

Nonparametric Control Chart for Monitoring Profiles Using Change Point Formulation and Adaptive Smoothing

Changliang Zou¹, Peihua Qiu² and Douglas Hawkins²

¹*Nankai University* and ²*University of Minnesota*

Abstract: In many applications, quality of a process is best characterized by a functional relationship between a response variable and one or more explanatory variables. Profile monitoring is used for checking the stability of this relationship over time. Control charts based on nonparametric regression are particularly useful when the in-control (IC) or out-of-control (OC) relationship is too complicated to be specified parametrically. This paper proposes a novel nonparametric control chart, using a sequential change-point formulation with generalized likelihood ratio tests. Its control limits are determined by a bootstrap procedure. This chart can be implemented without any knowledge about the error distributions, as long as a few IC profiles are available beforehand. Moreover, benefiting from certain good properties of the dynamic change-point approach and of the proposed charting statistic, the proposed control chart not only offers a balanced protection against shifts of different magnitudes, but also adapts to the smoothness of the difference between IC and OC regression functions. Consequently, it has a nearly optimal performance for various OC conditions.

Key words and phrases: Adaptive smoothing, Bandwidth selection, Change point, Generalized likelihood ratio test, Local linear kernel smoothing, Profile monitoring, Statistical process control.

1. Introduction

Because of recent progress in sensing and information technologies, automatic data acquisition is commonly used in industry. Consequently, a large amount of quality related data of certain processes become available. Statistical process control (SPC) of such data-rich processes is an important component for monitoring their performance. In many applications, quality of a process is characterized by the relationship between a response variable and one or more explanatory variables. At each sampling stage, one observes a collection of data points of these variables that can be represented by a curve (or, profile). In

some calibration applications, the profile can be described adequately by a linear regression model. In some other applications, more flexible models are necessary for describing profiles properly. This paper focuses on monitoring nonparametric profiles over time.

Most existing references on profile monitoring focus on cases in which profiles can be adequately described by a linear regression model. See, for instance, Kang and Albin (2000), Kim, Mahmoud and Woodall (2003), Mahmoud and Woodall (2004), Zou, Zhang and Wang (2006), and Mahmoud *et al.* (2007), among several others. Multiple and polynomial regression profile models are considered by Zou, Tsung and Wang (2007), and Kazemzadeh, Noorossana and Amiri (2007). Recently, nonlinear regression profile models have attracted much attention from statisticians. For instance, Williams, Woodall and Birch (2007) suggest three general approaches to nonlinear profile monitoring in Phase I analysis. Colosimo and Pacella (2007) propose methods for monitoring roundness profiles of manufactured items. Williams *et al.* (2007) apply the nonlinear regression approach by Williams, Woodall and Birch (2007) to monitoring nonlinear dose-response profiles. Lada, Lu and Wilson (2002) and Ding, Zeng and Zhou (2006) investigate a general category of nonlinear profiles, using dimension-reduction techniques, wavelet transformations, and independent component analysis. See Woodall *et al.* (2004) for an overview on this topic.

Recently, Zou, Tsung and Wang (2008) considered the nonparametric profile model

$$y_{ij} = g(x_{ij}) + \varepsilon_{ij}, \quad i = 1, \dots, n_j, \quad j = 1, 2, \dots, \quad (1.1)$$

where $\{x_{ij}, y_{ij}\}_{i=1}^{n_j}$ is the j th random sample, x_{ij} is the i th design point in the j th sample, g is a smooth nonparametric profile, and ε_{ij} are i.i.d. random errors with mean 0 and variance σ^2 . Zou, Tsung and Wang propose a control chart that integrates the classical multivariate exponentially weighted moving average procedure with the generalized likelihood ratio testing procedure that is based on nonparametric regression. This chart can monitor shifts in both g and σ^2 . It has been shown that this approach overcomes the fundamental limitation of parametric profile monitoring techniques that they cannot detect certain shifts due to misspecified out-of-control (OC) models.

Zou, Tsung and Wang's (2008) control chart, which is called the nonpara-

metric exponentially weighted moving averaging (NEWMA) chart hereafter, has a number of model assumptions. In certain applications, these assumptions may not all hold and performance may be unsatisfactory. This is briefly discussed below. First, the NEWMA chart makes explicit use of the IC true regression function g and the error variance σ^2 in model (1.1). In practice, both g and σ^2 are often unknown and need to be estimated from IC data. If such data are of small to moderate size, then there would be considerable uncertainty in the parameter estimates, which in turn would distort the IC run length distribution of the control chart. Even if the control limit of the chart is adjusted properly to attain a desired IC run length behavior, its OC run length would still be severely compromised (cf. Jones (2002)). Second, the NEWMA chart assumes that the error distribution, say F , is Normal, while in practice F is often unknown. In such cases, it remains challenging to design the NEWMA chart properly (see Remark 2 in Zou, Tsung and Wang (2008) for related discussion). Third, numerical examples in Zou, Tsung and Wang (2008) demonstrate that the NEWMA chart depends heavily on the choice of a bandwidth used in smoothing profile data. However, the proper choice of this parameter is not discussed thoroughly. Fourth, the NEWMA chart has another challenge that its best performance can be achieved only after it is “tuned” to the shift magnitude that is again often unknown in practice.

In this article, we propose a new control chart that addresses the issues raised above. The new control chart adopts the on-line change-point detection approach (cf. Hawkins Qiu and Kang (2003)). It handles sequential profile readings by simultaneously updating parameter estimates and checking for OC conditions. An adaptive procedure for selecting the bandwidth parameter is incorporated into the construction of the control chart so that it can adapt to the unknown smoothness of the difference between the IC and OC regression functions which, remarkably, improves its robustness to various OC profile conditions. Furthermore, a bootstrap procedure is used to determine the control limits of the proposed chart without any knowledge of F , as long as a few, say m_0 , IC profiles are available beforehand. These m_0 IC profiles are mainly used for estimating the error distribution. Thus, m_0 does not need to be large.

The proposed control chart is described in detail in Section 2. Its numeri-

cal performance is investigated in Section 3. In Section 4, we demonstrate the method using an example from the semiconductor manufacturing industry. Several remarks conclude the article in Section 5. Some technical details are provided in the Appendix.

2. Methodology

We describe the proposed control chart in four parts. In Subsection 2.1, the change-point formulation and the associated generalized likelihood ratio test are introduced. A Phase II control chart based on these techniques is described in Subsection 2.2. A bootstrap procedure for determining the control limits is presented in Subsection 2.3. Finally, some practical guidelines regarding design and implementation of the proposed control chart are provided in Subsection 2.4.

2.1. Change-point model and generalized likelihood ratio test

To ease presentation, we assume that model (1.1) has an unknown but stable error variance over time, the n_j 's all have the same value n , and design points $\{x_{1j}, \dots, x_{nj}\}$ are fixed for different j , denoted as $\mathbf{X} = \{x_1, \dots, x_n\}$. The assumption of stable error variance is often approximately valid in calibration applications of manufacturing industry, and is also consistent with the existing literature on profile monitoring. Possible extensions of the proposed control chart to cases in which both the regression function and the error variance have shifts, or the design points are random, are briefly discussed in Section 5.

The change-point model can be expressed as

$$y_{ij} = \begin{cases} g(x_i) + \varepsilon_{ij}, & i = 1, \dots, n, \quad \text{if } 1 \leq j \leq \tau \\ g_1(x_i) + \varepsilon_{ij}, & i = 1, \dots, n, \quad \text{if } \tau < j \leq t, \end{cases} \quad (2.1)$$

where τ is the unknown change point, $g \neq g_1$ are the unknown IC and OC regression functions, and ε_{ij} are i.i.d. errors with an unknown distribution F of mean 0 and unknown variance σ^2 .

