Computers & Industrial Engineering 85 (2015) 132–144

Contents lists available at ScienceDirect

Computers & Industrial Engineering

journal homepage: www.elsevier.com/locate/caie

Phase I analysis of multivariate profiles based on regression adjustment ☆

Jiajia Zhang [a], Haojie Ren [a,*], Rui Yao [b], Changliang Zou [a], Zhaojun Wang [a]

[a] Institute of Statistics, Nankai University, Tianjin, China
[b] Beijing Institute of Control Engineering, Beijing, China

ARTICLE INFO

Article history:
Received 3 November 2014
Received in revised form 23 February 2015
Accepted 25 February 2015
Available online 19 March 2015

Keywords:
Functional principal component analysis
Nonlinear profile
Nonparametric regression
Profile analysis
Statistical process control

ABSTRACT

The use of statistical process control in monitoring and diagnosis of process and product quality profiles remains an important problem in various manufacturing industries. Although the analysis of profile data has been extensively studied in the literature, the challenges associated with monitoring and diagnosis of multiple functional profiles are yet to be well addressed because it is usually difficult to properly model the inter-relationship of multiple profiles. Motivated by a real-data application in semiconductor industries, we develop a new modelling and monitoring framework for Phase-I analysis of multiple profiles. The proposed framework incorporates the regression-adjustment technique into the functional principal component analysis. In this framework, the multiple profiles are treated as multivariate functional observations and their regression-adjusted residuals are used for monitoring. Simulation results show that the proposed method could describe the major structure of profile variation well and effectively find outlying profiles in a historical dataset due to sufficiently utilizing the information on between-profiles correlation.

1. Introduction

Because of recent progress in sensing and information technologies, automatic data acquisition has become the norm in various industries. Consequently, a large amount of quality-related data of certain processes have become available. Statistical process control (SPC) based on such data is an important component of process monitoring and control. In many applications, the quality of a process is characterized by the relationship between a response variable and one or more explanatory variables. A collection of data points of these variables can be observed at each sampling stage, which can be represented by a curve (i.e., profile). In some calibration applications, the profile can be described adequately by a linear regression model. In other applications, however, more flexible models are necessary in order to describe the profiles properly. An extensive discussion of the related research problems has been given by Woodall, Spitzner, Montgomery, and Gupta (2004).

Profile monitoring has been extensively studied in SPC and several methods have been developed for monitoring linear and nonlinear profile data. Some examples include the use of multivariate control charts for monitoring linear and nonlinear regression coefficients (Kang & Albin, 2000; Mahmoud & Woodall, 2004; Zou, Tsung, & Wang, 2007; Williams, Woodall, & Birch, 2007), monitoring methods based on mixed-effect models (Jensen, Birch, & Woodall, 2008; Paynabar, Jin, Agapiou, & Deeds, 2012), dimension-reduction techniques (Lada, Lu, & Wilson, 2002; Ding, Zeng, & Zhou, 2006; Chicken, Pignatiello, & Simpson, 2009; Paynabar & Jin, 2011; Viveros-Aguilera, Steiner, & Mackay, 2014), and methods for monitoring roundness profiles (Colosimo & Pacella, 2007; Colosimo, Cicorella, Pacella, & Blaco, 2014). Extensive discussion about various research problems on profile monitoring can be found in Woodall (2007), Noorossana, Saghaei, and Amiri (2011) and Qiu (2014, chap. 10).

Most recent studies concentrated on the situation with a univariate profile that only contains one response variable. Although such profiles can characterize various applications as described in the literature, multivariate functional profiles in which multiple response variables are involved simultaneously may be even more representative of most industrial applications in certain real world practices. When the correlation structure between quality characteristics is ignored and profiles are monitored separately, then misleading results may be expected (c.f., Lowry, Woodall, Champ, & Rigdon (1992) and Hawkins (1991) for relevant discussions). However, research on the monitoring and diagnosis of multivariate general profiles is still scanty. Noorossana, Eyvazian, Amiri, and Mahmoud (2010) discussed multivariate linear profile monitoring

☆ This manuscript was processed by Area Editor Min Xie.

* Corresponding author.
E-mail address: renhaojie1991@163.com (H. Ren).

http://dx.doi.org/10.1016/j.cie.2015.02.025
0360-8352/

in Phase I analysis, mainly based on the ordinary least square estimation. Another relevant work is Zou, Ning, and Tsung (2012) which focused on a study of the Phase II method for monitoring a general multivariate linear profile by using the LASSO-based multivariate SPC techniques. More recently, Chou, Chang, and Tsai (2014) developed a process monitoring strategy for monitoring multiple correlated nonlinear profiles. Nevertheless, how to apply conduct Phase I monitoring for general multivariate profiles including nonlinear profiles, still remains a challenge and has not been thoroughly investigated in the literature. In practice, Phase I process control is crucially important to check the stability of historical profile data and to obtain accurate estimates of the baseline model parameters used for Phase II monitoring (Zhang & Albin, 2009). The main objective of this paper is to develop a new nonparametric method for Phase I monitoring of multivariate profile data.

In the literature of nonparametric profile monitoring, the nonparametric profile model (c.f. Zou, Tsung, & Wang, 2008) is usually considered

$$y_{ij} = g(x_{ij}) + \varepsilon_{ij}, \quad j = 1, \ldots, n_i, \quad i = 1, 2, \ldots, \tag{1}$$

where $\{x_{ij}, y_{ij}\}_{j=1}^{n_i}$ is the ith sample collected over time, x_{ij} is the jth design point in the ith profile, g is a smooth nonparametric profile, and ε_{ij}s are i.i.d. normal random errors with mean 0 and variance σ^2. Zou et al. (2008) developed a procedure based on the combination of local linear smooth and traditional SPC charting techniques. Furthermore, to account for the within-profile correlations, Qiu, Zou, and Wang (2010) proposed to use the nonparametric mixed-effects model which allows flexible variance–covariance structures. Recently, Hung, Tsai, Yang, Chuang, and Tseng (2012) introduced a technique called Support Vector Regression to model the profile relationship between the response variable and explanatory variables, while the within-profile correlation is accommodated by using a resampling technique called block bootstrap. Closely related to this idea, Chuang, Hung, and Yang (2013) provided an easy-to-implement and computationally cheaper framework for monitoring nonparametric profiles by taking into account the within-profile correlation.

It is not straightforward to extend (1) to multivariate settings. One technical challenge is how to model such multivariate profiles by taking the correlations between curves into account. Recent and representative work is Soleimani and Noorossana (2014). However, such traditional methods developed for a linear regression may not fully characterize the information from a complex profile datastream where between-curves correlations may be highly related to the covariates (design points). Although certain efforts have been made on the within profile autocorrelation in simple linear profiles or parametric profiles (see Soleimani, Noorossana, & Amiri, 2009; Soleimani, Noorossana, & Niaki, 2013; Wang & Tamirat, 2014; Khedmati & Niaki, 2015), the complexity of simultaneously handling the between-curves and within-profile correlations still challenges us. Engineering applications that give rise to profile data often lead to correlated error terms. As demonstrated by Qiu et al. (2010), neglecting the within-profile correlation will result in adverse effects on both in-control (IC) and out-of-control (OC) properties of control schemes. The situation may be more serious for multivariate profile processes due to the intricacy of the model and the large number of responses. Moreover, how to integrate an appropriate regression function nonparametric test with classical SPC techniques is not quite straightforward.

In this paper, we try to deal with all the aforementioned challenges, and to resolve the latent issue of existing nonparametric modelling and monitoring methods that are unable to efficiently utilize full information from a multivariate profile process. We start by introducing a manufacturing example taken from an etching process.

