Nonparametric Dynamic Curve Monitoring

Peihua Qiu, Xuemin Zi & Changliang Zou

Nonparametric Dynamic Curve Monitoring

Peihua Qiu[1], Xuemin Zi[2] and Changliang Zou[3]

[1] *Department of Biostatistics, University of Florida, USA;*

[2] *School of Science, Tianjin University of Technology and Education, China;*

[3] *Institute of Statistic and LPMC, Nankai University, China*

Abstract

Rapid sequential comparison between the longitudinal pattern of a given subject and a target pattern has become increasingly important in modern scientific research for detecting abnormal activities in many data-rich applications. This paper focuses on this problem when observations are collected sequentially with uncorrelated or correlated noise involved. A dynamic monitoring procedure is developed after connecting the curve monitoring problem to curve comparison. Under the framework of generalized likelihood ratio testing, we suggest a new exponentially weighted moving average (EWMA) control chart that can accommodate unequally spaced design points. An adaptive parameter selection feature is built in the proposed control chart so that the chart can detect a wide range of longitudinal pattern shifts effectively. To furnish fast computation, recursive formulas are derived for computing the charting statistic. Numerical studies show that the proposed method can deliver a satisfactory performance, and it outperforms existing methods in various cases. An example from the semiconductor manufacturing industry is used for the illustration of its implementation.

Keywords: Adaptive test; Curve comparison; Generalized likelihood ratio test; Model checking; Nadaraya-Watson kernel estimation; Statistical process control.

1   Introduction

Curve comparison is a fundamental research problem with broad applications. In practice, different curves often represent mean responses of different populations of subjects at varying levels of a covariate (e.g., time or concentration of a drug ingredient). Classical curve comparison methods compare parameter estimates of parametric regression models, and more recent research devotes much effort to various nonparametric settings. See, for instance, Hall and Hart (1990) and Kulasekera (1995). These traditional curve comparison methods are all retrospective in the sense that different curves are compared only after all data are collected. In many applications, however, observations of a subject are collected sequentially, and it is fundamentally important to detect the difference between its longitudinal mean response and the mean response of well-functioning subjects (called regular mean response hereafter) as early as possible so that some unpleasant consequences can be avoided. Therefore, it is ideal to compare the mean response of the given subject sequentially with the regular mean response. This dynamic curve monitoring problem is the focus of the current paper.

One example of dynamic curve monitoring is discussed by Qiu and Xiang (2014, 2015), which concerns the longitudinal pattern of people's total cholesterol level and other related medical indices. Early detection of the irregular longitudinal pattern of these indices can minimize the damage of stroke and other deadly cardiovascular diseases. In industrial applications, setting up a dynamic monitoring system could also be helpful. For example, Lee et al. (2011) described a semiconductor manufacturing process in which an etching chamber is equipped with many sensors to record observations of several quality variables of chips. In such examples, the curve observations are collected over time. If a curve already shows significantly irregular pattern during the process, there is no need to wait until the entire production process ends. A dynamic monitoring is particularly desirable for the processes with high production cost and limited resources (e.g., limited number of sensors and limited transmission and processing time). The engineers may be reluctant to wait for the collection of full curves if early detection can be made based on partial observations. Recently, continuous surveillance of a large amount of datastreams are greatly needed in many process monitoring applications. In mass production, when many units are monitored simultaneously, using a dynamic monitoring would be able to achieve cost saving.

The dynamic curve monitoring problem can be formulated nonparametrically as follows.

2

Assume that the response variable in question is z, and a subject's response values are obtained at some deterministic or random time points $\{t_1, t_2, \ldots\}$ over the design interval $[0, T]$. These observations follow the model

$$z(t_i) = \begin{cases} \mu(t_i) + \sigma(t_i)\varepsilon(t_i), & \text{for} \quad t_i \in [0, \tau], \\ \mu(t_i) + \delta(t_i) + \sigma(t_i)\varepsilon(t_i), & \text{for} \quad t_i \in (\tau, T], \end{cases} \tag{1}$$

where τ is an unknown change-point, $\sigma^2(t)$ is a smooth variance function, and $\varepsilon(t)$ is random noise with $E(\varepsilon(t_i)|t_i) = 0$ and $\text{Var}(\varepsilon(t_i)|t_i) = \zeta^2$. Without loss of generality, we let $\zeta^2 = 1$. In model (1), we further assume that the mean shift function $\delta(t)$ is a smooth function in $(\tau, T]$. This assumption is valid for the step shift and the drift, both of which are commonly considered in the statistical process control (SPC) literature (Qiu 2014). By model (1), different subjects are assumed to have the same mean function $\mu(t)$ and variance function $\sigma^2(t)$, while the change-point τ and the shift function $\delta(t)$ are allowed to vary from subject to subject. Our goal is to sequentially detect the occurrence of the mean response change as soon as possible *for each subject*, based on all available observations of the subject that have been collected up to the current time point.

It should be pointed out that the dynamic curve monitoring problem described above seems related to the nonparametric profile monitoring problem in SPC (e.g., Zou et al. 2008); but, the two problems are actually completely different, as described below. In the profile monitoring problem, observations obtained from a sampled product are a profile or a curve that describes the functional relationship between response variables and explanatory variables. The profiles are usually collected over time from different sampled products. The major goal of the profile monitoring is for detecting/identifing any abnormal functional relationship, based on the observed sequence of profile data. Thus, we monitor the sequence of profiles, and in some sense this problem is related to the curve comparison problem but in a sequential context. See, e.g., Chicken et al. (2009), Qiu and Zou (2010), Qiu et al. (2010) and the book by Noorossana et al. (2011) for related discussions. In contrast, in the dynamic curve monitoring problem, although observations of each subject look like a profile, each subject is actually treated as a separate process and we are mainly interested in sequentially monitoring the longitudinal behavior of each subject. In other words, in the latter problem, only one profile is involved and the monitoring of its longitudinal pattern is based on all its available observations up to the current time point.

Recently, Qiu and Xiang (2014) proposed a dynamic screening system (DySS) to solve

the dynamic curve monitoring problem. In the DySS method, the regular longitudinal pattern is first estimated from the observed longitudinal data of a group of well-functioning subjects, and then a cumulative sum (CUSUM) control chart is applied to the standardized observations of a new subject for sequentially monitoring its longitudinal behavior. The DySS method is simple and effective; however, when the observation times are unequally spaced, this method ignores the distribution of the observation times $\{t_1, t_2, \ldots\}$ completely since a regular control chart cannot accommodate this distribution. More importantly, the DySS is designed under the assumption that the mean response after the change occurs is a known constant. Our intuition suggests that if $g(t)$ changes drastically (e.g., the sine function considered in Figure 1 in Section 3), then a traditional chart, such like CUSUM or exponentially weighted moving average (EWMA), cannot be powerful because its reference value is likely to be mis-specified, and the chart would oversmooth the true function $g(t)$ and thus obscure the main feature of the out-of-control (OC) condition.

This paper suggests a novel dynamic curve monitoring procedure by making a connection between the dynamic curve monitoring problem and the curve comparison problem. Our proposed procedure is constructed under the framework of generalized likelihood ratio (GLR) testing which has become a commonly used framework for constructing nonparametric testing procedures in regression modeling (e.g., Fan et al. 2001). In that framework, we propose an exponentially weighted loss function that is tailored to our sequential curve monitoring problem. Based on that loss function, an exponentially weighted local constant kernel smoother is developed accordingly for sequentially estimating regression curves, resulting in a dynamic EWMA (DEWMA) control chart. The distribution information of the observation times $\{t_1, t_2, \ldots\}$ is incorporated naturally into such estimators. Then, a GLR test statistic, abbreviated as adaptive DEWMA (ADEWMA), is constructed after considering a set of admissible values for a smoothing parameter and after accommodating an adaptive parameter selection procedure. To provide fast computation, recursive formulas are developed for computing the hybrid GLR test statistic. Numerical examples in Section 3 will show that our proposed new method indeed performs better than the DySS method when the mean shift function $g(t)$, is not flat.

The remainder of the paper is organized as follows. In Section 2, we describe our proposed method and some of its extensions in detail. Its finite-sample performance is studied in Section 3. Section 4 contains a real-data example to illustrate the application of our proposed

4

method. Finally, several remarks conclude the paper in Section 5. Some technical details and additional simulation results are provided in the Supplementary Material.

2 Methodology

2.1 Model and hypotheses

In the literature, SPC is often divided into two phases. In Phase I, a set of process data is 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 set is then used for estimating the in-control (IC) parameters of the process distribution (e.g., $\mu(\cdot)$ and $\sigma(\cdot)$ in the model (1)). These parameters can be estimated from a sufficiently large IC dataset by either parametric or nonparametric methods under different assumptions. In Phase II, the estimated IC model from the IC dataset is used, and the major goal of this phase is to detect any changes in the model after an unknown time point. Because Phase II process monitoring is the focus of the current paper, for convenience we assume that $\mu(\cdot)$ and $\sigma(\cdot)$ are known, which is a convention in the Phase II SPC literature.

After a transformation $y(t_i) = \{z(t_i) - \mu(t_i)\}/\sigma(t_i)$, the problem (1) can be reformulated as

$$y(t_i) = \begin{cases} \varepsilon(t_i), & \text{for} \quad t_i \in [0, \tau], \\ g(t_i) + \varepsilon(t_i), & \text{for} \quad t_i \in (\tau, T], \end{cases} \tag{2}$$

where $g(t) = \delta(t)/\sigma(t)$ is a smooth function in $(\tau, T]$ by the smoothness assumption of $\delta(t)$ and $\sigma^2(t)$ made in Section 1. The related hypotheses become $H_0 : \tau > T$ versus $H_1 : \tau \in [0, T]$. The observations are collected over time, and thus we need to make a decision based on the standardized observations $\{y(t_i), t_i\}_{i=1}^m$ at each time point t_m.