To check whether a possible change point occurs at $\tau = k$, a two-sample generalized likelihood ratio (GLR) test for testing the null hypothesis that g and g_1 are the same can be derived in a way similar to that in Fan, Zhang and Zhang (2001). For ease of exposition, we can think of F as normal, although this is not necessary in either asymptotic theory or practical use of the proposed chart. The major idea in deriving the GLR test is to replace the unknown functions g

and g_1 by their nonparametric estimators constructed from profile data when we define the GLR statistic. To be specific, the generalized log-likelihood functions under the IC and OC conditions are, respectively,

$$\begin{aligned} l_0 &= -nt \ln(\sqrt{2\pi}\sigma) - \sum_{j=1}^t \left[\frac{1}{2\sigma^2} \sum_{i=1}^n (y_{ij} - \hat{g}^{(0)}(x_{ij}))^2 \right] \\ l_1 &= -nt \ln(\sqrt{2\pi}\sigma) - \sum_{j=1}^k \left[\frac{1}{2\sigma^2} \sum_{i=1}^n (y_{ij} - \hat{g}^{(1)}(x_{ij}))^2 \right] \\ &\quad - \sum_{j=k+1}^t \left[\frac{1}{2\sigma^2} \sum_{i=1}^n (y_{ij} - \hat{g}_1(x_{ij}))^2 \right], \end{aligned}$$

where $\hat{g}^{(0)}(\cdot)$ denotes the local linear kernel estimator (LLKE) of g based on the pooled t profiles, and $\hat{g}^{(1)}(\cdot)$ and $\hat{g}_1(\cdot)$ denote the LLKEs of g and g_1 based on the first k and the remaining $t - k$ profiles, respectively. After some mathematical manipulations, the GLR statistic up to time point t is defined by

$$T_{h,k,t} = -2(l_0 - l_1) = \frac{k(t-k)}{t\sigma^2} (\bar{\mathbf{Y}}_{0,k} - \bar{\mathbf{Y}}_{k,t})^T \mathbf{V}_h (\bar{\mathbf{Y}}_{0,k} - \bar{\mathbf{Y}}_{k,t}), \quad (2.2)$$

where

$$\begin{aligned} \bar{\mathbf{Y}}_{i,m} &= \frac{1}{m-i} \sum_{j=i+1}^m \mathbf{Y}_j, \quad \mathbf{Y}_j = (y_{1j}, \dots, y_{nj})^T, \\ \mathbf{V}_h &= \mathbf{W}_h^T + \mathbf{W}_h - \mathbf{W}_h^\otimes, \quad \mathbf{W}_h = (\mathbf{W}_n(x_1), \dots, \mathbf{W}_n(x_n))^T, \\ \mathbf{W}_n(x_i) &= (W_{n1}(x_i), \dots, W_{nn}(x_i))^T, \quad W_{nj}(x) = U_{nj}(x) / \sum_{i=1}^n U_{ni}(x), \\ U_{nj}(x) &= K_h(x_j - x) [m_{n2}(x) - (x_j - x)m_{n1}(x)], \\ m_{nl}(x) &= \frac{1}{n} \sum_{j=1}^n (x_j - x)^l K_h(x_j - x), \quad l = 1, 2, \end{aligned}$$

$K_h(\cdot) = K(\cdot/h)/h$, K is a symmetric density kernel function, h is a bandwidth, and $\mathbf{A}^\otimes = \mathbf{A}^T \mathbf{A}$. Obviously, the test statistic (2.2) is a two-sample counterpart of the one-sample GLR test statistic in Fan, Zhang and Zhang (2001). Since σ^2 is assumed unknown here, we replace it by the consistent nonparametric estimator

originally suggested by Hall and Marron (1990),

$$\hat{\sigma}_t^2 = \frac{1}{t(n - df)} \sum_{j=1}^t (\mathbf{Y}_j - \mathbf{W}_h \mathbf{Y}_j)^\otimes, \quad (2.3)$$

where $df = \text{tr}(\mathbf{V}_h)$.

Remark 1 The above testing problem is analogous to *nonparametric covariance analysis* or *comparisons of multiple curves* in the context of nonparametric regression testing. Many references in this area can be found in the literature under various settings and assumptions, including Hall and Hart (1990), Young and Bowman (1995), Kulasekera and Wang (1997), and Dette and Neumeyer (2001). A recent review on this topic is given by Neumeyer and Dette (2003). From an asymptotic viewpoint, $T_{h,k,t}$ in (2.2) is similar to the test statistic based on the difference of two variance estimators constructed from the data that are before and after the change-point, respectively. Dette and Neumeyer (2001) show that the latter statistic has good finite sample properties and is often more powerful than several alternative test statistics found in the literature, partly because certain good properties of the classical likelihood ratio test are inherited by the GLR method, as demonstrated in Fan, Zhang and Zhang (2001). \square

Like many other smoothing-based tests, performance of the test (2.2) depends upon the smoothing bandwidth h . Selection of h such that the testing power is optimal remains an open problem in this area. It is widely recognized that the optimal h for nonparametric curve estimation is generally not optimal for testing (see e.g., Hart (1997)). A uniformly most powerful test usually does not exist due to the fact that nonparametric regression functions have infinite dimensions. For the lack-of-fit testing problem, Horowitz and Spokoiny (2001) suggest choosing a single h based on the maximum of a studentized conditional moment test statistic over a sequence of smoothing parameters, and prove that the resulting test would have certain optimality properties. Based on these results, Guerre and Lavergne (2005) suggest choosing h from a sequence of pre-specified values, and demonstrate that their method has a number of favorable properties. Because this method is easy to use and has good performance in various cases, we use it here for choosing h . Let \mathcal{H}_n be a set of admissible smoothing parameters

defined as

$$\mathcal{H}_n = \{h_j = h_{\max} a^{-j} : h_j \geq h_{\min}, j = 0, \dots, J_n\}, \quad (2.4)$$

where $0 < h_{\min} < h_{\max}$ are the lower and upper bounds, and $a > 1$ is a parameter. Clearly, the number of values in \mathcal{H}_n is $J_n \leq \log_a(h_{\max}/h_{\min})$. Following Guerre and Lavergne (2005), we select h to be

$$\tilde{h} = \arg \max_{h \in \mathcal{H}_n} \{(T_{h,k,t} - \mu_h) - (T_{h_0,k,t} - \mu_{h_0}) - \gamma_n v_{h,h_0}\},$$

where $\gamma_n > 0$ is a chosen penalty parameter, μ_h is the mean of $T_{h,k,t}$, and v_{h,h_0}^2 is the variance of $T_{h,k,t} - T_{h_0,k,t}$. After accommodating the chosen bandwidth, the testing statistic becomes

$$\tilde{T}_{k,t} = (T_{\tilde{h},k,t} - \mu_{\tilde{h}})/v_{h_0}, \quad (2.5)$$

where $v_{h_0}^2$ is the variance of $T_{h_0,k,t}$, $\mu_{\tilde{h}} = \sum_{i=1}^n V_{\tilde{h}}^{(ii)}$, and $V_{\tilde{h}}^{(ij)}$ denotes the (i, j) -th element of the matrix $\mathbf{V}_{\tilde{h}}$. For given h and h_0 , it can be shown that consistent estimators of v_h^2 and v_{h,h_0}^2 are, respectively,

$$\hat{v}_h^2 = 2 \sum_{i=1}^n \sum_{j=1}^n [V_h^{(ij)}]^2, \quad \hat{v}_{h,h_0}^2 = 2 \sum_{i=1}^n \sum_{j=1}^n [V_h^{(ij)} - V_{h_0}^{(ij)}]^2.$$

Some statistical properties of the adaptive GLR test statistic $\tilde{T}_{k,t}$, including asymptotic null distribution and consistency under contiguous alternatives, are given in the Appendix.

Remark 2 Compared to the method of Horowitz and Spokoiny (2001), the parameter selection method of Guerre and Lavergne (2005) has a number of good properties. First, their selection criterion favors a baseline statistic under H_0 , which guarantees that the asymptotic distribution of $\tilde{T}_{k,t}$ is the same as that of $(T_{h_0,k,t} - \mu_{h_0})/v_{h_0}$. Hence, asymptotically speaking, the selected smoothing parameter based on data would not inflate the type-I error probability (i.e., test size); see Proposition 1 in the Appendix for more discussion. Second, this selection procedure allows us to use v_{h_0} in $\tilde{T}_{k,t}$, instead of $v_{\tilde{h}}$, which would asymptotically increase the test power. See Guerre and Lavergne (2005) for more detailed discussions about these issues.