2. A motivating example

We use an industrial etching process example taken from semiconductor manufacturing to illustrate the motivation for this research. The etching chamber is equipped with more than 50 sensors which record the values of several variables with time during a batch. For illustrative purposes, only 7 major variables $x_1, \ldots, x_7$ will be considered. Variable x_1 is related to spectral analysis of chamber gas, while x_2 to x_5 relate to plasma operations. Variables x_6 and x_7 are the chamber temperature and pressure, respectively. There are many steps involved in the batch operation, but here we only focus on the second step since engineers consider that this step is one of the most crucial steps which usually contains enough critical information to distinguish the out-of-control conditions in the process. The entire dataset contains 364 wafers which consist of 16 lots, while each lot represents a short-term operating cycle. We aim at using the Phase I monitoring schemes to identify any abnormal profile observations from the dataset.

The profile observations of each variable are synchronized so that the number and positions of covariate points (say, time, in this example) are equal (Akima, 1970). Fig. 1 shows the profiles of the first four profiles for 364 wafers. The x-axis represents totally 137 synchronized profile points. The profile curve of variable x_5 is quite similar to that of x_4 and thus is not presented here. The values of x_6 and x_7 for each wafer are almost identical and thus we choose to plot the sample mean of each profile against the wafer index (from 1 to 364). Clearly, the profile curves for each variable present a similar functional and smooth change along with the time point. Some of them significantly deviate from the population curves and should be regarded as outlying profiles which we would like to use some detection procedures to automatically screen out. The values of temperature x_6 and pressure x_7 vary within certain ranges and can be viewed as two environmental covariates in our following analysis.

From Fig. 1(a)–(d), we can vaguely see that there might be some interrelationships between different variables. To further confirm, we specifically present Fig. 2(a)–(c) to show the scatter plots of the raw data for the x_1 against x_3, x_2 against x_4 and x_5 against x_4, respectively, at the time point $t = 5$. It is now clear that the correlations between the variables are significant, which suggests that the profile curves have considerable interrelationships and consequently a multivariate modelling and monitoring framework is likely to be more appropriate than a univariate one.

Besides the between-curves correlations, since the measurements of all the first five variables in each profile (wafer) are taken in consecutive time intervals, the data exhibit a considerable amount of positive serial autocorrelation in each profile. For example, Fig. 3 depicts three estimated within-profile correlation curves $\hat{\rho}_j(t, t')$ against $t = 1, \ldots, 137$ for $j = 1, 3, 4$, where $\hat{\rho}_j(t, t')$ denotes the estimated correlation between the observations at t and t' time points for the jth variable. We chose $t' = 5$, 35 and 65 for illustration because similar results can also be observed for other time points (not reported here but available from the authors). We can see that within-profile correlation is substantial, and thus it should not be ignored.

Naturally, we may use the model (1) to describe the profile curves for variables x_1 to x_5. The local linear smoothing curve (Fan & Gijbels, 1996) of the average of the 364 wafers is obtained for each variable. They can be deemed as an estimate of the population profile models. Then, the corresponding residual curves are obtained by subtracting estimated population curves from the original observations. From those residual curves, some

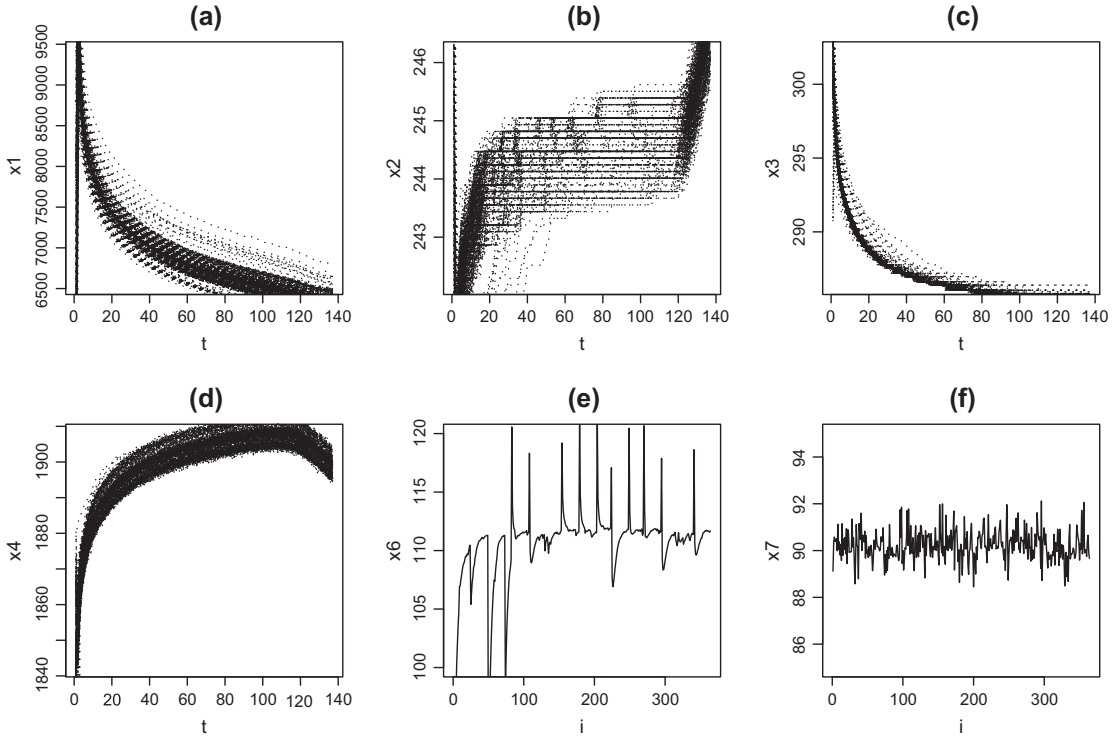

Fig. 1. (a–d): The synchronized profiles of sensor variables x_1 to x_4 over a 364 wafer batch and (e–f): the within-profile sample means of x_6 and x_7 for 364 wafers.

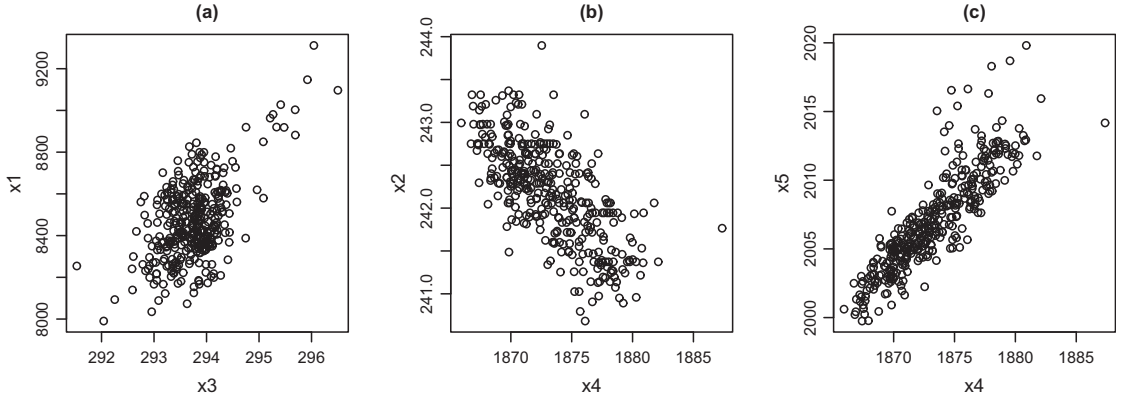

Fig. 2. The scatter plots of the raw data at $t = 5$: (a) the x_1 against x_3; (b) x_2 against x_4; and (c) x_5 against x_4.

