design points, and the variance of design points within a profile, respectively, in the random design case, which can be estimated from the IC data, and c_1 and c_2 are two constants. Empirically, c_1 and c_2 can be any values in the interval $[1.0, 2.0]$. By Equation (14), a smaller bandwidth is suggested for the random design case because the actual number of observations used in the MENPC chart at each time point is about $c_{0,t,\lambda}$ in such cases which is roughly $\tilde{n}(2-\lambda)/\lambda$.

On Choosing λ . Traditionally, a larger λ leads to a quicker detection of larger shifts (cf., e.g., Lucas and Saccucci 1990). However, in the mixed-effects modeling, efficient estimation of the common profile function g requires the use of observations across a number of different profiles, due mainly to the existence of random effects. From Theorems 1 and 2, we can see that the effective number of profiles used in the MENPC chart at each time point is asymptotically $(2-\lambda)/\lambda$. So, to estimate g properly, $(2-\lambda)h/\lambda$ should be large enough [cf. Equation (C8)-II in Appendix A]. Consequently, λ cannot be chosen to be too large. Otherwise, even for a shift of large magnitude, $T_{t,h,\lambda}$ may not be able to detect it quickly, due to a large bias in estimating g . Empirically, we suggest choosing $\lambda \in [0.02, 0.1]$ if h_E in Equation (14) is used.

On Choosing $\{s_k, k = 1, 2, \dots, n_0\}$. Based on our numerical experience, the selection of $\{s_k, k = 1, 2, \dots, n_0\}$ does not much affect the performance of the MENPC chart, as long as n_0 is not too small and s_k 's cover all the key parts of g_0 (e.g., peaks/valleys or oscillating regions) well. In our numerical examples presented in Section 3, we find that results will hardly change when $n_0 \geq 40$.

3. A SIMULATION STUDY

We present some simulation results in this section regarding the numerical performance of the proposed Phase II nonparametric profile monitoring chart MENPC. Throughout this section, the kernel function is chosen to be the Epanechnikov kernel function $K(x) = 0.75(1-x^2)I(-1 \leq x \leq 1)$, which is commonly used in the local smoothing literature due to some of its optimality properties (see chapter 2 of Fan and Gijbels 1996 for a related discussion). The IC ARL is fixed at 200. The error distribution is assumed to be standard Normal. For simplicity, we assume that $n_i = n = 20$ for all i , $x_{ij} \sim \text{Uniform}(0, 1)$, for $j = 1, \dots, n$, $s_k = (k - 0.5)/n_0$, for $k = 1, \dots, n_0$, and $n_0 = 40$. All ARL values reported in this section are averages of 10,000 replicated simulations. In addition, as suggested by Hawkins and Olwell (1998), here we focus on the steady-state OC ARL behavior of the chart, and assume that shifts can only occur after time $\tau = 30$. When computing the OC ARL values, any simulation run in which a signal occurs before the $(\tau + 1)$ th profile will be ignored.

It is challenging to compare the proposed method with alternative methods since there is no obvious comparable method in the literature. Here, we first consider the control chart based on fixed-effects modeling for monitoring nonparametric profiles as an alternative method, denoted as FENPC. In this approach, f_i in Equation (1) is assumed to be zero, and consequently, $v^2(x) = \sigma^2$ is used in the construction of $T_{t,h,\lambda}$ [cf. Equations (9)–(11)]. Note that the FENPC chart can be regarded as a generalization of the NEWMA chart by Zou, Tsung, and Wang

(2008); the latter assumes that design points in different profiles are deterministic and unchanged from one profile to another while the former can handle arbitrary designs. By comparing the MENPC chart with the FENPC chart, we can see what will happen if within-profile correlation exists but is ignored. Following the recommendations in Section 2.5, for both charts, we use $h = 1.5[n(2-\lambda)/\lambda]^{-1/5} \sqrt{\text{Var}(x)}$ in Phase II SPC, and use the CV method by Wu and Zhang (2002) for choosing bandwidths in modeling the IC data.

First, we study the possible effect of within-profile correlation on the IC run-length distributions of the two charts in the following four cases:

- (I) $f_i(x_{ij}) = 0$;
- (II) $f_i(x_{ij}) = b\alpha_i x_{ij}$;
- (III) $f_i(x_{ij}) = b\alpha_i \cos(2\pi x_{ij})$;
- (IV) $[f_i(x_{i1}), \dots, f_i(x_{in})]^T \sim b \cdot \text{MN}_n(\mathbf{0}, \Sigma)$,

where $\alpha_i, i = 1, 2, \dots$, are independent standard normal random variables, $\text{MN}_n(\mathbf{0}, \Sigma)$ denotes the n -dimensional multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix Σ , and b is a constant. In all cases we assume that $g_0(\cdot) = 0$. Obviously, in case (I), there is no within-profile correlation. In case (II), the random component $f_i(x)$ is a linear function of the covariate x . In case (III), it is a cosine function of the covariate x . In case (IV), the random component vector has a joint Normal distribution. In this case, we further assume that $\Sigma = (\rho_{jk})$ and $\rho_{jk} = 0.2^{|x_{ij} - x_{ik}|}$, for $j, k = 1, \dots, n$. In each case, a large IC sample with $m = 500$ and $n = 200$ is generated. By using estimated $\gamma(\cdot, \cdot)$ and σ^2 from this IC data, the control limits of the two control charts are computed, as described in Section 2.3. Then their IC ARL's and the corresponding standard deviations of the run length, denoted as ARL_0 and SDRL_0 , respectively, are summarized in Table 1.

From the table it can be seen that, in case (I) when the assumption of within-profile independence is valid, the ARL's and SDRL's of both charts are close to their nominal values 200, as expected. However, in cases (II)–(IV) when within-profile correlation is substantial, the FENPC chart has large biases in both ARL_0 and SDRL_0 , especially when b is large. As a comparison, our proposed MENPC chart still performs well in these cases.

Next, we investigate the OC performance of the two control charts. The following two representative OC models are considered here:

- (i) $g_1(x) = 2\theta(x - 0.5)$;
- (ii) $g_1(x) = \theta \sin(2\pi(x - 0.5))$.

In case (i), $\delta(x) = g_1(x) - g_0(x)$ is a straight line; it oscillates greatly in case (ii). The parameter θ controls the shift magnitude. For each control chart, two λ values 0.1 and 0.2 are used. With each λ value, the bandwidth in Equation (14) is used. In this comparison, we pretend that the IC model is known exactly, and the control limit of the FENPC chart is adjusted to attain the desired IC ARL value 200. Therefore, the difference between the MENPC and FENPC charts in this comparison is mainly in whether or not the within-profile correlation is taken into ac-

Table 1. IC ARL and SDRL values of the MENPC and FENPC charts

	<i>b</i>	Model (I)		Model (II)		Model (III)		Model (IV)	
		ARL ₀	SDRL ₀						
MENPC	0.25	205	203	196	197	198	199	206	208
	0.50	205	203	201	200	195	194	208	204
	1.00	205	203	193	190	194	194	206	205
FENPC	0.25	199	197	110	109	170	172	38.3	34.0
	0.50	199	197	29.8	29.2	105	104	21.5	20.0
	1.00	199	197	8.48	8.29	35.5	34.6	15.1	14.2

count in Phase II SPC. Obviously, for IC models (I) and (IV), $v^2(x)$ is independent of x ; thus, the two charts will be equivalent in such cases. For this reason, Table 2 presents the OC ARL values of the two charts for IC models (II) and (III) only.

From Table 2, we can have the following results. First, the MENPC chart outperforms the FENPC chart in all cases, which demonstrates the fact that the former chart is more effective because it explicitly incorporates within-profile heteroscedastic-

Table 2. OC ARL comparison of the MENPC and FENPC charts when $ARL_0 = 200$, $n = 20$, $n_0 = 40$, and $\lambda = 0.1$ or 0.2

IC model	θ	OC model (i)		OC model (ii)	
		MENPC	FENPC	MENPC	FENPC
$\lambda = 0.1$					
(II)	0.20	130 (1.36)	139 (1.48)	85.3 (0.83)	100 (0.98)
	0.30	80.5 (0.78)	98.0 (0.99)	40.5 (0.32)	52.2 (0.46)
	0.40	48.6 (0.42)	62.6 (0.59)	22.3 (0.15)	29.0 (0.21)
	0.60	20.7 (0.13)	28.4 (0.20)	10.6 (0.05)	13.1 (0.06)
	0.80	12.1 (0.06)	16.0 (0.09)	6.81 (0.03)	8.57 (0.03)
	1.20	6.64 (0.02)	8.43 (0.03)	4.06 (0.02)	5.14 (0.02)
	1.60	4.60 (0.02)	5.82 (0.02)	2.93 (0.01)	3.71 (0.01)
	2.00	3.51 (0.01)	4.49 (0.01)	2.33 (0.01)	2.96 (0.01)
(III)	0.20	131 (1.38)	162 (1.73)	68.3 (0.64)	121 (1.25)
	0.30	81.0 (0.79)	121 (1.26)	31.2 (0.24)	65.7 (0.60)
	0.40	48.1 (0.42)	81.2 (0.76)	17.6 (0.11)	34.2 (0.25)
	0.60	21.4 (0.14)	33.3 (0.24)	9.05 (0.04)	14.4 (0.06)
	0.80	12.4 (0.06)	17.7 (0.09)	6.02 (0.02)	9.14 (0.03)
	1.20	6.59 (0.03)	9.04 (0.03)	3.70 (0.01)	5.39 (0.02)
	1.60	4.51 (0.02)	6.10 (0.02)	2.68 (0.01)	3.92 (0.01)
	2.00	3.43 (0.01)	4.71 (0.01)	2.20 (0.01)	3.15 (0.01)
$\lambda = 0.2$					
(II)	0.20	162 (1.70)	167 (1.83)	137 (1.51)	136 (1.43)
	0.30	128 (1.36)	131 (1.38)	85.3 (0.89)	87.2 (0.92)
	0.40	93.8 (0.97)	97.1 (1.02)	47.4 (0.46)	52.5 (0.51)
	0.60	43.6 (0.40)	48.1 (0.47)	15.7 (0.11)	19.6 (0.15)
	0.80	19.5 (0.15)	24.4 (0.20)	7.94 (0.04)	9.81 (0.05)
	1.20	7.48 (0.03)	9.23 (0.05)	4.06 (0.01)	4.82 (0.02)
	1.60	4.59 (0.02)	5.49 (0.02)	2.79 (0.01)	3.28 (0.01)
	2.00	3.39 (0.01)	3.97 (0.01)	2.16 (0.01)	2.55 (0.01)
(III)	0.20	164 (1.77)	181 (1.94)	124 (1.32)	156 (1.66)
	0.30	133 (1.41)	156 (1.66)	69.2 (0.70)	113 (1.17)
	0.40	93.7 (0.98)	125 (1.33)	35.2 (0.32)	71.5 (0.71)
	0.60	41.6 (0.38)	70.1 (0.72)	12.0 (0.07)	24.3 (0.19)
	0.80	18.8 (0.14)	32.0 (0.27)	6.69 (0.03)	10.7 (0.05)
	1.20	7.25 (0.03)	9.94 (0.05)	3.60 (0.01)	5.04 (0.02)
	1.60	4.39 (0.02)	5.73 (0.02)	2.53 (0.01)	3.42 (0.01)
	2.00	3.23 (0.01)	4.13 (0.01)	2.03 (0.01)	2.70 (0.01)
	2.40	2.60 (0.01)	3.27 (0.01)	1.73 (0.01)	2.21 (0.01)

NOTE: Standard errors are in parentheses.

ity of observations into its charting statistic. Second, the control charts with $\lambda = 0.2$ do not have satisfactory performance in most cases, compared to the charts with $\lambda = 0.1$. That is because the charts use about $[(2 - \lambda)h/\lambda]n$ observations at each given point and the factor $(2 - \lambda)h/\lambda = 1.37$ is fairly small in the case when $\lambda = 0.2$. Consequently, the charts estimate the regression function g with large bias and its ability to detect profile shifts is thus greatly reduced. This result confirms our recommendation in Section 2.5 that λ should be chosen smaller for monitoring profiles with within-profile correlation than for monitoring profiles with independent observations. In addition, further simulations (not reported here) also show that, when n_0 is chosen larger than 40, the performance of either chart will not change much.

Next, we compare our proposed MENPC control chart, which is based on nonparametric mixed-effects modeling, with the control charts by Jensen, Birch, and Woodall (2008) and Jensen and Birch (2009) that are based on linear and nonlinear mixed-effects modeling. It should be pointed out that both charts by Jensen and co-authors focus on Phase I SPC only. To compare with the proposed Phase II MENPC chart, they need to be modified for online sequential profile monitoring. With the nonlinear profile monitoring chart by Jensen and Birch (2009), this modification turns out to be difficult for the following reason. As pointed out by Zou, Tsung, and Wang (2008) and Williams, Woodall, and Birch (2007), when nonlinear regression methods are used for constructing a control chart, non-convergence or slow convergence of numerical algorithms is often an issue because Newton–Raphson iterative algorithms are routinely used in such cases to obtain parameter estimates. For Phase I analysis, this issue may not be serious as long as the initial values of the iterative algorithm are properly chosen. However, for Phase II online process monitoring where a large number of tests are performed, it is usually difficult to find the proper initial values to guarantee the convergence of the iterative algorithm, especially when the profile model (instead of just the model parameters) changes after the process goes OC. For this reason, only the linear mixed-effects (LME) modeling approach by Jensen, Birch, and Woodall (2008) is considered here, and it is modified for Phase II profile monitoring as follows. Assume that the i th profile data follow the LME model

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{X}_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, 2, \dots,$$

where \mathbf{X}_i is the design matrix, $\mathbf{y}_i = (y_{i1}, \dots, y_{in_i})^T$, $\boldsymbol{\beta}$ is the coefficient vector of the fixed-effects term, $\mathbf{b}_i \sim N_p(0, \mathbf{D})$ the coefficient vector of the random-effects term, and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \dots, \varepsilon_{in_i})^T$. Then, $\boldsymbol{\beta}$ can be estimated by the following weighted least-squares estimator constructed from the i th profile data:

$$\widehat{\boldsymbol{\beta}}_i = (\mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i)^{-1} \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{y}_i,$$

where $\mathbf{V}_i = \mathbf{X}_i \mathbf{D} \mathbf{X}_i^T + \sigma^2 \mathbf{I}$. Following the framework of the MEWMA chart by Zou, Tsung, and Wang (2007), which is for online monitoring of general linear profiles using multivariate EWMA schemes, let us consider a sequence of EWMA working vectors

$$\mathbf{w}_i = (1 - \lambda)\mathbf{w}_{i-1} + \lambda(\widehat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_0),$$

Table 3. Six sets of parameters of the two OC models for comparing the MENPC and LMEP control charts

	Model (1)			Model (2)	
	β_0	β_1	β_2	β_4	β_5
(i)	1.3	2.0	3.0	0.1	1.5
(ii)	1.5	2.0	3.0	0.3	1.5
(iii)	1.0	2.3	3.0	0.5	1.5
(iv)	1.0	2.5	3.0	0.1	2.5
(v)	1.0	2.0	3.3	0.3	2.5
(vi)	1.0	2.0	3.5	0.5	2.5

where $\boldsymbol{\beta}_0$ denotes the IC value of $\boldsymbol{\beta}$. Then, the control chart triggers a signal if

$$Q_i = \frac{2 - \lambda}{\lambda} \mathbf{w}_i^T (\mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i) \mathbf{w}_i > L,$$

where $L > 0$ is a control limit chosen to achieve a specific IC ARL. This chart is called the linear mixed-effect profile (LMEP) monitoring chart hereafter.

