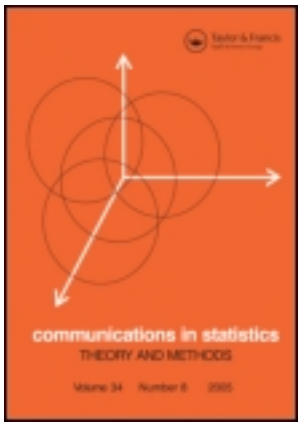


Communications in Statistics - Theory and Methods
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lsta20

# Monitoring Nonlinear Profiles with Random Effects by Nonparametric Regression

Jyh-Jen Horng Shiau [a], Hsiang-Ling Huang [a], Shuo-Hui Lin [a] & Ming-Ye Tsai [a]
[a] Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan
Published online: 24 Apr 2009.

To cite this article: Jyh-Jen Horng Shiau, Hsiang-Ling Huang, Shuo-Hui Lin & Ming-Ye Tsai (2009) Monitoring Nonlinear Profiles with Random Effects by Nonparametric Regression, Communications in Statistics - Theory and Methods, 38:10, 1664-1679, DOI: 10.1080/03610920802702535

To link to this article: http://dx.doi.org/10.1080/03610920802702535

*Communications in Statistics—Theory and Methods*, 38: 1664–1679, 2009

ISSN: 0361-0926 print/1532-415X online
DOI: 10.1080/03610920802702535

Taylor & Francis
Taylor & Francis Group



## Regression Analysis

# Monitoring Nonlinear Profiles with Random Effects by Nonparametric Regression

JYH-JEN HORNG SHIAU, HSIANG-LING HUANG, SHUO-HUI LIN, AND MING-YE TSAI

Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan

*The monitoring of process/product profiles is presently a growing and promising area of research in statistical process control. This study is aimed at developing monitoring schemes for nonlinear profiles with random effects. We utilize the technique of principal components analysis to analyze the covariance structure of the profiles and propose monitoring schemes based on principal component (PC) scores. The number of the PC scores used in constructing control charts is crucial to the detecting power. In the Phase I analysis of historical data, due to the dependency of the PC-scores, we adopt the usual Hotelling $T^2$ chart to check the stability. For Phase II monitoring, we study individual PC-score control charts, a combined chart scheme that combines all the PC-score charts, and a $T^2$ chart. Although an individual PC-score chart may be perfect for monitoring a particular mode of variation, a chart that can detect general shifts, such as the $T^2$ chart and the combined chart scheme, is more feasible in practice. The performances of the schemes under study are evaluated in terms of the average run length.*

**Keywords** Average run length; Control charts; Nonlinear profile monitoring; Principal components analysis; Profile-to-profile variation; Spline smoothing.

**Mathematics Subject Classification** 62P30; 62H25; 62G08.

## 1. Introduction

Statistical process control (SPC) has been widely applied in many areas, especially in industries. Classical methods for SPC are usually developed for applications in which the quality of the product or process can be measured by one or multiple quality characteristics. However, in many situations, the quality-related response of interest is not a single variable but a function of some independent variables. Such a functional response is referred to as a profile in the literature.

Kang and Albin (2000) presented an example of linear profiles from an etching process in semiconductor manufacturing, in which the etching quality depends on

Received June 30, 2008; Accepted December 18, 2008
Address correspondence to Jyh-Jen Horng Shiau, Institute of Statistics, National Chiao Tung University, Hsinchu 30010, Taiwan; E-mail: jyhjen@stat.nctu.edu.tw

the performance of the mass flow controller (MFC). If an MFC is in control, the measured pressure (the response $y$) in the chamber is approximately a linear function of the gas flow (the independent variable $x$). Another example of linear profiles regarding a calibration process was presented in Mahmoud and Woodall (2004).

An example of nonlinear profiles was described in Kang and Albin (2000) but not studied, regarding the dissolving process of aspartame (an artificial sweetener) in which the quality is characterized by the amount of aspartame dissolved per liter of water at different levels of temperature. For illustration, Fig. 1 shows the plot of four hypothetical aspartame profiles. Walker and Wright (2002) presented another example of nonlinear profiles, namely, vertical density profiles (VDP) of particle boards. The density of a particle board was measured with a profilometer that used a laser device to take measurements at fixed depths across the thickness of the engineered wood board. The data set is available at http://bus.utk.edu/stat/walker/VDP/Allstack.TXT.

Profile monitoring is a relatively new research area in quality control. Kang and Albin (2000) studied the problem of linear profile monitoring and proposed two control schemes by modeling the profiles with the simple linear regression model, $Y = A_0 + A_1 x + \epsilon$, where $Y$ is the response variable and $x$ is the independent variable; $A_0$ and $A_1$ are the parameters to be estimated; the noise variables $\epsilon$'s are independent and normally distributed with mean zero and common variance $\sigma^2$. By centering the $x$-values to make the least squares estimators of the $Y$-intercept ($A_0$) and slope ($A_1$) independent of each other, Kim et al. (2003) proposed

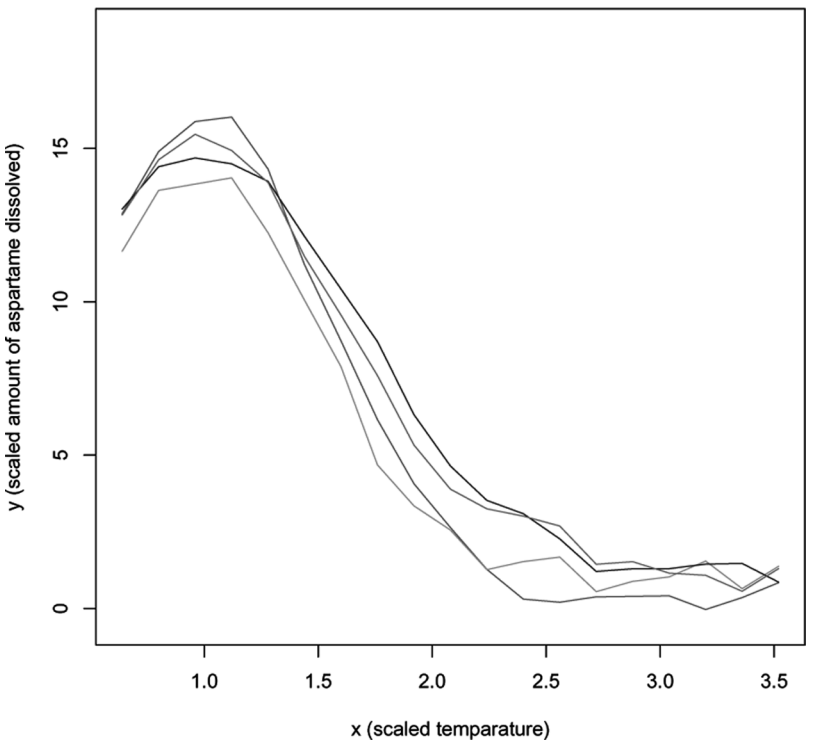


**Figure 1.** Four hypothetical aspartame profiles.

a combined-chart scheme in which three EWMA charts designed, respectively, for detecting shifts in intercept, slope, and standard deviation ($\sigma$) are used simultaneously. Mahmoud and Woodall (2004) presented and compared several control charts for Phase I analysis of linear profiles and applied some of the charts to a calibration application. For more discussions on linear profile monitoring, see the review paper by Woodall et al. (2004).