In practice, the true change-point τ is often unknown. To test whether there is a shift in the regression function g for profiles up to time point t , we consider the adaptive GLR test statistic

$$\tilde{T}_{\hat{\tau}} = \max_{1 \leq k < t} \tilde{T}_{k,t}, \quad (2.6)$$

which is the maximum over all possible split points (i.e., binary segmentations). The maximizer $\hat{\tau}$ is used as an estimator of τ . Consistency of this change-point estimator can be established as in Proposition 2 in Zou, Tsung and Wang (2008).

2.2. Phase II SPC

The change point detection procedure (2.2)-(2.6) is developed for a collection of profiles with fixed sample size (c.f. (2.1)). In this part, we adapt it for on-line Phase II SPC. In traditional cases when observations are univariate and normally distributed, Phase II applications of change-point detection with some parameters specified beforehand have been discussed by Pollak and Siegmund (1991), Siegmund and Venkatraman (1995), Gombay (2000), Lai (2001), and Pignatiello and Simpson (2002). For the setting in which none of the parameters are assumed known, Hawkins, Qiu and Kang (2003), Hawkins and Zamba (2005a, b) discuss Phase II SPC for detecting possible shifts in the mean, the variance, and both the mean and variance. Zamba and Hawkins (2006) discuss multivariate SPC in which the measurement mean can change but the measurement covariance remains constant. Phase II linear profile monitoring by the change-point approach has been discussed by Zou, Zhang and Wang (2006). In this paper, we discuss nonparametric profile monitoring using the change point detection procedure introduced in the previous subsection.

Assume that there are m_0 IC profiles available. In many applications, one could collect a larger number of observations for each profile in Phase I analysis than in Phase II monitoring, because of the additional care commonly taken in Phase I analysis, while the OC condition could usually be effectively captured by using relatively small n . To make a distinction, hereafter we use n_0 to denote the number of observations in each of the m_0 IC profiles and n ($n \leq n_0$) to denote the number of observations in each future profile. Moreover, let $\mathbf{X}_{\text{IC}} = \{x_1, \dots, x_n, x_{n+1}, \dots, x_{n_0}\}$ be the set of design points in the m_0 IC profiles with the corresponding response variables $\{y_{1j}, \dots, y_{nj}, y_{(n+1)j}, \dots, y_{n_0j}\}$,

for $j = 1, \dots, m_0$. Our proposed Phase II SPC procedure is described as follows.

- After the $(t - m_0)$ th Phase II profile has been obtained, where $t > m_0$, calculate

$$lr_{m_0,t} = \max_{m_0 \leq k < t} \tilde{T}_{k,t}. \quad (2.7)$$

- If $lr_{m_0,t} > h_{m_0,t,\alpha}$, where $h_{m_0,t,\alpha}$ is some suitable control limit (see next subsection for discussion about its selection), then an out-of-control signal is triggered. After the signal, the systematic diagnostic procedure described in Zou *et al.* (2008) can be used for locating the mean profile change.
- If $lr_{m_0,t} \leq h_{m_0,t,\alpha}$, then the monitoring process continues by obtaining the $(t + 1)$ th Phase II profile and by repeating the previous two steps.

This sequential scheme differs from the NEWMA chart in an obvious way in that the IC regression function and the error variance in each observed profile can be both unknown in the former scheme, while they are assumed known in the latter one. Furthermore, in the NEWMA chart, we need to choose the procedure parameter λ , besides the smoothing bandwidth, which is not necessary in the change-point approach (cf., Hawkins, Qiu and Kang (2003); Han and Tsung (2004)).

It should be noted that computing (2.7) involves estimation of the error variance σ^2 for each t . When t is given, (2.3) gives a consistent estimator of σ^2 . From the t -th to the $(t + 1)$ -st profile, this estimator can be easily updated as

$$\hat{\sigma}_{t+1}^2 = [(t + m_0)(n - df)\hat{\sigma}_t^2 + (\mathbf{Y}_t - \mathbf{W}_{h_b}\mathbf{Y}_t)^\otimes]/[(t + 1 + m_0)(n - df)], \quad (2.8)$$

where h_b is a pre-specified bandwidth. Zou *et al.* (2008) provide some practical guidelines about the selection of h_b . This parameter can also be selected before Phase II monitoring from the m_0 IC profiles by certain data-driven procedures, such as CV and GCV.

Remark 3 The statistic $lr_{m_0,t}$ is a little different from its counterpart used in Hawkins *et al.* (2003). Here, $lr_{m_0,t}$ is the maximum of $\tilde{T}_{k,t}$ on a constrained interval $m_0 \leq k < t$, instead of across all possible values of k . Since the first m_0 profiles are IC, this modification should be reasonable. \square

In practice, it might be more convenient to plot the normalized statistic $lr_{m_0,t}/h_{m_0,t,\alpha}$ over t in a control chart. In such cases, the normalized control limit is a constant 1. Besides $h_{m_0,t,\alpha}$, our proposed control chart has a number of other parameters. Selection of all these parameters is discussed in the next two subsections.

2.3. A Bootstrap Procedure for Determining Control Limits

From the construction of $lr_{m_0,t}$, it is easy to check that it does not depend on the IC regression function g . Therefore, $h_{m_0,t,\alpha}$ should not depend on g either. In this part, we propose a bootstrap procedure for determining $h_{m_0,t,\alpha}$ based on the m_0 IC profiles.

First the m_0 IC profiles are averaged, and then the averaged data are smoothed by local linear kernel smoothing. The resulting $m_0 \times n_0$ residuals are

$$\hat{e}_{ij} := y_{ij} - \mathbf{W}_{n_0}(x_i)\bar{\mathbf{Y}}_{0,m_0}, \quad j = 1, \dots, m_0, \quad i = 1, \dots, n_0,$$

where $\mathbf{W}_{n_0}(x_i)$ is the smoothing operator defined immediately below (2.2), using \mathbf{X}_{IC} (instead of \mathbf{X}) and h_b (which is mentioned below (2.8)). Second, generate a bootstrap profile $\{(x_i, y_i^*), i = 1, \dots, n\}$ by defining $y_i^* = e_i^*$, for $i = 1, \dots, n$, where e_i^* is drawn from $\{\hat{e}_{ij}, i = 1, \dots, n_0, j = 1, \dots, m_0\}$ with replacement. Third, for each given t value, by this resampling procedure, simulate the whole monitoring process, including generating the first m_0 IC profiles (at design points \mathbf{X}) and all future profiles, and then computing the corresponding $lr_{m_0,t}^*$ value. Fourth, repeat step three B times. Then, for a given false alarm probability α that corresponds to the IC average run lengths (ARL) $1/\alpha$, the control limits $h_{m_0,t,\alpha}$ can be approximated by values satisfying

$$\begin{aligned} \Pr\left(lr_{m_0,t}^* > h_{m_0,t,\alpha} \mid lr_{m_0,i}^* \leq h_{m_0,i,\alpha}, 1 \leq i < t\right) &= \alpha, \quad \text{for } t > 1, \\ \Pr\left(lr_{m_0,1}^* > h_{m_0,1,\alpha}\right) &= \alpha. \end{aligned}$$

Of course, the above probabilities should be interpreted as frequencies in B bootstrap replications.

Based on our numerical experience and the empirical results in Hawkins, Qiu and Kang (2003), which discusses similar simulation-based control limits, $h_{m_0,t,\alpha}$ gradually converges to a constant as t increases. Therefore, we suggest computing the first (about) $1/(2\alpha)$ control limits and then using the last one of this

sequence to approximate the remaining control limits. In addition, for computing each $h_{m_0,t,\alpha}$, about 10,000 bootstrap replications should be good enough to obtain reliable approximations. For instance, if IC ARL is 200, then we need to compute the first 100 control limits, and this requires about 16,500 bootstrap sequences such that there are about 10,000 sequences left for computing the 100th control limit $h_{m_0,100,\alpha}$. Numerical accuracy of this bootstrap procedure is further investigated in Section 3.