In the next example, we compare the LMEP and MENPC charts under the IC model

$$y_{ij} = 1 + 2x_{ij} + 3x_{ij}^2 + \alpha_i x_{ij} + \varepsilon_{ij},$$

where α_i , for $i = 1, 2, \dots$, are iid standard normal random variables. This model assumes that the fixed-effects part is a quadratic function of the predictor, and the random-effects part is a linear function. So it is a LME model. In the simulation, n_i , λ , x_{ij} , and ε_{ij} are chosen or generated in the same way as that in the example of Tables 1 and 2. The following two OC models are considered here:

- (1) $y_{ij} = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_{ij}$;
- (2) $y_{ij} = 1 + 2x_i + 3x_i^2 + \beta_4 \sin(2\pi\beta_5 x_i) + \varepsilon_{ij}$,

where β 's are deterministic coefficients. Obviously, OC Equation (1) is an LME model, and OC Equation (2) is not an LME model. Six sets of values of β 's are listed in Table 3, which correspond to OC models that differ from the IC model with different degrees. The OC ARL values of the LMEP and MENPC charts are presented in Table 4. From the table, we can see that, even for OC Equation (1) which is an LME model, the results of MENPC are compatible with the results of LMEP. For OC Equation (2), MENPC outperforms LMEP uniformly.

4. A REAL-DATA APPLICATION

In this section we demonstrate the proposed methodology by applying it to a dataset from a manufacturing process of aluminium electrolytic capacitors (AEC's). This process transforms raw materials, such as anode aluminum foil, cathode aluminum foil, guiding pin, electrolyte sheet, plastic cover, aluminum shell, and plastic tube into AEC's that are appropriate for use in low-leakage circuits and are well adapted to a wide range of environmental temperatures. The whole manufacturing process consists of a sequence of operations, including clenching, rolling, soaking, assembly, cleaning, aging, and classifying. Before packing, a careful quality monitoring step is required by sampling from a batch of products.

Table 4. OC ARL comparison of the MENPC and LMEP charts when $ARL_0 = 200, n = 20, \lambda = 0.1$ or 0.2

		OC model (1)		OC model (2)	
		MENPC	LMEP	MENPC	LMEP
$\lambda = 0.1$	(i)	18.9 (0.13)	18.3 (0.12)	73.1 (0.69)	154 (1.61)
	(ii)	8.93 (0.04)	8.67 (0.04)	24.4 (0.15)	109 (1.15)
	(iii)	63.3 (0.60)	91.7 (0.91)	12.5 (0.05)	75.1 (0.77)
	(iv)	29.0 (0.25)	44.1 (0.38)	107 (1.10)	166 (1.75)
	(v)	108 (1.08)	102 (1.03)	44.6 (0.35)	130 (1.38)
	(vi)	59.4 (0.54)	50.9 (0.46)	20.5 (0.10)	94.2 (0.98)
$\lambda = 0.2$	(i)	25.0 (0.21)	22.0 (0.19)	127 (1.20)	149 (1.57)
	(ii)	9.90 (0.06)	8.54 (0.05)	59.4 (0.57)	102 (1.07)
	(iii)	76.0 (0.73)	113 (1.17)	23.3 (0.18)	64.5 (0.68)
	(iv)	35.4 (0.35)	56.9 (0.57)	163 (1.74)	165 (1.75)
	(v)	121 (1.27)	123 (1.29)	114 (1.13)	124 (1.33)
	(vi)	69.6 (0.71)	67.5 (0.70)	68.8 (0.64)	87.9 (0.92)

NOTE: Standard errors are in parentheses.

Regarding the quality of AEC's, the most important characteristic is the dissipation factor (DF), which can be automatically measured by an electronic device. However, it is known that DF measurements will change significantly with environmental temperature, and there is a specific requirement about the adaptability of AEC's to the temperature. To monitor the adaptability, engineers put a sampled AEC in a container. Then the container's temperature is controlled, and the temperature is supposed to stably increase from -26°F to 78°F . In this process, measurements of DF and the actual temperature inside the container are taken at 53 equally spaced time points. The actual temperature inside the container is reported by a temperature sensor. So, for each sampled AEC, a set of 53 observations of the pair (temperature, DF), which corresponds to (x, y) in Equation (1), are obtained for monitoring the adaptability of the AEC to the temperature. Figure 1 shows three AEC profiles along with an NME estimate of the IC profile function (see related discussion in the following). It should be noted that the

actual temperature inside a container will fluctuate around its nominal level at each observation time.

Therefore, the actual temperature readings of different containers at a given observation time are all different, although the differences are usually small. For this dataset, profile monitoring charts requiring deterministic and fixed design points in different profiles [e.g., the one by Zou, Tsung, and Wang (2008)] will be difficult to use.

The entire AEC dataset contains 144 profiles and each profile has $n = 53$ observations. We use the first 96 profiles to calibrate the proposed model and the remaining ones to test the model. A calibration sample of this size might be smaller than one would like to fully determine the IC distribution, but suffices to illustrate the use of the method in a real-world setting. Since the DF measurements are taken in consecutive time intervals, the AEC data exhibit a considerable amount of positive serial autocorrelation, which is confirmed by our analysis described in the following.

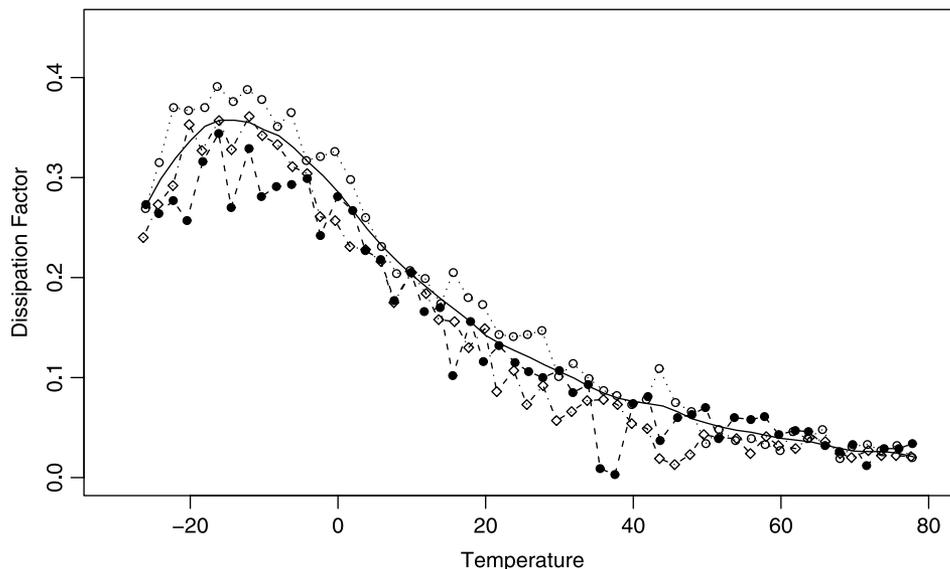


Figure 1. Three AEC profiles (lines connecting points with three different symbols) and the NME estimate (solid curve) of the IC profile function.

We first fit Equation (1) to the calibration sample of the first 96 profiles of the data by the iterative procedure in Equations (3) through (6) using the suggested initial values of \mathbf{D} and σ^2 given in Section 2.2 and the CV bandwidth selection procedure suggested by Wu and Zhang (2002). The resulting IC profile estimate \hat{g} is displayed in Figure 1 by the solid curve. From Equations (7) and (8), we can also compute the estimated correlation of two observations of the response variable y at any two points s_1 and s_2 in the design interval

$$\hat{\rho}(s_1, s_2) = \hat{\gamma}(s_1, s_2) / [\hat{v}(s_1)\hat{v}(s_2)],$$

where $\hat{\gamma}(s_1, s_2)$ is defined in Equation (7), $\hat{v}^2(s) = \hat{\gamma}(s, s) + \hat{\sigma}^2$ is the estimated variance of y at s , and $\hat{\sigma}^2$ is defined in Equation (8). Let $x_j^* = 2(j - 1) - 26$, for $j = 1, 2, \dots, 53$, be 53 equally spaced points in the design interval $[-26, 78]$, which denote the nominal temperature levels used in taking DF measurements of the sampled AEC's. The estimated correlations $\hat{\rho}(x_j^*, x_{j+1}^*)$, $\hat{\rho}(x_j^*, x_{j+3}^*)$, $\hat{\rho}(x_1^*, x_j^*)$, and $\hat{\rho}(x_j^*, x_{53}^*)$, for $j = 1, 2, \dots, 53$, are shown in Figure 2(a). From the plot, we can see that correlation within AEC profiles is substantial; thus, it should not be ignored. Figure 2(b) shows the estimated standard deviation $\hat{v}(x_j^*)$ of the response variable y at x_j^* , for $j = 1, 2, \dots, 53$, from which heteroscedasticity of the response variable y at different positions of x is clearly seen. Therefore, the proposed MENPC chart will be more appropriate to use in this case, compared to the FENPC chart discussed in the previous section which ignores the heteroscedasticity. In addition, we can obtain an estimate of the error standard deviation σ to be 0.016, by the formula in Equation (8), which is much smaller than $\hat{v}(x_j^*)$, especially when $j \in [12, 50]$. This result implies that the random-effects term in Equation (1) describes a substantial amount of random variation in the data.

Next, we construct the proposed MENPC chart for Phase II profile monitoring using the estimated IC parameters computed from the IC data. As in the simulation study discussed in the previous section, the IC ARL is fixed at 200, and λ is chosen to be 0.1. For simplicity, we choose $n_0 = n = 53$ and

$\{s_k, k = 1, 2, \dots, n_0\}$ to be equally spaced in the design interval $[-26, 78]$ of the explanatory variable. All other parameters are chosen to be the same as those used in the example of Table 1. The control limit is computed to be 18.24 by simulation. The charting statistics $T_{t,h,\lambda}$, for $t = 97, \dots, 144$ are shown in Figure 3 along with the control limit by the solid curve and solid horizontal line, respectively. The corresponding FENPC chart, using the same λ and h as those in the MENPC chart, is also presented in the figure along with its control limit 34.52, by the dashed curve and the dashed horizontal lines. From the plot, it can be seen that the MENPC chart gives a signal of profile shift around the 112th time point, and remains above the control limit for several profiles until the 120th profile. This result confirms a marked step-change which seems to have occurred around the 108th profile. The process seems to have been adjusted around the 119th profile; thus, the MENPC charting statistic goes back below its control limit afterward. As a comparison, the FENPC chart does not give a signal until the 118th profile.

5. SUMMARY AND CONCLUDING REMARKS

In this article we propose a Phase II control chart for monitoring nonparametric profiles. This chart is based on nonparametric mixed-effects modeling, local linear kernel smoothing, and EWMA process monitoring. It can accommodate within-profile correlation and arbitrary design. Numerical studies show that it is effective in detecting step profile shifts in various cases. Some numerical studies not reported in the article show that it is also effective in detecting certain drifts in profiles. The AEC example demonstrates that our method can be implemented conveniently in industrial applications.

As pointed out in Section 1, this article focuses on Phase II profile monitoring only. It requires much future research to extend our method to Phase I analysis, in which detection of outliers and spikes will also be of interest, besides detection of step shifts in profiles. For Phase II profile monitoring, we only consider possible step shifts in the fixed-effects term g of Equation (1). In some cases, the variance-covariance structure of the

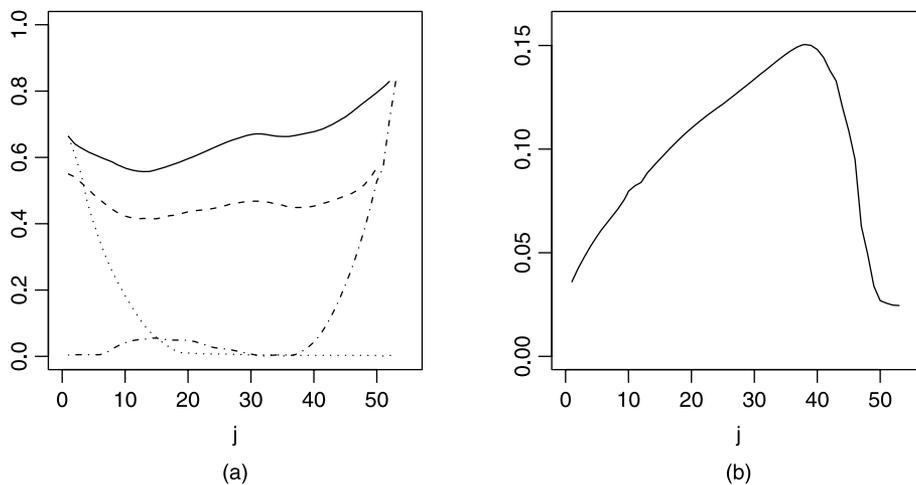


Figure 2. (a) Solid, dashed, dotted, and dash-dotted curves represent estimated within-profile correlations $\hat{\rho}(x_j^*, x_{j+1}^*)$, $\hat{\rho}(x_j^*, x_{j+3}^*)$, $\hat{\rho}(x_1^*, x_j^*)$, and $\hat{\rho}(x_j^*, x_{53}^*)$, for $j = 1, 2, \dots, 53$, where $\{x_j^*, j = 1, 2, \dots, 53\}$ are 53 equally spaced points in the design interval $[-26, 78]$. (b) Estimated standard deviation $\hat{v}(x_j^*)$ of the response variable y at x_j^* , for $j = 1, 2, \dots, 53$.

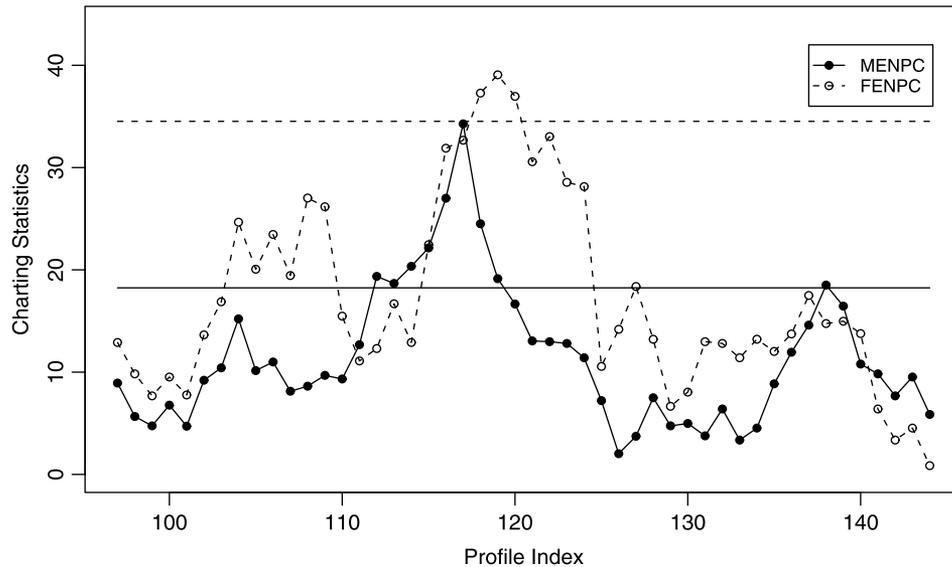


Figure 3. MENPC and FENPC control charts for monitoring the AEC process. The solid and dashed horizontal lines indicate their control limits, respectively.

profiles may also change over time. Such a change may or may not occur simultaneously with the shift in g . Online detection of possible changes in the variance–covariance structure of the profiles is not trivial, and we leave it for our future research. In addition, in some applications, we might be interested in monitoring a multivariate relationship between a response variable and several predictors over time. At this moment, we are not aware of any existing research on this topic, and we leave it to our future research to generalize the proposed control chart discussed in this article to multivariate cases.

SUPPLEMENTAL MATERIALS

Proofs: This pdf file provides certain technical details, including proofs of Proposition 1 and Theorems 1 and 2 in Section 2. (supplement.pdf)

ACKNOWLEDGMENTS

The authors thank the editor, the associate editor, and two referees for many constructive comments and suggestions, which greatly improved the quality of the article. This research is supported in part by grant DMS-0721204 from NSF of U.S.A. and grants 10771107 and 10711120448 from NNSF of China.