It is also worth pointing out that this sequential detection problem is essentially similar to the detection of unknown patterned mean shifts, which is particularly important in monitoring autocorrelated or feedback-controlled processes where dynamic patterns in mean shifts are usually observed. See Han and Tsung (2006) and Shu et al. (2008) for example.

More recently, Capizzi and Masarotto (2012) developed an adaptive GLR chart (AGLR) that combines an EWMA estimate and a wavelet smoother. They showed that the AGLR chart is able to deliver robust detection to various changes, especially for oscillatory mean patterns. By connecting the curve monitoring to curve comparison, we propose a simple charting approach to the monitoring of (2). It is not only effective for detecting various alternative models but also for accommodating the information of unequally spaced design that is not considered in the existing works mentioned above.

2.2   Exponentially weighted GLR testing

To sequentially test whether a possible mean change occurs by a given time point, a GLR test for testing the null hypothesis that $g(t) = 0$ can be derived in a similar way to that in Fan et al. (2001), provided that a loss function is well defined. To simplify exposition, we first assume that $\varepsilon(t_i)$'s are independent, and the correlated data case will be discussed in Section 2.4. Based on all observed data $\{y(t_1), \ldots, y(t_m)\}$ up to the current time t_m, we suggest using the following exponentially weighted loss function for statistical inference about the mean shift function $g(t)$:

$$Q(t_m; \lambda) \equiv \sum_{i=1}^{m} \{y(t_i) - g(t_i)\}^2 (1 - \lambda)^{t_m - t_i}, \tag{3}$$

where $\lambda \in (0, 1]$ is a smoothing parameter. Obviously, $Q(t_m; \lambda)$ combines the ideas of local smoothing and exponential weighting schemes used in the EWMA literature (Lucas and Saccucci 1990) through the term $(1 - \lambda)^{t_m - t_i}$. By (3), all observations up to t_m are used in $Q(t_m; \lambda)$, the ones farther away from t_m receive less weight, and the weights decrease exponentially fast as t_i moves away from t_m.

To define the test statistic of the exponentially weighted GLR test, we need to replace the unknown function $g(t)$ by its estimator based on the observed data. See Zou et al. (2009) for a related discussion. To this end, the local polynomial kernel smoothing method can be considered (see Fan and Gijbels 1996, chap. 2). For instance, if the local constant kernel smoothing method is used here, then we can consider the minimization problem

$$\arg\min_{a \in R} \sum_{i=1}^{m} \{y(t_i) - a\}^2 (1 - \lambda)^{t_m - t_i},$$

where $(1 - \lambda)^{t_m - t_i}$ can be regarded as weights defined by a special kernel function. The solution to a of this minimization problem can be defined as the local constant kernel estimator of $g(t_m)$, denoted as $\widehat{g}_\lambda(t_m)$, which has the expression

$$\widehat{g}_\lambda(t_m) = \sum_{i=1}^{m} w_i(t_m) y(t_i) / \sum_{i=1}^{m} w_i(t_m), \tag{4}$$

where $w_i(t_m) = (1 - \lambda)^{t_m - t_i}$. Because $\widehat{g}_\lambda(t_m)$ only uses observations on one side of t_m, this estimator is closely related to the one-sided local polynomial kernel estimators, defined with one-sided kernel functions, that are commonly used in jump regression analysis (cf., Qiu 2005). We would like to mention that although the local linear kernel estimator has certain advantages over the local constant kernel estimator as shown in the literature (c.f., Fan and Gijbels 1996, sec. 3.2.5), our simulation results show that these two estimators do not differ much in our current monitoring problem. Thus, the local constant kernel procedure (4) is chosen for simplicity.

After the mean shift function $g(t)$ is replaced by its estimator $\widehat{g}_\lambda(t)$, the exponentially weighted loss functions under H_1 and H_0 are respectively

$$Q_{\mathrm{H}_1}(t_m; \lambda) = \sum_{i=1}^{m} \{y(t_i) - \widehat{g}(t_i)\}^2 w_i(t_m), \quad \text{and}$$

$$Q_{\mathrm{H}_0}(t_m; \lambda) = \sum_{i=1}^{m} \{y(t_i)\}^2 w_i(t_m).$$

The weighted GLR test statistic (WGLR) is then defined by

$$\begin{aligned} W_\lambda(t_m) &= Q_{\mathrm{H}_0}(t_m; \lambda) - Q_{\mathrm{H}_1}(t_m; \lambda) \\ &= \sum_{i=1}^{m} w_i(t_m) \{2y(t_i) - \widehat{g}(t_i)\} \widehat{g}(t_i). \end{aligned}$$

By the WGLR test, an alarm would be triggered at t_m if $W_\lambda(t_m)$ is large. By noticing that the sequence $\{W_\lambda(t_m), \widehat{g}(t_m)\}$ forms a two-dimensional Markov chain given the design points, the test statistic $W_\lambda(t_m)$ can be computed recursively. More specifically, we have

$$W_\lambda(t_m) = w_{m-1}(t_m) W_\lambda(t_{m-1}) + \{2y(t_m) - \widehat{g}(t_m)\} \widehat{g}(t_m), \tag{5}$$

$$\widehat{g}(t_m) = \{\alpha_{m-1} \widehat{g}(t_{m-1}) + y(t_m)\} / \alpha_m, \tag{6}$$

where $\alpha_m = \sum_{i=1}^{m} w_i(t_m) = w_{m-1}(t_m) \alpha_{m-1} + 1$, and all the initial values α_0, $\widehat{g}(t_0)$ and $W_\lambda(t_0)$ are zero. These recursive formulas can simplify the computation of the WGLR test greatly.

7

At different time points, the distributions of $W_\lambda(t_m)$ could differ, especially at the beginning of the process monitoring (i.e., m is small). The sequence $\{W_\lambda(t_m), m = 1, 2, \ldots\}$ would require quite a long time to attain its steady state (i.e., the state that the distribution of $W_\lambda(t_m)$ stabilizes when m increases). To overcome this challenge, let us consider

$$W_\lambda^*(t_m) = \{W_\lambda(t_m) - E_\lambda(t_m)\}/\sqrt{V_\lambda(t_m)}, \tag{7}$$

where $E_\lambda(t_m)$ and $V_\lambda(t_m)$ are the mean and variance of $W_\lambda(t_m)$. The recursive formulas to compute these two quantities are given in Appendix A of the Supplementary Material. By the standardization, $W_\lambda^*(t_m)$ would have zero mean and unit variance for all λ and t_m when the process is IC. Then, we can use a constant control limit L, and our proposed dynamic EWMA (DEWMA) gives a signal of process mean shift when the $W_\lambda^*(t_m)$ in (7) exceeds L.

Though the $W_\lambda(t_m)$ is derived in the spirit of GLR testing, it can be connected to the DySS method in the following way and such a connection could help us better understand the benefits of using $W_\lambda(t_m)$. To facilitate the discussion, let $d > 0$ be a basic time unit in a given application, which is the largest time unit that all (unequally spaced) observation times are its integer multiples. Define $n_i = t_i/d$, for $i = 0, 1, \ldots$, where $n_0 = t_0 = 0$. Then, $t_i = n_i d$, for all i, and n_i is the ith observation time in the basic time unit. Note that if t_i's are assumed equally spaced (i.e., they are simply $d, 2d, \ldots$), $\widehat{g}_\lambda(t_m)$ can reduce to

$$\begin{aligned}
\tilde{g}_\lambda(t_m) &= \sum_{i=1}^{m}(1-\lambda)^{(m-i)d}y(t_i) \bigg/ \sum_{i=1}^{m}(1-\lambda)^{(m-i)d}, \\
&= \sum_{i=1}^{m}(1-\tilde\lambda)^{(m-i)}y(t_i) \bigg/ \sum_{i=1}^{m}(1-\tilde\lambda)^{(m-i)} \tag{8}
\end{aligned}$$

which is a standard EWMA sequence of the observations $y(t_i)$ with smoothing parameter $\tilde\lambda = 1 - (1-\lambda)^d$. Under the GLR framework, if we use $\tilde{g}_\lambda(t_m)$ to replace $\widehat{g}(t_i)$, for all $i = 1, \ldots, m$, the WGLR test statistic $W_\lambda(t_m)$ becomes $\tilde{g}_\lambda^2(t_m)\sum_{i=1}^{m}(1-\tilde\lambda)^{(m-i)}$. The DEWMA chart would be (asymptotically) equivalent to the chart that signals a change when

$$\frac{|\tilde{g}_\lambda(t_m)|\sum_{i=1}^{m}(1-\tilde\lambda)^{m-i}}{\{\sum_{i=1}^{m}(1-\tilde\lambda)^{2(m-i)}\}^{1/2}} > L. \tag{9}$$

Chart (9) is the EWMA control chart with time varying control limits (Steiner 1999). In the literature, it has been well demonstrated that EWMA and CUSUM charts have similar performance if their parameters are appropriately chosen (see Qiu 2014). Therefore, the

performance of the chart (9) (a variant of DEWMA) would be similar to that of the DySS method in cases when observation times are equally spaced.