Shiau and Weng (2004) extended the above linear profile monitoring schemes to a scheme suitable for profiles of more general forms via nonparametric regression. No assumptions are made for the form of the profiles except the smoothness. The nonparametric regression model considered is $Y = g(x) + \epsilon$, where $g(x)$ is a smooth function and $\epsilon$ is the random error as before. Spline regression was adopted as the curve fitting/smoothing technique for its simplicity. They proposed an EWMA chart for detecting mean shifts, an R chart for variation changes, and an EWMSD (standard deviation) chart for variation increases.

Note that the models described above all consist of a deterministic line/curve plus random noises. It does not account for some allowable profile-to-profile variations that we often observe in many profile data, e.g., the aspartame example and VDP example, where these profile-to-profile variations should be considered as caused by common causes. A monitoring scheme constructed based on the aforementioned "fixed-effect" model may interpret these common-cause variations as caused by some special causes and signal many false alarms. Thus, we need a suitable model that can cope with these common-cause variations and construct a monitoring scheme accordingly.

For this, Shiau et al. (2006a) considered a random-effect linear model to develop monitoring schemes for linear profiles. Similarly, Jensen et al. (2006a) proposed a linear mixed (effects) model for linear profiles. Williams et al. (2003) fitted nonlinear profiles by nonlinear parametric regression and then monitored profiles with some $T^2$ statistics of the estimated parameters. Later, Williams et al. (2007) extended this methodology to nonlinear profiles with a non-constant variance at set points to analyze a set of heteroscedastic dose-response profiles. Adopting a random-effect parametric nonlinear regression model for profiles, Shiau et al. (2006b) proposed a robust nonlinear profile monitoring scheme. Jensen et al. (2006b) proposed using nonlinear mixed models to model nonlinear profiles. Note that the parametric approaches mentioned above all need to pre-specify a parametric functional form for profiles, a task often not so easy for practitioners.

In this article, we extend the nonparametric fixed-effect model of Shiau and Weng (2004) to a random-effect model in order to incorporate some profile-to-profile variability as caused by common causes. With the random-effect model, we focus on the covariance structure and use the principal components analysis (PCA) to analyze it. Ding et al. (2006) also considered modeling profiles nonparametrically for a Phase I analysis, but proposed using ICA (independent components analysis) instead of PCA for monitoring profiles that are in clusters, a situation PCA may fail to preserve the clustering feature of the original data.

PCA is very useful in summarizing and interpreting a set of profile data with the same equally spaced $x$-values for each profile. We remark that the smoothing step described above can relax this requirement for profile data since the equally spaced data can be obtained from the smoothed profiles easily. Some pioneer works on analyzing curves with PCA include Castro et al. (1986), Rice and Silverman (1991), Jones and Rice (1992), and others. For applications, Shiau and Lin (1999) analyzed

a set of accelerated LED degradation profiles to estimate the mean lifetime of the product with the techniques of nonparametric regression and PCA.

In this article, for Phase I profile monitoring, we propose using the usual Hotelling $T^2$ chart, a commonly used control chart designed for multivariate process data, by treating the principal component (PC) scores of a profile obtained from PCA as the multivariate data. For Phase II process monitoring, we propose and study three monitoring schemes constructed by utilizing the eigenvalues and eigenvectors obtained from PCA to compute the PC-scores of each incoming profile, including individual PC-score charts, a combined chart that combines all of the PC-score charts and a $T^2$ chart (different from the $T^2$ chart of Phase I). The performances of these monitoring schemes are evaluated in terms of the average run length (ARL).

The rest of the article is organized as follows. Section 2 describes the proposed monitoring schemes in details. Section 3 shows some simulation results of Phase I study and a comparative study of the proposed schemes based on ARL for Phase II monitoring. Section 4 presents a case study using the VDP data from Walker and Wright (2002). Finally, Sec. 5 concludes the article with a brief summary and some remarks.



## 2. Proposed Monitoring Schemes

### 2.1. *A Motivated Example*

This study was motivated by the aspartame example given in Kang and Albin (2000). Since no data are available, a profile of the form $Y = I + Me^{N(x-1)^2} + \epsilon$ is used to mimic an aspartame profile. Then the idea is to perturb the parameters $I, M, N$ randomly to create allowable profile-to-profile variations for an in-control process.

Thus, the following random-effect model *was* considered to generate aspartame profiles:

$$Y_j = I + Me^{N(x_j-1)^2} + \epsilon_j, \quad j = 1, \ldots, p, \tag{1}$$

where $I \sim N(\mu_I, \sigma_I^2)$, $M \sim N(\mu_M, \sigma_M^2)$, $N \sim N(\mu_N, \sigma_N^2)$, $\epsilon \sim N(0, \sigma_\epsilon^2)$, and all the random components are independent of each other. Unfortunately, the response profile $\boldsymbol{Y} = (Y_1, \ldots, Y_p)'$ of model (1) has a complicated distribution with mean $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)'$ and covariance matrix $\boldsymbol{\Sigma}$ as follows. For $i, j = 1, \ldots, p$,

$$\begin{aligned} \mu_j &= E(Y_j) = \mu_I + \mu_M e^{\mu_N (x_j-1)^2 + \frac{\sigma_N^2 (x_j-1)^4}{2}}, \\ \mathrm{Cov}(Y_i, Y_j) &= \sigma_I^2 + (\mu_M^2 + \sigma_M^2)\left[ e^{\mu_N[(x_i-1)^2+(x_j-1)^2] + \frac{\sigma_N^2[(x_i-1)^2+(x_j-1)^2]^2}{2}} \right] \\ &\quad - \mu_M^2 e^{\mu_N (x_i-1)^2 + \frac{\sigma_N^2 (x_i-1)^4}{2} + \mu_N (x_j-1)^2 + \frac{\sigma_N^2 (x_j-1)^4}{2}} + \sigma_\varepsilon^2 \delta_{ij}, \end{aligned} \tag{2}$$

where $\delta_{ij} = 1$ if $i = j$; and 0 otherwise. Note that, by (2), the covariance matrix $\boldsymbol{\Sigma}$ will be changed if the mean of $M$ or $N$ shifts, a situation too complicated to analyze the performance of the control charts under study.