[Received October 2008. Revised August 2009.]

REFERENCES

- Altman, N. S. (1990), “Kernel Smoothing of Data With Correlated Errors,” *Journal of the American Statistical Association*, 85, 749–759. [267]
- Colosimo, B. M., and Pacella, M. (2007), “On the Use of Principal Component Analysis to Identify Systematic Patterns in Roundness Profiles,” *Quality and Reliability Engineering International*, 23, 707–725. [265]
- Davidian, M., and Giltinan, D. M. (1995), *Nonlinear Models for Repeated Measurement Data*, London: Chapman & Hall. [267]
- Diggle, P. J., Liang, K. Y., and Zeger, S. L. (1994), *Analysis of Longitudinal Data*, Oxford: Oxford University Press. [266]
- Ding, Y., Zeng, L., and Zhou, S. (2006), “Phase I Analysis for Monitoring Non-linear Profiles in Manufacturing Processes,” *Journal of Quality Technology*, 38, 199–216. [265]
- Fan, J., and Gijbels, I. (1996), *Local Polynomial Modeling and Its Applications*, London: Chapman & Hall. [267,268,271]
- Fan, J., and Marron, S. (1994), “Fast Implementation of Nonparametric Curve Estimators,” *Journal of Computational and Graphical Statistics*, 3, 35–56. [270]
- Fan, J., and Zhang, J. (2000), “Two-Step Estimation of Functional Linear Models With Applications to Longitudinal Data,” *Journal of the Royal Statistical Society, Ser. B*, 62, 303–322. [267]
- Fan, J., Zhang, C., and Zhang, J. (2001), “Generalized Likelihood Ratio Statistics and Wilks Phenomenon,” *The Annals of Statistics*, 29, 153–193. [268]
- Hall, P., and Marron, J. S. (1990), “On Variance Estimation in Nonparametric Regression,” *Biometrika*, 77, 415–419. [268]
- Hart, J. D. (1991), “Kernel Regression Estimation With Time Series Errors,” *Journal of the Royal Statistical Society, Ser. B*, 53, 173–187. [267]
- (1997), *Nonparametric Smoothing and Lack-of-Fit Tests*, New York: Springer. [266,270]
- Hawkins, D. M., and Olwell, D. H. (1998), *Cumulative Sum Charts and Charting for Quality Improvement*, New York: Springer-Verlag. [271]
- Hoover, D. R., Rice, J. A., Wu, C. O., and Yang, L. P. (1998), “Nonparametric Smoothing Estimates of Time-Varying Coefficient Models With Longitudinal Data,” *Biometrika*, 85, 809–822. [267]
- Jensen, W. A., and Birch, J. B. (2009), “Profile Monitoring via Nonlinear Mixed Models,” *Journal of Quality Technology*, 41, 18–34. [266,273]
- Jensen, W. A., Birch, J. B., and Woodall, W. H. (2008), “Monitoring Correlation Within Linear Profiles Using Mixed Models,” *Journal of Quality Technology*, 40, 167–183. [266,273]
- Kang, L., and Albin, S. L. (2000), “On-Line Monitoring When the Process Yields a Linear Profile,” *Journal of Quality Technology*, 32, 418–426. [265]
- Kazemzadeh, R. B., Noorossana, R., and Amiri, A. (2008), “Phase I Monitoring of Polynomial Profiles,” *Communications in Statistics: Theory and Methods*, 37, 1671–1686. [265]
- Kim, K., Mahmoud, M. A., and Woodall, W. H. (2003), “On the Monitoring of Linear Profiles,” *Journal of Quality Technology*, 35, 317–328. [265]
- 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, 79–90. [265]
- Laird, N. M., and Ware, J. H. (1982), “Random Effects Models for Longitudinal Data,” *Biometrics*, 38, 963–974. [266]
- Lin, X., and Carroll, R. J. (2000), “Nonparametric Function Estimation for Clustered Data When the Predictor Is Measured Without/With Error,” *Journal of the American Statistical Association*, 95, 520–534. [267,268]
- Lucas, J. M., and Saccucci, M. S. (1990), “Exponentially Weighted Moving Average Control Scheme Properties and Enhancements,” *Technometrics*, 32, 1–29. [271]
- Mahmoud, M. A., and Woodall, W. H. (2004), “Phase I Analysis of Linear Profiles With Calibration Applications,” *Technometrics*, 46, 380–391. [265]

- Mahmoud, M. A., Parker, P. A., Woodall, W. H., and Hawkins, D. M. (2007), "A Change Point Method for Linear Profile Data," *Quality and Reliability Engineering International*, 23, 247–268. [265]
- Rice, J. A., and Wu, C. O. (2001), "Nonparametric Mixed Effects Models for Unequally Sampled Noisy Curves," *Biometrics*, 57, 253–259. [266]
- Seifert, B., Brockmann, M., Engel, J., and Gasser, T. (1994), "Fast Algorithms for Nonparametric Estimation," *Journal of Computational and Graphical Statistics*, 3, 192–213. [270]
- Shi, M., Weiss, R. E., and Taylor, J. M. G. (1996), "An Analysis of Paediatric CD4 Counts for Acquired Immune Deficiency Syndrome Using Flexible Random Curves," *Applied Statistics*, 45, 151–163. [266]
- Walker, E., and Wright, S. P. (2002), "Comparing Curves Using Additive Models," *Journal of Quality Technology*, 34, 118–129. [266]
- Wang, Y. (1998), "Smoothing Spline Models With Correlated Random Errors," *Journal of the American Statistical Association*, 93, 341–348. [267]
- Williams, J. D., Birch, J. B., Woodall, W. H., and Ferry, N. M. (2007), "Statistical Monitoring of Heteroscedastic Dose-Response Profiles From High-Throughput Screening," *Journal of Agricultural, Biological, and Environmental Statistics*, 12, 216–235. [265]
- Williams, J. D., Woodall, W. H., and Birch, J. B. (2007), "Statistical Monitoring of Nonlinear Product and Process Quality Profiles," *Quality and Reliability Engineering International*, 23, 925–941. [265,266,273]
- Woodall, W. H. (2000), "Controversies and Contradictions in Statistical Process Control," *Journal of Quality Technology*, 32, 341–350. [265]
- Woodall, W. H., Spitzner, D. J., Montgomery, D. C., and Gupta, S. (2004), "Using Control Charts to Monitor Process and Product Quality Profiles," *Journal of Quality Technology*, 36, 309–320. [265]
- Wu, H., and Zhang, J. (2002), "Local Polynomial Mixed-Effects Models for Longitudinal Data," *Journal of the American Statistical Association*, 97, 883–897. [266-268,270,271,275]
- Zhang, D., Lin, X., Raz, J., and Sowers, M. (1998), "Semiparametric Stochastic Mixed Models for Longitudinal Data," *Journal of the American Statistical Association*, 93, 710–719. [267]
- Zou, C., Tsung, F., and Wang, Z. (2007), "Monitoring General Linear Profiles Using Multivariate EWMA Schemes," *Technometrics*, 49, 395–408. [265, 266,273]
- (2008), "Monitoring Profiles Based on Nonparametric Regression Methods," *Technometrics*, 50, 512–526. [265,271,273,274]
- Zou, C., Zhang, Y., and Wang, Z. (2006), "Control Chart Based on Change-Point Model for Monitoring Linear Profiles," *IIE Transactions*, 38, 1093–1103. [265]

Comment

Daniel W. APLEY

Department of Industrial Engineering
and Management Sciences
Northwestern University
Evanston, IL 60208-3119
(apley@northwestern.edu)

I congratulate the authors on a welcome addition to the profile monitoring literature, particularly regarding how to account for within-profile correlation. I cannot imagine a set of real profile data that does not have within-profile correlation, at least not for $\{x_{ij}: j = 1, 2, \dots, n_i\}$ that are densely enough spaced to be consistent with the spirit of profile data. Although within-profile correlation may have little adverse effect when fitting linear or simple parametric nonlinear models, as noted by Jensen, Birch, and Woodall (2008), it is inherently more insidious when fitting nonparametric models that are based on local smoothing. Profile variation due to a random component that is correlated within-profile looks deceptively like local, non-random changes in the mean if the model assumes no within-profile correlation. It is, therefore, likely to cause excessive false alarms. Although many authors noted the pervasiveness of within-profile correlation, surprisingly few developed algorithms that take this into account. Notable exceptions are the mixed model approach of Jensen, Birch, and Woodall (2008) and Jensen and Birch (2009) and the spatial autoregressive approach of Colosimo, Semeraro, and Pacella (2008). In light of this, I think the present work will be a welcome addition for practitioners who wish to monitor nonlinear profiles that are too irregularly shaped to be modeled parametrically.

I can find very little to criticize, but I would like to direct further scrutiny to three issues. The first two regard choosing the control limits to avoid excessive false alarms. These are hardly criticisms because I believe the authors' resampling procedure (with suitable modification) presents a nice solution to

this dilemma. The third is a broader issue that regards the nature of the assignable causes that are typically assumed in the profile monitoring literature.

BWARE OF ASYMPTOTIC RESULTS

In general, asymptotic results are often very useful in statistics. Take the central limit theorem, for example. One reason it is so useful is that the conditions under which we can approximate nonasymptotic reality using the asymptotic results of the central limit theorem are often met in practice: Usually only moderate sample sizes are needed to approximate the distribution of the average of a random sample as normal, at least for the level of accuracy required in many applications.

Regarding the asymptotic results of Theorem 1, on the other hand, I have doubts that the conditions required for their approximate validity are satisfied for typical profile monitoring applications. One conclusion that the authors draw following Theorem 1 is that the asymptotic distribution of $T_{t,h,\lambda}$ is independent of the "nuisance" parameter $\gamma(x, x')$ when the condition of bounded $n_i h$ is met (which is always the case in practice). The implication seems to be that one may choose the

control limits independent of $\gamma(x, x')$ and achieve a false alarm probability that is insensitive to uncertainty in this nuisance parameter (just prior to Theorem 1, the authors state that it can “shed some light on practical design of the chart”). But $\gamma(x, x')$ is the covariance function of $f_i(x)$, which represents the component of profile variation that accounts for within-profile correlation. It seems almost obvious that within-profile correlation can strongly affect the false alarm probability, for the simple reason that I state in the first paragraph of this discussion. Tighter arguments follow by noting that Equation (9) implies that the estimated profile mean $\widehat{g}_{t,h,\lambda}(s)$ is a locally weighted average of the observed profile values y_{ij} within some kernel neighborhood of s . Certainly, the variance of a weighted average depends strongly on the correlation between the variables that are being averaged. Because the weights are likely to be mostly positive, the variance of $\widehat{g}_{t,h,\lambda}(s)$ will be much larger if $\gamma(x, x')$ represents high positive correlation between $f(x)$ and $f(x')$ for x and x' both within the neighborhood of s than if $\gamma(x, x')$ represents no within-profile correlation. A larger variance of $\widehat{g}_{t,h,\lambda}(s)$ will, in turn, increase the mean of $T_{t,h,\lambda}$, perhaps substantially. Indeed, if the distribution of $T_{t,h,\lambda}$ truly were approximately independent of $\gamma(x, x')$, then the primary motivation for this work (handling within-profile correlation) will disappear.

So what explains this seeming contradiction that the in-control distribution of $T_{t,h,\lambda}$ depends strongly on within-profile correlation in most practical scenarios but has no dependence whatsoever asymptotically? The explanation is that, generally speaking, it can be very difficult to take asymptotic results for complex models involving a long list of convoluted conditions and render them down to their practical implications. In the present case, a careful inspection of the list of regularity conditions in Appendix A reveals some unrealistic ones that explain the contradiction at hand and account for other questionable conclusions that one might be tempted to draw from Theorem 1. For example, part of Condition C8 is that the kernel bandwidth $h \rightarrow 0$. In other words, the neighborhoods over which the kernel-weighted smoothing for $\widehat{g}_{t,h,\lambda}(s)$ takes place must shrink down to infinitesimally small neighborhoods. This condition is clearly at the heart of the invariance of the asymptotic distribution of $T_{t,h,\lambda}$ to $\gamma(x, x')$. For infinitesimally small neighborhoods, more of the y_{ij} in Equation (9) come from different i [i.e., more averaging across *different* profiles—see the next paragraph for why the exponentially weighted moving average (EWMA) time window is infinitely long in Theorem 1] and fewer from different j (i.e., less averaging *within* profiles). Hence, the asymptotic conditions of Theorem 1 do not even correspond to using local smoothing. They correspond to simply averaging y_{ij} over the time index i at a fixed spatial index j when estimating $g(\cdot)$. This is why the asymptotic results of Theorem 1 have the distribution of $T_{t,h,\lambda}$ independent of $\gamma(x, x')$.

Of course, this has little practical relevance because one will never choose infinitesimally small spatial neighborhoods and infinitely large time windows. It will be interesting to see how $\gamma(x, x')$ affects the distribution of $T_{t,h,\lambda}$ for typical values of h and λ commensurate with those recommended in Section 2.5.

A related concern is that Condition C8-II implies that both $\lambda \rightarrow 0$ and $t \rightarrow \infty$, if n_i is bounded (which is always the case in practice, no matter how large it is). This is entirely unrealistic because it implies that the EWMA is using an infinitely

long time window. In practice, to remain responsive to sudden large shifts, moderate length time windows (e.g., corresponding to $0.05 \leq \lambda \leq 0.2$) are usually chosen. As I discuss in the preceding paragraphs, the infinitely long time window (coupled with the infinitesimally small spatial neighborhood) explains why the asymptotic distribution of $T_{t,h,\lambda}$ in Theorem 1 is invariant to within-profile correlation. It also explains the asymptotic normality of $T_{t,h,\lambda}$ in Theorem 1. I suspect that for typical choice of λ , the distribution of $T_{t,h,\lambda}$ is quite positively skewed (as are distributions of many quadratic forms).

My main point here is that readers will be well advised not to assign too much significance to Theorem 1 and certainly not to select control limits based on the asymptotic distribution of $T_{t,h,\lambda}$ that Theorem 1 implies. Fortunately, readers do not have to because the authors included a very attractive resampling approach for calculating appropriate control limits. My next comment is related to this.

TEMPORAL AUTOCORRELATION AND FALSE ALARMS

The availability of spatially dense (i.e., large n_i) profile data is often the result of sophisticated, automated measurement technology. But this also often results in temporally dense profile data, for which a large number of profiles are collected over a relatively short period of time. Temporally dense profile data are likely to have temporal autocorrelation, in addition to the spatial autocorrelation represented by $\gamma(x, x')$.

It is well known that (positive) temporal autocorrelation causes a dramatic increase in the false alarm rate of many univariate control charts for detecting mean shifts. This is especially true for charts that have longer memory, such as a cumulative sum (CUSUM) with a small reference value or an EWMA with small λ (Apley and Lee 2003, 2008). Because of the manner in which the EWMA is involved in Equation (9), positive temporal autocorrelation will tend to increase the variance of $\widehat{g}_{t,h,\lambda}(s)$, thereby increasing the mean of $T_{t,h,\lambda}$. This follows from reasoning similar to that discussed in the preceding section in the context of spatial autocorrelation. An increase in the mean of $T_{t,h,\lambda}$ will, in turn, increase the false alarm rate, perhaps substantially.

Consequently, for many practical profile monitoring applications, I suspect that the control limits will have to be altered (widened) to account for the autocorrelation. We have simple time-series models for effectively representing temporal autocorrelation in the univariate control charting case, and these models offer means of appropriately altering the control limits. However, it seems doubtful that the model in Equation (1) can be augmented in a tractable yet realistic manner to represent temporal autocorrelation in the profile monitoring case. The authors' resampling approach in its present form would not result in properly widened control limits when autocorrelation is present, because the completely randomized profile resampling destroys any temporal structure in the data.