From the connection described above between the DEWMA chart (7) and the DySS method, the former should have at least two benefits described below. First, in cases when the observation times are unequally spaced, the CUSUM chart used in the DySS method or its EWMA counterpart (9) does not take into account the distribution of the observation times at all, while this information has been accommodated automatically in (3) through the weight $(1-\lambda)^{t_m-t_i}$. To appreciate this design adaptive feature, consider the following special case. Assume that $n_i - n_{i-1} = 1$ when $t_i \leq \tau$; $n_i - n_{i-1} = q$ and $g(t_i) = \delta > 0$ when $t_i > \tau$, where $q \gg 1$ is an integer. After τ, the weights of the observations $\{y(t_i), t_i \leq \tau\}$ in $\tilde{g}_\lambda(t_m)$ would decay much slower than those in $\widehat{g}_\lambda(t_m)$. Consequently, $\tilde{g}_\lambda(t_m)$ would be less sensitive to the shift, compared to $\widehat{g}_\lambda(t_m)$. Second, the chart (9) may be more effective than the DEWMA (7) when $g(t)$ is roughly a constant (i.e., the mean shift is roughly a step shift), while $W_\lambda^*(t_m)$ should be more robust and efficient in cases when $g(t)$ values change much over time. This can be explained as follows. At the current time t_m, the chart (9) makes a decision based on $\tilde{g}_\lambda(t_m)$ only, which is an estimator of $g(t_m)$. In the case when $t_m > \tau$, $g(t)$ decreases quite fast from a large value in $[\tau, t_m)$, and $g(t_m)$ is small, the detection power of (9) would be bad at t_m. As a comparison, when we define $W_\lambda^*(t_m)$, we use nonparametric estimates $\{\widehat{g}(t_i), 1 \leq i \leq m\}$ to approximate $\{g(t_i), 1 \leq i \leq m\}$, respectively. In the case discussed above, $W_\lambda^*(t_m)$ should be still sensitive to the mean shift because all $\{\widehat{g}(t_i), \tau \leq t_i \leq t_m\}$ are used. The above two benefits will be demonstrated numerically in Section 3.

The issue of unequally spaced observations has already been discussed in the literature concerning variable sampling interval (VSI) schemes (e.g., Ou et al. 2011). The sampling interval in a VSI chart is varied as a function of the charting statistic. In contrast, in our problem the mechanism of varying sampling interval is not controlled by our monitoring scheme but is determined by the process itself. The DEWMA statistic accommodates the distribution information of the time points by borrowing the strength of the local smoother.

2.3   Adaptive test

In the DEWMA chart, we need to determine the value of the weighting parameter λ. To further enhance the robustness of this chart, we propose using an adaptive procedure for

choosing this parameter. Note that λ can also be regarded as a smoothing parameter in nonparametric regression and could be chosen using cross-validation, AIC, BIC, or other model selection criteria; however it is well demonstrated that the optimal smoothing parameter for nonparametric curve estimation is generally not optimal for testing (cf., Hart 1997). Intuitively, a large λ would make the DEWMA chart effective in detecting shifts with large curvature and a small λ should be good for detecting shifts that are flat or with small curvature. For the lack-of-fit testing problem, Horowitz and Spokoiny (2001) suggested a hybrid method that combines a sequence of values of the smoothing parameter to attain a robust performance. Because this method has been demonstrated to have a good performance in various cases (cf., Qiu and Zou 2010; Capizzi and Masarotto 2011), we adopt it here.

Let $\Lambda = \{\lambda_j, j = 1, \ldots, q\}$ be a set of admissible values for the smoothing parameter λ. Then, at each time point t_m, we calculate $W^*_{\lambda_j}(t_m)$ for all $\lambda_j \in \Lambda$, and define

$$W_m = \max_{\lambda_j \in \Lambda} W^*_{\lambda_j}(t_m). \tag{10}$$

The maximum norm is used here for choosing the smoothing parameter in an adaptive way, so that the resulting chart can detect different types of mean shifts efficiently. Hereafter, the chart using W_m in (10) as the charting statistic is called the adaptive DEWMA (ADEWMA) chart. It is well known that optimal selection of the weighting parameter λ used in the EWMA control chart depends on the targeted shift. One benefit of the ADEWMA chart with W_m is that it can provide a balanced protection against various shift sizes. See Section 3 for related numerical justifications.

2.4 Practical guidelines

We provide some practical guidelines on implementing the proposed method.

On choosing Λ: In the standard EWMA chart (i.e., the case when $d = 1$ in (9)), it is well demonstrated that it is reasonable to choose $\lambda \in [0.02, 0.2]$. It can be checked that the effective number of observations used in the standard EWMA at each time point is asymptotically $(2 - \lambda)/\lambda$ (i.e., the reciprocal of the asymptotic variance of its charting statistic). For instance, if $\lambda = 0.2$, then the chart uses about 9 observations at each point. In the current ADEWMA chart, $\widehat{g}_\lambda(\cdot)$ is a major component, and it can be shown (see Appendix

B) that under some regularity conditions, its asymptotic variance is

$$\frac{1 - (1 - \lambda)^{dn_A}}{1 + (1 - \lambda)^{dn_A}} \equiv \gamma_\lambda^{-1}(d, n_A),$$

where dn_A is the mean interval length between two consecutive observation time points (see Appendix B for the definition of n_A). So, the effective number of observations used in $\widehat{g}_\lambda(t_m)$ is $\gamma_\lambda(d, n_A)$. Based on this argument and our empirical study, we find that it would result in a robust detection performance for the ADEWMA chart if we choose Λ so that $\{\gamma_\lambda(d, n_A), \lambda \in \Lambda\}$ contains $\{5, 10, 20, 40, 80\}$. Also, for the DEWMA chart (8), its performance will not change much if λ is chosen such that $\gamma_\lambda(d, n_A)$ is in the range $[5, 80]$.

On determining the control limit L: As discussed by Qiu and Xiang (2014), the commonly used performance measure ARL is inappropriate to use in cases when the observation time points are unequally spaced. In such cases, the average time to signal (ATS) is more relevant. If a signal is given at the sth observation time, then the ATS is defined as $E(t_s)$. It is obvious that the values of ATS are just constant multiples of the corresponding values of ARL in cases when the observation time points are equally spaced. After the desired in-control ATS is fixed, the value of L can be determined using Monte Carlo simulation by repeatedly generating random series $\{\varepsilon(t_i), t_i\}$ from the IC distribution. In practice, the IC distribution of $\{\varepsilon(t_i), t_i\}$ is usually unavailable and needs to be replaced by its empirical distribution or density estimator obtained from an IC dataset (Chatterjee and Qiu 2009).

On control of false positive rates: In many engineering or service applications, we need to monitor a large number of data streams simultaneously. In such cases, proper control of the (global) false positive rate (FPR) is important. It is reasonable to require that, for an IC process, the probability that a control chart does not trigger any alarm in the entire monitoring process is not too small. For the ADEWMA chart, this requirement becomes

$$\Pr\left(W_m < L, \text{ for all } t_m \in [0, T]\right) \geq 1 - p_0,$$

where p_0 is a pre-specified FPR. If l IC curves are monitored simultaneously, then on average lp_0 curves could be falsely declared OC. Though we used the classical ARL or ATS as the performance criterion in this paper, controlling ARL for our proposed chart could provide a good control of the FPR. It is well known that the geometric distribution can provide a decent approximation to the IC run length distributions of the CUSUM or EWMA charts, except in cases with short runs. This result is still valid for our proposed DEWMA and

11

ADEWMA charts based on our numerical experience (cf., Table 2). By this result, we have $p_0 \approx 1 - (1 - 1/\mathrm{ARL}_0)^n$, where n is the number of total time points and ARL_0 is the in-control ARL value. Hence, to reach the pre-specified FPR p_0, we can choose the control limit L such that $\mathrm{ARL}_0 = 1/\{1 - (1 - p_0)^{1/n}\}$. The simulation results in Table 2 show that this method for controling the overall FPR has a reasonably good performance.

Cases when the autocorrelation exists: In the foregoing discussion, we assume that the random errors $\{\varepsilon(t_j), j = 1, 2, \ldots\}$ are independent. In practice, they are often correlated. In the SPC literature, much recent research has considered SPC for autocorrelated processes. See, for instance, Tsung and Apley (2002), and the references therein. There are two general approaches to develop control charts in such cases. The first method fits a proper time series model to the process data and obtains the process forecast residuals. Then, traditional SPC charts can be applied to the residuals. Another method is to monitor the original observations using adjusted control limits for a given desired in-control ATS (cf., Wardell et al. 1994). In this paper, we use the first method. After obtaining the (roughly) uncorrelated residuals from time series modeling, the proposed charting statistic W_m can be constructed accordingly. It is expected that our proposed procedure is still applicable in cases when autocorrelation exists, which is confirmed by the numerical examples in Section 3.

2.5   Extensions

2.5.1   Data with a heavy-tailed distribution

Because of their computational simplicity and good asymptotic properties, the local polynomial kernel estimators have been commonly used in nonparametric model checking. However, their performance may be affected by data with heavy-tailed distributions or by outliers. A considerable effort has been devoted to the construction of robust nonparametric polynomial kernel smoothers. See Feng et al. (2012) and the references therein. A robust version of our DEWMA chart can be constructed to accommodate data with heavy-tailed distributions. Let $y^*(t_i) = \Phi^{-1}(F(y(t_i)))$, where $F(\cdot)$ is the cumulative distribution function (cdf) of $y(t_i)$ in model (2) and $\Phi^{-1}(\cdot)$ is the inverse of the cdf of the standard normal distribution. This transformation is commonly used in the literature to normalize the observed data so that the distribution of the transformed data would have lighter tails and become more symmet-

12

ric. Then, our DEWMA chart is applied to $y^*(t_i)$ instead of $y(t_i)$. In practice, $F(\cdot)$ can be replaced by the empirical distribution function obtained from an IC dataset. Capizzi and Masarotto (2012) provided more detailed discussions on using the empirical cdf.