So, instead, we model the aspartame profiles as realizations of a Gaussian stochastic process with the mean function

$$\mu(x) = \mu_I + \mu_M e^{\mu_N (x-1)^2} \tag{3}$$

and a covariance function $G(s, t)$, where $s, t$ are in the domain of $x$. To retain a similar profile-to-profile variation as it would be in the random-effect model (1), we let the in-control profiles follow $MVN(\boldsymbol{\mu}_0, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}_0 = (\mu_{01}, \ldots, \mu_{0p})'$ with

$$\mu_{0j} = \mu_I + \mu_M e^{\mu_N (x_j-1)^2}, \quad j = 1, \ldots, p, \tag{4}$$

and $\boldsymbol{\Sigma}$ is the covariance matrix given by equation (2).

When the mean function (3) is shifted, say, $\mu_I$ to $\mu_I + \alpha\sigma_I$, $\mu_M$ to $\mu_M + \beta\sigma_M$, and $\mu_N$ to $\mu_N + \gamma\sigma_N$, $\mu_{0j}$ is shifted from $\mu_I + \mu_M e^{\mu_N (x_j-1)^2}$ to

$$\tilde{\mu}_j \equiv (\mu_I + \alpha\sigma_I) + (\mu_M + \beta\sigma_M) e^{(\mu_N + \gamma\sigma_N)(x_j-1)^2}, \quad j = 1, \ldots, p.$$

Let $\tilde{\boldsymbol{\mu}} = (\tilde{\mu}_1, \ldots, \tilde{\mu}_p)'$. Then the shift on the mean of $\boldsymbol{Y}$ is $\boldsymbol{\delta} \equiv \tilde{\boldsymbol{\mu}} - \boldsymbol{\mu}_0$.

### 2.2. *Data Smoothing*

In order to extend nonlinear profiles of a fixed parametric form to smooth profiles of a flexible nonparametric form, a smoothing technique is needed for de-noising sample profiles. The idea of smoothing is to fit a smooth function whose final form is determined by the data and the chosen level of smoothness for the curve. One popular approach is to fit noisy data by splines. Frequently, cubic splines (i.e., piecewise cubic polynomials with continuous second derivatives) are used for such approximations. Two commonly used spline smoothing techniques are smoothing splines and *B*-spline regression, both are available in popular statistical packages like *R*, Splus, and others. Other smoothing techniques such as local polynomial smoothing and wavelets can be used as well. We remark based on our experiences that, by filtering out noises, the actual signals can be better extracted from the data and PCA can explore the variation among profiles a lot better. In particular, smoothing tends to be more advantageous as the noise level ($\sigma_\epsilon^2$) gets larger.

### 2.3. *Phase I Monitoring*

Assume that a set of $n$ historical profiles is available for Phase I analysis. We first apply a smoothing technique to each of the $n$ profiles to filter out the noise, and then apply PCA to the smoothed profiles as follows. Denote the $(p \times 1)$ data vector of the $i$th profile by $\boldsymbol{y}_i$ and the usual sample covariance matrix of $\{\boldsymbol{y}_i, i = 1, \ldots, n\}$ by $\boldsymbol{S}$. Apply the eigenanalysis to $\boldsymbol{S}$. The eigenvector $\boldsymbol{v}_r$ corresponding to the $r$th largest eigenvalue $\lambda_r$ is the $r$th principal component and $S_{ir} \equiv \boldsymbol{v}_r' \boldsymbol{y}_i$ is called the score of the $r$th principal component of the $i$th profile, $r = 1, \ldots, p$, $i = 1, \ldots, n$.

We select the number of the "effective" principal components by considering the total variation explained by the chosen principal components along with the principle of parsimoniousness that we often use in the variable selection problem. Denote this number by $K$ and the $(K \times 1)$ score vector $(S_{i1}, \ldots, S_{iK})'$ by $\boldsymbol{s}_i$.

For Phase I monitoring, due to the dependency of the $K$ PC-scores, we adopt the usual Hotelling $T^2$ statistic described below. For the $i$th profile, $i = 1, \ldots, n$, the $T^2$ statistic is defined as

$$T_i^2 = (\boldsymbol{s}_i - \bar{\boldsymbol{s}})' \boldsymbol{B}^{-1} (\boldsymbol{s}_i - \bar{\boldsymbol{s}}), \tag{5}$$

where $\bar{\boldsymbol{s}} = \sum_{i=1}^{n} \boldsymbol{s}_i / n$ and $\boldsymbol{B} = \sum_{i=1}^{n} (\boldsymbol{s}_i - \bar{\boldsymbol{s}})(\boldsymbol{s}_i - \bar{\boldsymbol{s}})'/(n-1)$, the usual sample mean and sample covariance matrix of the score vectors.

Since score vectors are distributed as multivariate normal asymptotically (Anderson, 2003), according to Tracy et al. (1992) and also Sullivan and Woodall (1996), we have:

$$\frac{n}{(n-1)^2} T_i^2 \sim Beta\left(\frac{K}{2}, \frac{n-K-1}{2}\right) \quad \text{approximately.}$$

Thus, an approximate $\alpha$-level upper control limit can be set at the $100(1-\alpha)$ percentile of the beta distribution with $K/2$ and $(n-K-1)/2$ as parameters.

For Phase I analysis, perform control-charting with the $T^2$ statistic of the score vectors in (5) to detect the out-of-control profiles in the historical data set. If there are any, remove them and redo PCA and control-charting with the remaining profiles. Repeat this procedure until all the remaining profiles are within the control limit. These remaining profiles are considered as "in-control" profiles and can be used to characterize the in-control process. The resulting principal components and eigenvalues can then be used to set up the control limit for Phase II on-line monitoring.

### 2.4. *Phase II Monitoring*

As in most of Phase II studies, we assume the in-control process distribution of the profiles after de-noising has been characterized as $N_p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$, either from prior experiences or estimated from the Phase I analysis.

Our Phase II monitoring schemes are also based on PCA. Apply PCA to $\boldsymbol{\Sigma}_0$ to obtain eigenvalues, $\lambda_1 \geq \cdots \geq \lambda_p \geq 0$, and the corresponding eigenvectors, $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p$. Similar to that in Phase I analysis, choose the number of effective principal components $K$ based on the parsimoniousness and the total variation that the first $K$ PCs account for. More specifically, since the $r$th PC accounts for $\lambda_r / \sum_{r=1}^{p} \lambda_r$ of the total variation, we can simply choose the first $K$ such that $\sum_{r=1}^{K} \lambda_r / \sum_{r=1}^{p} \lambda_r$ reaches a desired level.