Fortunately, bootstrap resampling procedures can be modified to take into account temporal autocorrelation. Two main approaches for accomplishing this are the Markov bootstrap procedure of Paparoditis and Politis (2002) and the block bootstrap

procedure of Künsch (1989). In the Markov bootstrap procedure, one assumes the autocorrelation can be described in terms of a vector of random variables that follows a vector Markov process with unknown transition probability function. Essentially, one uses a form of kernel density estimation to fit the transition probability function, which one then uses to govern the sequential resampling in a manner that attempts to preserve the autocorrelation. In the block bootstrap procedure, which entirely avoids the need to model the autocorrelation, one will resample the profiles in blocks, instead of one-at-a-time. Each block will consist of a temporally contiguous set of profiles, the first of which is chosen randomly. Shan and Apley (2008) provided a more detailed description of the two procedures for a related problem.

The main drawback of the Markov bootstrapping procedure is that one must identify a Markov vector of variables that can represent the autocorrelation. The vector must have a dimension low enough to allow kernel density estimation of the transition probability function (e.g., one- or two-dimensional), and the transition probability function must be reasonably close to Markov. It is doubtful that the nonparametric structure of Equation (1) will yield a suitable low-dimensional Markov vector that can realistically account for the temporal autocorrelation in typical profile data. The main drawback of the block bootstrap procedure is that the total number of profiles [m in Equation (1)] must be relatively large. The length of each individual block should be only a fraction of m (e.g., less than 10%), while each individual block should be long enough to allow the dynamics due to the autocorrelation within each block to dominate the discontinuities between blocks. The value $m \geq 500$ recommended in Section 2.5 might be sufficient for the block bootstrap procedure in most cases.

In light of this, the block bootstrap procedure will most likely be more appropriate than the Markov bootstrap procedure for setting the control limits when monitoring temporally autocorrelated profiles. For the authors' real-data application, it will be interesting to see if temporal autocorrelation were present. This can be easily assessed graphically by choosing a few values of s , and then constructing time series charts of $\widehat{g}(s) + \widehat{f}_i(s)$ versus i for $i = 1, 2, \dots, m$. If temporal autocorrelation appears substantial, it will also be interesting to see if the block bootstrap procedure results in wider control limits than the regular bootstrapping procedure.

FOR WHAT SHOULD WE BE LOOKING?

The objective in this article, as well as in most of the profile monitoring literature, is to detect a change in the profile mean function $g(\cdot)$. To borrow statistical process control (SPC) terminology, in the model $y_{ij} = g(x_{ij}) + f_i(x_{ij}) + \varepsilon_{ij}$, the authors view changes in $g(\cdot)$ as assignable causes of variation and $f_i(x_{ij}) + \varepsilon_{ij}$ as common causes of variation. For many applications, I imagine changes in $g(\cdot)$ are important (perhaps the most important) indicators of assignable causes, and more generally, of problems with the process that should be detected and corrected. On the other hand, there are many applications in which assignable

causes are manifested as something other than a change in the mean.

Within the realm of the model in Equation (1), the next obvious characteristic to monitor is a change in the covariance $\gamma(x, x')$. Jin and Shi (1999) considered stamping tonnage profile monitoring (each profile is the press tonnage signature for one cycle, corresponding to one stamped part) and provide an illuminating discussion of a number of typical assignable causes and the effects they have on the profiles. Although some are manifested as changes in the profile mean, many others will be more reasonably represented as increased variation in $f_i(x)$ at one or more x , which corresponds to an increase in $\gamma(x, x)$. The authors of the present article mention detecting changes in $\gamma(x, x')$ in their conclusions, but leave it as future work due to its nontrivial nature. However, detecting certain changes in $\gamma(x, x')$ may, in fact, require only a trivial modification of their algorithm. Because of the quadratic nature of $T_{t,h,\lambda}$, I suspect that their algorithm will be reasonably effective at detecting variance changes [i.e., changes in $\gamma(x, x)$] if we forego the EWMA by using $\lambda = 1$ when estimating $\widehat{g}_{t,h,\lambda}(s)$. In addition to using $\lambda = 1$ when estimating $\widehat{g}_{t,h,\lambda}(s)$, we might impart memory in a different way by incorporating exponential weighting directly into the equation for $T_{t,h,\lambda}$. Specifically, using the authors' notation, we might consider

$$\sum_{i=0}^{t-1} (1 - \rho)^i T_{t-i,h,1},$$

as a control chart statistic, where $0 < \rho \leq 1$ denotes another EWMA parameter, and the third subscript on $T_{t,h,1}$ indicates that we use $\lambda = 1$ when estimating $\widehat{g}_{t,h,\lambda}(s)$. We can view this as analogous to an exponentially weighted moving variance.

I doubt that monitoring for generic changes in the covariance structure $\gamma(x, x')$ beyond the variance $\gamma(x, x)$ will be fruitful. In analogy with T^2 control charts for high-dimensional multivariate data (viewing each profile as a vector), for real datasets I envision such a chart plagued by alarms caused by innocuous changes in the correlation between $f_i(x)$ and $f_i(x')$ with no appreciable change in their variances. Even though these represent legitimate changes in the within-profile covariance, I believe most practitioners will prefer to view them as nuisance alarms.

Regarding looking outside the realm of Equation (1) for assignable causes, from one perspective it is unnecessary, because the nonparametric nature of Equation (1) makes it almost completely generic. Indeed, any set of profiles can be represented as a mean function plus a deviation from the mean that is zero-mean (by definition) with some covariance function. Likewise, almost any change in the profiles that we might care about will result in a change in the mean and/or covariance. However, using an algorithm based on the generic model in Equation (1) may be far from the most effective way of detecting such changes. Referring again to the stamping example of Jin and Shi (1999), some of their identified assignable causes did result in changes in the profile mean, but they were very specific changes of known form (e.g., a reduction in the peak tonnage) that can be premodeled. This is analogous to knowing in advance the direction of a change in the mean vector in multivariate control charting. Taking such information into account will greatly enhance the power to detect such changes.

Other assignable causes in Jin and Shi (1999) resulted in oscillation of the tonnage signature that amounted to highly structured changes in $\gamma(x, x')$. Again, an algorithm that incorporates knowledge of the specific structure of the change can result in much more powerful detection. With limited, incomplete knowledge of the structure of the change, monitoring coefficients of a Fourier or wavelet representation of the profiles can sometimes be useful (see Chicken, Pignatiello, and Simpson 2009, and the references therein).

In general, premodeling potential assignable causes and their effects on the profiles may be quite difficult for many applications, requiring advanced engineering knowledge and resources. It will be useful to have better “Phase I” exploratory data analysis tools for discovering and empirically modeling the effects of typical assignable causes based on large historical sets of profiles, over which various assignable causes occurred. It will also be useful to have an approach that looks specifically for a small set of patterns that might be easily premodeled, while simultaneously monitoring for more general profile changes via a $T_{t,h,\lambda}$ -like statistic. Apley and Lee (2010) developed a related approach for multivariate process data, but this will be difficult to extend to profile data.

I will close by thanking the authors for a thought-provoking article and a useful approach that I hope will find its way into SPC practitioners’ toolboxes. I would also like to thank the editor, David Steinberg, for recognizing the merit of their work and inviting these discussions.

Comment

Hugh A. CHIPMAN

Department of Mathematics and Statistics
Acadia University
Wolfville, NS B4P 2R6
Canada
(hugh.chipman@acadiau.ca)

We thank the editor for the opportunity to be discussants and congratulate the authors on a stimulating article.

Profile monitoring is an area of growing interest and importance. The authors develop a methodology that meets many of the needs of practitioners. They propose a flexible model based on a solid statistical foundation. Nonparametric local regression methods and random effects form the core of their approach. The random effects provide a convenient way of modeling covariance between responses observed at different points along the curve, a common feature of functional data. The procedure is quick in Phase II and appears to readily adapt to a variety of profile shapes.

To organize our discussion, we attempt to outline a list of desirable attributes and questions we can ask of a profile monitoring methodology. After describing each, we examine Qiu, Zou, and Wang in the context of that attribute or those ques-

ACKNOWLEDGMENT

This work was partially supported by the National Science Foundation under grant CMMI-0826081.

ADDITIONAL REFERENCES

- Apley, D. W., and Lee, H. C. (2003), “Design of Exponentially Weighted Moving Average Control Charts for Autocorrelated Processes With Model Uncertainty,” *Technometrics*, 45 (3), 187–198. [278]
- (2008), “Robustness Comparison of Exponentially Weighted Moving Average Charts on Autocorrelated Data and on Residuals,” *Journal of Quality Technology*, 40 (4), 428–447. [278]
- Apley, D. W., and Lee, H. Y. (2010), “Simultaneous Identification of Premodeled and Unmodeled Variation Patterns,” *Journal of Quality Technology*, 42, 36–51. [280]
- Chicken, E., Pignatiello, J. J., and Simpson, J. R. (2009), “Statistical Process Monitoring of Nonlinear Profiles Using Wavelets,” *Journal of Quality Technology*, 41 (2), 198–212. [280]
- Colosimo, B. M., Semeraro, Q., and Pacella, M. (2008), “Statistical Process Control for Geometric Specifications: On the Monitoring of Roundness Profiles,” *Journal of Quality Technology*, 40 (1), 1–18. [277]
- Jin, J., and Shi, J. (1999), “Feature-Preserving Data Compression of Stamping Tonnage Information Using Wavelets,” *Technometrics*, 41 (4), 327–339. [279,280]
- Künsch, H. R. (1989), “The Jackknife and the Bootstrap for General Stationary Observations,” *The Annals of Statistics*, 17 (3), 1217–1241. [279]
- Papadimitis, E., and Politis, D. N. (2002), “The Local Bootstrap for Markov Processes,” *Journal of Statistical Planning and Inference*, 108, 301–328. [278]
- Shan, X., and Apley, D. W. (2008), “Blind Identification of Manufacturing Variation Patterns by Combining Source Separation Criteria,” *Technometrics*, 50 (3), 332–343. [279]

R. Jock MACKAY and Stefan H. STEINER

Department of Statistics and Actuarial Science
University of Waterloo
Waterloo, ON N2L 3G1
Canada

tions. Before presenting our list, we briefly discuss a motivating example.

Example. To help fix ideas and provide a broader basis for discussing desirable attributes, we briefly describe a profile monitoring problem familiar to us. Mosesova (2007) provides additional details. The data arise from a manufacturing process in which a ram force-fits a steel valve seat into an aluminum cylinder head. Every insertion yields a force–time profile, three of which are displayed in Figure 1. In this particular process, a feedback controller adjusts the force in an attempt to maintain constant ram velocity during insertion. After an initial rise in

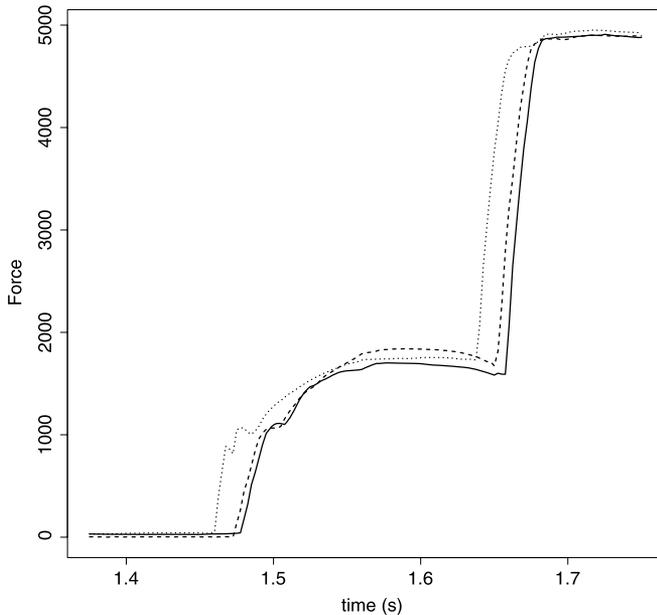


Figure 1. Three force–time profiles.

force corresponding to contact of the ram and the valve seat, the insertion force remains roughly constant as the seat is inserted. Once fully inserted, the force is increased in an attempt to maintain constant velocity. Every head has four cylinders, each with an intake and exhaust valve. The insertions displayed in Figure 1 correspond to three consecutive insertions of the intake valve in the same cylinder. Data are available on all eight valves for thousands of heads, ordered by time and date of manufacture.

1. FLEXIBILITY OF PURPOSE

The general goal of any process monitoring methodology is to detect changes that stand out above the common cause variation. Profiles can change in many ways, and ideally, the methodology can be adapted to be sensitive to prescribed changes. We may want to:

(a) Detect changes in particular features of the profile such as the maximum value, the location of the maximum, or the time point at which a specified event occurs (e.g., force begins increasing from 0 in Figure 1).

(b) Detect changes away from the “normal” profile toward one of several prespecified “bad” profiles. This might be accomplished via measures of closeness to representative profiles, or specification of a model in which some parameters identify departures toward the bad profiles.

(c) Detect unspecified changes in the mean profile.

(d) Detect changes in the variation (or covariance) of the residual profiles. This variation can be either functional (“wiggle”) or noise (background randomness).

(e) Detect both persistent changes and single outlying profiles.

Purposes (c), (d), and (e) arise in conventional monitoring applied to a single response variate. Purposes (a), (b), and the

idea of functional variation in (d) are unique to profile monitoring and arise from the functional nature of the data. Qiu, Zou, and Wang focus on purpose (c). Their methodology is designed to detect persistent changes of the mean profile. In Phase I, the authors obtain an estimate of the in control (IC) mean profile. In Phase II, at each observation point, they obtain an estimate of the current mean profile using an exponentially weighted moving average (EWMA) scheme combined with a local linear kernel estimate that allows for nonconstant variance at each point along the profile as estimated in Phase I. The monitoring statistic in Equation (11) is based on the differences between the estimates of the current and the IC mean profiles.

Although all the ingredients of process monitoring are present in the proposed chart, they are assembled in a non-standard way. A more conventional approach is to calculate a discrepancy measure for each profile, and then use EWMA (or another charting method) to combine the discrepancies. For example, if we have a new profile \mathbf{y}_i then we can define the discrepancy $(\mathbf{y}_i - \mathbf{g}_0(x_i))^T \Sigma_i^{-1} (\mathbf{y}_i - \mathbf{g}_0(x_i))$, where Σ_i is the covariance matrix for \mathbf{y}_i calculated using the results from Phase I. The embedding of the EWMA in the estimation process will make it more difficult to swap EWMA for other kinds of charting, such as cumulative sum (CUSUM) or Shewart charts. We feel it will be difficult to adapt the Qiu, Zou, and Wang approach to detecting single outlying profiles. In addition, practitioners may be more willing to use a new monitoring method if the elements of that method resemble existing strategies.

The need for multiple charts for different purposes is common in process monitoring where we are looking to detect changes other than persistent shifts in the mean, e.g., \bar{X} and s charts for a single characteristic. Shewart charts have greater power to detect a single outlying observation, while EWMA or CUSUM’s are good for quickly detecting relatively small persistent changes or drifts. Ideally, in any methodology, there should be flexibility to detect a variety of possible process changes. The statistic being charted can be designed for specified departures from the IC condition and simultaneous charting used for combinations of departures. An important question (beyond the scope of this discussion) is to determine when the unconventional method proposed by the authors is better than the more conventional approach we describe.

Although the proposed method focuses on purpose (c), one may ask whether it can be modified to detect changes in (a) specific features, (b) departures in specified directions, or (d) unspecified changes in variation. Note, as suggested by the authors, inclusion of the weight function $\Gamma_1(s)$ in the monitoring statistic $T_{t,h,\lambda}$ in Equation (11) allows for increased sensitivity to detect changes in specified sections of the mean profile that correspond to features of interest. To detect specified departures, the mean Phase I curve \mathbf{g}_0 can be replaced by “bad” curves in the calculation of Equation (11), although this will require a change in control limits. An out-of-control process will be flagged by profiles close to the “bad” baseline. It is not clear how to adapt the proposed methodology to detect changes in the covariance structure (d).