2.5.2 Intrinsic level shifts

In some applications, the longitudinal patterns of different subjects may have level shifts due to some intrinsic reasons and such level shifts are not our concern. One such example will be discussed in Section 4, where the downward/upward level shifts of the observed data shown in Figure 6(a) should not worry us. For such applications, our control chart should be made flexible enough to accommodate intrinsic level shifts, which is the focus of this part.

The IC model that accommodates intrinsic level shifts can be described as follows:

$$y(t) = \mu(t) + h(t; \boldsymbol{\nu}) + \sigma(t)\varepsilon(t),$$

where $h(t; \boldsymbol{\nu})$ is a parametric function with the parameter $\boldsymbol{\nu} \in \mathbb{R}^p$ for describing the level shifts, and other terms are similar to those in model (1). Models for different subjects may have different values of $\boldsymbol{\nu}$, which is similar to mixed-effects modeling (Rice and Wu 2001). After data centralization (i.e., replacing $y(t)$ by $y(t) - \mu(t)$), the corresponding model for online monitoring becomes

$$y(t) = \begin{cases} h(t; \boldsymbol{\nu}) + \sigma(t)\varepsilon(t), & \text{for} \quad t \in [0, \tau], \\ h(t; \boldsymbol{\nu}) + \delta(t) + \sigma(t)\varepsilon(t), & \text{for} \quad t \in (\tau, T]. \end{cases} \tag{11}$$

Because different subjects are allowed to have different values of $\boldsymbol{\nu}$, we need to require enough observations collected before the change-point to obtain a rough estimate of $\boldsymbol{\nu}$ for the subject under monitoring. Otherwise, the IC and OC models for that subject may not be identifiable. This is similar to the so-called "masking-effect" in the self-starting monitoring context (Hawkins and Olwell 1998). Thus, we suggest starting to monitor a subject after collecting m_0 IC observations. How large m_0 is required depends on the complexity of $h(t; \boldsymbol{\nu})$ and the dimension of $\boldsymbol{\nu}$, which merits more future investigation.

The monitoring problem based on model (11) is related to the goodness-of-fit model checking problem for parametric/semi-parametric models (González-Manteiga and Crujeiras 2013). The GLR test is still applicable in this case. The only modification needed is that

13

the nuisance parameter vector $\boldsymbol{\nu}$ needs to be "profiled out" because it is unknown before process monitoring. More details can be found in the Supplementary Material.

3   Performance Evaluation

We present some numerical results in this section regarding the performance of the proposed method and compare it with some alternative methods. All results in this section are obtained from 10,000 replicated simulations. In all control charts, the in-control ATS, denoted as ATS_0, is fixed at 200. Results for some other ATS_0 values are similar. Both the zero-state (i.e., $\tau = 0$ in model (1)) and the steady-state ATS performances of the related charts are considered. Again, because the comparison results are similar in the two cases, only the steady-state ATS (SSATS) results are presented in this section. To evaluate the SSATS behavior of each chart, any series in which a signal occurs before the time τ is discarded (see Hawkins and Olwell 1998). The τ is fixed as 50 throughout this section and the results with $\tau > 50$ are quite similar.

Comparing the proposed method with the DySS method turns out to be not straightforward because the former uses an EWMA chart, the latter uses a CUSUM chart, and it is difficult to compare the two different charts in a fair manner. To overcome this difficulty, we modify the DySS method by replacing its CUSUM chart with an EWMA chart and keeping all other aspects of the method unchanged. This modified method will be abbreviated as EWMA which should not cause any confusion. Then, the two methods can be compared fairly if we choose their smoothing parameters in the way that the effective numbers of observations used in their charting statistics are equal. The adaptive EWMA chart (AEWMA) proposed by Capizzi and Masarotto (2003) is included as a benchmark for evaluating the performance of ADEWMA because the former is shown to be robust to various shift magnitudes when the mean function is a constant. We also compare our ADEWMA with Capizzi and Masarotto's (2012) AGLR chart which is known to be effective for various patterned mean shifts. The AGLR is implemented using the "AGLR" R package. The Fortran codes for implementing the proposed method are available in the Supplementary Material.

Let the design interval $[0, T]$ be $[0, 1000]$. The following two sampling scenarios are considered: (i) the observation time points are $t_i = 1, 2, 3, \ldots$ (i.e., they are equally spaced);
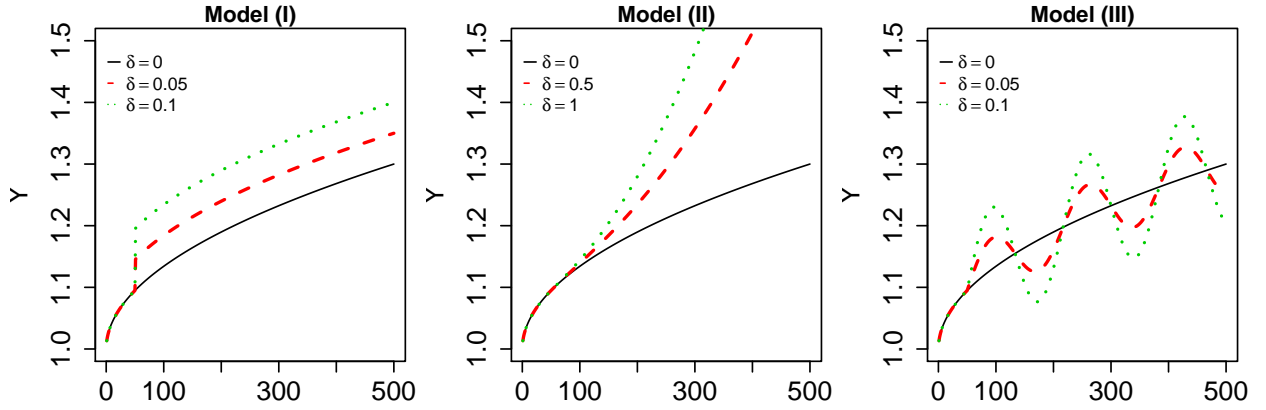
Figure 1: The OC functions of Models (I)-(III). In Model (III), $\eta = 0.003$.

and (ii) $t_i = t_{i-1} + Z_i + 1$, where $\{Z_i, i = 1, 2, \ldots\}$ are i.i.d. from the Poisson(θ) distribution. To make the results comparable, we set the basic time unit d to be 1 and $1/(\theta + 1)$ in cases (i) and (ii), respectively. Then, the average interval lengths between two consecutive time points are the same in the two cases. To better understand the usefulness of the weights $(1 - \lambda)^{t_m - t_i}$, we also consider a special sampling scenario, scenario (iii), which is the same as the sampling scenario (i) except that $n_i - n_{i-1} = 5$ when $\tau - 3 < i \leq \tau + 3$. By this sampling scheme, the sampling interval is much larger around the change point than those elsewhere.

The IC mean and variance functions are chosen to be $\mu(t) = 1 + 0.3t^{1/2}$ and $\sigma^2(t) = \mu^2(t)$. The errors $\varepsilon(t_i)$ are generated from the $N(0, 1)$ distribution. The following three different OC models are considered for demonstrating the effectiveness of our proposed method:

(I) Step Shift : $g(t) = \delta$, for $t > \tau$,

(II) Quadratic Drift : $g(t) = \dfrac{\delta}{4T}(t - \tau)^2$, for $t > \tau$,

(III) Sine-Oscillation Drift : $g(t) = \sin(\eta\pi(t - \tau))\delta$, for $t > \tau$,

where $\delta > 0$ is a parameter that controls the shift size. Model (I) considers a step change that occurs at τ, while Models (II) and (III) consider two more realistic scenarios that once a shift occurs, the shift size would change over time in a smooth or oscillating way. The three OC models with different values of δ are depicted in Figures 1(a)-(c). Figures S1(a)-(c) (in the Supplementary Material) show the local constant estimator $\widehat{g}(t)$ (EWMA), its local linear counterpart, the true function $g(t)$, and one set of observations $\{y(t_i), i = 1, 2, \ldots, 1000\}$ for the three models considered, respectively. The values of λ are respectively chosen to be 0.05,

15

0.1 and 0.2 for Models (I)-(III). We can see that both estimators of $g(t)$ can roughly track the dynamic changes of $g(t)$ and their difference is small. Also, to assess the robustness of ADEWMA, we consider another OC model, Model (IV), which is adopted from Capizzi and Masarotto (2012)

$$\text{(IV) Mixture-x} : g(r) = (-1)^{1-u(r)} g(r - 1) \text{ for } rd > \tau,$$