Now for each of the incoming profiles in Phase II monitoring, first smooth and then project it onto the first $K$ PCs to obtain $K$ PC-scores. Denote these scores by $S_1, \ldots, S_K$. Since these scores are independent and $S_r$ follows a normal distribution with mean $\boldsymbol{v}_r' \boldsymbol{\mu}_0$ and variance $\lambda_r$ when the process is in control, it is easy to construct a control chart for each of the $K$ PC-scores accordingly. Denote the desired in-control false-alarm rate by $\alpha$. Then the control limits for the $r$th PC-score chart, which monitors the statistic $S_r$, is $\boldsymbol{v}_r' \boldsymbol{\mu}_0 \pm Z_{\alpha/2} \sqrt{\lambda_r}$, $r = 1, \ldots, K$.

If a particular mode of process change can be caught by one of the first $K$ principal components, then we can use that particular PC-score chart to monitor it. However, very often a process shift is reflected in more than one principal component. When this happens, we can consider a combined chart scheme by

combining all $K$ PC-score charts. A combined chart scheme signals out-of-control when any of the $K$ individual charts signals. Thus, the proposed combined chart is equivalent to monitoring the statistic

$$\max_{1\le r\le K}\left|\frac{S_r-\boldsymbol{v}_r'\boldsymbol{\mu}_0}{\sqrt{\lambda_r}}\right|.$$

This chart signals out-of-control when $\max_{1\le r\le K}|(S_r-\boldsymbol{v}_r'\boldsymbol{\mu}_0)/\sqrt{\lambda_r}|>Z_{\alpha'/2}$, where the individual false-alarm rate $\alpha'$ should be chosen at the level of $1-(1-\alpha)^{1/K}$ so that the overall false-alarm rate is at the desired level $\alpha$.

We can also consider a $T^2$ chart by monitoring the statistic

$$T^2=\sum_{r=1}^{K}\frac{(S_r-\boldsymbol{v}_r'\boldsymbol{\mu}_0)^2}{\lambda_r}, \tag{6}$$

which follows the chi-square distribution with $K$ degrees of freedom (denoted by $\chi^2_K$) when the process is in control. Thus, the upper control limit is the $100(1-\alpha)$ percentile of $\chi^2_K$.

### 2.5. *ARL of the Proposed Schemes*

We evaluate the performances of the proposed Phase II monitoring schemes described above in terms of ARL, the average run length. The ARL values of the individual PC-score chart can be computed as follows. Assume that the mean of the profile has been shifted from $\boldsymbol{\mu}_0$ to $\boldsymbol{\mu}_0+\boldsymbol{\delta}$. The probability of detecting the shift by the $r$th PC-score chart is

$$\begin{aligned}p&=1-P\left(\left|\frac{S_r-\boldsymbol{v}_r'\boldsymbol{\mu}_0}{\sqrt{\lambda_r}}\right|\le Z_{\alpha/2}\right)=1-P\left(-\frac{\boldsymbol{v}_r'\boldsymbol{\delta}}{\sqrt{\lambda_r}}-Z_{\alpha/2}\le Z\le-\frac{\boldsymbol{v}_r'\boldsymbol{\delta}}{\sqrt{\lambda_r}}+Z_{\alpha/2}\right)\\&=1-\Phi\left(-\frac{\boldsymbol{v}_r'\boldsymbol{\delta}}{\sqrt{\lambda_r}}+Z_{\alpha/2}\right)+\Phi\left(-\frac{\boldsymbol{v}_r'\boldsymbol{\delta}}{\sqrt{\lambda_r}}-Z_{\alpha/2}\right),\end{aligned}$$

where $\Phi$ is the cumulative distribution function of the standard normal variate $Z$ and $Z_{\alpha/2}$ is the $100(1-\alpha/2)$ percentile of $Z$. Then the value $1/p$ is the ARL of the $r$th PC-score chart.

Since the PC-scores $S_1,\ldots,S_K$ are independent, the ARL of the combined chart also can be computed easily by the reciprocal of

$$\begin{aligned}p&=1-P\left(\max_{1\le r\le K}\left|\frac{S_r-\boldsymbol{v}_r'\boldsymbol{\mu}_0}{\sqrt{\lambda_r}}\right|\le Z_{\alpha'/2}\right)=1-\prod_{r=1}^{K}P\left(\left|\frac{S_r-\boldsymbol{v}_r'\boldsymbol{\mu}_0}{\sqrt{\lambda_r}}\right|\le Z_{\alpha'/2}\right)\\&=1-\prod_{r=1}^{K}\left[\Phi\left(-\frac{\boldsymbol{v}_r'\boldsymbol{\delta}}{\sqrt{\lambda_r}}+Z_{\alpha'/2}\right)-\Phi\left(-\frac{\boldsymbol{v}_r'\boldsymbol{\delta}}{\sqrt{\lambda_r}}-Z_{\alpha'/2}\right)\right],\end{aligned}$$

where $\alpha'=1-(1-\alpha)^{1/K}$ is the individual false-alarm rate.

Since $T^2$ statistic in (6) follows a noncentral chi-square distribution with $K$ degrees of freedom and noncentrality $\xi=\sum_{r=1}^{K}(\boldsymbol{v}_r'\boldsymbol{\delta})^2/\lambda_r$ (denoted by $\chi^2_K(\xi)$).

Then the detecting power of the $T^2$ chart can be easily calculated by

$$p = P(T^2 > \chi^2_{K,\alpha}) = P(\chi^2_K(\xi) > \chi^2_{K,\alpha}),$$

where $\chi^2_{K,\alpha}$ denotes the $100(1-\alpha)$ percentile of the central chi-square distribution $\chi^2_K$.

## 3. Simulation and Comparative Studies

### 3.1. *Settings for Simulation*

In our simulation study, we generate profiles from $MVN(\boldsymbol{\mu}_0, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}_0$ is given in (4) and $\boldsymbol{\Sigma}$ is given in (2) with $\mu_I = 1$, $\sigma_I = 0.2$, $\mu_M = 15$, $\sigma_M = 1$, $\mu_N = -1.5$, $\sigma_N = 0.3$, and $x = 0.64, 0.8, \ldots, 3.52$. Both of $x$ and $y$ values are scaled variables, not the actual temperature levels and the amount of aspartame dissolved in the dissolving process. Denote the in-control ARL by $ARL_0$. All charts are designed to have the same $ARL_0 = 370.3704$, which corresponds to the false-alarm rate of $\alpha = 0.0027$.

### 3.2. *A Study on the Number of Principal Components*

To study how the choice of the number of effective principal components affects the detecting power of the monitoring scheme, we conduct a simulation study. In this study, the detecting power is measured by the ability of the monitoring scheme in detecting the real out-of-control profiles. For example, in a data set of 50 simulated profiles with 3 out-of-control profiles, if the scheme catches 2 of the 3, then the detecting power measured is $2/3$. The false-alarm rate can be measured in a similar way. Let the number of the principal components used be $k$. Then for each data set, compute the detecting power and the percentage of the total variation explained by $k$ principal components for various values of $k$.