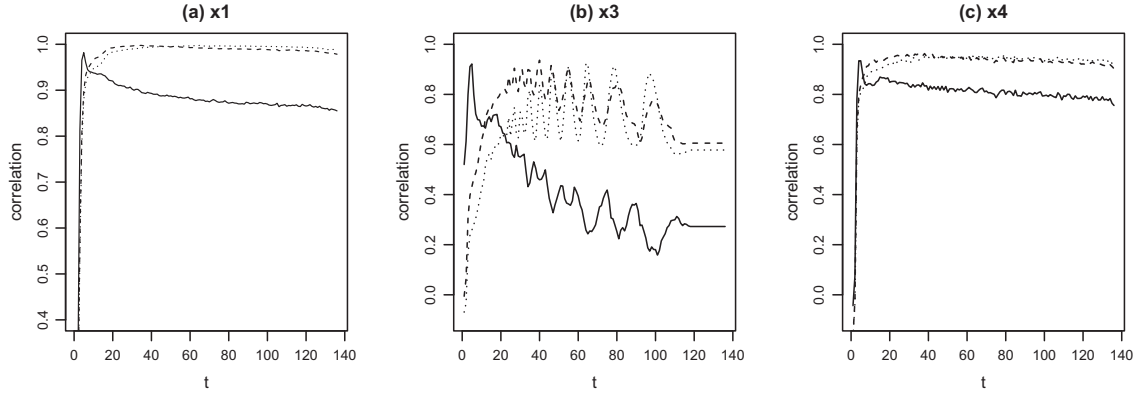

Fig. 3. Solid, dashed and dotted curves represent the estimated within-profile correlations $\hat{\rho}_j(t, t')$ when $t' = 5,\ 35,\ 65$, respectively; (a): $j = 1$, (b): $j = 3$, and (c): $j = 4$.

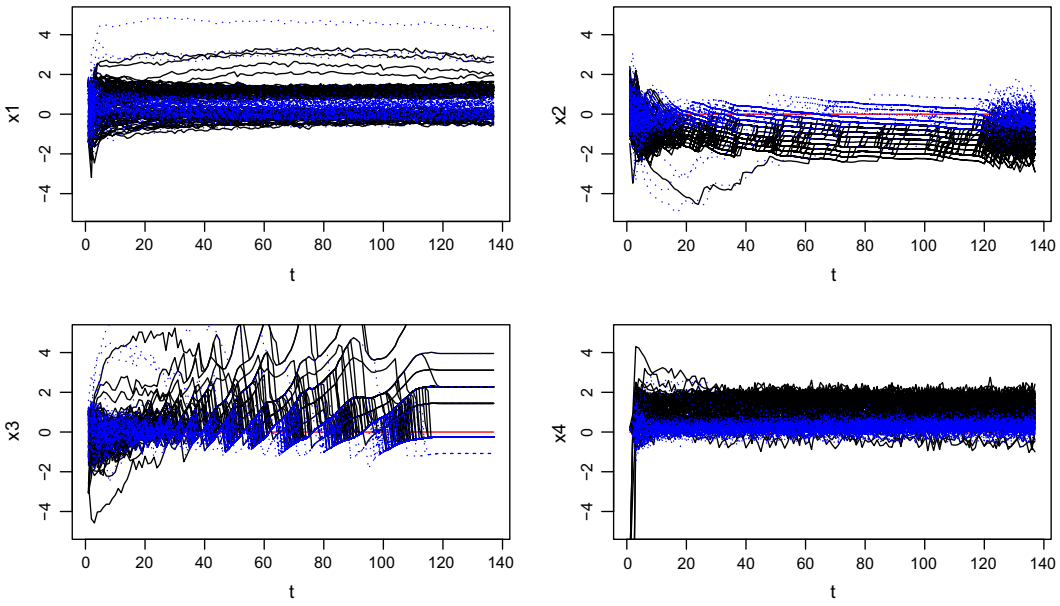

Fig. 4. The first 100 residual curves (black) and the second 100 residual curves (blue). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

interesting phenomenon can be observed. Fig. 4 shows the first 100 and the second 100 residual curves with black and blue color respectively for variables x_1 to x_4, according to the actual start time of the batch. From Fig. 4 we can see a significant aging trend. Actually, this may be explained by the fact that the temperature x_6 is apparently lower for the first 100 wafers as shown in Fig. 1(e) which represents a change of equipment status. Certainly, it may also affect the other variables, however, such aging phenomena for x_1 and x_3 are masked and cannot be easily distinguished from Fig. 4.

Moreover, Fig. 5 shows the residual curves of the variables x_1, x_3 and x_4 for the two of 16 lots. Clearly, the errors are far away from our desired mean zero stochastic processes. This figure suggests that these curves exhibit some strong "pattern" which reflects some intrinsic short term trends are included in the residuals. Such short term fluctuations, mainly due to the between-curves and within-profile correlations, are regarded as part of normal operations but they must be accounted for before real faults can be detected. For the x_2 variable, short-term fluctuations are not very significant and thus the associated curves are not presented. In addition, it can be seen that the processing of the first wafer of a lot is usually different from that of later wafers in the lot. This is known as the so-called "first-wafer effect". A good modelling and monitoring scheme should also be able to indicate these outlying profiles automatically. This serve as a fundamental benchmark for assessing the effectiveness of various procedures. In the remainder of this paper, we propose a Phase I monitoring scheme which consists of a modelling procedure and a Shewhart-type control chart for detecting outlying profiles in such datasets. We will give a step-by-step demonstration of how to implement the proposed scheme in practice in a later section.

3. Regression-adjustment-based profile monitoring

Our proposed methodology is termed as regression-adjusted-based multivariate profile monitoring (RAMP) hereafter and is described in three parts. In Section 3.1, we introduce a regression-adjusted varying coefficient model to handle the multivariate nonparametric profile monitoring problem. Its model estimation is discussed in Section 3.2. In Section 3.3, we develop a Phase I profile monitoring method based on the proposed model.

3.1. Problem and modelling strategy

Consider a vector-valued profile observations set $\{\mathbf{X}_i(t), i = 1, \ldots, N\}$, where the sample size N denotes the number of profiles. Without loss of generality, we assume that time $t \in \mathcal{T} = [a, b]$, $-\infty < a < b < \infty$. Moreover, the observations

$$\mathbf{X}_i(t) = \{X_{i1}(t), \ldots, X_{ip}(t)\}^T,$$

are assumed to be mutually independent, where p is the number of variables in the i-th profile. Besides these profile observations, we also observe some other non-profile observations, such as temperature or pressure, denoted as $\mathbf{Z}_i$. It is required to set up some appropriate model for describing the multivariate process $\mathbf{X}_i(t)$ before any further efficient monitoring, which turns out to be rather difficult due to its complexity.