2. FLEXIBILITY OF APPLICATION

Process monitoring is an inherently applied discipline. A successful profile monitoring method will see widest application if it can be adapted to a wide variety of contexts:

- (a) Does the method require retooling for different profile shapes?
- (b) What happens if the data are collected in subgroups?
- (c) Within each profile, must observations be made at equally spaced time points? Must different profiles be observed at the same time points?
- (d) Do the profiles need registration? For instance, the curves in Figure 1 cannot be easily monitored until they are aligned by an affine transformation of the “time” axis. Similar registration of the vertical (e.g., force) axis may also be required. In some circumstances, nonlinear time warping functions (Ramsay and Silverman 2005) may be required to align multiple points of interest along curves.
- (e) How much of the procedure can be automated? Is computation in Phase II quick? In some applications, the data stream might be huge and fast, and even setting up the chart might require automation.
- (f) Are there automatic or semiautomatic choices of tuning constants (e.g., EWMA weight λ or kernel bandwidth h)?
- (g) Are covariates observed that will affect each curve? For instance, in our application, there can be cylinder and valve effects. While eight separate analyses (four cylinders by two valve types) can be carried out, a combined model with covariate effects (e.g., additive shifts for cylinder number and valve type) may increase power by borrowing strength across multiple data streams. In general, covariates can be fixed for each curve (as in our valve seat insertion example), or vary over time as the curve is observed.

The methodology of Qiu, Zou, and Wang does well at (a) and (c). The nonparametric curve estimation is very flexible, does not require equally spaced data, and should be applicable to any shape of curves. Registration (d) is not discussed in the article, although we suspect the authors are implicitly assuming curves are registered. The authors pay special attention to the design of Phase II modeling, gaining computation speed (e) by dropping random effects from the model and employing quick updating formulae. The choice of tuning parameters (f) is discussed, though fine tuning may still be somewhat of an art form. Tuning constants are difficult to set automatically since they will depend on the nature of the out-of-control condition one wishes to detect. For instance, in Figure 1, the out-of-control condition might be the shape of “wobble” near time = 1.5 (requiring a small smoothing bandwidth) or the height of the flat section around time = 1.6 (requiring a large smoothing bandwidth). Practitioners may have little information about such a condition.

Subgrouping (b) is a common technique employed in univariate control charting. It also may be an issue in the AEC example in Qiu, Zou, and Wang where there was sampling of profiles from batches of AEC’s. In the Phase I modeling or the Phase II charting, there is no explicit recognition that within batch variation may be different than between batch variation.

Qiu, Zou, and Wang did not consider adjustment for covariates (f). Such adjustments are not common in conventional

monitoring. Extensions to this case will require that the locally linear model be augmented to include regression terms for the covariates with either fixed or random effects.

3. MODELING ASSUMPTIONS

All modeling requires assumptions, often to simplify computation or theoretical derivations, or to focus attention on aspects of the problem that are particularly relevant. In profile monitoring, three key questions are:

- (a) Is there heteroscedasticity at different time points within a curve?
- (b) Do correlations exist between measurements made at different time points on the same curve?
- (c) Do dependencies exist between different curves?

Qiu, Zou, and Wang model heteroscedasticity (a) in both Phases I and II. In Phase I, they allow for within-profile correlation (b) via a random effects term. However, the correlations are not used explicitly for monitoring, as random effects are dropped from the Phase II model. Also, the weighted local likelihood before Equation (9) uses only variances (i.e., diagonals of the covariance matrix). Will it be straightforward to replace the sum in Equation (11) by a quadratic form that includes an inverted covariance matrix? We believe detection power might be enhanced by explicitly accounting for such covariances in Phase II.

The authors make the standard assumption that profiles are independent over time (c). However, autocorrelation is common, especially if 100% inspection is employed. Profiles sampled within the same batch or close together in time are apt to be more alike than profiles sampled from different batches or far apart in time.

4. PHASE I ISSUES

The availability of in-control data for Phase I modeling is a key component of any monitoring methodology since it enables calibration of statistics that are to be used for detection of process changes in Phase II. Considerations in Phase I include:

- (a) Phase I calculations are done off-line providing plenty of modeling and computation time.
- (b) The data used Phase I must be sampled from an IC process to enable accurate calibration. Methods are needed to check the Phase I data for outliers or other anomalies that should be removed before calculating the control limits.
- (c) A combination of theory and analysis of Phase I data must provide control limits for use in Phase II.

Qiu, Zou, and Wang effectively exploit the availability of off-line IC data (a) for estimating the IC mean profile and for developing Phase II control limits. They assume that IC data are available, but as noted in their discussion, provide no methodology to identify anomalies (b). Without such tools, it is difficult to imagine implementing the proposed chart in Phase II.

Calculation of control limits (c) with specified IC average run lengths (ARL’s) is a key component of any monitoring procedure. Qiu, Zou, and Wang adopt an empirical approach that

requires a large IC Phase I dataset. In determining the IC ARL's, the authors need to remove the effect of the initial conditions for the EWMA. As the EWMA weight for τ profiles in the past is $(1 - \lambda)^\tau$ using $\tau = 30$ is unlikely to be sufficient. With $\lambda = 0.02$ (the smallest recommended value), $(1 - \lambda)^{30} = 0.55$, which is the weight of the initial value in the EWMA statistic.

In Table 1, the authors compare random and fixed effect modeling. The fixed effect model does not have the desired in-control ARL. To make the comparison fair, we see no reason why the control limits for the fixed effects approach cannot be set to achieve the desired IC ARL.

5. PROPERTIES OF THE PHASE II ALGORITHM

In Phase II, we see the fruits of our labor with a method that will signal when the process goes out-of-control. We require:

- (a) Simple and quick calculations as new profiles arrive.
- (b) Good detection properties for relevant departures (as described previously under flexibility of purpose).
- (c) Interpretability.

The authors demonstrate promising indications on all these criteria. The absence of random effect terms in Phase II of the model (as noted earlier) means that the Phase I and II models are different. We wonder whether such a difference will have

any impact on detection properties (b). The proposed method is interpretable, in that the EWMA-smoothed curve that signalled the departure can be directly compared to the IC mean profile. However, the complex form of the model will make it difficult to pinpoint the cause of a signal if it is not evident in the displayed curve.

6. CLOSING THOUGHTS

The need for profile monitoring is increasing due to the availability of high-resolution data from many processes. This stimulating article shows how flexible nonparametric statistical methods can be used in a specific profile monitoring framework. The approach of Qiu, Zou, and Wang has many essential attributes that we feel a profile monitoring methodology should have and promises extensions in many directions.

ADDITIONAL REFERENCES

- Mosesova, S. A. (2007), "Flexible Mixed-Effect Modelling of Functional Data, With Applications to Process Monitoring," Ph.D. dissertation, University of Waterloo, Dept. of Statistics and Actuarial Science, Waterloo, Canada. Available at <http://hdl.handle.net/10012/3104>. [280]
- Ramsay, J. O., and Silverman, B. W. (2005), *Functional Data Analysis*, New York: Springer. [282]

Comment

F. TSUNG

Department of Industrial Engineering
and Logistics Management
Hong Kong University of Science and Technology
Clear Water Bay, Kowloon
Hong Kong
(season@ust.hk)

I would like to congratulate Qiu, Zou, and Wang on an interesting and innovative article that addresses a fundamental profile monitoring problem in statistical process control (SPC). I think this is a timely discussion because there is an urgent need for SPC techniques in various industries (not only in manufacturing, but also in service) that can handle complex functional monitoring and surveillance on a real-time basis. The proposed methodology focuses on the single covariate case, but it should be possible to extend it to a more practical case with multiple covariates. In my discussion I will focus on the profile monitoring cases with high-dimensional multiple covariates.

Due to the technological progress in hardware and software, most companies and organizations record and process huge amounts of data about production, business transactions, and service operations. These data streams contain very useful information that can be extracted through data modeling, characterization, monitoring, and forecasting. To remain competitive,

it is important for organizations to develop enterprise systems that allow managers to characterize relationships among performance and variables and to detect and prevent abnormal activities in operation.

Statistical monitoring and surveillance was widely recognized as an important and critical tool for detecting and identifying abnormal behavior (Tsung, Zhou, and Jiang 2007). Conventional approaches such as using statistical process control (SPC) techniques for system monitoring and surveillance often assume that the state of a system can be represented by a single random variable or a random vector of low dimensionality. However, many systems are far more complicated and

their states may be characterized by high-dimensional profiles over time or by multiple predictor variables instead of by a single random variable (Wang and Tsung 2005). For example, in a telecommunications case provided by Wei Jiang, customer profiles usually consist of different types of information—geographical, contractual, products/services purchased, transactions, and so on—which are recorded either off-line or on-line. To manage various market risks, from customer churn to product migration, it is important to monitor customer profiles continuously to identify critical customer behaviors promptly so that marketing and managerial decisions can be made to mitigate emerging risks. Since customer profiles often have dozens of attributes, monitoring such profiles usually relies on domain knowledge to keep track of only a few key performance indexes (KPI's) or on other dimension-reduction techniques to capture the simple relationships of the attributes. While such relationships are difficult to capture due to human variability and critical information may be lost when reducing dimensions of the profiles, industries are interested in applying statistical monitoring and surveillance tools with full access to all profile information simultaneously. Similar examples can be found in credit card and insurance fraud detection in which a collection of thousands of variables and transactions is closely monitored daily, and also in public health surveillance where statistical tools were developed for timely detection and prevention of various types of adverse health events so that health care policies and tactics can be initiated promptly.

In most of these examples, conventional SPC monitoring techniques are not directly applicable for the following reasons: The high-dimensionality and the large scale of the data being monitored make theories based on conventional estimation and testing methods inapplicable or of limited use for real-time, high-dimension statistical computing. Second, the basic hypothesis testing theory in SPC methods is to detect simple shifts in a process mean or variance under normality assumptions, which is reasonable in many manufacturing applications. However, in monitoring complex, high-dimensional data, the challenge is to detect complicated systematic changes from a huge number of data streams that may include many discrete or nonnormal variables. Moreover, many monitoring methods are developed based on the assumption that historical data are able to fit a statistical model that captures the relationships among variables. Due to the high dimensionality, it is often prohibitive to fit a reasonably good model to characterize the relationships among variables. And the relationships may shift and drift over time. Thus, it is essential yet challenging in practice to have an effective monitoring and surveillance system that can make use of dynamic incoming data streams to update anticipated drifts of the system and at the same time detect unanticipated shifts for corrective action.

To address such a challenge, we may consider extending modern variable selection techniques such as the least absolute shrinkage and selection operator (LASSO; Tibshirani 1996) and least angle regression (LARS; Efron et al. 2004) to deal with high-dimensional profile monitoring. These variable selection methods were applied extensively and successfully to various high-dimensional regression problems in genomics research such as gene expression and proteomics studies, biomedical imaging, functional MRI, tomography, tumor classification,

signal processing, and image analysis. They can improve estimation accuracy by effectively identifying the subset of important predictors and also enhance model interpretability with parsimonious representation. Recently, Wang and Jiang (2009) and Zou and Qiu (2009) independently proposed two variable-selection-based control charts. The authors developed the charts based on the following assumption: in a high-dimensional process, the probability that all variables shift simultaneously is rather low. Instead, an alarm is more likely to be caused by a hidden source, which affects one or a small set of observable variables. Both of them consider the penalized likelihood functions based on the conventional multinormality assumption. Their fundamental difference is that Wang and Jiang (2009) used an L_0 penalty term, but Zou and Qiu (2009) considered a type of L_1 penalty function.

To further elaborate the idea, we propose a variable-selection-based nonparametric process monitoring approach based on an ongoing joint project with Wei Jiang and Changliang Zou. While parametric methods are useful in certain applications, questions will always arise about the adequacy of the distributional assumptions and about the potential impact of misspecification of distributions on charting performance. This problem is particularly severe in telecommunications and financial customer profile applications. To deal with this problem, we develop a nonparametric SPC methodology based on empirical likelihood (Owen 1988, 2001) and incorporate the variable selection feature into it. Here we only focus on monitoring the multiple covariates,

$$\mathbf{x}_j \stackrel{\text{iid}}{\sim} \begin{cases} F_0(\mathbf{x} - \boldsymbol{\mu}_0) & \text{for } j = 1, \dots, \tau \\ F_1(\mathbf{x} - \boldsymbol{\mu}_1) & \text{for } j = \tau + 1, \dots, \end{cases} \quad (1)$$

where $F_0 \neq F_1$ are the unknown in-control and out-of-control distribution functions that have unequal location parameters $\boldsymbol{\mu}_0 \neq \boldsymbol{\mu}_1$. Denoting $\boldsymbol{\delta}_j \equiv E(\mathbf{x}_j) - \boldsymbol{\mu}_0$, the monitoring problem is essentially equivalent to

$$\boldsymbol{\delta}_j = \begin{cases} \mathbf{0} & \text{for } j = 1, \dots, \tau \\ \boldsymbol{\delta} & \text{for } j = \tau + 1, \dots, \end{cases} \quad (2)$$

where $\boldsymbol{\delta}$ is an unknown shift vector. Without loss of generality, assume that $\boldsymbol{\mu}_0 = \mathbf{0}$.

We aim to introduce the weighted version of empirical likelihood (EL) and then use it to formulate the charting statistic by incorporating the exponentially weighted moving average (EWMA) scheme. At any time point t , consider the following weighted empirical likelihood (WEL) for $\boldsymbol{\delta}$ in Equation (2), evaluated at $\boldsymbol{\delta}_0 = \mathbf{0}$, which combines some prespecified weights with the EL framework

$$L_t(\mathbf{0}) = \sup \left\{ \prod_{i=1}^t p_i^{w_i} \mid \sum_{i=1}^t p_i = 1, \sum_{i=1}^t p_i \mathbf{x}_i = \mathbf{0}, p_i \geq 0 \right\},$$

where w_i 's are prechosen positive weights. Specifically, an ideal choice for w_i can be the exponential weighting scheme used in EWMA at different time points, i.e., $(1 - \lambda)^{t-i}$ where $0 \leq \lambda \leq 1$ is a smoothing parameter. Then, the following weighted empirical log-likelihood ratio (WELR) can be used as the charting statistic

$$l_t(\mathbf{0}) = -2b_t \log[L_t(\mathbf{0})/L_t(\hat{\boldsymbol{\delta}})], \quad (3)$$

where

$$L_t(\hat{\delta}) = \sup \left\{ \prod_{i=1}^t q_i^{w_i} \mid \sum_{i=1}^t q_i = 1, q_i \geq 0 \right\}, \quad (4)$$

and $b_t = \sum_{i=1}^t w_i / \sum_{i=1}^t w_i^2$ is a scaling constant to normalize the variation of $l_t(\mathbf{0})$. The Lagrange multiplier method leads to

$$l_t(\mathbf{0}) = 2b_t \sum_{i=1}^t w_i \log(1 + \theta^T \mathbf{x}_i),$$

where θ satisfies the score equation $\sum_{i=1}^t w_i \mathbf{x}_i / (1 + \theta^T \mathbf{x}_i) = \mathbf{0}$. Since it is a data-driven scheme, we can expect it to be more robust to various multivariate nonnormal data than the conventional multivariate exponentially weighted average (MEWMA; Lowry et al. 1992) chart in the in-control situation. Moreover, it avoids the estimation of the covariance matrix from the historical data by studentizing internally, and hence its in-control performance deteriorates less when the number of reference samples is relatively small, and its ability to detect location shifts is robust to changes in the variance–covariance matrix.