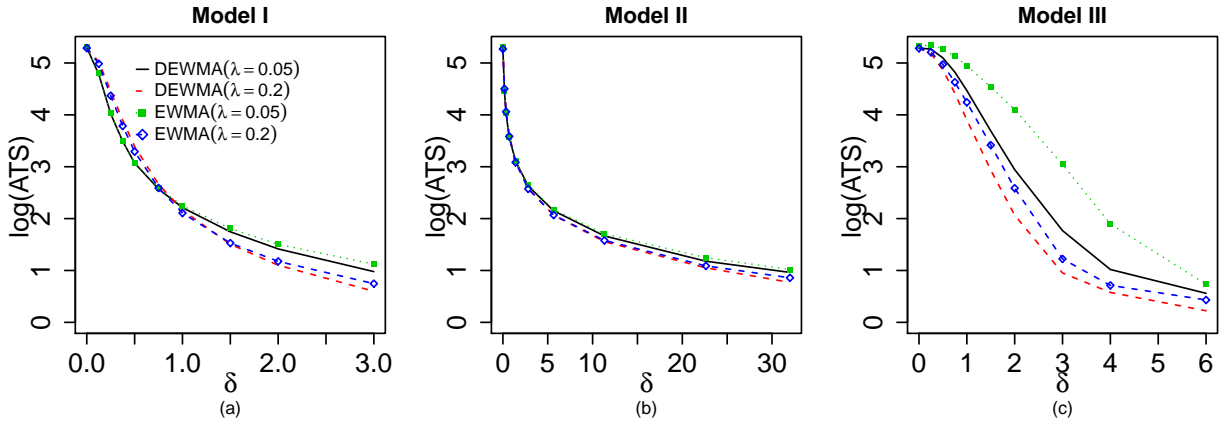where $u(r) \sim \text{Binomial}(1, x/100)$, $g(\tau + 1) = \delta$ and we set $x = 80$.

Figure 2: Out-of-control ATS comparison of the DEWMA and EWMA charts under the sampling scenario (i) consisting in the equally spaced observations. Natural logarithm is being used for vertical axis. For Model (II), the x-axis is scaled to $\delta^{0.5}/2$ for better exposition.

To appreciate the effectiveness of our proposed method, we first compare the DEWMA and EWMA methods when they use the same value of λ. Two λ values of 0.05 and 0.2 are considered. In order to give a clear picture of the difference between the two methods, we consider using $\eta = 0.25$ in Model (III) which yields a rather oscillating shift function $g(t)$ (much more oscillating than that illustrated in Figure 1). The parameter θ in the sampling scenario (ii) is chosen to be 3. The results for the three OC models under the sampling scenarios (i) and (ii) are presented in Figure 2 and Figure S2, respectively. Comparing results for equal spacing (Figure 2) with those for random spacing (Figure S2), we can see that the results for the two sampling scenarios are similar. For Model (I) in which a step change occurs, the smaller λ leads to a quicker detection of smaller shifts. The two methods with the same value of λ have a comparable performance in detecting small shifts (actually, EWMA is slightly better in such cases), and the DEWMA method outperforms the EWMA method in detecting large shifts. This can be explained intuitively as follows. Recall that

16

one main difference between the two methods is that EWMA approximates all $\{g(t_i), i \leq m\}$ by $\widehat{g}(t_m)$ in (3) while DEWMA approximates them by $\{\widehat{g}(t_i), i \leq m\}$. (3) can be written as

$$\sum_{i=1}^{\tau} \{y(t_i) - g(t_i)\}^2 (1 - \lambda)^{t_m - t_i} + \sum_{i=\tau+1}^{m} \{y(t_i) - g(t_i)\}^2 (1 - \lambda)^{t_m - t_i}. \tag{12}$$

When the shift is small, a chart needs to accumulate a relatively large number of observations to trigger a signal after the change-point; thus, the second part in (12) will play a major role. In such cases, the two types of approximates of $\{g(t_i), i \leq m\}$ described above would have similar impact on the chart performance. However, when the shift is large, any effective charts would give signals after only a few OC observations are collected. In such cases, to approximate $g(t_i)$ by $\widehat{g}(t_m)$ when $t_i \leq \tau$ would contaminate the first summation in (12), and consequently the performance of the EWMA method would be compromised. For Model (II), the above comparison conclusions still hold.

For Model (III), the story seems quite different from that of Models (I) and (II). First, a larger λ would lead to a much smaller ATS for all δ values and for both methods. This result is not surprising because Model (III) includes an oscillating shift function and thus a monitoring method using a smaller effective number of observations would be more efficient. In practice, however, we usually do not have much information about the OC model and thus we cannot choose an appropriate λ value beforehand. Such a difficulty leads to the use of our proposed ADEWMA chart. From Figure 2(c), it can also be noticed that DEWMA outperforms EWMA by a considerable margin, which confirms our previous analysis.

When all the other settings are the same as those in Figures 2(a)-(c), the simulation results under scenario (iii) are shown in Figure S3. The advantage of the DEWMA over EWMA in this case is more prominent than that in the case of scenario (i) shown in Figures 2(a)-(c). That is because the DEWMA chart accommodates well the unequally spaced observation times using the weights $(1 - \lambda)^{t_m - t_i}$, while the EWMA method does not.

Next, we compare the ADEWMA chart with the EWMA, AEWMA and AGLR. To evaluate their overall performance in detecting different shifts, besides ATS, we also compute their relative mean index (RMI) values (Han and Tsung 2006), defined as

$$\text{RMI} = \frac{1}{K} \sum_{l=1}^{K} \frac{\text{ATS}_{\delta_l} - \text{MATS}_{\delta_l}}{\text{MATS}_{\delta_l}},$$

17

Table 1: Out-of-control ATS comparison for Models (II)-(III) when the sampling scenarios (ii) and (iii) are considered. For EWMA, two λ values 0.05 and 0.2 are considered.

| Model | δ | Scenario (ii) | | | | | Scenario (iii) | | | | |
		ADEWMA	AGLR	AEWMA	EWMA 0.05	EWMA 0.2	ADEMA	AGLR	AEWMA	EWMA 0.05	EWMA 0.2
	0.125	88.0	86.9	85.0	86.6	90.1	88.7	93.7	89.4	91.8	96.1
	0.5	58.1	57.3	57.0	56.3	58.6	58.6	61.5	61.6	60.6	62.2
	2	36.4	35.8	37.3	36.1	36.1	37.7	40.0	41.8	40.1	39.8
	8	22.4	22.3	23.8	22.5	21.8	27.1	28.5	30.1	29.7	28.4
(II)	32	13.4	13.5	14.9	14.1	13.2	18.5	20.6	21.0	23.3	20.5
	128	7.97	8.14	8.87	8.85	7.99	11.0	12.2	11.8	15.0	12.5
	512	4.72	4.94	5.13	5.59	4.89	6.90	6.99	6.96	9.74	7.35
	2048	2.87	3.01	2.97	3.53	3.05	5.53	6.59	6.26	6.85	6.61
	4096	2.25	2.36	2.30	2.82	2.43	2.34	3.69	2.81	6.30	4.79
	RMI	0.014	0.028	0.059	0.10	0.032	0.007	0.121	0.100	0.363	0.207
	0.25	173	170	176	203	180	171	177	178	216	191
	0.5	121	121	141	181	135	123	139	150	199	155
	0.75	76.1	74.7	105	150	93.7	83.6	99.3	120	175	117
	1	44.7	45.7	72.4	120	60.5	55.0	69.6	93.3	152	88.7
(III)	1.5	16.5	17.2	32.3	68.5	24.9	28.8	41.7	52.4	112	53.7
	2	7.19	8.03	14.5	38.7	11.7	18.1	30.8	32.0	82.2	37.0
	3	2.63	3.10	4.17	13.4	3.64	6.78	18.5	13.1	46.8	24.3
	4	1.74	2.00	2.07	5.49	2.19	2.80	8.55	4.55	30.7	17.4
	6	1.29	1.40	1.31	2.31	1.53	2.00	2.10	2.01	19.6	5.36
	RMI	0.004	0.066	0.443	1.995	0.305	0.000	0.343	0.505	3.871	1.420

where K is the total number of shifts considered, ATS_{δ_l} is the ATS of a given chart when detecting shift $\boldsymbol{\delta}_l$, and MATS_{δ_l} is the smaller ATS between the ATS values of the two charts considered when detecting shift $\boldsymbol{\delta}_l$. A control chart with a smaller RMI value is considered better in their overall performance. Simulation results for Models (II) and (III) under scenarios (ii) and (iii) are presented in Table 1. A more complete version of Table 1 with standard deviations of the time to signal (SDTS) and the results for Model (IV) is provided by Tables S1-S2. In the AEWMA chart, a weight score function and two parameters need to be specified. For simplicity, we choose the Huber function

$$\phi_{hu}(e) = \{e + (1 - \lambda_A)k\} I(e < -k) + \{e - (1 - \lambda_A)k\} I(e > k) + \lambda_A e I(|e| < k)$$

and set $\lambda_A = 0.015$ and $k = 3.03$ which give good protection against a range of shifts