To account both of the between-curves and within-curve correlations, we circumvent jointly modelling $\mathbf{X}_i(t)$ by considering the following functional multiple regression of $X_{ij}(t)$, the jth component of $\mathbf{X}_i(t)$, on all other components of $\mathbf{X}_i(t)$ and $\mathbf{Z}_i$'s,

$$X_{ij}(t) = \mu_j(t) + \boldsymbol{\beta}_j^T(t)\mathbf{X}_{i(-j)}(t) + \boldsymbol{\alpha}_j^T\mathbf{Z}_i + Y_{ij}(t), \tag{2}$$

where $\mu_j(t)$ is the mean (intercept) function and $Y_{ij}(t)$ is the stochastic error with $E\{Y_{ij}(t)\} = \mathbf{0}$ and covariance function $\gamma_i(s,t) = \mathrm{Cov}(Y_{ij}(s), Y_{ij}(t))$. Here $\mathbf{X}_{i(-j)}(t)$ denotes the observation vector function without the jth profile observations, say $\mathbf{X}_{i(-j)}(t) = \{X_{i1}(t), \ldots, X_{i,j-1}(t), X_{i,j+1}(t), \ldots, X_{ip}(t)\}^T$. On one hand, for any fixed t, this model is just Hawkins's (1991, 1993) regression-adjusted model for multivariate process control if $\mathbf{Z}_i$'s are not involved. On the other hand, this model can be also viewed as a functional semi-parametric varying coefficient model. Since introduced by Hastie and Tibshirani (1993), the varying coefficient model has been widely applied in many scientific areas, such as economics, finance, politics, epidemiology, medical science, ecology and so on. Due to its flexibility and interpretability, in the past ten years, it has experienced rapid developments in both theory and methodology; see Fan

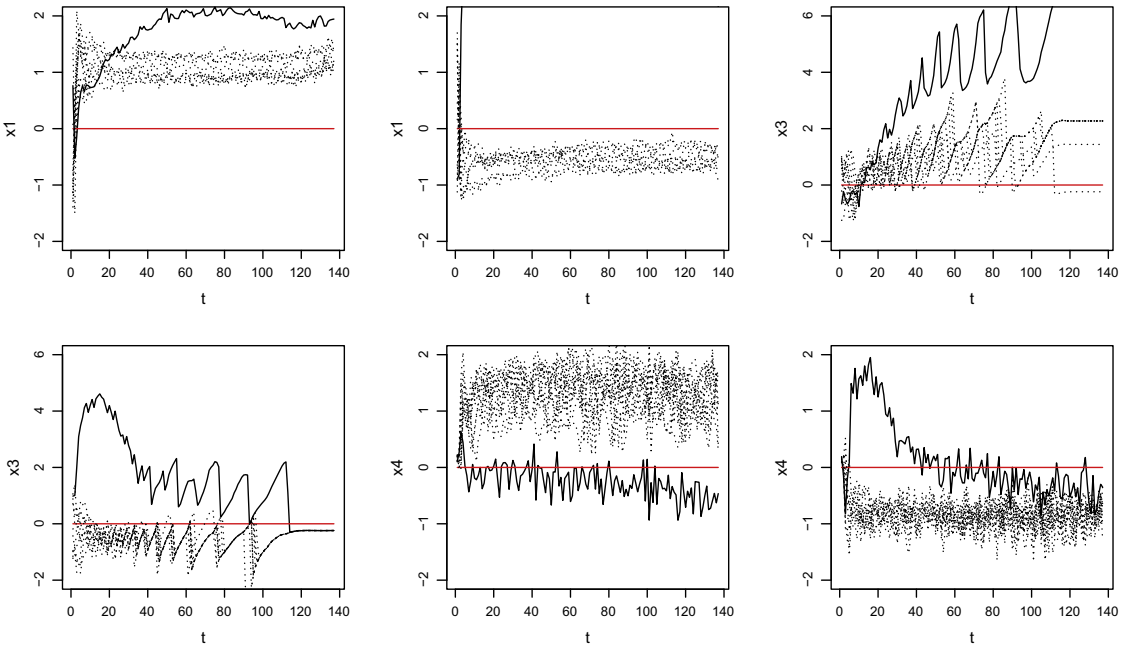

Fig. 5. The residual curves in two lots. The first wafer in each lot is represented by solid line to distinguish with the other wafers (dotted line).

and Zhang (2008) for a comprehensive survey. By using this model, we expect that the curve associations could be largely explained by the functional regression coefficient $\boldsymbol{\beta}_j^T(t)$, while the within-curve correlation would be mainly reflected through the stochastic process $Y_{ij}(t)$ and $\mathbf{Z}_i$.

3.2. Model estimation

In this part, we discuss the estimation of the model (2), including $\mu_j(t)$, $\boldsymbol{\beta}_j$, $\boldsymbol{\alpha}_j$ and $\gamma_i(s,t)$ from a reference dataset. These quantities will be used in constructing a Phase I or Phase II semi-parametric profile control chart when between-curves and within-curve correlations are present and can be described by the model (2). As our model is a semi-parametric varying coefficient model, we adapt the two-step estimation procedure proposed by Fan and Zhang (2000) for model set up. Fan and Zhang's (2000) procedure is asymptotically optimal and also computationally efficient, and therefore has become a commonly used methodology for estimating the varying coefficient models.

Let $\{t_k, k=1,\ldots,n\}$ be the distinct time point belong to $\mathcal{T}$, where n is the number of observations in each profile. From the model, the data collected at time t_j follow the linear model

$$X_{ij}(t_k) = \mathbf{U}_{ijk}^T \boldsymbol{\zeta}_j(t_k) + Y_{ij}(t_k),$$

where $\mathbf{U}_{ijk} = \{1, \mathbf{X}_{i(-j)}^T(t_k), \mathbf{Z}_i^T\}^T$ and $\boldsymbol{\zeta}_j(t_k) = \{\mu_j(t_k), \boldsymbol{\beta}_j^T(t_k), \boldsymbol{\alpha}_{jk}^T\}^T$. Note that $E\{Y_{ij}(t_k)\} = 0$ and $\mathrm{Var}\{Y_{ij}(t_k)\} = \gamma(t_k,t_k)$. Clearly, this is a standard linear model. Thus, standard least squares theory leads to

$$\check{\boldsymbol{\zeta}}_{jk} = (\tilde{\mathbf{U}}_{jk}^T \tilde{\mathbf{U}}_{jk})^{-1} \tilde{\mathbf{U}}_{jk}^T \tilde{\mathbf{X}}_{jk}.$$

$\check{\boldsymbol{\zeta}}_{jk}$ is an estimator of $\boldsymbol{\zeta}_j(t_k)$ with

$$E(\check{\boldsymbol{\zeta}}_{j,k}) = \boldsymbol{\zeta}_j(t_k), \quad \mathrm{Cov}(\check{\boldsymbol{\zeta}}_{j,k}) = \gamma(t_k,t_k)(\tilde{\mathbf{U}}_{jk}^T \tilde{\mathbf{U}}_{jk})^{-1},$$

where $\tilde{\mathbf{U}}_{jk} = (\mathbf{U}_{1jk}, \ldots, \mathbf{U}_{Njk})^T$ and $\tilde{\mathbf{X}}_{jk} = \{X_{1j}(t_k), \ldots, X_{Nj}(t_k)\}^T$.

The raw estimates are generally not smooth. Also, they are inefficient since they have not used the information from the neighboring time points and hence their efficiency can be improved. Moreover, we may want to estimate the values of the profile curves at non-design points. A natural way to refine the raw estimates is to smooth them over time. We now describe briefly how to smooth the raw estimates $\{(t_k, \check{\boldsymbol{\zeta}}_{jk}), k=1,\ldots,n\}$ for obtaining the smooth coefficient function $\boldsymbol{\zeta}_j(t)$ via one of the existing smoothing techniques, local linear kernel smoothing. Assume that $\boldsymbol{\zeta}_j(t)$ is twice continuously differentiable. Then a typical local linear estimator is given by

$$\hat{\mu}_j(t) = \sum_{k=1}^n w_k(t) \check{\mu}_{jk}, \quad \text{and} \quad \hat{\boldsymbol{\beta}}_j(t) = \sum_{k=1}^n w_k(t) \check{\boldsymbol{\beta}}_{jk}, \tag{3}$$

where $\check{\mu}_{jk}$ and $\check{\boldsymbol{\beta}}_{jk}$ are a part of the estimator $\check{\boldsymbol{\zeta}}_{jk}$, and $w_k(t) = V_k(t) \big/ \sum_{j=1}^n V_j(t)$,

$$V_j(t) = K_h(t_j - t)\big[m_2(t) - (t_j - t)m_1(t)\big],$$

$$m_l(t) = \frac{1}{n}\sum_{j=1}^n (t_j - t)^l K_h(t_j - t), \quad l = 1,\ 2,$$

and $K_h(\cdot) = K(\cdot/h)/h$ with K being a symmetric probability density function and h a bandwidth. For $\boldsymbol{\alpha}$, we may directly take the overall average, say

$$\hat{\boldsymbol{\alpha}}_j = \frac{1}{n}\sum_{k=1}^n \check{\boldsymbol{\alpha}}_{jk}. \tag{4}$$