In cases with high dimensionality where it is reasonable to assume that only a few components in the shift vector $\delta = \mu_1$ will be nonzero when a shift occurs, there is potential for improving the foregoing WELR chart by integrating the empirical likelihood with some additional constraints. This amounts to redefining the WEL under H_1 by imposing an adaptive-LASSO (ALASSO) penalty in Equation (4),

$$L_t(\hat{\delta}) = \sup \left\{ \prod_{i=1}^t q_i^{w_i} \mid \sum_{i=1}^t q_i = 1, \sum_{i=1}^t q_i (\mathbf{x}_i - \delta) = \mathbf{0}, \sum_{k=1}^p \frac{1}{|\mathbf{U}_t^{(k)}|} |\delta^{(k)}| \leq s, q_i \geq 0 \right\}. \quad (5)$$

The corresponding WELR statistic can be defined in a similar fashion to Equation (3). Although the computation involved

in Equation (3) is trivial partly due to the fact that the modified Newton–Raphson algorithm used in the computation is efficient, this does not appear to be the case for Equation (5). Hence, that may require efforts to solve the optimization problem [Equation (5)] efficiently for on-line uses.

In summary, given the fact that billions of dollars are lost every year due to telecommunications fraud and credit card forgery, monitoring of complex profiles and quick detection of fraudulent events and abnormal activities have become critical in a variety of industries. The proposed nonparametric profile monitoring methodology is timely, and more research efforts are needed for monitoring business processes with large numbers of high-dimensional transactions and detecting fraudulent records among them. Finally, I would like to thank Qiu, Zou, and Wang, and the editor of *Technometrics*, David Steinberg, for the opportunity to comment.

ADDITIONAL REFERENCES

- Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004), “Least Angle Regression,” *The Annals of Statistics*, 32, 407–489. [284]
- Lowry, C. A., Woodall, W. H., Champ, C. W., and Rigon, S. E. (1992), “A Multivariate Exponentially Weighted Moving Average Control Chart,” *Technometrics*, 34, 46–53. [285]
- Owen, A. B. (1988), “Empirical Likelihood Ratio Confidence Intervals for a Single Functional,” *Biometrika*, 75, 237–249. [284]
- (2001), *Empirical Likelihood*, New York: Chapman & Hall/CRC. [284]
- Tibshirani, R. (1996), “Regression Shrinkage and Selection via the Lasso,” *Journal of the Royal Statistical Society, Ser. B*, 58 (1), 267–288. [284]
- Tsung, F., Zhou, Z. H., and Jiang, W. (2007), “Applying Manufacturing Batch Techniques to Fraud Detection With Incomplete Customer Information,” *IIE Transactions*, 39, 671–680. [283]
- Wang, K., and Jiang, W. (2009), “High-Dimensional Process Monitoring and Fault Isolation via Variable Selection,” *Journal of Quality Technology*, 41, 247–258. [284]
- Wang, K., and Tsung, F. (2005), “Using Profile Monitoring Techniques for a Data-Rich Environment With Huge Sample Size,” *Quality and Reliability Engineering International*, 21, 677–688. [284]
- Zou, C., and Qiu, P. (2009), “Multivariate Statistical Process Control Using LASSO,” *Journal of the American Statistical Association*, 104, 1586–1596. [284]

Comment

William H. WOODALL, Jeffrey B. BIRCH, and Pang DU

Department of Statistics
Virginia Tech
Blacksburg, VA 24061-0439
(bwoodall@vt.edu)

1. GENERAL COMMENTS

The authors are to be congratulated on their valuable contribution to the rapidly growing profile monitoring field. They offer the most general modeling approach in the area and address a number of Phase I and Phase II issues. Woodall (2007) provided a review of the profile monitoring area, but a number of articles have appeared since. In particular, Shiau et al. (2009)

also incorporated the use of nonparametric regression methods when there are random effects in Phase I and Phase II, but used an approach quite different from the one presented here.

2. PHASE I/PHASE II ISSUES

The authors focus on Phase II methods, although some estimation issues are addressed using Phase I data. We believe, however, that it is often much more difficult than the authors imply to obtain a “clean” set of profiles in Phase I to use as the in-control data. This difficulty will be compounded by the fact that a very large Phase I dataset of at least 500 profiles is recommended for use with their approach. We note that there was no Phase I analysis reported in their example with the dataset from the aluminum electrolytic capacitor manufacturing process. The first 96 of 144 profiles were used as the in-control sample, but without reporting the results of any screening for outlying profiles. It will be very informative for the authors to provide more information on the Phase I data, and if at all possible, make the entire dataset available online.

Woodall et al. (2004) pointed out that a fundamental issue in profile monitoring is to decide how much of the profile-to-profile variation is to be considered common cause variation. Use of the nonparametric mixed-effects model in Equation (1) incorporates profile-to-profile variation as common cause variation automatically. This becomes relevant in the results of the simulation study reported in Table 1 in Section 3. The mixed-effects nonparametric profile control (MENPC) approach incorporates the profile-to-profile variation into the model whereas the fixed-effects nonparametric profile control (FENPC) method does not. If one ignores any sources of variation in setting up the control chart, then there will be an increased number of signals. This is true even in the univariate case in which one monitors the mean of a process as shown, for example, by Woodall and Thomas (1995). This phenomenon was illustrated in a profile monitoring situation by Jensen and Birch (2009).

In general, when the in-control parameters are estimated in Phase I the performance of the chart in Phase II is not the same as it would be if the in-control parameters were known. The literature on this topic was reviewed by Jensen et al. (2006). The authors recommend a large Phase I dataset and determine the control limits using a large number of simulations, however, so the effect of estimation error may be negligible in their case.

In discussing their algorithm for Phase I estimation, the authors correctly point out that the frequency of nonconvergence can be reduced with good initial estimates of \mathbf{D} and σ^2 . Yet their initial estimator $\sigma_{(0)}^2$ of σ^2 given in their Step 1 (Section 2.2) utilizes the residuals resulting from using an initial fit of the reference profile based on local linear kernel regression (LLR). This clearly yields inflated estimates of σ^2 as the given expression is actually estimating the combined within-profile and between-profile variability and not just the within-profile variability. An improved estimator of σ^2 can be obtained by replacing $\widehat{g}^{(P)}(x_{ij})$ in the formula for $\sigma_{(0)}^2$ with the nonparametric estimator of the i th profile, sometimes referred to as the estimated i th cluster specific (\widehat{CS}_i) curve, where each \widehat{CS}_i is estimated using LLR.

3. EXPONENTIAL WEIGHTING IN PHASE II

The authors use exponential weighting of the profiles in their weighted negative log likelihood expression $WL(a, b;$

$s, \lambda, t)$. The monitoring statistic itself, $T_{t,h,\lambda}$, is of the Shewhart type without any weighting of past monitoring statistics. In $WL(a, b; s, \lambda, t)$, the weighting can be seen more easily to be the same as usually given in an exponentially weighted moving average (EWMA) chart if one simply multiplies the entire expression by λ . The authors recommend the use of λ values in the range $[0.02, 0.1]$. With the standard EWMA chart, however, use of low values of the weighting constant can lead to potential delays in detecting shifts in the process due to inertial effects. See, for example, Woodall and Mahmoud (2005). It seems that this can also be an issue with the proposed profile method if undetected profile shifts in one direction were to be followed by shifts in the opposite direction, where “direction” is used loosely here.

An alternative Phase II method can involve obtaining estimates of the t profiles by some appropriate mixed nonparametric method. Then one can combine these t estimated profiles via a typical EWMA weighting scheme with weight λ given to the most recent profile, $\lambda(1 - \lambda)$ to the previous profile, $\lambda(1 - \lambda)^2$ to the profile before that, and so on. This combined profile, say \widehat{g} , can then be compared with g_0 using a metric similar to the one in Equation (11) with the current $\widehat{\xi}_{ij}$ values replaced by $\widehat{\xi}(s) = [\widehat{g}(s) - g_0(s)]$. As a second alternative, metrics reflecting the difference between fitted curves and g_0 at each sampling stage can be monitored using an EWMA scheme.

4. TWO NONPARAMETRIC PHASE I APPROACHES

Some of our research showed that if \widehat{g} is the nonparametric estimator of the reference profile, sometimes referred to as the estimated population average (\widehat{PA}) profile, at n_0 locations of the regressors, and if $\widehat{g} + \widehat{f}_i$ is the estimated i th cluster specific (\widehat{CS}_i) curve, at the same n_0 locations of the regressors, then the Hotelling’s T^2 statistics of the form $T_i^2 = \widehat{\mathbf{f}}_i^t \widehat{\Sigma}^{-1} \widehat{\mathbf{f}}_i$ are quite adequate for use in a screening test for even relatively small values of n_i , m , and n_0 (see Abdel-Salam and Birch 2009). The matrix $\widehat{\Sigma}$, the n_0 by n_0 estimated variance matrix of $\widehat{\mathbf{f}}_i$, may be obtained in a variety of ways, including use of the standard moment estimator or the successive difference estimator (recommended for detecting sustained step shifts in profiles) as discussed in Jensen and Birch (2009), for example.

The Phase I approach taken by the authors is to adopt the local linear mixed effects (LLME) technique of Wu and Zhang (2002). The model is based on the assumption that the within-profile random errors are iid with constant variance. They assume the random effects have a more general covariance structure.

Two other approaches for nonparametric regression modeling of cluster-correlated data across m subjects (or profiles) were proposed and studied at Virginia Tech. One method, called the conditional mixed linear model, allows for local estimation of the Laird–Ware model by kernel weights for each profile (Waterman 2002 and Waterman, Birch, and Schabenberger 2007). The method is more general than the one considered by the authors in that it allows for different model matrices for the fixed and random effects while also allowing for fully correlated random errors and random effects for which local restricted maximum likelihood (REML) estimates can be obtained. This method was only applied to the estimation aspects

during Phase I, but our work shows excellent integrated mean square error properties for n and m as small as 10 and 20, respectively. Computational inefficiency results, however, for this method are due to choosing the bandwidth using an adjusted cross-validation procedure.

A second, more computationally efficient approach was studied extensively for Phase I profile monitoring. This procedure results by replacing the local linear mixed model using kernel weights as the nonparametric regression method for estimating the Laird–Ware model with the use of penalized splines (Abdel-Salam and Birch 2009). As with the Waterman (2002) method, this penalized spline approach is completely general and results in local REML estimates for all variance–covariance components. Profile screening for Phase I monitoring is accomplished through the Hotelling’s T_i^2 statistic computed using the estimated random effects (the eblups) for the i th profile that result from the penalized spline fit. Our simulation for Phase I screening shows excellent probability-of-signal results for this method when detecting sustained step shifts, especially when compared to the misspecified linear mixed model parametric approach of Jensen, Birch, and Woodall (2008). This method is directly comparable to the estimation method presented by the authors for Phase I analysis.

5. SOME FINAL COMMENTS

There are a few aspects of the proposed method for which we would like some additional explanation. In Phase I, the authors estimated the profile curve using the method developed in Wu and Zhang (2002) with bandwidth selected by their cross-validation procedure. In Phase II, due to the concern of computational cost, the authors used a different method similar to Lin and Carroll (2000) with an empirical bandwidth selection procedure based on Equation (14). It is not clear to us how to make sure that the estimates from these two different methods are consistent with each other. In other words, if Phase II also deals with IC data, do we expect the same convergence rate for the estimates under the same regularity and smoothness conditions?

In Proposition 1, the authors derive the point-wise convergence, or local convergence, for their Phase I estimator, which is a nice result. In general, however, a more interesting type of convergence in curve estimation is global convergence such as the convergence in mean integrated square error (MISE) in theorem 1 of Wu and Zhang (2002). Given the similarity of the Phase I estimation to Wu and Zhang (2002), we wonder why the authors chose not to derive a similar global convergence result. Particularly, the conditions, such as conditions (A.3) and (A.4) in Wu and Zhang (2002), required for such a result may help the authors to justify their choice of $n_i \geq 20$ and $m \geq 500$ for information criterion (IC) data in Section 2.5.

The authors stated that their local polynomial approach can be extended to cases with multiple predictors. Our limited experience seems to suggest that the extension of local polynomial estimation to the multivariate case, particularly when interactions are included, can be rather computationally involved. On the other hand, other smoothing techniques such as the smoothing spline analysis of variance (ANOVA) models in Gu (2002)

can handle multivariate estimation more conveniently. Can the authors shed some light on this issue?

Our final comment is directed toward the second simulation presented in the article. Here, the authors’ method is compared to the linear mixed-effect (LME) modeling approach of Jensen, Birch, and Woodall (2008) after modification by the authors for Phase II profile monitoring and referred to as the LMEP method. From Table 4, it is seen that for operating characteristic (OC) model (1), that the linear mixed-effect profile (LMEP) is very competitive with the authors’ method but uniformly worse for OC model (2). The comparison of average run length (ARL) values for OC model (2) is very misleading, however, in that the estimation procedure used by LMEP is a procedure that requires correct specification of the parametric model. That is clearly not true for this case as the OC model (2) is nonlinear. Consequently, the ARL values for LMEP reflect the adverse effect of model misspecification and strongly reflect the need to use a nonparametric regression method if the model used by the parametric procedure is in doubt. If the authors wished to compare their method to a parametric procedure, they should consider modifying the method suggested by Jensen and Birch (2009) to Phase II profile monitoring and use OC model (2) as their parametric model. In this case, the confounding effect of model misspecification will be removed and a better comparison with the authors’ procedure can be made.

ACKNOWLEDGMENT

This research was supported in part by National Science Foundation grant CMMI-0927323.

ADDITIONAL REFERENCES

- Abdel-Salam, A.-S., and Birch, J. B. (2009), “Profile-Monitoring Analysis With Fixed and Random Effects Using Nonparametric and Semiparametric Methods,” contributed paper presented at the *Joint Statistical Meetings*, August 2009, Washington, DC. [286,287]
- Gu, C. (2002), *Smoothing Spline ANOVA Models*, New York: Springer-Verlag. [287]
- Jensen, W. A., Jones-Farmer, L. A., Champ, C. W., and Woodall, W. H. (2006), “Effects of Parameter Estimation on Control Chart Properties: A Literature Review,” *Journal of Quality Technology*, 38 (4), 349–364. [286]
- Shiau, J. J. H., Huang, H.-L., Lin, S.-H., and Tsai, M.-Y. (2009), “Monitoring Nonlinear Profiles With Random Effects by Nonparametric Regression,” *Communications in Statistics—Theory and Methods*, 38, 1664–1679. [285]
- Waterman, M. J. (2002), “Linear Mixed Model Robust Regression,” Ph.D. dissertation, Dept. of Statistics, Virginia Polytechnic Institute and State University, VA. [286,287]
- Waterman, M. J., Birch, J. B., and Schabenberger, O. (2007), “Linear Mixed Model Robust Regression,” Technical Report 07-3, Virginia Tech, Dept. of Statistics, available at http://www.stat.org.vt.edu/dept/web-el/tech_reports/TechReport07-3.pdf. [286]
- Woodall, W. H. (2007), “Current Research in Profile Monitoring,” *Produção*, 17 (3), 420–425. [285]
- Woodall, W. H., and Mahmoud, M. A. (2005), “The Inertial Properties of Quality Control Charts,” *Technometrics*, 47 (4), 425–436. [286]
- Woodall, W. H., and Thomas, E. V. (1995), “Statistical Process Control With Several Components of Common Cause Variability,” *IIE Transactions*, 27 (6), 757–764. [286]

Rejoinder

Peihua Qiu,

School of Statistics
University of Minnesota
Minneapolis, MN 55455
(qiu@stat.umn.edu)

Changliang ZOU and Zhaojun WANG

LPMC
and
Department of Statistics
Nankai University
Tianjing 300071
China

Profile monitoring is a relatively new research area, but it has a profound application background (cf., Woodall et al. 2004; Wang and Tsung 2005). Due to the fact that the data structure in profile monitoring is much more complicated than that in conventional process monitoring problems, it is also a challenging task. In our method (called QZW hereafter), we try to apply some recent statistical tools developed in some other research areas, including longitudinal data and functional data analysis (e.g., Liang and Zeger 2002; Fitzmaurice et al. 2008), to the area of profile monitoring. As pointed out by the discussants, there are still some issues in our proposed method that need to be addressed in future research. In the next several sections, we provide our thoughts about some of the main issues raised by the discussants.