2.4. Practical Guidelines

On choosing m_0 , n_0 and n : From the description of the proposed SPC procedure, we can see that its control limits are computed from the m_0 IC profiles (each with n_0 observations) by the bootstrap procedure. Therefore, both m_0 and n_0 should not be too small. If the IC error distribution F is known, then the control limits can be computed directly from F using the method in Hawkins, Qiu and Kang (2003). In practice, F is often unknown and most practitioners would accumulate a few IC profiles before starting monitoring, which is a major motivation for us to propose the current SPC procedure. Based on our numerical experience, to describe the error distribution reasonably well in various cases, use $n_0 > 50$ and $m_0 \geq 8$. Note that the recommended m_0 here is much smaller than the recommended m_0 in Zou, Tsung and Wang (2008), where $m_0 \geq 40$ is suggested. That is mainly because the m_0 IC profiles are used for estimating both the IC regression function g and the error variance σ^2 in the latter case, while g does not need to be estimated in the former case. Regarding the choice of n and design points positions, one may refer to Zou, Tsung and Wang (2008) for a detailed discussion.

On choosing γ_n , a , h_{\max} and h_{\min} in (2.5): Theoretically speaking, these quantities should satisfy certain conditions to obtain the corresponding asymptotic results. See the Appendix for detailed discussion. In our simulations, we found that performance of the proposed control chart was hardly affected by these parameters, consistent with the findings in Guerre and Lavergne (2005). By both theoretical arguments and numerical studies, we recommend the choices $a = 1.4$, $\gamma_n = 2.5\sqrt{\ln(J_n + 1)}$, J_n could be 4, 5 or 6, $h_{\max} = n^{-1/7}$, and $h_j = h_{\max}a^{-j}$, $j = 1, \dots, J_n$. Note that $h_{\max} = n^{-1/7}$ is recommended partially due to Condition (C5) that $nh^8 \rightarrow 0$.

On computation: To implement the proposed method efficiently, we suggest recording a recursive array of the running total sums $\mathbf{S}_t = \mathbf{S}_{t-1} + \mathbf{Y}_t$. Then, computation in (2.2) can be simplified by using $\bar{\mathbf{Y}}_{0,k} - \bar{\mathbf{Y}}_{k,t} = \mathbf{S}_k/k - (\mathbf{S}_t - \mathbf{S}_k)/(t-k)$. Considering that \mathbf{V}_h can be calculated before monitoring and the estimators of the error variance are also calculated in a recursive way (cf., (2.8)), the computational task involved in our proposed procedure is actually quite simple. For instance, when IC ARL=200 and $n = 35$, it spends about five minutes in searching for the control limits based on 20,000 simulations, using a Pentium 2.4MHz CPU. Computer codes in Fortran are available from the authors upon request.

3. Simulation Study

We present some simulation results in this section regarding the numerical performance of the proposed SPC procedure, called the adaptive change point (ACP) procedure hereafter. In the procedure, parameters were chosen according to the practical guidelines discussed in Subsection 2.4. More specifically, we chose $m_0 = 8$, $n_0 = 50$, $n = 25$, $J_n = 5$, $a = 1.4$, $\gamma_n = 2.5\sqrt{\ln(J_n + 1)}$, $h_{\max} = n^{-1/7}$, and $h_j = h_{\max}a^{-j}$, for $j = 1, \dots, J_n$. The kernel function used in (2.2) was the Epanechnikov kernel function $K(x) = 0.75(1 - x^2)I(-1 \leq x \leq 1)$, which has certain optimality properties (cf., Fan and Gijbels (1996)). The IC ARL was fixed at 200.

We first investigated the performance of IC run-length distribution of the proposed procedure. As mentioned in Section 2.3, the IC distribution of our charting statistic does not depend on the IC regression function g . Hence, we used $g = 0$ in this example. Equally spaced design points $x_i = (i - 0.5)/n$, for $i = 1, \dots, n$, were used in all profiles. For the m_0 IC profiles, another n points i/n , for $i = 1, \dots, n$, were used as extra design points x_{n+1}, \dots, x_{n_0} . The following four error distributions were considered: (I) $\varepsilon_{ij} \sim N(0, 1)$; (II) $\varepsilon_{ij} \sim U(0, 1) - 0.5$; (III) $\varepsilon_{ij} \sim t(5)$; and (IV) $\varepsilon_{ij} \sim \chi^2(5) - 5$, where $U(0, 1)$ denotes the Uniform distribution on $[0, 1]$, $t(5)$ and $\chi^2(5)$ denote the Student- t and chi-squared distributions with degrees of freedom 5, respectively. For comparison purposes, besides the proposed bootstrap procedure for determining the control limits, we also investigated the run-length behavior of the ACP chart when its control limits were approximated by the method used in Hawkins, Qiu and Kang (2003) under the normal error distribution. The sample averages and sample

Table 1. ARL and SDRL values of the ACP chart with and without bootstrap approximations of the control limits

	with bootstrap		without bootstrap	
	ARL	SDRL	ARL	SDRL
(I)	203.1	205.5	200.2	199.7
(II)	205.8	203.6	275.2	269.7
(III)	207.3	218.4	77.5	78.8
(IV)	195.4	212.9	54.6	55.2

standard deviations of the run length, denoted as ARL and SDRL respectively, over 1000 replications are summarized in Table 1.

From Table 1, it can be seen that in case (I), the ARL and SDRL of the ACP chart without using bootstrap approximations are both close to the nominal values 200, as expected. However, in cases (II)-(IV) when F is respectively light-tailed, heavy-tailed and right-skewed, the ACP chart without using bootstrap approximations produced large biases in both IC ARL and SDRL. In comparison, the actual IC ARL and SDRL values of the ACP chart with bootstrap approximations are close to 200 in all cases considered.

Next, we investigated the OC performance of the proposed control chart. To compare it with alternative methods turns out to be difficult, due to lack of an obvious comparable method. One possible alternative method is the NEWMA chart proposed by Zou, Tsung and Wang (2008), although a minor modification is necessary because the original NEWMA chart assumes that the IC regression function g and error variance σ^2 are known and both of them are assumed unknown in the current setting. The NEWMA charting statistic is

$$\mathbf{E}_j = \lambda \mathbf{Z}_j + (1 - \lambda) \mathbf{E}_{j-1}, \quad j = 1, 2, \dots,$$

where $\mathbf{Z}_j = (\mathbf{Y}_j - \mathbf{G})/\sigma$, $\mathbf{G} = (g(x_1), \dots, g(x_n))^T$, \mathbf{E}_0 is a n -dimensional starting vector, and λ is a smoothing parameter. The chart signals if

$$Q_j = \mathbf{E}_j' \mathbf{V}_h \mathbf{E}_j > L \frac{\lambda}{2 - \lambda},$$

where $L > 0$ is a control limit chosen to achieve a specific IC ARL. A natural modification of the NEWMA procedure is to replace g and σ^2 by their sequential estimators, referred to as the self-starting NEWMA procedure (SSN) hereafter. To be specific, if the chart does not signal after the t -th observed profile, then replace \mathbf{G} and σ by $\hat{\mathbf{G}} = \mathbf{W}_h \bar{\mathbf{Y}}_{0,t}$ and $\hat{\sigma}_t$ (cf., (2.8)), respectively. In addition, to

illustrate effectiveness of the adaptive selection of the bandwidth, we also obtain the OC ARLs of the change-point formulation with a fixed h . That is, in this procedure, $T_{h,k,t}$ in (2.2), instead of $\tilde{T}_{k,t}$, is used in (2.7). This procedure is called the fixed change point (FCP) chart hereafter. Another possible method to compare is the “naive” multivariate EWMA approach that treats \mathbf{Z}_j as a long multivariate vector (cf., Lowry *et al.* 1992), referred to as the MEW chart. For all the FCP, SSN and MEW charts, we followed the recommendation of Zou, Tsung and Wang (2008) to choose the bandwidth h to be $h_E = 1.5n^{-1/5}[\sum_{i=1}^n (x_i - \bar{x})^2/n]^{1/2}$.