We choose some settings of shifts. For each setting, we generate fifty profiles within which some profiles are generated from the shifted population. Then repeat each setting 20,000 times to get the averaged detecting power and the averaged percentage of the total variation explained.

As one would expect, the result (not shown) of the study indicates that having more principal components does explain more variation, but not necessarily has more detecting power. In fact, the power of the $T^2$ statistic starts to drop when $k$ gets to a certain level, which usually is a fairly small number. So it is necessary to choose an appropriate number of principal components.

### 3.3. *A Simulated Aspartame Example – Phase I Monitoring*

We now demonstrate Phase I analysis with the aspartame example described before. In Phase I, we generate 200 in-control historical profiles, each with 19 set points. First de-noise these profiles by smoothing splines using statistical package R. Figures 2(a) and (b) display the 200 simulated profiles and their smoothed profiles, respectively. Then we apply PCA to the sample covariance matrix of these smoothed profiles and get the 19 eigenvalues, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{19} \geq 0$ and their associated eigenvectors. The first 4 principal components account for 75.16%, 19.41%, 2.60%, and 0.92% of the total variation in the profiles, respectively. For profile monitoring,

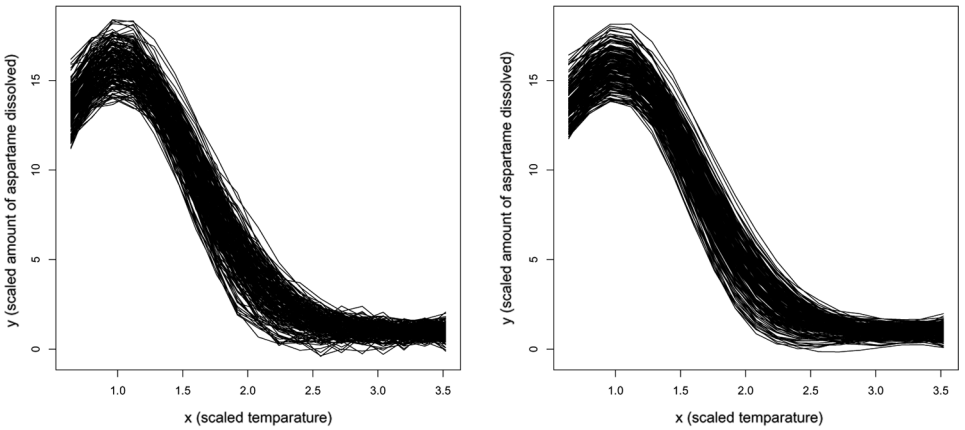


**Figure 2.** (a) 200 generated and (b) smoothed in-control aspartame profiles.



we decide to choose $K = 3$ for parsimoniousness since with three PCs, it has already accounted for 97.17% of the total variation. In practice, it is also fine to choose $K = 4$. Now project each (smoothed) profile onto the first three eigenvectors to get the scores and then compute $T^2$ by (5). The resulting $T^2$ control chart (not shown) indicates that the process is in control.

### 3.4. *An ARL Comparison Study – Phase II Monitoring*

To compare the performances of the proposed schemes for Phase II monitoring described in Subsec. 2.4, we compute the ARL values of each scheme as derived in Subsec. 2.5.

Let the $(i, j)$th entry of the in-control covariance matrix $\mathbf{\Sigma}_0$ be (2) without the $\sigma_\epsilon^2 \delta_{ij}$ term. Apply PCA to this "population" covariance matrix. It is found that the first four principal components, respectively, account for 74.82%, 22.58%, 2.30%, and 0.29%, which totals 99.99%, of the total variation; and other components practically explain nothing. This is mainly because we have only three degrees of freedom in varying profiles (without measurement error), namely, the values of $I$, $M$, and $N$ in model (1).

Figure 3(a) depicts the first three eigenvectors of $\mathbf{\Sigma}_0$. To see the effect of a particular principal component, say $\boldsymbol{v}_r$, Ramsay and Silverman (2005) presented a visualizing tool that plots $\boldsymbol{\mu}_0 \pm L\boldsymbol{v}_r$, where $L$ is a suitable multiple. Figures 3(b)–(d) illustrate, respectively, the corresponding features captured by the first three principal components with $L = 3$. From the figures, we can see that the PC1 captures the variation of the vertical shifting of the profiles except for the right tail; PC2 reflects the variation in the height of the peak and the declining rate of the curve; PC3 captures mainly the vertical shifting in the tail and slightly the variation at the peak.

For ARL comparison, consider the $I$-shift from $\mu_I$ to $\mu_I + \delta\sigma_I$, $M$-shift from $\mu_M$ to $\mu_M + \delta\sigma_M$, and $N$-shift from $\mu_N$ to $\mu_N + \delta\sigma_N$, for $\delta = 0, 0.25, \ldots, 3$. Figures 4(a)–(c) display the ARL curves for the shifts in $I, M, N$, respectively.