Finally, we consider the estimation of covariance function. The covariance function of $Y_{ij}(\cdot)$ is

$$\gamma_j(t,s) = E\big[Y_{ij}(t), Y_{ij}(s)\big].$$

Under mild conditions, $Y_{ij}(t)$ admits the following expansions (Bosq, 2000)

$$Y_{ij}(t) = \sum_{1 \le k < \infty} \xi_{ijk} \upsilon_{jk}(t), \tag{5}$$

where the sequences $\{\xi_{ijk}, i=1,\ldots,N, k=1,2,\ldots\}$ are i.i.d. multinormal random variables with mean 0 and variance σ_{jk}^2. Denote λ_{jk} and $\upsilon_{jk}(\cdot)$ as the eigenvalues and eigenfunctions of the covariance operator $\gamma_j(t,s)$ respectively, i.e., they are defined by

$$\int_a^b \gamma_j(t,s)\upsilon_{jk}(s)ds = \lambda_{jk}\upsilon_{jk}(t), \quad t \in \mathcal{T}, \quad k = 1,2,\ldots$$

By (5), we have

$$\gamma_j(t,s) = \sum_{k=1}^{\infty} \sigma_{jk}^2 \upsilon_{jk}(t)\upsilon_{jk}(s).$$

Similar to the classic functional data analysis (FDA), which is the same as profile analysis in the quality engineering field, $\gamma(t,s)$ is estimated by

$$\hat{\gamma}_j(t_k,t_l) = \frac{1}{N}\sum_{1\leqslant i\leqslant N}[X_{ij}(t_k) - \mathbf{U}_{ijk}^T\hat{\boldsymbol{\xi}}_j(t_k)][X_{ij}(t_l) - \mathbf{U}_{ijl}^T\hat{\boldsymbol{\xi}}_j(t_l)],$$

where $\hat{\boldsymbol{\xi}}_{jk} = (\hat{\mu}_j(t_k), \hat{\boldsymbol{\beta}}_j^T(t_k), \hat{\boldsymbol{\alpha}}_j^T)^T$. The corresponding estimators of λ_{jk} and $\upsilon_{jk}(\cdot)$ are $\hat{\lambda}_{jk}$ and $\hat{\upsilon}_{jk}(\cdot)$, defined by

$$\int_a^b \hat{\gamma}_j(t,s)\hat{\upsilon}_{jk}(s)ds = \hat{\lambda}_{jk}\hat{\upsilon}_{jk}(t), \quad t \in \mathcal{T}, \quad k = 1,2,\ldots. \tag{6}$$

Remark 1. Suppose the model (2) is correct and under some mild conditions, the consistency of $\mu_j(t)$, $\boldsymbol{\beta}_j$ and $\boldsymbol{\alpha}_j$ can be established in a similar fashion to Theorems 1–3 in Fan and Zhang (2000). In addition, two technical conditions given in Appendix A and B are sufficient to guarantee that $\hat{\gamma}_j(t,s)$, $\hat{\lambda}_{jk}$ and $\hat{\upsilon}_{jk}(\cdot)$ are consistent estimators of $\gamma_j(t,s)$, λ_{jk} and $\upsilon_{jk}(\cdot)$, respectively.

3.3. Phase I profile monitoring

The Phase I analysis is usually divided into two types, the outlier detection and change-point detection. Here we focus on the former and discuss potential extension to the latter in Section 6. In this setting, we want to test the null hypothesis

$$H_0 : \mu_{1j}(t) = \cdots = \mu_{Nj}(t), \quad t \in \mathcal{T}, \quad j = 1,\ldots,p,$$

against the alternative

H_1: There is a subset $\mathcal{A}_N$ of $\{1,\ldots,N\}$ such that $\mu_{kj}(t) \neq \mu_{lj}(t)$, for each $k \in \mathcal{A}_N$ and $l \notin \mathcal{A}_N$,

where $\mathcal{A}_N$ is the outlying profile set. Or equivalently, write the model as

$$E[X_{ij}(t)|\mathbf{X}_{i(-j)}(t), \mathbf{Z}_i] = \begin{cases} \mu_{ij}(t) + \boldsymbol{\beta}_{ij}^T(t)\mathbf{X}_{i(-j)}(t) + \boldsymbol{\alpha}_{ij}^T\mathbf{Z}_i, & i \in \mathcal{A}_N, \\ \mu_{0j}(t) + \boldsymbol{\beta}_{ij}^T(t)\mathbf{X}_{i(-j)}(t) + \boldsymbol{\alpha}_{ij}^T\mathbf{Z}_i, & i \notin \mathcal{A}_N, \end{cases}$$

where $\mu_{0j}(\cdot)$ is the in-control common function for the jth variable. Of course, we do not specify the value of $\mu_{0j}(\cdot)$ in the hypothesis H_0 since this is the most common case in the practice of Phase I study.

Denote $\Delta_{ij}(t) = X_{ij}(t) - U_{ijk}^T\hat{\boldsymbol{\xi}}_{jk}$, $i = 1, \ldots, N$. If there is no outlying profile in the data, the mean is constant and we can expect that $|\Delta_{ij}(t)|$ is small for all $1 \leqslant i \leqslant N$ and all $t \in \mathcal{T}$. Contrarily, if there are some outliers in the sample, $\max_{1\leqslant i\leqslant N}|\Delta_{ij}(t)|$ would become large due to the shift of the mean of the outliers. We must bear in mind that the observations considered in this paper are profile data which may be in an infinite dimensional space. The covariance function would be difficult to interpret and does not give a fully comprehensible presentation of the structure of the variability in the observed data directly.

To this end, we consider to use the functional principal component analysis (FPCA) (Ramsay & Silverman, 2005) to reduce the dimension and to construct a test by the projections of the functions $\Delta_{ij}(t)$ on the principal components of the profile observations. The projections are all linear combination of $\{\hat{\upsilon}_{jk}(t), k = 1,2,\ldots\}$. The coefficients corresponding to the largest d_j eigenvalues are

$$\hat{\eta}_{ijk} = \int_a^b \{X_{ij}(t) - U_{ijk}^T\hat{\boldsymbol{\xi}}_{jk}\}\hat{\upsilon}_{jk}(t)dt, \quad i = 1,\ldots,N, \quad j = 1,\ldots,p, \quad k = 1,\ldots,d_j. \tag{7}$$

These coefficients are ideal indicators which reflect the difference between the jth profile curve and the overall mean function. In particular, $\hat{\eta}_{ijk}$ shows the deviation degree of the jth profile curves in ith sample on the kth mode of variation. Therefore, we propose the following charting statistics

$$S_{jd_j} = \max_{1\leqslant i\leqslant N}\sum_{1\leqslant k\leqslant d_j}\frac{\hat{\eta}_{ijk}^2}{\hat{\lambda}_{jk}}, \quad j = 1,\ldots,p. \tag{8}$$

With respect to the choice of the number d of the eigenfunction υ_k, there are several approaches proposed in the literature. The cumulative percentage variance approach may be the most straightforward one. The other widely like the pseudo Akaike information criterion and cross-validation (cf., Yao, Müller, & Wang, 2005) could also be used and the performance comparison of all these criterions deserves future research. In our numerical studies and real-data analysis, we choose d based on the cumulative percentage variance method which explains 85% of the variance. We shall see that this method is not only convenient but also quite effective in capturing shifts.