Because the SSN chart cannot be implemented without knowing the error distribution, the standard normal error distribution was used in this example, and was assumed known to the SSN chart. In both SSN and MEW charts, we chose $\lambda = 0.2$. For ACP, all parameters were chosen as in the previous example. Control limits of all the charts considered here were searched by simulations to attain the nominal IC ARL 200. The IC model used was $g(x_i) = 1 - \exp(-x_i)$, $i = 1, \dots, n$, and the following four OC models were considered: (I) $g_1(x_i) = 1 - \exp(-x_i) + \delta x_i$; (II) $g_1(x_i) = 1 - (1 + \delta) \exp(-x_i)$; (III) $g_1(x_i) = 1 - \exp(-x_i) + \delta \cos(\pi x_i)$; (IV) $g_1(x_i) = 1 - \exp(-x_i) + 0.75 \sin(\delta \pi(x_i - 0.5))$. By changing δ , these models can cover various cases with different smoothness of $g - g_1$ and different shift magnitudes from g to g_1 . We considered the shift time to be $\tau = 20$, and the first m_0 profiles in the sequence of observed profiles were used as IC profiles. Table 2 presents the OC ARLs over 10,000 replications of the ACP, FCP, SSN and MEW charts. In addition, the OC ARLs of the NEWMA chart with known g and σ and with $\lambda = 0.2$ are included in the last column. The control limits L of the EWMA-type control charts are included in the last row of the table.

From the table, we can observe the following results. First, effectiveness of the change-point formulation can be clearly seen by comparing the OC ARLs of the FCP and SSN charts. The FCP chart performed almost uniformly better than the SSN charts, consistent with the findings in Hawkins, Qiu and Kang (2003) about SPC of univariate normal processes, and with the theoretical and empirical results in Han and Tsung (2004) as well about comparison of the EWMA and GLR charts. Second, the ACP chart performed better than the FCP chart

Table 2. OC ARL comparison of ACP, FCP, SSN, NEWMA and MEW charts when IC
 ARL=200, $m_0 = 8$, $n = 25$ and $\tau = 20$.

δ	ACP	FCP	SSN	MEW	NEWMA	
(I)	0.20	132 (1.79)	136 (1.77)	151 (1.95)	177 (2.07)	35.4 (0.32)
	0.30	66.5 (1.25)	78.3 (1.39)	97.1 (1.63)	149 (1.98)	16.5 (0.13)
	0.40	24.8 (0.61)	30.2 (0.63)	50.6 (1.19)	112 (1.77)	9.82 (0.06)
	0.60	6.11 (0.05)	7.31 (0.11)	8.13 (0.18)	37.8 (1.00)	5.20 (0.03)
	0.80	3.44 (0.02)	3.88 (0.02)	4.04 (0.02)	8.53 (0.25)	3.58 (0.01)
	1.00	2.33 (0.01)	2.59 (0.01)	2.96 (0.01)	4.44 (0.02)	2.76 (0.01)
	1.50	1.29 (0.00)	1.39 (0.01)	1.90 (0.01)	2.55 (0.01)	1.88 (0.01)
(II)	0.20	119 (1.81)	123 (1.72)	141 (1.92)	171 (2.06)	28.3 (0.25)
	0.30	45.8 (0.96)	58.1 (1.20)	83.5 (1.59)	134 (1.88)	12.8 (0.09)
	0.40	14.9 (0.36)	19.2 (0.48)	34.5 (0.94)	88.8 (1.58)	7.92 (0.04)
	0.60	4.64 (0.03)	5.30 (0.03)	5.65 (0.05)	18.5 (0.56)	4.43 (0.02)
	0.80	2.72 (0.01)	3.03 (0.02)	3.45 (0.02)	5.33 (0.04)	3.07 (0.01)
	1.20	1.46 (0.01)	1.59 (0.01)	2.13 (0.01)	2.85 (0.01)	2.05 (0.01)
	2.00	1.00 (0.00)	1.00 (0.00)	1.34 (0.01)	1.72 (0.01)	1.31 (0.00)
(III)	0.20	102 (1.55)	111 (1.57)	121 (1.72)	166 (2.06)	24.7 (0.21)
	0.30	32.4 (0.73)	39.3 (0.74)	54.2 (1.14)	123 (1.85)	11.4 (0.08)
	0.40	10.5 (0.13)	13.4 (0.23)	17.5 (0.46)	75.4 (1.49)	7.02 (0.04)
	0.50	5.93 (0.05)	6.94 (0.07)	7.28 (0.13)	33.1 (0.90)	5.10 (0.02)
	0.75	2.70 (0.01)	2.97 (0.02)	3.29 (0.01)	5.25 (0.05)	3.07 (0.01)
	1.00	1.69 (0.01)	1.86 (0.01)	2.34 (0.01)	3.28 (0.01)	2.26 (0.01)
	1.50	1.07 (0.00)	1.12 (0.00)	1.59 (0.01)	2.06 (0.01)	1.57 (0.01)
(IV)	0.25	70.9 (1.32)	85.4 (1.48)	109 (1.72)	152 (1.98)	17.6 (0.14)
	0.50	7.56 (0.07)	8.90 (0.09)	12.4 (0.41)	51.9 (1.20)	5.84 (0.03)
	0.75	3.71 (0.02)	4.20 (0.03)	4.37 (0.03)	9.96 (0.29)	3.78 (0.02)
	1.00	2.70 (0.01)	3.01 (0.02)	3.34 (0.02)	5.28 (0.06)	3.07 (0.01)
	2.00	3.21 (0.02)	3.10 (0.02)	3.36 (0.02)	5.41 (0.10)	3.11 (0.01)
	4.00	4.51 (0.02)	4.16 (0.02)	4.25 (0.03)	6.54 (0.43)	3.69 (0.01)
	6.00	5.81 (0.03)	16.1 (0.28)	15.3 (0.24)	20.2 (1.82)	8.24 (0.04)
L			15.93	45.98	15.77	

NOTE: Standard errors are in parentheses.

in most cases, especially for detecting small and moderate shifts, which demonstrates the benefit of adaptive bandwidth selection. It is worth mentioning that, in case (IV) when $\delta = 6$, $g - g_1$ oscillated dramatically, and the ACP chart performed substantially better. This demonstrates the fact that the adaptive GLR statistic can adapt to the unknown smoothness of $g - g_1$ and pick up smaller bandwidths automatically when detecting more irregular alternatives. See Proposition 2 and Remark A3 in Appendix for more detailed discussion. Of course, one may want to use a relatively small h in the FCP or SSN chart as well in such cases. However, the OC model is often unknown in advance. The FCP or SSN chart with a fixed h can outperform the ACP chart for certain OC models, but they can also be much worse for other OC models. As a comparison, the ACP chart can always be nearly the best. Third, we found that the ACP chart can even outperform the NEWMA chart constructed with known g and σ for detecting moderate and large shifts. In fact, our simulations (not reported here) showed that, when $\tau \geq 120$, the ACP chart outperformed the NEWMA chart nearly uniformly, which is mainly due to the joint benefits of the adaptive GLR test and the change-point approach. Regarding the MEW chart, as expected, it did not perform well in all cases considered, because it completely ignores the profile structure of the data. We conducted some other simulations with various F , n and τ , to check whether the above conclusions would change in other cases. These simulation results, not reported here but available from the authors, showed that the ACP chart works well for other error distributions as well in terms of its OC ARL, and its good performance still holds for other choices of n and τ .

4. Application to Monitoring a Deep Reactive Ion Etching Process

In this section, we demonstrate the ACP chart by applying it to a dataset obtained from the semiconductor manufacturing industry for monitoring a deep reactive ion etching (DRIE) process that is critical to the output wafer quality and that requires careful monitoring. In the DRIE process, the desired profile is the one with smooth and straight sidewalls and flat bottoms, and ideally the sidewalls should be perpendicular to the bottom of the trench with certain degrees of smoothness around the corners. Various shapes of profiles, such as positive and negative profiles, which are due to underetching and overetching, are considered to be unacceptable (cf., Figure 1). More detailed introduction about the DRIE

process can be found in Rauf *et al.* (2002) and Zhou *et al.* (2004).