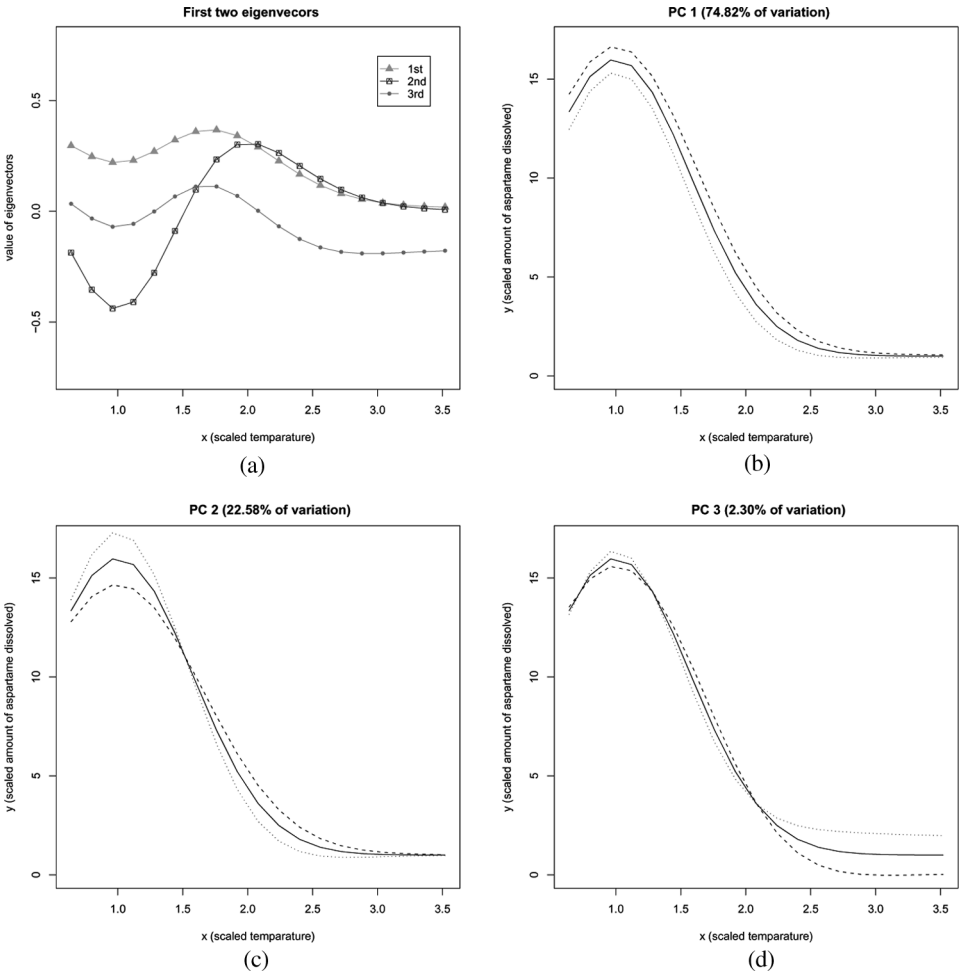


**Figure 3.** Aspartame example. (a) Plot of $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$; (b)–(d) $\boldsymbol{\mu}_0 \pm 3\boldsymbol{v}_r$, $r = 1, 2, 3$.

We observe the following.

- Both PC1 and PC3 have some power in catching the shift in $I$ because both represent the mode of variation in vertical shifting (but in different areas) of the profile as shown in Figs. 3(a) and (c). We are a little bit surprised to see that PC1 is less powerful than PC3. This may be explained by: (i) PC3 explains almost all the variation in vertical shifting in the tail area; (ii) although PC1 can explain the vertical shifting for $x < 2.5$, the other two PCs also pick up some, especially around the peak area. PC2 hardly has any power in detecting $I$-shift. The power of the $T^2$ and combined chart are between that of PC1 and PC3, and as the shift gets larger, the difference in ARL between PC3 and the $T^2$/combined chart gets smaller. Also, the $T^2$ and combined chart are very close with $T^2$ slightly better for the shift size $\alpha \leq 1.5$ and the combined chart better for $\alpha \geq 1.75$.

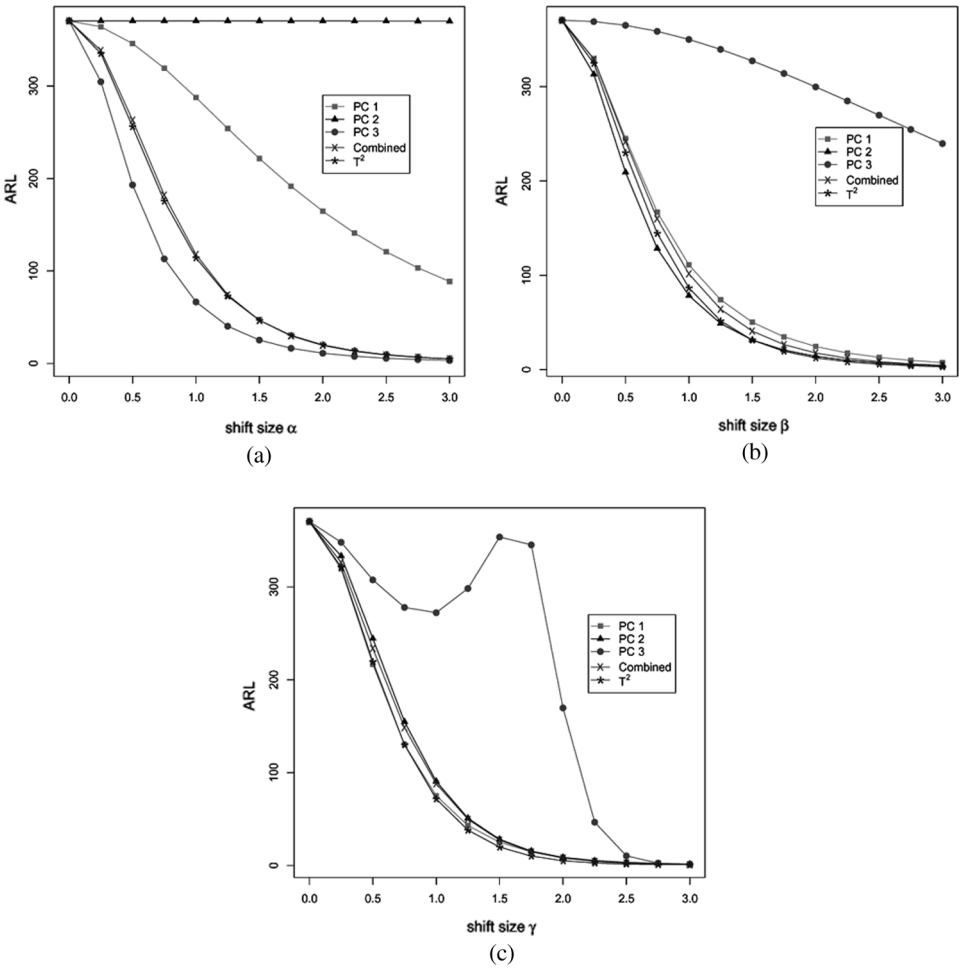


**Figure 4.** ARL comparisons of aspartame example. (a) $I$; (b) $M$; and (c) $N$ shifts.

- As to the $M$-shifts, except for PC3 that does not have much power, the other four charts are not too far from each other. The order of the performance is about PC2 > $T^2$ > combined > PC1, with the exception that $T^2$ finally beats PC2 for larger shifts. Here ">" means "performs better than".
- For detecting $N$-shift, PC3 has a strange ARL curve (see Fig. 4(c)), which may be caused by the fact that the shift in the mean vector when projected onto PC3, $\boldsymbol{v}_3'\boldsymbol{\delta}$, is not monotone in the shift multiple $\delta$. PC1 performs the best for small shifts but gets worse as the shift gets larger and eventually becomes the worst one for large shifts. $T^2$ is the second best for small shifts and then quickly becomes the best. PC2 and the combined chart are fairly close for small shifts; PC2 wins when the shift size is small and soon loses it to the combined chart for moderate to large shifts.