Now, we obtain p test statistics S_{jd_j}, $j = 1,\ldots,p$. To identify the outlier set $\mathcal{A}_N$, we may construct p control charts in which the charting statistics $\sum_{1\leqslant k\leqslant d_j}\hat{\eta}_{ijk}^2/\hat{\lambda}_{jk}$ are plotted against the observation index i and control limit $L_j(a)$ is naturally the upper a quantile of S_{j,d_j} given a false alarm rate a. Here the cut-off value or threshold, $L_j(a)$, plays an important role in dividing anomalous and non-anomalous profiles. Therefore, the basis for the decision on outlier identification lies on finding a proper threshold value. We may have the following result.

Proposition 1. *Suppose the conditions (C1)–(C6) in the Appendix A and B hold. Then, under null hypothesis H_0, for each $x \in \mathbb{R}$, we have*

$$P\left\{\frac{S_{j,d_j}}{2} - \log N - \left(\frac{d_j}{2} - 1\right)\log\log N + \log\Gamma\left(\frac{d_j}{2}\right) \leqslant x\right\} \to e^{-e^{-x}}, \quad \text{as} \quad N \to \infty. \tag{9}$$

where $e^{-e^{-x}}$ is the density of Gumbel distribution.

The proof of this proposition is sketched in the supplemental file. The asymptotic null distribution of S_{j,d_j} is independent of the nuisance parameter $\gamma_j(t,s)$ and thus S_{j,d_j} is asymptotically pivotal. By this result, we can obtain the approximate control limit

$$L_j(a) = 2c_a + 2\log N + (d_j - 2)\log\log N - 2\log\Gamma(d_j/2), \tag{10}$$

where c_a is the upper a quantile of Gumbel distribution with density $e^{-e^{-x}}$. Then, we can construct a Shewhart-type control chart for the jth variable, plotting the charting statistic S_{jd_j} along with the control limit $L_j(a)$. The false alarm rate is approximately a.

Finally, it remains to decide the false alarm rate a for each S_{j,d_j} given an overall type-I error a_0. The probability, $\Pr\{\text{at least one } S_{j,d_j} \text{ exceeds } L_j(a)\}$, seems rather difficult to exactly derive because S_{j,d_j}'s are usually correlated and the correlations depend on some unknown parameters. A simple and natural method to tackle this problem is by using the classical Bonferroni procedure, that is let $a = a_0/p$. We can expect that this will maintain an overall false alarm rate smaller than but close to its nominal analog a_0 when the number of parameters, p, is not large.

Remark 2. It is worth pointing out that in the above hypothesis test we do not consider possible changes of functional coefficients $\boldsymbol{\beta}_j^T(t)$ and non-functional $\boldsymbol{\alpha}_j$. Actually, as discussed before, the two terms $\boldsymbol{\beta}_j^T(t)\mathbf{X}_{i(-j)}(t)$ and $\boldsymbol{\alpha}_j^T\mathbf{Z}_i$ are used for accounting the between-curves and within-curve correlations and thus an implicit assumption we make here is just those correlations would not be likely to have a change. Extending the proposed procedure to monitoring $\boldsymbol{\beta}_j(t)$ and $\boldsymbol{\alpha}_j$ is highly non-trivial and is a question left for future research.

Remark 3. In the context of nonparametric regression estimation, the optimal bandwidth, h, is usually determined by minimizing the asymptotic mean square error of the estimator. Frequently used bandwidth selection techniques are data-driven methods, such as the least square cross-validation and generalized cross-validation (see Fan & Gijbels, 1996). We have observed in our simulations that the observed significance changes mildly over a wide range of value. Based on the asymptotic results in Fan and Zhang (2000) and our numerical experience, we recommend the empirical bandwidth formula,

$$h_E = c \times \left\{ n^{-1}\sum_{k=1}^{n}(t_k - \bar{t})^2 \right\}^{1/2} n^{-1/5}, \tag{11}$$

where $\bar{t} = \sum_{k=1}^{n} t_k$ and c is a constant. Empirically, c can be any value in the interval [1,2]. Here $n^{-1/5}$ is the order of the optimal bandwidth for varying coefficient model estimation and $n^{-1}\sum_{k=1}^{n}(t_k - \bar{t})^2$ is a measure of the sparseness of the design points, which also is involved in an asymptotically optimal h formula (see Fan & Zhang (2000) for details). In the next section we only use $c = 1$ and 2 to conduct our simulations. This formula works well for a wide range of models and sample sizes as shown in Section 4.

To end this section, we summarize the detailed steps for implementing the proposed RAMP procedure as follows. The *R* code for implementing the proposed procedure is available from the authors upon request.

RAMP procedure.

- Step 1. Given a dataset, setup up the model (2). Obtain parameter estimates through Eqs. (3), (4) and (6);
- Step 2. Compute the projected coefficients $\hat{\eta}_{ijk}$ based on (7);
- Step 3. Obtain the charting statistics S_{jd_j}'s using (8);
- Step 4. Choose an overall false alarm rate a_0 and let $a = a_0/p$. Calculate the control limits $L_j(a),\ j = 1,\ldots,p$ using (10);
- Step 5. Construct p control charts in which the charting statistics $\sum_{1\leqslant k\leqslant d_j}\hat{\eta}_{ijk}^2/\hat{\lambda}_{jk}$ are plotted against the observation index i, along with the control limit $L_j(a)$.

4. Simulation study

First, we study the RAMP's in-control performance under the empirical false alarm rate. The number and variety of models are too large to allow a comprehensive, all-encompassing comparison. Our goal is to show the effectiveness, robustness and sensitivity of the proposed RAMP method, and thus we only choose certain representative models for illustration. Specifically, $p = 5$ is considered and the following five functions are used for $\boldsymbol{\mu}_0 = (\mu_{01},\ldots,\mu_{0p})$:

$$\mu_{01}(t) = 1,\ \mu_{02}(t) = 2t,\ \mu_{03}(t) = t^2,$$
$$\mu_{04}(t) = \sin(2\pi t),\ \mu_{05}(t) = \log(1+t).$$

To illustrate the effectiveness of our proposed monitoring scheme, the simulated data $\mathbf{X}_i(t)$ is generated from the following two models:

Model (I):

$$X_{ij}(t) = \mu_j(t) + t\sum_{k=1}^{j-1}X_{ik}(t) + Y_{ij}(t),\ i = 1,\ldots,N,\ j = 1,\ldots,p, \tag{12}$$

where $Y_{ij}(t)$ comes from standard Brownian motion and $Y_{ij}(t)$'s are mutually independent for each j. To simulate a standard Brownian Motion, we repeatedly generated independent Gaussian random variables with mean 0 and standard deviation $1/\sqrt{n}$. The value of the Brownian Motion at time i/n is the first i increments. Clearly, this model belongs to the model (2).

Model (II):

$$\mathbf{X}_i(t) = \boldsymbol{\mu}_0(t) + \mathbf{Y}_i(t),\quad i = 1,\ldots,N, \tag{13}$$

where $\mathbf{Y}_i(t)$ is a multivariate standard Brownian motion with $E\mathbf{Y}_i(t) = \mathbf{0}$ with the covariance matrix $\boldsymbol{\Sigma}_0 = (\sigma_{j_1j_2})$ is chosen to be $\sigma_{jj} = 1$ and $\sigma_{j_1j_2} = 0.75^{|j_1-j_2|}$, for $j_1, j_2 = 1, \ldots, p$. To simulate a multivariate Brownian Motion, we repeatedly generated independent multivariate normal vectors with mean 0 and covariance $1/n\boldsymbol{\Sigma}_0$. The random vectors of the multivariate Brownian Motion at time i/n is the sum of the first i vectors.