1. PHASE I AND PHASE II PROFILE MONITORING

Our article focuses on Phase II profile monitoring. Instead of assuming the in-control (IC) profile mean function and other parameters and functions (i.e., g , γ , and σ^2) to be known, we assume that there is an IC dataset from which g , γ , and σ^2 can be estimated. In practice, however, it is still a challenging task to obtain the IC dataset. In that regard, we agree with Woodall, Birch, and Du completely, and think that much future research is required on Phase I analysis of profile data.

In the limited literature on Phase I analysis of profile data, Jensen, Birch, and Woodall (2008) and Jensen and Birch (2009) developed procedures for monitoring linear and nonlinear profiles. Their nonlinear profile monitoring procedure in Jensen and Birch (2009) can be easily generalized to the nonparametric setup by using the nonparametric mixed-effects modeling described in Section 2.2 of QZW. Let

$$T_i^2 = \widehat{\mathbf{f}}_i^T \widehat{\Sigma}^{-1} \widehat{\mathbf{f}}_i, \quad (1)$$

where $\widehat{\mathbf{f}}_i = (\widehat{f}_i(s_1), \dots, \widehat{f}_i(s_{n_0}))^T$, $\{s_1, s_2, \dots, s_{n_0}\}$ are n_0 given points in the design interval of x [cf., Equation (1) in QZW], and $\widehat{\Sigma}$ is an estimated covariance matrix of $\widehat{\mathbf{f}}_i$. Then, the i th profile is an outlier if T_i^2 is larger than a threshold value. In the aluminium electrolytic capacitor (AEC) data example discussed in Section 4 of QZW, the first 96 profiles are treated as an IC dataset. We agree with Woodall, Birch, and Du that, in practice, it still needs to be checked whether there are any outliers in this data. To this end, we apply the above Phase I outlier detection procedure to this IC dataset, with $\widehat{\Sigma}$ chosen to be the successive difference estimator, as recommended by Jensen and Birch

(2009) for detecting sustained step shifts in profiles. Namely,

$$\widehat{\Sigma} = \frac{1}{2(m-1)} \sum_{i=1}^{m-1} (\widehat{\mathbf{f}}_{i+1} - \widehat{\mathbf{f}}_i)(\widehat{\mathbf{f}}_{i+1} - \widehat{\mathbf{f}}_i)^T.$$

Figure 1 presents T_i^2 , for $1 \leq i \leq 96$, along with a control limit corresponding to the significance level of 0.05 that is computed by a bootstrap procedure as follows. Each time we draw 96 \widehat{f}_i 's with replacement from $\{\widehat{f}_i, 1 \leq i \leq 96\}$ that are computed beforehand by the procedure described in Section 2.2 of QZW. Then, a bootstrap version of $\widehat{\Sigma}$ is computed from the resampled \widehat{f}_i 's, and 96 bootstrap observations of T_i^2 are computed from the resampled \widehat{f}_i 's and the corresponding bootstrap version of $\widehat{\Sigma}$. This process is repeated 10,000 times, and the control limit is defined to be the 95th percentile of all bootstrap observations of T_i^2 computed. From the plot, it can be seen that no outliers are detected by this procedure.

It should be pointed out that, for detecting outliers in Phase I profile monitoring, the T^2 control chart defined in Equation (1) may not be the most powerful one. In the context of longitudinal data analysis, a similar issue was investigated by Fung et al. (2002) who proposed certain influence diagnostics and outlier detection procedures using semiparametric mixed-effects modeling. For Phase I monitoring of nonparametric profiles, discussion about *nonparametric covariance analysis* and *comparison of multiple curves* in the context of nonparametric regression testing (cf., Young and Bowman 1995; Dette and Neumeyer 2001; Neumeyer and Dette 2003) might also be relevant.

We appreciate the comment made by Chipman, MacKay, and Steiner that the absence of random-effects terms in our Phase II modeling [cf., the expression of $WL(a, b; s, \lambda, t)$ in the second paragraph of Section 2.3 in QZW] may affect the efficiency of our proposed Phase II profile monitoring chart. Woodall, Birch, and Du raised a similar issue and they asked why we use the method by Wu and Zhang (2002) in Phase I modeling and the different method by Lin and Carroll (2000) in Phase II profile monitoring. As explained in Section 2.3 of QZW, the major reason for us to use two different methods in Phase I and Phase II analysis is that the computation involved in the iterative algorithm of the method by Wu and Zhang (2002) is quite substantial. For Phase I analysis in which the sample size is fixed that

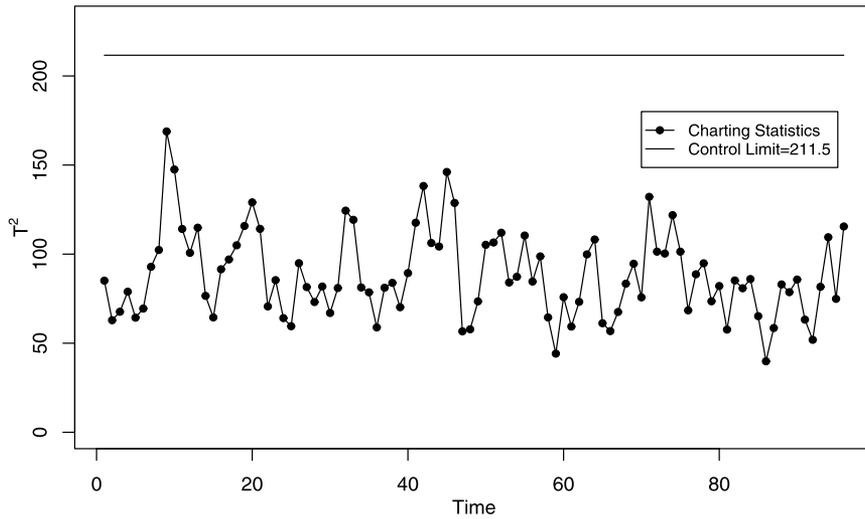


Figure 1. Phase I T^2 control chart defined in Equation (1) for monitoring the first 96 profiles of the AEC dataset.

algorithm is still feasible. But, for online Phase II monitoring, this method is cumbersome and may not be feasible. It is true that by using the method of Lin and Carroll (2000) in Phase II analysis [cf., Equations (9) and (10) in QZW], it appears that only the heteroscedasticity of the within-profile observations is taken care of and the within-profile correlation was not fully accommodated. However, according to Lin and Carroll (2000), under some regularity conditions, it will not have much of an effect on the estimated profile mean function to only accommodate the heteroscedasticity properly without specifying the complete correlation structure of the within-profile observations. To further investigate this issue, we run a simulation in cases when the IC model (II), the out-of-control (OC) models (i) and (ii), and $\lambda = 0.1$ are considered (cf., Section 3 of QZW for their definitions). In this example, besides our proposed chart mixed-effects nonparametric profile control (MENPC), we also consider the chart constructed as follows. Let $\hat{g}_{i,h,\lambda}^*(s)$ be the estimator of $g(s)$, obtained by the algorithm described in Section 2.2 of QZW, except that the expression in Equation (2) in QZW is replaced by

$$\sum_{i=1}^t \left\{ \frac{1}{\sigma^2} \sum_{j=1}^{n_i} [y_{ij} - \mathbf{z}_{ij}^T(\boldsymbol{\beta} + \boldsymbol{\alpha}_i)]^2 K_h(x_{ij} - s) + \boldsymbol{\alpha}_i^T \mathbf{D}^{-1} \boldsymbol{\alpha}_i + \ln |\mathbf{D}| + n_i \ln(\sigma^2) \right\} (1 - \lambda)^{t-i},$$

where $\lambda \in [0, 1]$ is a weighting parameter. Obviously, the above expression combines different profiles for Phase II monitoring using the exponentially weighted moving average (EWMA) weighting scheme. Then, a charting statistic can be constructed in a similar way to Equation (11) in QZW, after $\{y_{ij}\}$ are replaced by $\{\xi_{ij} = y_{ij} - g_0(x_{ij})\}$ in the above expression. This control chart is denoted as MENPC1, and it is based on the method by Wu and Zhang (2002) in both the Phase I and Phase II analysis. With all the procedure parameters chosen in the same way as those in Table 2 of QZW, the OC average run length (ARL) values of the charts MENPC and MENPC1 are presented in

Table 1. From the table, it can be seen that the two charts perform similarly in all cases considered, and the chart MENPC1 is slightly better in cases with OC model (ii).

Regarding the initial estimator $\sigma_{(0)}^2$ used in the iterative algorithm in Section 2.2 of QZW, Woodall, Birch, and Du suggested replacing $\hat{g}^{(P)}(x_{ij})$ in the formula for $\sigma_{(0)}^2$ given in the paragraph immediately before the expression in Equation (7) of QZW by $\hat{g}_i(x_{ij})$ where \hat{g}_i is the local linear kernel estimator of g that is constructed from the i th profile data alone. We tried this idea in some numerical examples and find that the modified initial estimator is indeed better.

2. TEMPORAL AUTOCORRELATION

In QZW, we only consider the possible correlation among within-profile observations, and assume that observations between profiles are independent of each other. We appreciate the comment made by Apley that temporal autocorrelation among profiles collected at consecutive time points might also be common in practice. We agree with Apley completely on this issue, and believe that it is an important future research problem to propose profile monitoring charts that can accommodate both

Table 1. OC ARL comparison of the charts MENPC and MENPC1 when $ARL_0 = 200$, $n = 20$, $n_0 = 40$, $\lambda = 0.1$ and IC model (II) is used

θ	OC model (i)		OC model (ii)	
	MENPC	MENPC1	MENPC	MENPC1
0.20	130 (1.36)	134 (1.22)	85.3 (0.83)	84.8 (0.86)
0.30	80.5 (0.78)	77.2 (0.80)	40.5 (0.32)	37.4 (0.33)
0.40	48.6 (0.42)	46.5 (0.44)	22.3 (0.15)	20.6 (0.16)
0.60	20.7 (0.13)	19.8 (0.11)	10.6 (0.05)	10.1 (0.05)
0.80	12.1 (0.06)	11.8 (0.07)	6.81 (0.03)	6.62 (0.03)
1.20	6.64 (0.02)	6.60 (0.03)	4.06 (0.02)	3.95 (0.02)
1.60	4.60 (0.02)	4.64 (0.02)	2.93 (0.01)	2.88 (0.01)
2.00	3.51 (0.01)	3.54 (0.01)	2.33 (0.01)	2.33 (0.01)
2.40	2.88 (0.01)	2.85 (0.01)	1.95 (0.01)	1.96 (0.01)

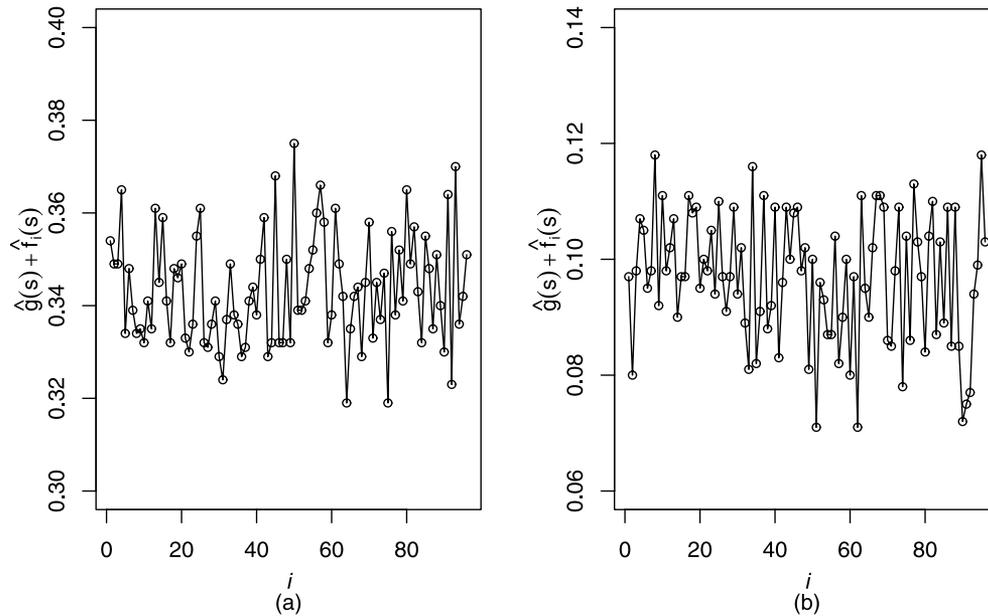


Figure 2. Plots of the estimated profiles $\widehat{g}(s) + \widehat{f}_i(s)$ over i , for $1 \leq i \leq 96$, at two specific positions $s = x_{10}^*$ [plot (a)] and $s = x_{30}^*$ [plot (b)].

within-profile and between-profile correlation. A number of useful references on handling temporal autocorrelation in some conventional process monitoring problems have been cited in Apley's discussion. Another relevant article by Noorossana, Amiri, and Soleimani (2008), tried to handle autocorrelated linear profiles using certain time series models. The method proposed in that article has the potential to be generalized for handling autocorrelated nonparametric profiles, which needs to be further studied.

By the suggestion of Apley, in the AEC data example, we present the estimated profiles $\widehat{g}(s) + \widehat{f}_i(s)$ over i , for $1 \leq i \leq 96$, at two specific positions $s = x_{10}^*$ and $s = x_{30}^*$ in the two panels of Figure 2. We also compute the lag-1 and lag-2 autocorrelations of the time series shown in each panel. They are 0.104 and 0.058 in the case of panel (a) and 0.154 and 0.074 in the case of panel (b). From the plots and the computed autocorrelation values, we can see that temporal autocorrelation is not evident in this data, which can be explained by the fact that all the profiles in this example are actually collected over a relatively long time period.

In his discussion, Apley proposed two possible approaches for accommodating between-profile correlation. One is the Markov bootstrap procedure and the other one is the block bootstrap procedure. He thought that the block bootstrap procedure might be more appropriate to use for monitoring profiles because of the relatively complicated structure of the profile data. In the AEC example, we compute the control limit using the block bootstrap procedure with the bootstrap sample size 10,000 and the block size 9 which is about 10% of the IC data. The computed control limit is 19.37. Compared to the control limit 18.24 reported in QZW, this control limit is marginally larger and it does not change the signal time at the 112th time point (cf., Figure 3 in QZW).

3. ALTERNATIVE CHARTING STATISTICS

Besides our proposed charting statistic $T_{t,h,\lambda}$ defined by Equation (11) in QZW, the discussants propose two alternative charting statistics. For the purpose of detecting a shift in the covariance function $\gamma(x_1, x_2)$ (cf., its definition in Section 2.1 of QZW), Apley suggested using the charting statistic

$$\widetilde{T}_t^{(1)} = \sum_{i=0}^{t-1} (1-\rho)^i T_{t-i,h,1},$$

where $\rho \in [0, 1]$ denotes an EWMA weighting parameter. From the definition of $T_{t,h,\lambda}$, we can see that $T_{t-i,h,1}$ is a quadratic measure of the difference between the estimated profile mean function from the $(t-i)$ th profile data alone and the IC profile mean function g_0 . Therefore, $\widetilde{T}_t^{(1)}$ which is an EWMA statistic constructed from $\{T_{t-i,h,1}\}$ tries to combine information from different profiles about the difference between individual profiles and g_0 . In their discussion, Woodall, Birch, and Du proposed this charting statistic as well.