We learn from this study that a mode of variation often cannot be captured by a single PC-score chart. Even for the variation as simple as the vertical shifting

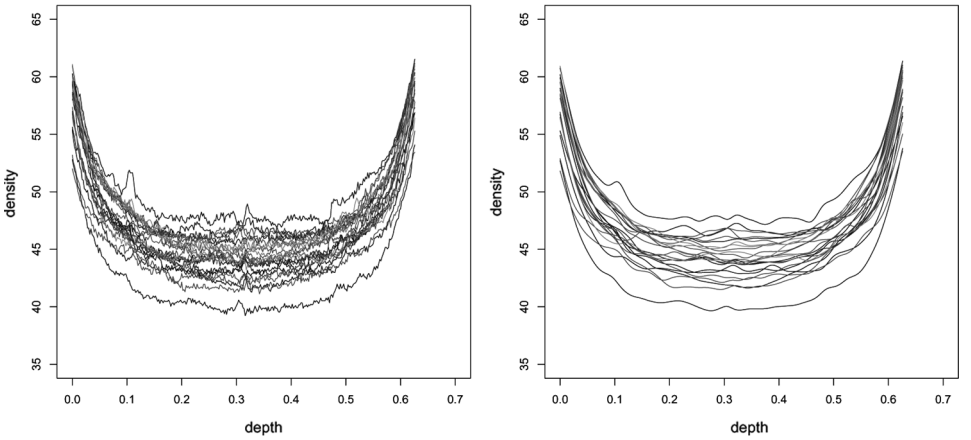


**Figure 5.** (a) 24 VDP profiles; (b) 24 smoothed VDP profiles.

like the *I*-shift, it takes more than one principal component to represent this effect. Also, the $T^2$ and combined chart are fairly close to each other and comparable with the best PC-score chart.

## 4. A Case Study – VDP Example

The VDP data set described in Sec. 1 contains $n = 24$ profiles, each was measured at $p = 314$ set points. Figure 5(a) is the plot of the VDP data. First de-noise these profiles by smoothing splines using statistical package R; see Fig. 5(b) for the plot of the smoothed profiles. Apply PCA to the sample covariance matrix of the smoothed profiles. And the first four principal components account for 85.26%, 10.83%, 1.90%, and 0.84% of the total variation in the profiles, respectively. We select $K = 3$ principal components for Phase I process monitoring because the total variation explained by the first three PCs is already as high as 97.99%. Figure 6(a) is the plot of the first three eigenvectors. Figures 6(b)–(d) show the modes of variation they capture by plotting the mean vector $\boldsymbol{\mu}$ and $\boldsymbol{\mu} \pm 10\boldsymbol{v}_r$, $r = 1, 2, 3$. The figures show that: (i) PC1 represents the variation in the ground level of the VDP profiles, especially the bottom part of the "bathtub"; (ii) PC2 can catch some variation in the roundness of the bottom part of the bathtub; and (iii) PC3 may be able to describe the variation in the roundness on the two ends and the central part of the bathtub bottom. All these interpretations can also be seen from the three eigenvectors shown in Fig. 6(a). As to Phase I monitoring, the $T^2$ control chart (not shown) indicates no out-of-control profiles in the VDP data.

## 5. Concluding Remarks

In this study, we propose and discuss monitoring schemes for nonlinear profiles. We use the principal components analysis to analyze the covariance matrix of the profiles and then utilizing the principal component scores that capture the main features of the profile data for process monitoring.

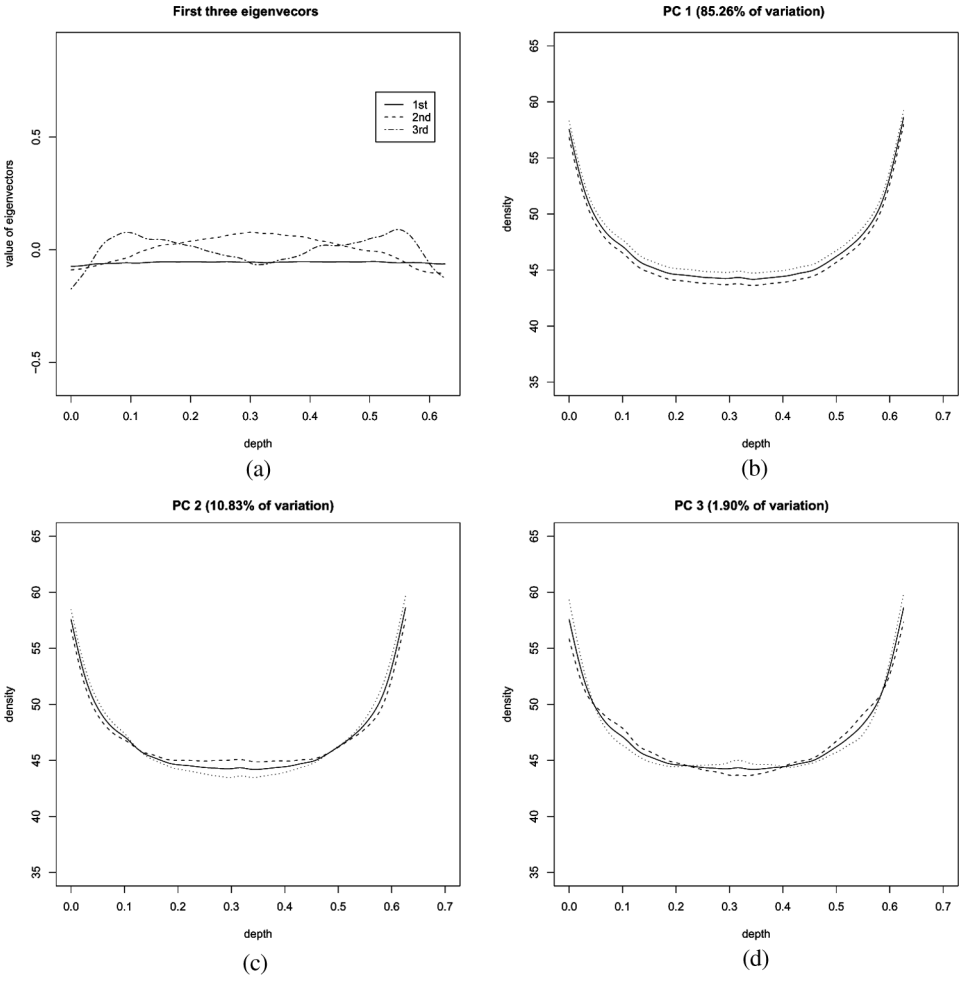


**Figure 6.** VDP example. (a) Plot of $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$; (b)–(d) $\boldsymbol{\mu} \pm 10\boldsymbol{v}_r$, $r = 1, 2, 3$.

In addition to the individual PC score charts, we study two charts that utilize the overall information contained in the $K$ effective principal components, namely the combined chart and $T^2$ chart. The $T^2$ chart performs somewhat better than the combined chart in terms of the average run length, but not too far off. However, by providing charts for all of the effective components, the combined chart gives more clues for finding assignable causes than the $T^2$ chart.