Both the two processes were realized on a grid of n equispaced points in $\mathcal{T} = [0,1]$. To simulate a standard Brownian motion, we repeatedly generated independent Gaussian random variables with mean 0 and standard deviation $1/\sqrt{n}$. The value of the Brownian motion at time $t_k = k/n$ is the first k increments. We have studied many cases with different profile sample sizes, but we only report the results of the above two models when the sample sizes N are chosen to be 50, 100, 200 and 400. In each scenario, the empirical size is computed based on 2500 replications. In each replication, the number of the eigenfunctions d was chosen automatically by the cumulative percentage variance approach mentioned before.

Table 1 gives the observed type I error rates (empirical sizes) based on the critical values computed from the limiting distribution established by Proposition 1. The bandwidth h is chosen through the formula (11). $c = 1$ and $c = 2$ are considered in this table. The nominal significance level is chosen to be 0.05 for a representative illustration. The results indicate that the empirical sizes approximate the nominal significance levels as the sample size increases. The empirical sizes are not sensitive to the values of c and the number of observations in each profile, n. When N is relatively large, the suggested simple approximation is basically reasonable and close to the nominal one. In a majority of cases, the approximation of critical value seems to result in conservative tests. This partially stems from the reason that the use of Bonferroni procedure.

Next, we study the signal abilities of different tests in rejecting null hypothesis. Comparing the proposed procedure with alternative nonparametric methods turned out to be difficult due to the lack of an obvious comparable method. This is because most of

Table 1
Observed type I error rates (%) of the proposed Phase I monitoring procedure using the threshold values computed from the limiting distribution established by Proposition 1; The nominal size is 5%. c is the constant in the empirical bandwidth formula (11).

N	Model (I)				Model (II)			
	$n = 75$		$n = 150$		$n = 75$		$n = 150$	
	$c = 1$	$c = 2$	$c = 1$	$c = 2$	$c = 1$	$c = 2$	$c = 1$	$c = 2$
50	1.6	1.5	1.7	1.4	1.6	1.4	1.8	1.6
100	3.1	3.0	3.2	3.1	2.9	2.8	3.1	2.9
200	3.9	3.9	3.8	3.9	3.7	3.6	3.7	3.5
400	4.3	4.1	4.2	4.0	4.4	4.2	4.5	4.5

the approaches in the literature were designed for univariate profile curves. Thus, we consider to extend the functional outlier detection (abbreviated as FOD hereafter) method suggested by Yu, Zou, and Wang (2012) to the present multivariate setting. To be more specific, we apply Yu et al.'s (2012) detection procedure for each univariate profile and trigger a signal if any one test statistic out of p statistics is larger than its critical value. Note that the FOD method also makes use of FPCA and thus all the settings and choices of parameters in RAMP are adopted for FOD as well for a relatively fair comparison.

To get a more fair and clearer picture, we perform a size-corrected power comparison in the sense that the actual critical values are found through simulations so that both tests have accurate sizes, 0.05, in each case. Figs. 6 and 7 plot the empirical power against the shift size δ under Models (I) and (II) respectively, when there are five outlying profiles whose third profile functions change from $\mu_{03}(t)$ to $\mu_{03}(t)+\delta\sin(2\pi(t-1/2))$. In both figures, n is fixed as 100 and N is chosen as 100 or 200. The simulation results indicate that although both methods have satisfactory detection performance when δ is large, our proposed RAMP method outperforms FOD in certain degrees. When N increases, both methods get better as we can expect but the RAMP still performs better than FOD. This can be well understood as FOD completely ignores the correlation between profiles. In contrast, a method like RAMP, which properly utilizes the correlation information contained in the data, would be more powerful if between-profiles correlations are strong. Moreover, the comparison between the simulation results of different values of n (are not shown here but available at authors' institutions) indicates that our proposed procedure turns to be quite robust when the number of observations along each curve varies.

Finally, Fig. 8 shows the empirical power against the shift size δ under Model (II) when there are five outlying profiles whose the fourth profile observations are generated by

$$X_{i4}(t)=\mu_{04}(t)+t(X_{i1}+\delta X_{i2}(t)+X_{i3}(t))+Y_{ij}(t),\ i=1,\ldots,N,\ j=1,\ldots,p. \tag{14}$$

This out-of-control model represents both the mean function of $X_{i4}(t)$ and the relationship between the second and fourth profile components have changed. Apparently, the proposed RAMP still works better in this case. Of course, the FOD is effective as well because the mean function has changed. Specially, our simulation results indicate that if only the relationship changed, the FOD would not be effective any more as we can expect. We conducted some other simulations with various combinations of N, the number of outliers and shift scenarios to check whether the above observations still hold in other settings. The simulation results show that our method has quite satisfactory performance in other cases as well. In general, the proposed method makes sufficient information about between-profiles correlation and thus is more robust and powerful than the multi single-profile-based detection methods.

5. Example revisited

Here, we revisit the etching profile monitoring case presented in Section 2 and use that example to demonstrate how to implement the proposed scheme step by step in practice. Recall that we have $N=364$ wafers and equally spaced time points of size $n=137$ in this example. Following the proposed procedure, we first set up the model (2) by taking the two environmental variables x_6 and x_7 as the parametric part $\mathbf{Z}$ and the other five variables as $\mathbf{X}$. Hence, five sperate regression-adjusted varying coefficient models are constructed. All the observations are firstly centered and standardized.

By using the two-step estimation method given in Section 3.2, we can obtain the estimates of $\mu_j(t)$, $\boldsymbol{\beta}_j(t)$, $\boldsymbol{\alpha}_j$ and $\gamma_j(t,s)$. The estimates of coefficient functions, $\hat{\boldsymbol{\beta}}_j(t)$, are plotted in Fig. 9. Obviously, this figure suggests that those coefficient functions are unlikely to