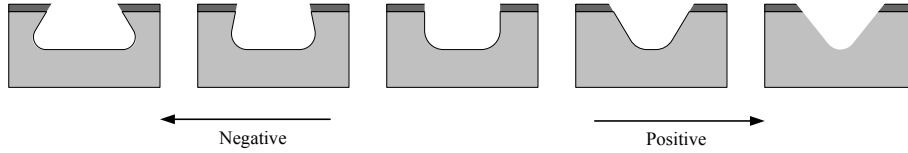


Figure 1. Illustrations of various etching profiles from a DRIE process.

Table 3. Results of various charts for monitoring the DRIE dataset. Note that profile monitoring starts at $t = 9$ since the first 8 profiles are treated as IC ones.

t	ACP					FCP	SSN	MEW
	$lr_{m_0,t}$	$\hat{\sigma}_t$	$h_{m_0,t,\alpha}$	$lr_{m_0,t}/h_{m_0,t,\alpha}$	$\hat{\tau}$			
9	0.740	0.431	6.526	0.113	8	0.342	2.849	18.64
10	0.924	0.420	6.816	0.136	8	0.357	4.832	23.30
11	2.493	0.423	7.112	0.351	8	0.639	13.29	47.84
12	1.957	0.420	7.194	0.272	8	0.611	11.76	37.34
13	1.126	0.418	7.437	0.151	8	0.440	6.755	30.00
14	1.573	0.416	7.741	0.203	8	0.320	2.490	24.23
15	2.423	0.411	8.136	0.298	14	0.581	5.512	29.39
16	6.771	0.411	8.487	0.798	14	0.816	15.32	36.90
17	9.053	0.408	8.702	1.040	14	1.109	16.89	40.49
							20.05	61.93

This dataset has been analyzed by Zou, Tsung and Wang (2008), and its details can also be found in Wang and Tsung (2007) and the references cited therein. In the dataset, the first 18 profiles are known to be IC. In order to demonstrate the effectiveness of our proposed approach, we only take the first eight IC profiles for implementing the bootstrap procedure and discard the remaining ten. In each of these eight IC profiles, profile dimensional readings are collected at seventy design points (i.e., $n_0 = 70$), which satisfies the requirements of the bootstrap procedure. Hence, the desired IC performance could be well approximated even if we do not make assumptions on the error distribution. As in Zou, Tsung and Wang (2008), IC ARL is fixed at 370. All other parameters of the ACP chart are chosen as in the simulation examples. Then the ACP chart was used for monitoring Phase II profiles, each of which has $n = 35$ observations. As detailed in Zou, Tsung and Wang (2008), there are nine Phase II profiles in the dataset, and the last three are classified as inferior profiles based on engineering knowledge. Table 3 tabulates the statistics $lr_{m_0,t}$, $\hat{\sigma}_t$, the bootstrap

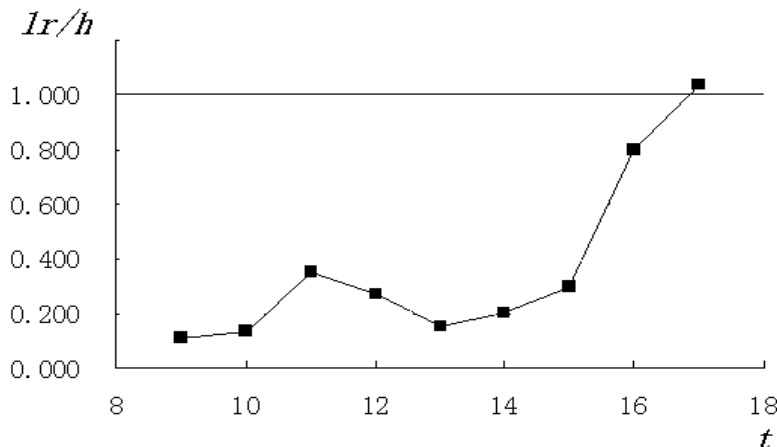


Figure 2. The ACP control chart for Phase II monitoring of the DRIE process. Note that profile monitoring starts at $t = 9$ since the first 8 profiles are treated as IC ones.

approximated control limits $h_{m_0,t,\alpha}$, the ratio $lr_{m_0,t}/h_{m_0,t,\alpha}$, and the estimated change-point $\hat{\tau}$. The corresponding control chart based on the ratio charting statistic $lr_{m_0,t}/h_{m_0,t,\alpha}$ is shown in Figure 2, where the control limit is 1. As a comparison, corresponding results of the FCP, SSN, and MEW charts are also presented in this table. As in the simulation examples, we chose $\lambda = 0.2$ in the SSN and MEW charts. Their control limits are given in the last row of the table.

From Table 3 and Figure 2, it can be seen that the ACP chart signals a shift after the 17-th profile is monitored, which matches the NEWMA chart of Zou, Tsung and Wang (2008). As a by-product, the current chart gives a change-point estimate of $\hat{\tau} = 14$. In Zou, Tsung and Wang (2008), a separate diagnostic method needs to be used. But it gives exactly the same result about the shift position. From the table, we can see that the FCP chart also gives a signal at $t = 17$, but the SSN and MEW charts do not give any signal by that time.

5. Concluding Remarks

In this paper, we propose a control chart for monitoring nonparametric profiles. Our proposed control chart integrates the dynamic change-point approach with the adaptive generalized likelihood ratio test. A bootstrap procedure is suggested for determining its control limits without specifying the error distribution. This method does not assume the IC regression function g or the error variance

σ^2 to be known, which could substantially shorten the period of Phase I analysis. The proposed control chart offers not only a balanced protection against shifts of different magnitudes, but also adapts to different smoothness of the IC and OC regression functions. Consequently, it has a nearly optimal performance for various OC conditions. As demonstrated by the DRIE example, the proposed approach can be implemented conveniently in industrial applications. It should be quite effective as long as a few IC profiles are available. Besides these properties, the ACP chart can be easily extended to cases of unequally spaced or even random design points, or cases in which monitoring of both the regression function and the error variance is of interest. These generalizations would require certain modifications in the GLR statistic (2.2).

Acknowledgments

The authors thank the editor, an associate editor, and two referees for many constructive comments that greatly improved the article. This research is supported in part by a grant from NSF of USA. Changliang Zou's research is also supported in part by the grant No. 10771107 from NNSF of China.

Appendix: Some Statistical Properties of the GLR Test $\tilde{T}_{k,t}$

We now present some statistical properties of the GLR test statistic $\tilde{T}_{k,t}$. First, a set of conditions is presented for later use. Without loss of generality, we take $0 \leq x_1 \leq x_2 \leq \dots \leq x_n \leq 1$.

Conditions:

(C1) There exists a positive density f that is Lipschitz continuous and bounded away from zero such that $\int_0^{x_i} f(u) du = i/n$, $i = 1, \dots, n$.

(C2) Both g and g_1 have continuous second derivatives in $[0, 1]$.

(C3) $K(u)$ is symmetric and bounded, $u^3K(u)$ and $u^3K'(u)$ are bounded, and $\int u^4K(u) du < \infty$.

(C4) $E(|\varepsilon_{11}|^4) < \infty$.

(C5) $h = h_n$ satisfies $h \rightarrow 0$, $nh^3 \rightarrow \infty$, and $nh^8 \rightarrow 0$.

(C6) The penalty sequence γ_n is of order $\sqrt{\ln \ln n}$.

Remark A1 By condition C5, the set of bandwidths \mathcal{H}_n (cf., (2.4)) should roughly satisfy $h_{\max} = O(n^{-\frac{1}{8}-s_1})$ and $h_{\min} = O(n^{-\frac{1}{3}+s_2})$, where s_1 and s_2 are two small positive constants. \square

Proposition 1 will establish the asymptotic null distribution of $\tilde{T}_{k,t}$, its proof requires the following lemmas.

Lemma 1 *If C1-C5 hold, then $\hat{\sigma}_t^2 = \sigma^2 [1 + O_p((tn)^{-1/2}) + O_p((nh)^{-1})]$.*

This lemma can be proved easily by combining proofs in Hall and Marron (1990), about the Nadaraya-Watson estimator, with certain \mathcal{L}_r -convergence properties of LLKE (Fan (1993); Fan and Gijbels (1996)). The proof is omitted.