18

from $[0.25, 4]$ under Model (I). Also, as suggested by Capizzi and Masarotto (2012), we use $M = 128$, $\lambda = 0.12$, $k_q = 2$ and $k_w = 3.5$ in the AGLR chart. We acknowledge that better performances might be achieved by adjusting parameters under the considered scenarios, but our numerical results indicate that those parameters work reasonably well and thus serve the purpose of comparison here.
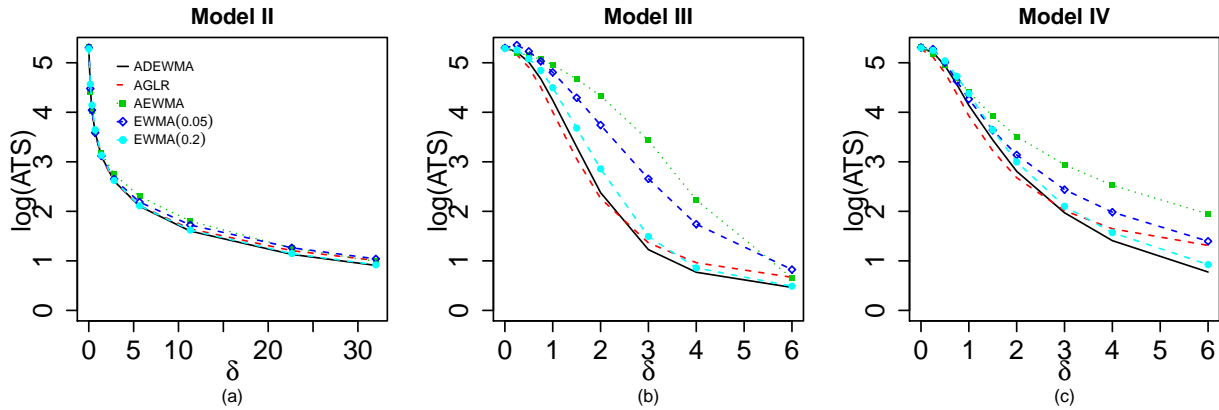

Figure 3: Out-of-control ATS comparison of the modified ADEWMA, AGLR, AEWMA and standard EWMA charts with adjusted control limits when the sampling scenario (ii) is used, consisting in the unequally spaced observations and the errors come from the Student's t distribution.

From Tables 1 and S1-S2, we can make the following conclusions. First, in cases with Model (II) when the shift function $g(t)$ is sufficiently smooth, the EWMA method with $\lambda = 0.05$ performs well only in detecting small shifts, and it performs well only in detecting large shifts when $\lambda = 0.2$. As a comparison, ADEWMA performs reasonably well in detecting small shifts, and it performs better than the two EWMA charts in detecting large shifts, which is mainly due to the benefit of including a larger value of $\lambda = 0.333$ in Λ. Second, in cases with Model (III) when the shift function $g(t)$ oscillating much over time, the EWMA method with $\lambda = 0.2$ performs better than itself with $\lambda = 0.05$, which is consistent with the results shown in Figures 2(c). In such cases, ADEWMA outperforms the EWMA charts with quite large margins in detecting all shifts. This reflects the fact that ADEWMA can adapt to the smoothness of the shift function $g(t)$ and offer a robust detection ability. Third, the three "adaptive" charts, ADEWMA, AGLR and AEWMA, have similar performances under Model (II), while the former two charts perform considerably better than the latter one under Models (III) and (IV). This is not surprising to us because both the ADEWMA and AGLR are "local" in nature but the AEWMA is developed under the assumption that

$\mu(t)$ is a constant mean function (but with unknown size). Fourth, the ADEWMA and AGLR are comparable in most cases, though the ADEWMA seems to perform a little better in terms of the RMI index, especially under Models (III) and (IV). The AGLR outperforms the ADEWMA under the sampling scenario (ii) and Models (II)-(IV) in terms of ATS and SDTS when δ is small. Fifth, the advantage of ADEWMA under the sampling scenario (iii) is quite remarkable which again demonstrates the benefit of accommodating the feature of unequally spaced design points.

All the above results are under the assumption that the random errors $\varepsilon(t_i)$ are normally distributed. As a matter of fact, the advantages of ADEWMA over EWMA hold for cases with non-normal random errors as well, as can be seen from the following example. Figure 3 and Figure S4 shows the out-of-control ATS values of the modified version of ADEWMA discussed in Section 2.5.1, and the AGLR, AEWMA and EWMA charts, in cases when the sampling scenario (ii) is used and the errors come from the Student's t-distribution with five degrees of freedom and the chi-squared distribution with five degrees of freedom, respectively. Results for the sampling scenario (iii) are reported in Figure S5. In the modified version of ADEWMA, the original data are first transformed so that the distribution of the transformed data is closer to normal. In EWMA, again two λ values 0.05 and 0.2 are considered. Because it is not easy to determine the parameters used in AEWMA and AGLR in such a case, we adopted the same strategy to implement these two charts, as for ADEWMA. Namely, we first transform the original data in exactly the same way as that for ADEWMA. Then, their parameters are chosen in the same way as in the above example. The modified ADEWMA chart is able to adapt to the unknown smoothness and magnitude of the mean shift function $g(t)$ in this case as well. It can pick an appropriate λ value from Λ, and yield an overall satisfactory performance in detecting the shifts.

The results in cases when observations at different time points are correlated and their error terms follow an AR(1) model with the coefficient 0.6 (see Sec. 2.3 in Qiu and Xiang 2014) are presented in Figures 4 and S6, under scenarios (ii) and (iii), respectively. All the other settings are the same as those in Table 1. The advantages of ADEWMA over the EWMA and AEWMA are still quite obvious in this case. The AGLR also performs quite well under the sampling scenario (ii) but is not as good as the ADEWMA under the sampling scenario (iii), as we can expect. Certainly, it is challenging to decorrelate the observed data with an arbitrary autocorrelation structure and certainly warrants much future research.
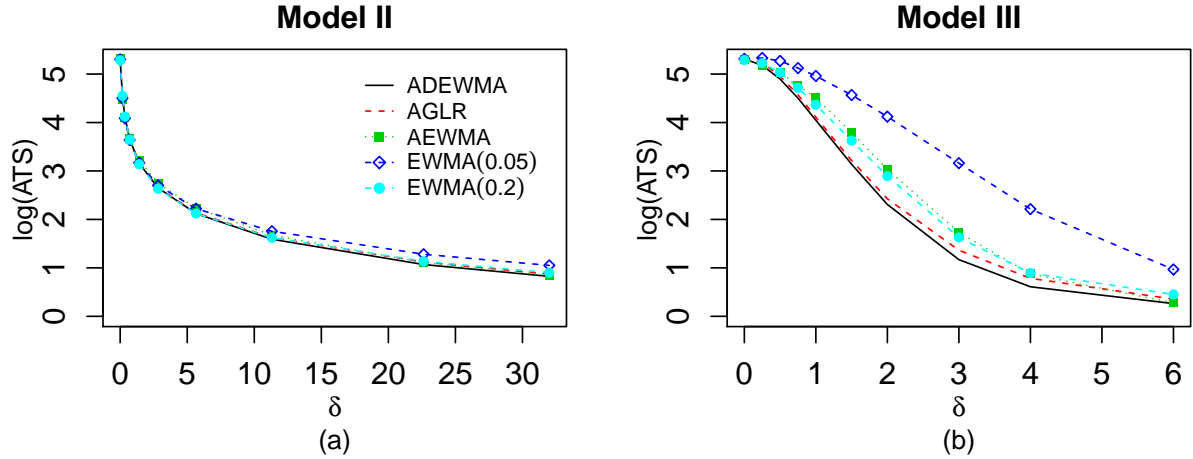
20

Figure 4: Out-of-control ATS values in cases when the error terms follow an AR(1) model with the coefficient 0.6. The sampling scenario (ii) is used, consisting in the unequally spaced observations.

Considering its convenience and robustness in various circumstances, our empirical results shown above and in the Supplementary Material may suggest that the ADEWMA should be a reasonable alternative for dynamic curve monitoring. It should also be pointed out that the competitors AEWMA and AGLR have been designed for the sampling scenario (i) and thus their performances may largely be affected by unequally spaced observations. Essentially, the local smoothing part of our method could be used as a component of these two methods to deal with missing values and better performances can be expected after accommodating certain modifications.

Finally, Table 2 presents the averaged false positive rates (FPRs) of DEWMA and ADEWMA charts under the sampling scenario (ii) for various values of p_0 when $T = 100$ or 200. As suggested in Section 2.4, the control limit is decided so that the in-control ATS equals $1/\{1 - (1 - p_0)^{1/T}\}$. From the table, it can be seen that the empirical FPRs are quite close to the nominal levels in most cases.

4 An Application

In this section, we apply the proposed methodology to a real dataset taken from an industrial etching process in semiconductor manufacturing. In this example, the etching chamber is equipped with many sensors to record observations of several quality variables of chips over

Table 2: Averaged false positive rates under the sampling scenario (ii).

	$T = 100$			$T = 200$		
	p_0			p_0		
	0.05	0.1	0.2	0.05	0.1	0.2
DEWMA (0.05)	0.048	0.092	0.192	0.044	0.104	0.188
DEWMA (0.2)	0.046	0.094	0.192	0.047	0.110	0.208
ADEWMA	0.046	0.091	0.190	0.051	0.098	0.195

time. For illustration purposes, only the variable related to plasma operations is considered here. Engineers think that this variable is one of the most important ones that contains enough information to distinguish the OC conditions in the etching process. The sensor observations are collected every second. A recorded observation is an average of sensor observations within a minute. Each chip was monitored for 98 minutes. See Figure 5(a) for an illustration of the observations of 5 chips. In this example, engineers are concerned about significant changes in the longitudinal pattern of the observations that may indicate some assignable causes in the etching process. From Figure 5(a), we can see that observations of different chips demonstrate a similar longitudinal pattern. However, different chips seem to have different starting points in their observations so that their longitudinal patterns exhibit downward/upward shifts from one another. Such level-shifts are mainly due to the "on-off" feature in the etching process and they should be regarded as intrinsic variations rather than out-of-control. See Lee et al. (2011) for a detailed explanation about this phenomenon. Figures 5(b)-(c) show the centralized observations (i.e., $y(t) - \widehat{\mu}(t)$) of two chips. No obvious non-random patterns can be noticed in the two scatter plots.