When the shift corresponds to a mode of variation that a particular principal component represents, then it would be ideal to use the individual PC-score chart for process monitoring because this particular individual chart will have the best power among the charts under study. Unfortunately, this ideal situation is rare in practice. Moreover, by just monitoring one individual PC chart, one is running a risk of not being able to detect other types of process changes. One may argue that we should monitor the process with all $K$ individual charts simultaneously in order to catch all potential changes. But with the same false-alarm rate $\alpha$ for each of the individual charts, the overall false-alarm rate is greatly increased to

$1-(1-\alpha)^K$, which is about $K$ times of the original false-alarm rate. Thus, for being more practical and conservative, we recommend using the $T^2$ chart or the combined chart scheme, because they still have comparably good power to monitor these particular types of shifts and have a lot better power than the individual chart for general types of shifts.

It is noted that the degree of smoothness used in the data smoothing step has a great impact on the result of the subsequent PCA step. The high total explanation power of the first few principal components demonstrated in this article is in fact caused by the high degree of smoothness in the smoothed curves. This argument is supported by our finding that if B-spline regression is used for smoothing, the number of B-spline bases used is exactly the number of the principal components with nonzero eigenvalues. So if the underlying profiles (i.e., with no noises) are fairly smooth as what we have in the hypothetical aspartame example (in which a data profile is a three-parameter exponential function plus noises), then the data dimension can be well reduced by applying PCA to the smoothed curves. The situation in the VDP data is similar. On the other hand, if the underlying profiles are not very smooth and data profiles are not too much over-smoothed, then it might take quite a few principal components to explain good enough proportion of variation. We remark that, regardless of which $K$, the number of effective principal components, is chosen, the false-alarm rate for each of our schemes stays at $\alpha$. However, the diagnosis of out-of-control conditions would likely become more complicated when $K$ is large.

The degree of smoothness may affect the effectiveness of the Phase II process monitoring as well. If the noise levels are about the same for both Phase I and Phase II profile data, we suggest applying the same degree of smoothness to them. In this way, the results of Phase II monitoring are somehow not that sensitive to the extent of smoothing. However, when profiles are over-smoothed to the extent that some local features are lost, then the process changes associated with these vanished local features cannot be detected. On the other hand, when profiles are under-smoothed to an extent, some spurious features may appear in the fitted curves. Unfortunately, these spurious features may not appear at the same place and may not have the same form across profiles. Then the estimated in-control model obtained from Phase I data may not suitable for effective Phase II monitoring. We remark that even for the case that the in-control process is known or appropriately characterized, the spurious features in the "smoothed" Phase II profiles caused by under-smoothing will cause more false alarms to signal.

In practice, when one employs the $T^2$ chart or the combined chart scheme and detects significant shifts, it is desirable to find the sources responsible for the shifts. For this, we suggest to rank the standardized PC-scores and investigate the corresponding principal components in order, starting from the largest PC score. With the help of the plots like Figs. 3(a)–(d) and 6(a)–(d) and engineers' expertise, the characteristics of the principal components sometimes can reveal potential root causes of the shifts.

The monitoring of process or product profiles has become a popular and promising area of research in statistical process control in recent years. At the same time, functional data analysis (FDA) is also gaining lots of attentions and applications. We believe many techniques developed for FDA may be extended to developing new profile monitoring techniques in SPC.

## Acknowledgments

The authors would like to express their gratitude to the Editor, an Associate Editor, and two referees for the careful review and constructive suggestions. This work was supported in part by the National Research Council of Taiwan, Grant No. NSC93-2118-M009-007 and NSC95-2118-M-009-006-MY2.

## References

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*. 3rd ed. New York: Wiley.

Castro, P. E., Lawton, W. H., Sylvestre, E. A. (1986). Principal modes of variation for processes with continuous sample curves. *Technometrics* 28:329–337.

Ding, P. L., Zeng, L., Zhou, S. Y. (2006). Phase I analysis for monitoring nonlinear profiles in manufacturing processes. *Qual. Eng.* 5:619–625.

Jensen, W. A., Birch, J. B., Woodall, W. H. (2006a). Profile monitoring via linear mixed models. Technical Report No. 06-2, Department of Statistics, Virginia Polytechnic Institute and State University, Virginia.

Jensen, W. A., Birch, J. B., Woodall W. H. (2006b). Profile monitoring via nonlinear mixed models. Technical Report No. 06-4, Department of Statistics, Virginia Polytechnic Institute and State University, Virginia. *Journal of Quality Technology*. To appear.

Jones, M. C., Rice, J. A. (1992). Displaying the important features of large collections of similar curves. *Amer. Statistician* 46:140–145.

Kang, L., Albin, S. L. (2000). On-line monitoring when the process yields a linear profile. *J. Qual. Technol.* 32:418–426.

Kim, K., Mahmoud, M. A., Woodall, W. H. (2003). On the monitoring of linear profiles. *J. Qual. Technol.* 35:317–328.

Mahmoud, M. A., Woodall, W. H. (2004). Phase I analysis of linear profiles with calibration applications. *Technometrics* 46:380–391.

Ramsay, J. O., Silverman, B. W. (2005). *Applied Functional Data Analysis*. 2nd ed. New York: Springer.

Rice, J. A., Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J. Roy. Statist. Soc. Ser. B* 53:233–243.

Shiau, J.-J. H., Lin, H.-H. (1999). Analyzing accelerated degradation data by nonparametric regression. *IEEE Trans. Reliab.* 48:149–158.

Shiau, J.-J. H., Lin, S.-H., Chen, Y.-C. (2006a). Monitoring linear profiles based on a random-effect model. Technical Report. Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan.

Shiau, J.-J. H., Yen, C.-L., Feng, Y.-W. (2006b). A new robust method for Phase I monitoring of nonlinear profiles. Technical Report. Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan.

Shiau, J.-J. H., Weng, Z.-P. (2004). Profile monitoring by nonparametric regression. Technical Report. Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan.

Sullivan, J. H., Woodall, W. H. (1996). A comparision of multivariate control charts for individual observations. *J. Qual. Technol.* 28:398–408.

Tracy, N. D., Young, J. C., Mason, R. I. (1992). Multivariate control charts for individual observations. *J. Qual. Technol.* 24:88–95.

Walker, E., Wright, S. P. (2002). Comparing curves using additive models. *J. Qual. Technol.* 34:118–129.

Williams, J. D., Birch, J. B., Woodall, W. H., Ferry, N. M. (2007). Statistical monitoring of heteroscedastic dose-response profiles from high-throughput screening. *J. Agricult. Biolog. Environ. Statist.* 2:216–235.

Williams, J. D., Woodall, W. H., Birch, J. B. (2003). Phase I analysis of nonlinear product and process quality profiles. Technical Report No. 03-5, Department of Statistics, Virginia Polytechnic Institute and State University, Virginia.

Woodall, W. H., Spitzner, D. J., Montgomery, D. C., Gupta, S. (2004). Using control charts to monitor process and product quality profiles. *J. Qual. Technol.* 36:309–320.