Lemma 2 *Under C1-C5 and the null hypothesis, $(T_{h,k,t} - \check{\mu}_h)/\check{\sigma}_h \xrightarrow{\mathcal{L}} N(0, 1)$, where*

$$\check{\mu}_h = \frac{2}{h} \left(K(0) - \frac{1}{2} \int K^2(t) dt \right), \quad \check{\sigma}_h^2 = \frac{8}{h} \int \left(K(t) - \frac{1}{2} K * K(t) \right)^2 dt.$$

Proof. Note that $[k(t-k)/t]^{\frac{1}{2}} (\bar{Y}_{0,k} - \bar{Y}_{k,t})$ can be rewritten as $\boldsymbol{\xi}_{k,t} = (\xi_{k,t,1}, \dots, \xi_{k,t,n})^T$, where $\xi_{k,t,i} = [k(t-k)/t]^{\frac{1}{2}} (\bar{\varepsilon}_{0,k,i} - \bar{\varepsilon}_{k,t,i})$ and $\bar{\varepsilon}_{k,t,i} = \sum_{j=k+1}^t \varepsilon_{ij}/(t-k)$. Obviously, $\xi_{k,t,i}$ satisfies $E(\xi_{k,t,i}) = 0$, $E(\xi_{k,t,i}^2) = \sigma^2$, $E(|\xi_{k,t,i}|^4) < \infty$, and $\xi_{k,t,i}$ and $\xi_{k,t,j}$ are independent of each other when $i \neq j$. Thus, this lemma follows by the technical arguments in the proof of Theorem 5 in Fan, Zhang and Zhang (2001), and by the arguments on modification of conditions in Proposition 1 of Zou, Tsung and Wang (2008). \square

Lemma 3 *Under C1-C5, we have $\hat{v}_h^2 = v_h^2 + o_p(h^{-1})$ and $\hat{v}_{h,h_0}^2 = v_{h,h_0}^2 + o_p(h^{-1})$.*

This lemma can be easily proved by a direct calculation using Lemma 7.1 in Fan, Zhang and Zhang (2001).

Lemma 4 *Let $T_{h,k,t}^c = T_{h,k,t} - \mu_h$, and $\boldsymbol{\nu}_h$ be a $n \times n$ matrix with diagonal elements zero and (i, j) -th off-diagonal elements $V^{(ij)}$. Under conditions C1-C6, if $\gamma_n > (1+c)\sqrt{2 \ln J_n}$ for some $c > 0$, then*

$$(i) \hat{v}_{h,h_0} = O_p(h^{-1} - h_0^{-1})^{1/2}; \quad (ii) \frac{\boldsymbol{\xi}_{k,t}^T (\boldsymbol{\nu}_h - \boldsymbol{\nu}_{h_0}) \boldsymbol{\xi}_{k,t}}{v_{h,h_0}} \xrightarrow{\mathcal{L}} N(0, 1); \text{ and}$$

$$(iii) \max_{h \in \mathcal{H}_n \setminus \{h_0\}} \left| \frac{T_{h,k,t}^c - T_{h_0,k,t}^c}{\hat{v}_{h,h_0}} \right| = (1 + o_p(1)) \times \max_{h \in \mathcal{H}_n \setminus \{h_0\}} \left| \frac{\boldsymbol{\xi}_{k,t}^T (\boldsymbol{\nu}_h - \boldsymbol{\nu}_{h_0}) \boldsymbol{\xi}_{k,t}}{v_{h,h_0}} \right| + o_p(1).$$

Results (i) and (ii) above can be shown by similar arguments to those in the proof of Lemma 2; (iii) is a direct conclusion of (i).

Proposition 1 *Under C1-C6 and the null hypothesis, if $\gamma_n > (1+c)\sqrt{2 \ln J_n}$ for some $c > 0$, then $\tilde{T}_{k,t} \xrightarrow{\mathcal{L}} N(0, 1)$.*

Proof: By (i) and (iii) of Lemma 4,

$$\begin{aligned} Pr(\tilde{h} \neq h_0) &= Pr\left(\max_{h \in \mathcal{H}_n \setminus \{h_0\}} \left| \frac{T_{h,k,t}^c - T_{h_0,k,t}^c}{\widehat{v}_{h,h_0}} \right| > \gamma_n\right) \\ &\leq Pr\left(\max_{h \in \mathcal{H}_n \setminus \{h_0\}} \left| \frac{\boldsymbol{\xi}_{k,t}^T (\boldsymbol{\nu}_h - \boldsymbol{\nu}_{h_0}) \boldsymbol{\xi}_{k,t}}{v_{h,h_0}} \right| \geq \frac{\gamma_n}{1+c}\right) + o_p(1). \end{aligned}$$

It can be easily checked that $\boldsymbol{\nu}_h - \boldsymbol{\nu}_{h_0}$ satisfies the conditions of Lemma 2 in Guerre and Lavergne (2005). Using (ii) of that lemma, and Lemmas 1-4 above, the remaining part of the proof can be completed by the same arguments as those in the proof of Theorem 1 in Guerre and Lavergne (2005). \square

Proposition 2 below considers the consistency of $\tilde{T}_{k,t}$ under local alternatives, in which we use the notation

$$\begin{aligned} \theta &= k/t, \quad \delta(x) = g_1(x) - g(x), \quad \eta_1 = \int K(t)t^2 dt, \\ \eta_2 &= 8 \int \left(K(t) - \frac{1}{2} K * K(t) \right)^2 dt, \quad \zeta_\delta = \frac{t\theta(1-\theta)}{\sigma^2} \int \delta^2(u) f(u) du, \\ \zeta_1 &= \frac{t\theta(1-\theta)\eta_1^2}{4\sigma^2} \int [\delta''(u)]^2 f(u) du. \end{aligned}$$

Proposition 2 *Assume that C1-C6 hold, and that ζ_1 is of an order within $[n^{-7/16}\gamma_n, n^{1/2}\gamma_n]$. Then the asymptotic power of the adaptive GLR test is at least*

$$\Phi\left[n^{8/9}\zeta_\delta\left(\frac{\gamma_n\eta_2^{1/2}}{8\zeta_1}\right)^{1/9} - \frac{9}{8}\gamma_n\eta_2^{1/2}\right];$$

thus, the test is consistent when $\zeta_\delta \geq c\zeta_1^{1/9}\left(\gamma_n\eta_2^{1/2}/n\right)^{8/9}$ and $c > 0$ is large enough.

Proof. First, the power of the test satisfies

$$P(\tilde{T}_{k,t} \geq v_{h_0} z_\alpha) \geq P(T_{h,k,t} - \mu_h \geq v_{h_0} z_\alpha + \gamma_n \widehat{v}_{h,h_0}), \quad (\text{A.1})$$

where z_α is the $(1 - \alpha)$ -quantile of the standard normal distribution. Thus, the adaptive GLR test inherits the power properties of $T_{h,k,t}$ up to $\gamma_n \widehat{v}_{h,h_0}$. See Guerre and Lavergne (2005) for related discussions.