The dataset contains observations of a total of $n = 338$ chips. Among them, 22 chips are classified as inferior based on physical testing by engineers, and the remaining 316 chips as IC. Although this monitoring problem has a balanced design since the sensor observations are taken at 98 equally spaced time points, the data contain missing values. More specifically, 62 chips have missing values, and the missing proportions within these chips are around 0.05-0.1. The suggested weighting scheme in (3) is thus particularly useful for this case. For illustration, we use all observations in the IC group as the IC dataset and the others for testing. A nonparametric estimator of the population mean function $\mu(t)$ is first obtained. Then, we work on the centralized data $y(t) - \widehat{\mu}(t)$, and compute the sample autocorrelation and partial autocorrelation functions for each chip. The averages of these functions for all the
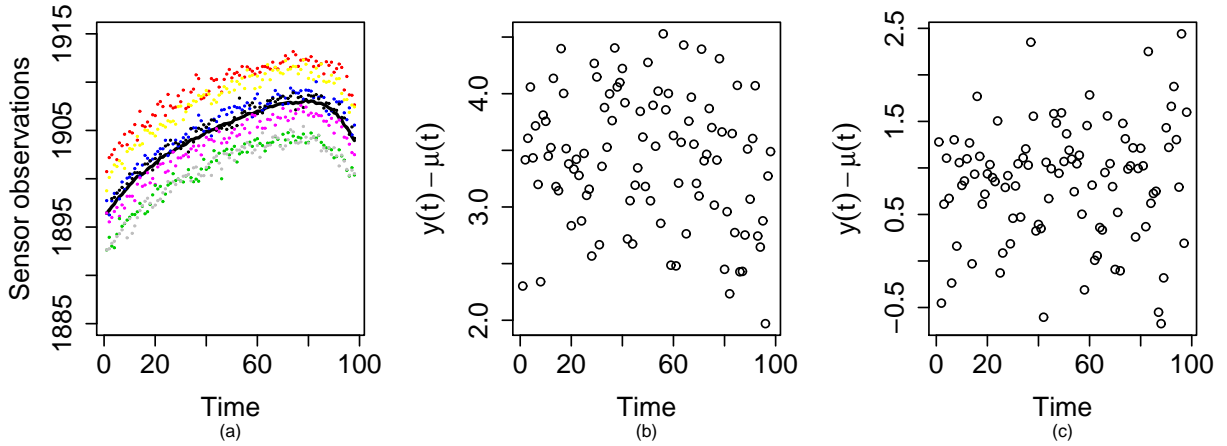
22

Figure 5: (a) Longitudinal sensor observations of five chips in the IC dataset, along with the nonparametric estimate $\widehat{\mu}(t)$ (solid line); (b) and (c): Scatter plots of sensor observations of two IC subjects after the transformation $y(t) - \widehat{\mu}(t)$.

chips in the IC dataset are shown in Figures 6(a)-(b). The estimated 1-lag autocorrelation coefficients are about 0.12, and no other autocorrelation coefficients exceed 0.2. Therefore, we can reasonably assume that the centralized data are independent over time. Figure 6(c) shows the estimated variances of the response variable y, denoted as $\widehat{\sigma}(t_i)$, at all observation times t_i, for $j = 1, \ldots 98$. It can be seen that variances of y at different time points are quite different. We further carry out the Shapiro-Wilk test for checking the normality of the standardized observations $\{y(t_i) - \widehat{\mu}(t_i)\}/\widehat{\sigma}(t_i)$, and the results are that only 6 IC chips have their p-values smaller than 0.05 and none of them are smaller than 0.02. Therefore, the normal assumption seems approximately valid and therefore we do not need to use the robust procedure proposed in Section 2.5.1 in this example.

To account for the level-shift, the model (11) and the related monitoring method suggested in Section 2.5.2 seem appropriate to use in this example. Also, it should be reasonable to assume that $h(t; \boldsymbol{\nu}) = \nu$ in (11). The weighted least-square estimator of ν at time t_m is

$$\widehat{\nu}_{m-1} = \frac{\sum_{i=1}^{m-1} \{y(t_i) - \mu(t_i)\} \sigma^{-2}(t_i)}{\sum_{i=1}^{m-1} \sigma^{-2}(t_i)}.$$

We set $m_0 = 20$ in this example and start the monitoring procedure from the 21st time point. Also, we set $d = 1$ and fix ATS$_0$ at 1,000. The set Λ is chosen to be $\Lambda = \{0.333, 0.181, 0.095, 0.049, 0.025\}$ such that the effective numbers of observations are approximately $\{5, 10, 20, 40, 80\}$, respectively, as suggested in Section 2.4. In such cases, the control limit L is computed to be 3.28. After adjusting the level shifts, as discussed in Section 2.5.2,
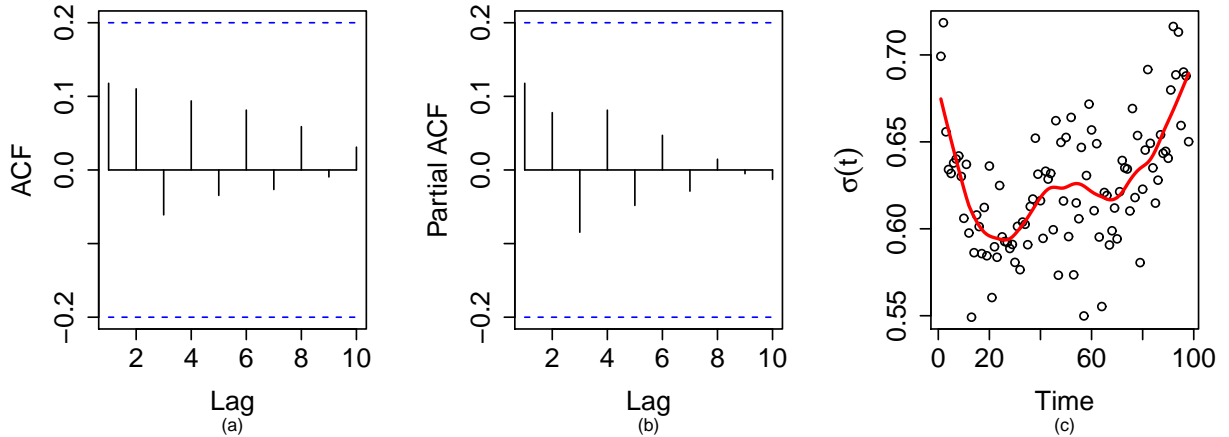
23

Figure 6: (a) and (b): The averages of sample autocorrelation and partial autocorrelation functions for all the chips in the IC dataset; (c): Estimated variances of observations in the IC dataset at each time point along with the local linear kernel estimator of their mean function (the solid line).

the ADEWMA charts for monitoring four chips in the test dataset are shown in the four right panels of Figure 7. From the plots, all four test chips receive signals. For all 22 chips in the test data, 20 of them receive signals. Note that the original paper Lee et al. (2011) on this dataset was more interested in a retrospective analysis of the data, instead of the prospective online monitoring problem focused in this paper. More specifically, by using our proposed dynamic monitoring method, observations of a chip are sequentially monitored and our monitoring procedure will make a decision each time when a new observation is obtained. If a signal is given at a specific time point, then we can conclude that the mean longitudinal pattern of the chip has a shift from the IC pattern, without collecting more data. Therefore, our method could lead to cost reduction and early detection of mean shift. We also apply the ADEWMA chart to the 316 IC chips, and 47 of them get signals, resulting in an observed FPR of 0.149. This is larger than the ideal one of $1 - (1 - 1/1000)^{98} = 0.093$ (see the related discussion in Section 2.4), but is still in an acceptable range.