The statistic $\widetilde{T}_t^{(1)}$ is natural to use. As a matter of fact, we also thought of it at the beginning of this research project. It was finally given up and replaced by $T_{t,h,\lambda}$ for the following reason. The estimator of g from a single profile is relatively noisy, especially when the profile contains only a small number of observations. The relatively large variability of such estimators of g will be inherited by $\widetilde{T}_t^{(1)}$ and make it less sensitive to shifts in the profile mean function. As a comparison, $\widehat{g}_{t,h,\lambda}(s)$ defined in Equation (9) of QZW is obtained from multiple profiles through the EWMA weighting scheme. Its variability is therefore smaller than the variability of $\widehat{g}_{t,h,1}(s)$ which is constructed from the t th profile alone. Consequently, the control chart based on $T_{t,h,\lambda}$ is expected to be more powerful for detecting a shift in the profile mean function, compared to the chart based on $\widetilde{T}_t^{(1)}$.

Instead of $T_{t,h,\lambda}$, Chipman, MacKay, and Steiner thought that it is more convenient to use the charting statistic

$$\tilde{T}_t^{(2)} = \sum_{i=0}^{t-1} (1 - \rho)^i (\mathbf{y}_i - \mathbf{g}_{0,i})^T \Sigma_i^{-1} (\mathbf{y}_i - \mathbf{g}_{0,i}),$$

where $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{ini})^T$ and $\mathbf{g}_{0,i} = (g_0(x_{i1}), g_0(x_{i2}), \dots, g_0(x_{ini}))^T$. Obviously, $\tilde{T}_t^{(2)}$ is an EWMA statistic constructed from quadratic discrepancies between \mathbf{y}_i and $\mathbf{g}_{0,i}$. From its construction, we think that $\tilde{T}_t^{(2)}$ would not be as sensitive to possible profile mean shifts as $T_{t,h,\lambda}$ because of the large variability in $\mathbf{y}_i - \mathbf{g}_{0,i}$. However, this chart might be good for detecting shifts in the covariance function of $\gamma(x_1, x_2)$ because $\mathbf{y}_i - \mathbf{g}_{0,i}$ is just $\mathbf{f}_i + \boldsymbol{\varepsilon}_i$ when the process is IC, where $\mathbf{f}_i = (f_i(x_{i1}), f_i(x_{i2}), \dots, f_i(x_{ini}))^T$ and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{ini})^T$.

To investigate the performance of the alternative charts based on $\tilde{T}_t^{(1)}$ and $\tilde{T}_t^{(2)}$, which are denoted as $ALT^{(1)}$ and $ALT^{(2)}$, respectively, we consider the following example, where the IC models (II) and (III) and the OC model (i) used in QZW are considered. In addition, we consider the following OC model in which the shift is in variances:

$$y_{ij} = g(x_{ij}) + f_i(x_{ij}) + (1 + \theta^*)\varepsilon_{ij}$$

for $j = 1, 2, \dots, n_i, i = 1, 2, \dots,$ (2)

where θ^* is a constant. In the charts MENPC, $ALT^{(1)}$, and $ALT^{(2)}$, $\lambda = \rho = 0.1$ and all other parameters are chosen to be the same as those used in the example of Table 2 of QZW. The OC ARL values of the three charts are presented in Table 2.

From Table 2, we can see that chart $ALT^{(1)}$ does not work well in all cases considered in this example in comparison with the other two charts. As explained earlier, this chart will not be efficient for detecting profile mean shifts, which is confirmed here. From the table, it seems that this chart is not good for

detecting shifts in variances either. This latter result is not surprising because the quantity $T_{t-i,h,1}$ that it uses does not contain much information about the covariance function $\gamma(x_1, x_2)$. Chart $ALT^{(2)}$, on the other hand, does perform reasonably well for detecting shifts in variances. But it is not powerful for detecting small to moderate profile mean shifts, as expected. Our proposed chart MENPC is designed for detecting profile mean shifts. So, it performs reasonably well in cases with OC model (i), especially when the mean shift is small or moderate (i.e., θ value in the table is between 0.20 and 1.20). However, this chart does not perform well for detecting shifts in variances. So, in practice, if shifts in both mean and variances are our concern, then we probably want to use the charts MENPC and $ALT^{(2)}$ simultaneously.

4. ARE ASYMPTOTIC RESULTS RELEVANT?

We appreciate the comment made by Apley regarding the asymptotic results included in QZW, and we agree with him completely that readers should not assign too much significance to Theorem 1 and other asymptotic results in QZW when designing the proposed control chart. Generally speaking, asymptotic results are valid only when the sample size is large. In reality, the related sample size is always finite. Therefore, asymptotic results are always a certain distance away from reality, and that distance depends on the sample size and how all the conditions and assumptions required by the asymptotic results are satisfied in a practical situation. In statistical process control (SPC), this is especially true because whenever we get a signal of shift from a control chart, the process (e.g., a production line) should be stopped immediately for people to find the root cause of the shift and then make certain appropriate adjustments of the process. Therefore, the sample size is hardly large in such cases. This may be the reason why asymptotic results are not often seen in the SPC literature.

Table 2. OC ARL values of the charts MENPC, $ALT^{(1)}$, and $ALT^{(2)}$ in cases when $ARL_0 = 200, n = 20, n_0 = 40,$ and $\lambda = \rho = 0.1$

	OC model (i)			OC model defined in Equation (2)				
	θ	MENPC	$ALT^{(1)}$	$ALT^{(2)}$	θ^*	MENPC	$ALT^{(1)}$	$ALT^{(2)}$
IC model (II)	0.20	130	197	196	0.05	173	183	74.1
	0.30	80.5	196	165	0.10	156	177	33.1
	0.40	48.6	195	142	0.15	134	174	16.6
	0.60	20.7	191	82.5	0.20	129	160	9.66
	0.80	12.1	192	45.8	0.30	101	156	4.26
	1.20	6.64	194	13.2	0.40	77.6	138	2.40
	1.60	4.60	197	4.42	0.50	62.3	124	1.77
	2.00	3.51	190	2.10	0.75	35.7	111	1.17
	2.40	2.88	185	1.37	1.00	21.9	92.4	1.05
IC model (III)	0.20	131	199	186	0.05	179	187	75.9
	0.30	81.0	197	153	0.10	157	182	34.7
	0.40	48.1	197	119	0.15	142	162	16.5
	0.60	21.4	196	68.4	0.20	138	168	9.83
	0.80	12.4	195	34.9	0.30	106	146	4.31
	1.20	6.59	193	8.77	0.40	82.7	140	2.45
	1.60	4.51	193	3.02	0.50	65.4	127	1.80
	2.00	3.43	192	1.60	0.75	34.5	111	1.17
	2.40	2.81	190	1.17	1.00	21.4	97.5	1.04

However, if we check the asymptotic results and their associated conditions and assumptions carefully, then we can still get some helpful information about the related SPC procedure. For instance, Apley already provided a thorough explanation why Theorem 1 in QZW concludes that $T_{t,h,\lambda}$ is independent of $\gamma(x_1, x_2)$ and σ^2 when $n_i h$ is bounded for each i and when other conditions hold. We agree with Apley that this result should not be used directly for choosing the control limit of our proposed chart. One major reason is that this result describes the IC behavior of the charting statistic $T_{t,h,\lambda}$ only; it does not take any profile mean shift into account. More specifically, the result holds when h tends to 0 and when other conditions are valid. But, a too small h will not be appropriate to use for detecting a profile mean shift effectively. On the other hand, this result together with the result (ii) of Theorem 1 in QZW also implies that h should be chosen small if it is desirable to have a chart that is less affected by the correlation among within-profile observations. In the case when $n_i h$ are large, the result (ii) reveals how the asymptotic distribution of $T_{t,h,\lambda}$ depends on $\gamma(x_1, x_2)$, which might be helpful in future research to modify $T_{t,h,\lambda}$ such that the modified version would incorporate the correlation function $\gamma(x_1, x_2)$ more effectively. As another example, according to result (i) of Theorem 2 in QZW, after the profile mean function changes from $g_0(x)$ to $g_1(x)$, the asymptotic distribution of the charting statistic $T_{t,h,\lambda}$ will depend on $\delta(x) = g_1(x) - g_0(x)$ and $\delta''(x)$. Therefore, if the curvature of $\delta(x)$ is bigger, then the corresponding shift is easier to detect, which has been confirmed in the numerical examples presented in Section 3 of QZW. See, for instance, Table 2 in QZW, where the curvature of $\delta(x)$ is much larger with OC model (ii) than the curvature of $\delta(x)$ with OC model (i).

5. GENERALIZATIONS TO MULTIVARIATE CASES

We appreciate Tsung’s comments on possible generalization of our proposed method to multivariate cases. He provided a nice description about several potential applications of multivariate profile monitoring and about some related research. We believe that his discussion provides us a strong motivation to study profile monitoring in multivariate cases in future research. In their discussion, Woodall, Birch, and Du also provided some comments on this topic, and they thought that other smoothing techniques such as the smoothing spline analysis of variance (ANOVA) might be more convenient to use, compared to the local polynomial kernel smoothing used in QZW. Frankly, we do not have much experience in multivariate cases, but would still like to share with readers some of our initial impressions described below.

Multivariate profile monitoring can be roughly classified into the following three categories:

- (i) each profile has one response and multiple covariates,
- (ii) each profile has multiple responses and one covariate,
- (iii) each profile has multiple responses and multiple covariates.

For category (i), some semiparametric modeling methods might be useful to describe the complicated high-dimensional profiles. For instance, the single-index and partial linear models (cf.,

Ruppert, Wand, and Carroll 2003), which has a relatively simple interpretation of the effect of each covariate on the response, might be appropriate in certain cases for describing multivariate profiles. As an example, one type of multivariate profile monitoring problem can be described using partial linear modeling as follows:

$$y_{ij} = \begin{cases} g_0(t_{ij}) + \mathbf{X}_i \boldsymbol{\beta} + f_i(t_{ij}) + \varepsilon_{ij} & \text{for } j = 1, 2, \dots, n_i, i = 1, \dots, \tau \\ g_1(t_{ij}) + \mathbf{X}_i \boldsymbol{\beta} + f_i(t_{ij}) + \varepsilon_{ij} & \text{for } j = 1, 2, \dots, n_i, i = \tau + 1, \dots, \end{cases}$$

where t denotes a univariate covariate that has a nonparametric relationship with the response y , \mathbf{X} denotes multiple covariates that affect y linearly, τ is an unknown change-point, $\boldsymbol{\beta}$ is a co-efficient vector, and other quantities are the same as those used in QZW. This model describes cases when the nonparametric relationship between y and t has a shift at τ . Obviously, similar models can be formulated for cases when the linear relationship between y and \mathbf{X} has a shift. By combining existing semiparametric model estimation methods and process control schemes, we believe that appropriate control charts can be constructed for monitoring such multivariate profiles in a way that is similar to the chart MENPC.

For category (ii), assume that we have p responses, and the observed IC data are from the following multivariate nonparametric mixed-effects model:

$$\mathbf{y}_{ij} = \mathbf{g}(x_{ij}) + \mathbf{f}_i(x_{ij}) + \boldsymbol{\varepsilon}_{ij} \quad \text{for } j = 1, 2, \dots, n_i, i = 1, \dots, m,$$

where $\mathbf{g}(x) = (g_1(x), \dots, g_p(x))^T$ is the fixed-effects term, $\mathbf{f}_i(x) = (f_{i1}(x), \dots, f_{ip}(x))^T$ is the random-effects term, $\mathbf{y}_{ij} = (y_{ij1}, \dots, y_{ijp})^T$, and $\text{Cov}(\boldsymbol{\varepsilon}_{ij}) = \boldsymbol{\Sigma}$. For a given point $s \in [0, 1]$, $\mathbf{g}(s)$ and $\mathbf{f}_i(s)$ can be estimated by minimizing the following penalized local linear likelihood function which is similar to the expression in Equation (2) in QZW

$$\sum_{i=1}^m \left\{ \text{tr} \left\{ [\mathbf{Y}_i - \mathbf{Z}_i(\tilde{\boldsymbol{\beta}} + \tilde{\boldsymbol{\alpha}}_i)] \boldsymbol{\Sigma}^{-1} [\mathbf{Y}_i - \mathbf{Z}_i(\tilde{\boldsymbol{\beta}} + \tilde{\boldsymbol{\alpha}}_i)]^T \mathbf{K}_i \right\} + [\text{vec}(\tilde{\boldsymbol{\alpha}}_i)]^T \mathbf{D}^{-1} \text{vec}(\tilde{\boldsymbol{\alpha}}_i) + \ln |\mathbf{D}| + n_i \ln |\boldsymbol{\Sigma}| \right\},$$

where

$$\begin{aligned} \mathbf{Y}_i &= (\mathbf{y}_{i1}, \dots, \mathbf{y}_{in_i})^T, & \tilde{\boldsymbol{\beta}} &= (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_p), \\ \tilde{\boldsymbol{\alpha}}_i &= (\boldsymbol{\alpha}_{i1}, \dots, \boldsymbol{\alpha}_{ip}); \\ \mathbf{Z}_i &= (\mathbf{z}_{i1}, \dots, \mathbf{z}_{in_i})^T, & \mathbf{z}_{ij}^T &= (1, x_{ij} - s), \end{aligned}$$

each $\boldsymbol{\beta}_j$ is a deterministic two-dimensional coefficient vector, each $\boldsymbol{\alpha}_{ij}$ is a two-dimensional vector of the random effects with mean $\mathbf{0}$ and covariance \mathbf{D}_j , $\mathbf{D} = \text{diag}\{\mathbf{D}_1, \dots, \mathbf{D}_p\}$, and \mathbf{K}_i are defined in Equation (4) of QZW. Then, a similar iterative algorithm to the one described in Section 2.2 of QZW can be developed for Phase I model estimation. The local weighted negative log-likelihood estimation and a corresponding Phase II control chart can also be developed in a similar way to that described in Section 2.3 of QZW.

Category (iii) is much more complicated than the previous two categories. Probably certain appropriate combinations of the models described previously can handle some special cases. This is an important and challenging area and serious research will be required to develop effective monitoring schemes.

ACKNOWLEDGMENTS

The authors thank the editor Professor David Steinberg for organizing this stimulating discussion. We are also grateful to all the discussants for their constructive comments about the method proposed in our article. This research is supported in part by the grant DMS-0721204 from NSF of U.S.A. and the grants 10771107 and 10711120448 from NNSF of China.

ADDITIONAL REFERENCES

- Dette, H., and Neumeier, N. (2001), "Nonparametric Analysis of Covariance," *The Annals of Statistics*, 29, 1361–1400. [288]
- Fung, W., Zhu, Z., Wei, B., and He, X. (2002), "Influence Diagnostics and Outlier Tests for Semiparametric Mixed Models," *Journal of Royal Statistical Society, Ser. B*, 64, 565–579. [288]
- Fitzmaurice, G., Davidian, M., Verbeke, G., and Molenberghs, G. (2008), *Longitudinal Data Analysis*, London: Chapman & Hall/CRC. [288]
- Liang, K. Y., and Zeger, S. (2002), *Analysis of Longitudinal Data* (2nd ed.), New York: Oxford University Press. [288]
- Neumeier, N., and Dette, H. (2003), "Nonparametric Comparison of Regression Curves: An Empirical Process Approach," *The Annals of Statistics*, 31, 880–920. [288]
- Noorossana, R., Amiri, A., and Soleimani, P. (2008), "On the Monitoring of Autocorrelated Linear Profiles," *Communications in Statistics—Theory and Methods*, 37, 425–442. [290]
- Ruppert, D., Wand, M. P., and Carroll, R. J. (2003), *Semiparametric Regression*, Cambridge: Cambridge University Press. [292]
- Wang, K., and Tsung, F. (2005), "Using Profile Monitoring Techniques for a Data-Rich Environment With Huge Sample Sizes," *Quality and Reliability Engineering International*, 21, 677–688. [288]
- Young, S. G., and Bowman, A. W. (1995), "Non-Parametric Analysis of Covariance," *Biometrics*, 51, 920–931. [288]