Under a local alternative δ , by similar mathematical manipulations to those in the proof of Theorem 7 in Fan, Zhang and Zhang (2001), we have

$$T_{h,k,t} - \mu_h = h^{-\frac{1}{2}}w + n\zeta_\delta(1 + o_p(1)) - nh^4\zeta_1, \quad (\text{A.2})$$

where $w = w_n$ is asymptotically normal with variance η_2 . Since \widehat{v}_{h,h_0} and \widehat{v}_h are of order $h^{-\frac{1}{2}}$, we can find an appropriate h in \mathcal{H}_n , say h_n , such that $n\zeta_\delta - nh^4\zeta_1 - \gamma_n\widehat{v}_{h,h_0}$ attains its maximum value asymptotically. Such an h_n can be defined by $h_n = h_0a^{-j_n}$, where j_n is the integer part of $(2/(9\ln a))\ln[8n\zeta_1h_0/(\gamma_n\eta_2^{1/2})]$. Note that h_n is indeed in \mathcal{H}_n when ζ_1 satisfies the condition imposed. After substituting h_n into (A.2), Proposition 2 follows from (A.1). \square

Remark A2 By (A.2), we observe that the asymptotic power of the test statistic $T_{h,k,t}$ depends not only on $\delta(\cdot)$, but also on $\delta''(\cdot)$. The term $nh^4\zeta_1$ in (A.2) explains, intuitively, the major reason why choosing the smoothing parameter h properly would gain test power. Practically, a smaller h is often more effective in detecting shifts with sharp or oscillating δ (i.e., δ'' is large), and a larger h often performs better when δ is flat or smooth (i.e., δ'' is small). This observation motivates us to use the adaptive selection procedure when conducting the GLR test. From the proof of Proposition 2, we can see that, to attain the stated asymptotic power, the order of “optimal” h should be $[\gamma_n\eta_2^{1/2}/(n\zeta_1)]^{2/9}$. Thus, the test based on $\widetilde{T}_{k,t}$ would adapt to different magnitudes of δ'' . Consequently, $\widetilde{T}_{k,t}$ would be more robust to unknown alternatives than the test based on $T_{h,k,t}$ with a fixed bandwidth. \square

References

- Colosimo, B. M. and Pacella, M. (2007). On the use of principle component analysis to identify systematic patterns in roundness profiles. *Qual. Reliab. Eng. Int.* **23**, 707-725.
- Dette, H. and Neumeier, N. (2001). Nonparametric analysis of covariance. *Ann. Statist.* **29**, 1361-1400.
- Ding, Y., Zeng, L. and Zhou, S. (2006). Phase I analysis for monitoring nonlinear profiles in manufacturing processes. *J. Qual. Technol.* **38**, 199-216.
- Fan, J. (1993). Local linear regression smoothers and their minimax efficiencies. *Ann. Statist.* **21**, 196-216.
- 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,” *Ann. Statist.* **29**, 153-193.
- Gombay, E. (2000). Sequential change-point detection with likelihood ratios. *Statist. Probab. Lett.* **49**, 195-204.
- Guerre, E. and Lavergne, P. (2005). Data-driven rate-optimal specification testing in regression models. *Ann. Statist.* **33**, 840-870.
- Hall, P. and Hart, J. D. (1990). Bootstrap test for difference between means in nonparametric regression. *J. Amer. Statist. Assoc.* **85**, 1039-1049.
- Hall, P. and Marron, J. S. (1990). On variance estimation in nonparametric regression. *Biometrika* **77**, 415-419.
- Han, D. and Tsung, F. (2004). A generalized EWMA control chart and its comparison with the optimal EWMA, CUSUM and GLR schemes,” *Ann. Statist.* **32**, 316-339.
- Hart, J. D. (1997). *Nonparametric Smoothing and Lack-of-Fit Tests*, Springer, New York.
- Hawkins, D. M., Qiu, P. and Kang, C. W. (2003). The changepoint model for statistical process control. *J. Qual. Technol.* **35**, 355-366.

- Hawkins, D. M. and Zamba, K. D. (2005a). Change point model for statistical process control with shift in variance. *J. Qual. Technol.* **37**, 21-31.
- (2005b). Statistical process control for shift in mean or variance using the change point formulation. *Technometrics* **47**, 164-173.
- 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.
- Jensen, W. A., Jones, L. A., Champ, C. W. and Woodall, W. H. (2006). Effects of parameter estimation on control chart properties: a literature review. *J. Qual. Technol.* **38**, 349-364.
- Jones, L. A. (2002). The statistical design of EWMA control charts with estimated parameters. *J. Qual. Technol.* **34**, 277-288.
- Kang, L. and Albin, S. L. (2000). On-line monitoring when the process yields a linear profile. *J. Qual. Technol.* **32**, 418-426.
- Kazemzadeh, R. B., Noorossana, R. and Amiri, A. (2008). Phase I monitoring of polynomial profiles. *Commun. Stat. Theory Methods*. **37**, 1671-1686.
- Kim, K., Mahmoud, M. A. and Woodall, W. H. (2003). On the monitoring of linear profiles. *J. Qual. Technol.* **35**, 317-328.
- Kulasekera, K. B. and Wang, J. (1997). Smoothing parameter selection for power optimality in testing of regression curves. *J. Amer. Statist. Assoc.* **92**, 500-511.
- Lada, E. K., Lu, J. -C. and Wilson, J. R. (2002). A wavelet-based procedure for process fault detection. *IEEE Trans. Semicond. Manuf.* **15**, 79-90.
- Lai, T. L. (2001). Sequential analysis: some classical problems and new challenges. *Statist. Sinica*. **11**, 303-350.
- Lowry, C. A., Woodall, W. H., Champ, C. W. and Rigdon, S. E. (1992). Multivariate exponentially weighted moving average control chart. *Technometrics* **34**, 46-53.
- Mahmoud, M. A., Parker, P. A., Woodall, W. H. and Hawkins, D. M. (2007). A change point method for linear profile data. *Qual. Reliab. Eng. Int.* **23**, 247-268.
- Mahmoud, M. A. and Woodall, W. H. (2004). Phase I analysis of linear profiles with calibration applications. *Technometrics* **46**, 380-391.
- Neumeyer, N. and Dette, H. (2003). Nonparametric comparison of regression curves: an empirical process approach. *Ann. Statist.* **31**, 880-920.
- Pignatiello, J. J. Jr. and Simpson, J. R. (2002). A magnitude-robust control chart for monitoring and estimating step changes for normal process means. *Qual. Reliab. Eng. Int.* **18**, 429-441.
- Pollak, M. and Siegmund, D. (1991). Sequential detection of a change in normal mean when the initial value is unknown. *Ann. Statist.* **19**, 394-416.
- Rauf, S., Dauksher, W. J., Clemens, S. B. and Smith, K. H. (2002). Model for a multiple-step deep Si etch process. *J. Vac. Sci. Technol. A* **20**, 1177-1190.
- Siegmund, D. and Venkatraman, E. S. (1995). Using the generalized likelihood ratio statistics for sequential detection of a change-point. *Ann. Statist.* **23**, 255-271.
- Wang, K. and Tsung, F. (2007). Run-to-run process adjustment using categorical observations. *J. Qual. Technol.* **39**, 312-325.
- 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. *J. Agric. Biol. Environ. Stat.* **12**, 216-235.
- Williams, J. D., Woodall, W. H. and Birch, J. B. (2007). Statistical monitoring of nonlinear product and process quality profiles. *Qual. Reliab. Eng. Int.* **23**, 925-941.
- Woodall, W. H., Spitzner, D. J., Montgomery, D. C. and Gupta, S. (2004). Using control charts to monitor process and product quality profiles. *J. Qual. Technol.* **36**, 309-320.
- Young, S. G. and Bowman, A. W. (1995). Non-parametric analysis of covariance. *Biometrics* **51**, 920-931.
- Zamba, K. D. and Hawkins, D. M. (2006). A multivariate change-point for statistical process control. *Technometrics* **48**, 539-549.
- Zhou, R., Zhang, H., Hao, Y. and Wang, Y. (2004). Simulation of the Bosch process with a string-cell hybrid method. *J. Micromech. Microeng.* **14**, 851-858.

- Zou, C., Tsung, F. and Wang, Z. (2007). Monitoring general linear profiles using multivariate EWMA schemes. *Technometrics* **49**, 395-408.
- Zou, C., Tsung, F. and Wang, Z. (2008). Monitoring profiles based on nonparametric regression methods. *Technometrics*, **50**, 512-526.
- Zou, C., Zhang, Y. and Wang, Z. (2006). Control chart based on change-point model for monitoring linear profiles. *IIE Trans.* **38**, 1093-1103.

Department of Statistics, Nankai University, Tianjin, 300071, China

E-mail: (chlzou@yahoo.com.cn)

School of Statistics, University of Minnesota, Minneapolis, MN 55455, USA

E-mail: (qiu@stat.umn.edu)

School of Statistics, University of Minnesota, Minneapolis, MN 55455, USA

E-mail: (doug@stat.umn.edu)