5   Concluding Remarks

Online curve comparison is a challenging problem and it has not been fully investigated in the literature yet. There is a strong need of reliable statistical methods for robust monitoring of curves by sequentially analyzing their longitudinal observations over time. We proposed a dynamic monitoring procedure using an exponentially weighted loss function and the GLR
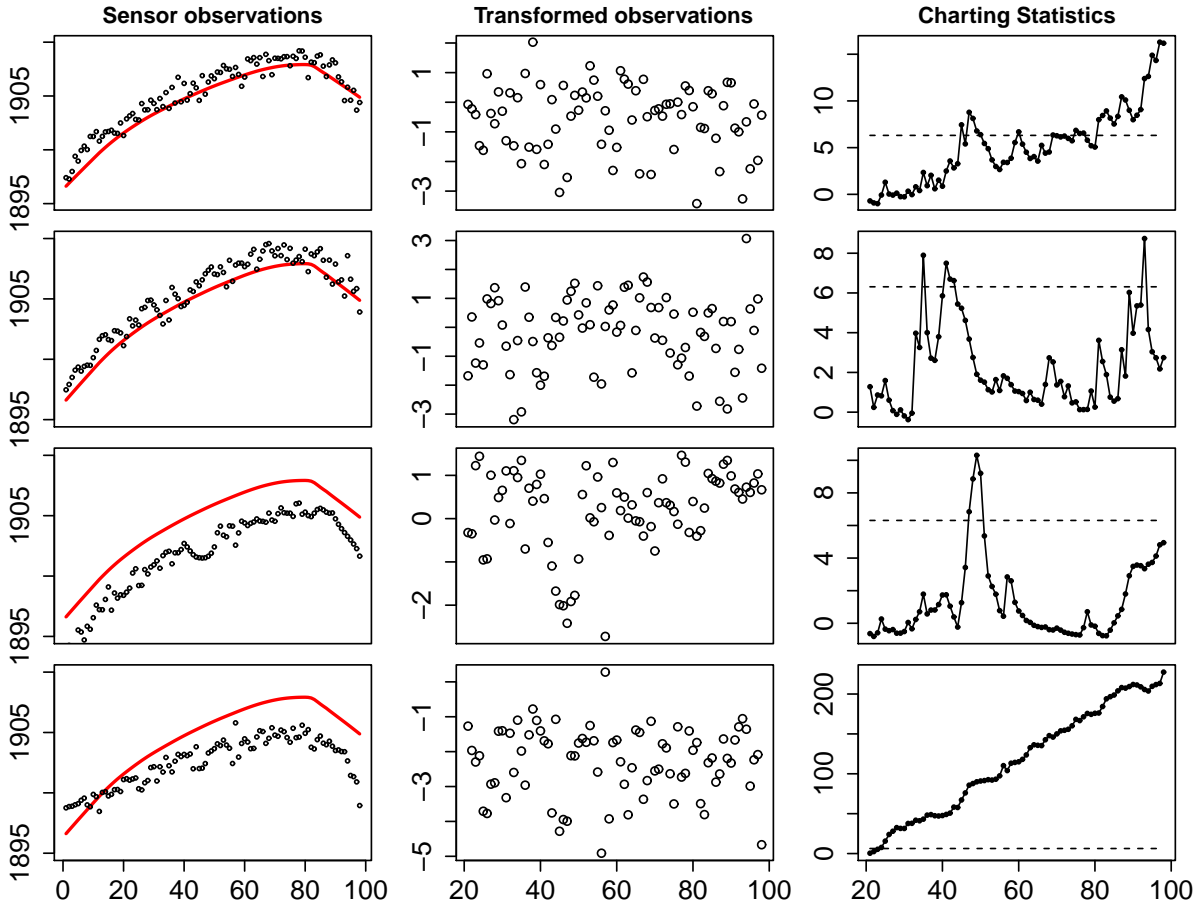
24

Figure 7: Left four panels: The raw sensor observations of four test chips along with $\widehat{\mu}(t)$; Middle four panels: Transformed observations $\widehat{y}(t_m) = (y(t_m) - \widehat{\nu}_{m-1})/\widehat{\sigma}(t_m)$; Right four panels: ADEW-MA charts for monitoring the four test chips. The horizontal lines denote the control limit.

testing method. An adaptive test statistic is suggested, which can accommodate different magnitudes and patterns of the shift function $g(t)$. The proposed method is easy to compute, and it could provide robust detection against various alternative models. In particular, it can integrate the distribution information of the observation time points with the information of the observations of the response variable. Thus, it offers an effective solution to the online curve comparison problem in cases when observation time points are unequally spaced.

In the practice of simultaneously monitoring many data streams, we are often not so much concerned with controlling the probability of obtaining at least one false alarm out of all the data streams being monitored. But, it is important to estimate and control the proportion of all false alarms which has been termed as the false discovery rate (FDR). There have been

25

some recent articles on sequential process monitoring based on the FDR consideration. See, for instance, Spiegelhalter et al. (2012) and the references therein. Combining our proposed method and the FDR idea deserves our future research.

Acknowledgement

Supplementary Material

The Supplementary Material contains some technical details, additional simulation results.

References:

Capizzi, G., and Masarotto, G. (2003), "An Adaptive Exponentially Weighted Moving Average Control Chart," *Technometrics*, 45, 199–207.

Capizzi, G., and Masarotto, G. (2011), "A Least Angle Regression Control Chart for Multidimensional Data," *Technometrics*, 53, 285–296.

Capizzi, G., and Masarotto, G. (2012), "Adaptive Generalized Likelihood Ratio Control Charts for Detecting Unknown Patterned Mean Shifts," *Journal of Quality Technology*, 44, 281–303.

Chatterjee, S., and Qiu, P. (2009). "Distribution-Free Cumulative Sum Control Charts Using Bootstrap-Based Control Limits," *The Annals of Applied Statistics*, 3, 349–369.

Chicken, E., Pignatiello, J. J. Jr., and Simpson, J. (2009), "Statistical Process Monitoring of Nonlinear Profiles Using Wavelets," *Journal of Quality Technology*, 41, 198–212.

Fan, J., and Gijbels, I. (1996), *Local Polynomial Modeling and Its Applications*, Chapman and Hall, London.

Fan, J., Zhang, C., and Zhang, J. (2001), "Generalized Likelihood Ratio Statistics and Wilks Phenomenon," *The Annals of Statistics*, 29, 153–193.

Feng, L., Zou, C., and Wang, Z. (2012), "Local Walsh-Average Regression," *Journal of Multivariate Analysis*, 106, 36–48.

González-Manteiga, W., and Crujeiras, R. M. (2013), "An Updated Review of Goodness-of-Fit Tests for Regression Models," *Test*, 22, 361–411.

Han, D., and Tsung, F. (2006), "A Reference-Free Cuscore Chart for Dynamic Mean Change Detection and a Unified Framework for Charting Performance Comparison," *Journal of the American Statistical Association*, 101, 368–386.

Hawkins, D. M., and Olwell, D. H. (1998), *Cumulative Sum Charts and Charting for Quality Improvement*, New York: Springer-Verlag.

Hall, P., and Hart, J. D. (1990), "Bootstrap Test for Difference Between Means in Nonparametric Regression," *Journal of the American Statistical Association*, 85,1039–1049.

Hart, J. D. (1997). *Nonparametric Smoothing and Lack-of-Fit Tests*, Springer, New York.

26

Horowitz, J. L., and Spokoiny, V. G. (2001), "An Adaptive, Rate-Optimal Test of a Parametric Mean-Regression Model Against a Nonparametric Alternative," *Econometrica*, 69, 599–631.

Kulasekera, K. B. (1995), "Comparison of Regression Curves Using Quasi-Residuals," *Journal of the American Statistical Association*, 90, 1085–1093.

Lee, S. P., Chao, A. G., Tsung, F., Wong, D. S. H., Tseng, S. T., and Jang, S. S. (2011), "Monitoring Batch Processes with Multiple On-Off Steps in Semiconductor Manufacturing," *Journal of Quality Technology*, 43, 142–157.

Lucas, J. M., and Saccucci, M. S. (1990), "Exponentially Weighted Moving Average Control Scheme Properties and Enhancements," *Technometrics*, 32, 1–29.

Noorossana, R., Saghaei, A., and Amiri, A. (2011), *Statistical Analysis of Profile Monitoring*, Wiley, New Jersey.

Ou, Y., Wu, Z., Yu, F., and Shamsuzzaman M. (2011), "An SPRT Control Chart with Variable Sampling Intervals," *The International Journal of Advanced Manufacturing Technology*, 56, 1149–1158.

Qiu, P. (2014), *Introduction to Statistical Process Control*, Boca Raton, FL: Chapman & Hall/CRC.

Qiu, P. (2005), *Image Processing and Jump Regression Analysis*, Wiley, New York.

Qiu, P., and Xiang, D. (2014), "Univariate Dynamic Screening System: An Approach For Identifying Individuals With Irregular Longitudinal Behavior," *Technometrics*, 56, 248–260.

Qiu, P., and Xiang, D. (2015), "Surveillance of Cardiovascular Diseases Using a Multivariate Dynamic Screening System," *Statistics in Medicine*, 34, 2204–2221.

Qiu, P., and Zou, C. (2010), "Control Chart for Monitoring Nonparametric Profiles with Arbitrary Design," *Statistica Sinica*, 20, 1655–1682.

Qiu, P., Zou, C., and Wang, Z. (2010), "Nonparametric Profile Monitoring By Mixed Effects Modeling (with discussions)," *Technometrics*, 52, 265–277.

Rice, J. A., and Wu, C. O. (2001), "Nonparametric Mixed Effects Models for Unequally Sampled Noisy Curves," *Biometrics*, 57, 253–259.

Shu, L., Jiang, W., and Tsui, K. L. (2008), "A Weighted CUSUM Chart for Detecting Patterned Mean Shifts," *Journal of Quality Technology*, 40, 194–213.

Spiegelhalter, D., Sherlaw-Johnson, C., Bardsley, M., Blunt, I., Wood, C., and Grigg, O. (2012), "Statistical Methods for Healthcare Regulation: Rating, Screening, and Surveillance," *Journal of the Royal Statistical Society, Series A*, 175, 1–47.

Steiner, S. H. (1999), "Exponentially Weighted Moving Average Control Charts with Time Varying Control Limits and Fast Initial Response," *Journal of Quality Technology*, 31, 75–86.

Tsung, F., and Apley, D. W. (2002), "The Dynamic T-squared Chart for Monitoring Feedback-Controlled Processes," *IIE Transactions*, 34, 1043–1053.

Wardell, D. G., Moskowitz H., and Plante, R. D. (1994), "Run-length Distributions of Special Cause Control Charts for Correlated Processes," *Technometrics*, 36, 3–17.

Zou, C., Qiu, P., and Hawkins, D. M. (2009), "Nonparametric Control Chart for Monitoring Profile Using the Change Point Formulation," *Statistica Sinica*, 19, 1337–1357.

Zou, C., Tsung, F., and Wang, Z. (2008), "Monitoring Profiles Based on Nonparametric Regression Methods," *Technometrics*, 50, 512–526.