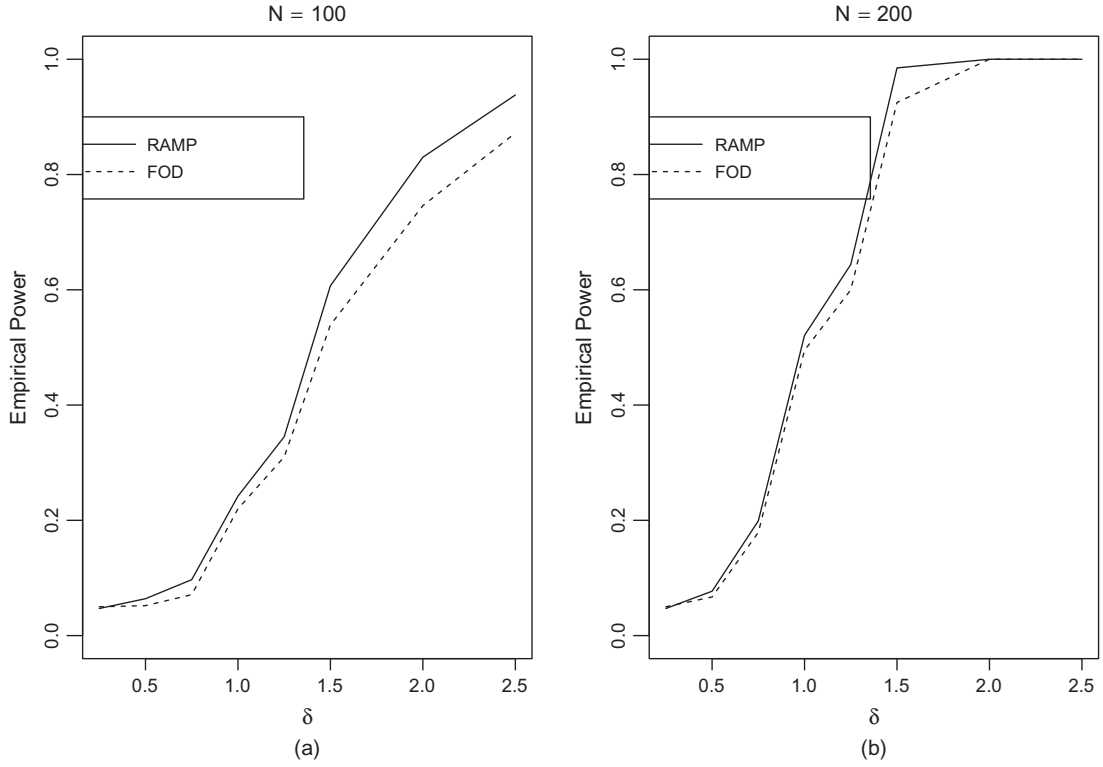

Fig. 6. Size-corrected power comparison between RAMP and FOD under Model (I) with five outlying profiles whose the third profile components have shifted from $\mu_{03}(t)$ to $\mu_{03}(t)+\delta\sin(2\pi(t-1/2))$ when $n=100$: (a) $N=100$ and (b) $N=200$. In each plot, the empirical power is plotted against δ.

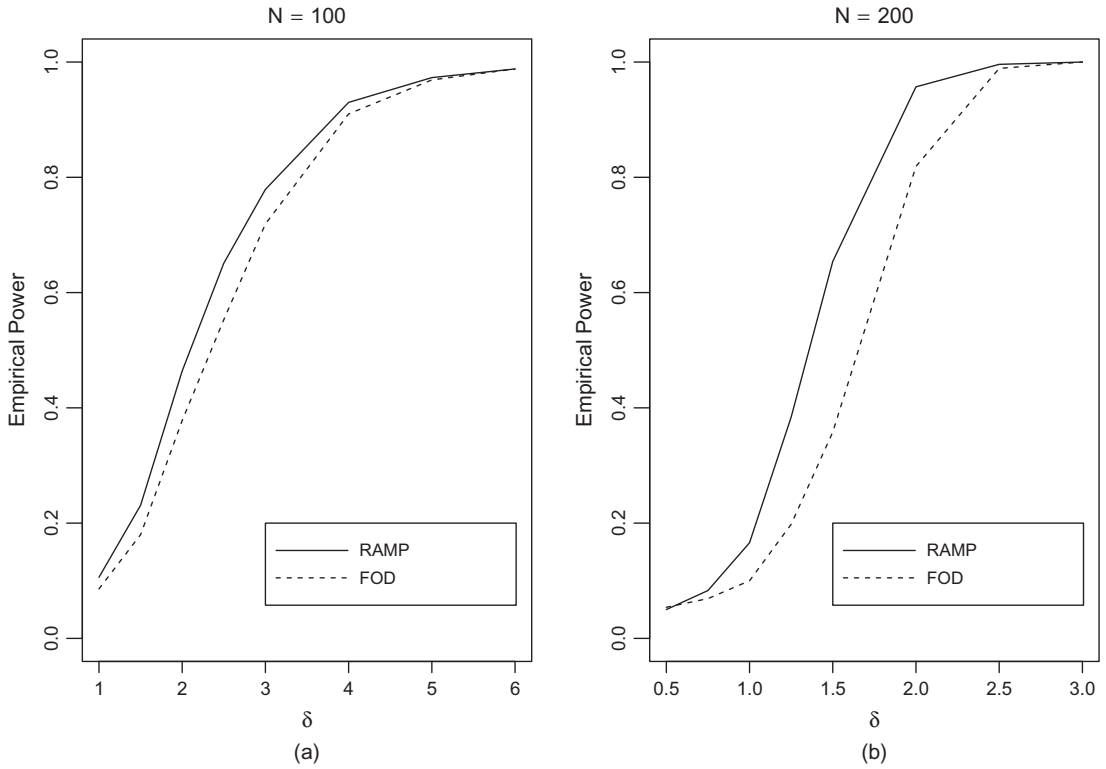

Fig. 7. Size-corrected power comparison between RAMP and FOD under Model (II) with five outlying profiles whose the third profile components have shifted from $\mu_{03}(t)$ to $\mu_{03}(t) + \delta \sin(2\pi(t - 1/2))$ when $n = 100$: (a) $N = 100$ and (b) $N = 200$. In each plot, the empirical power is plotted against δ.

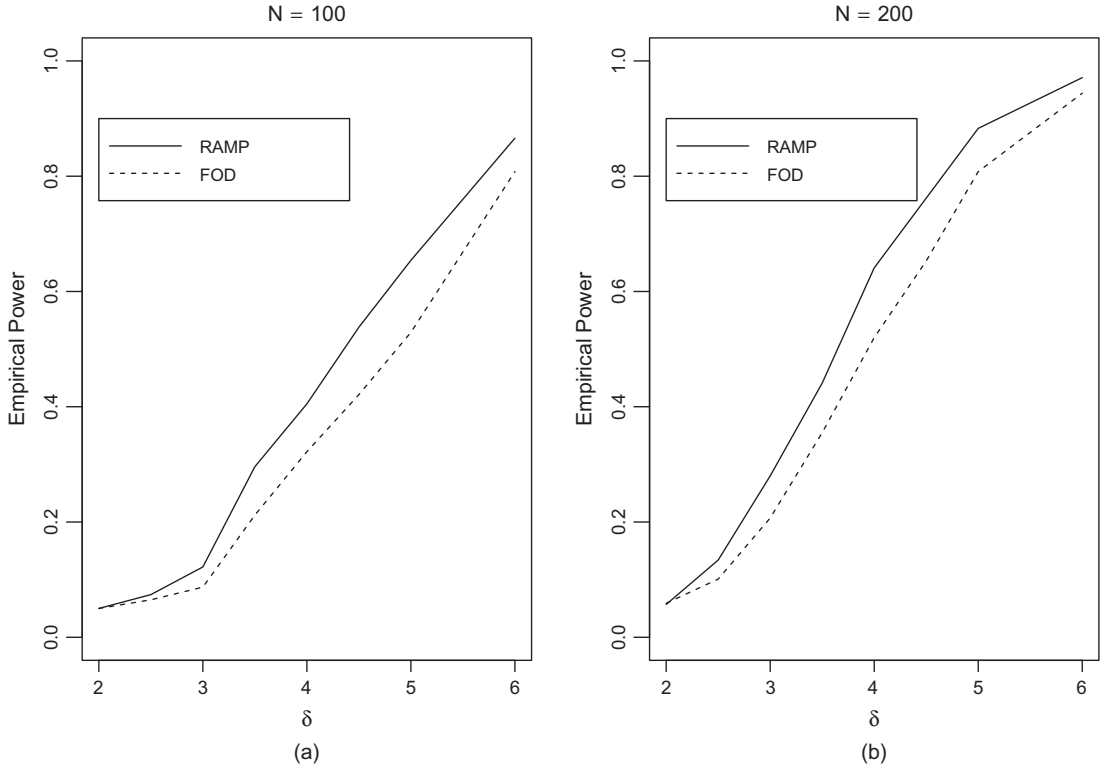

Fig. 8. Size-corrected power comparison between RAMP and FOD under Model (II) with five outlying profiles when the relationship between the second and fourth profile components have changed: (a) $N = 100$ and (b) $N = 200$. In each plot, the empirical power is plotted against δ.

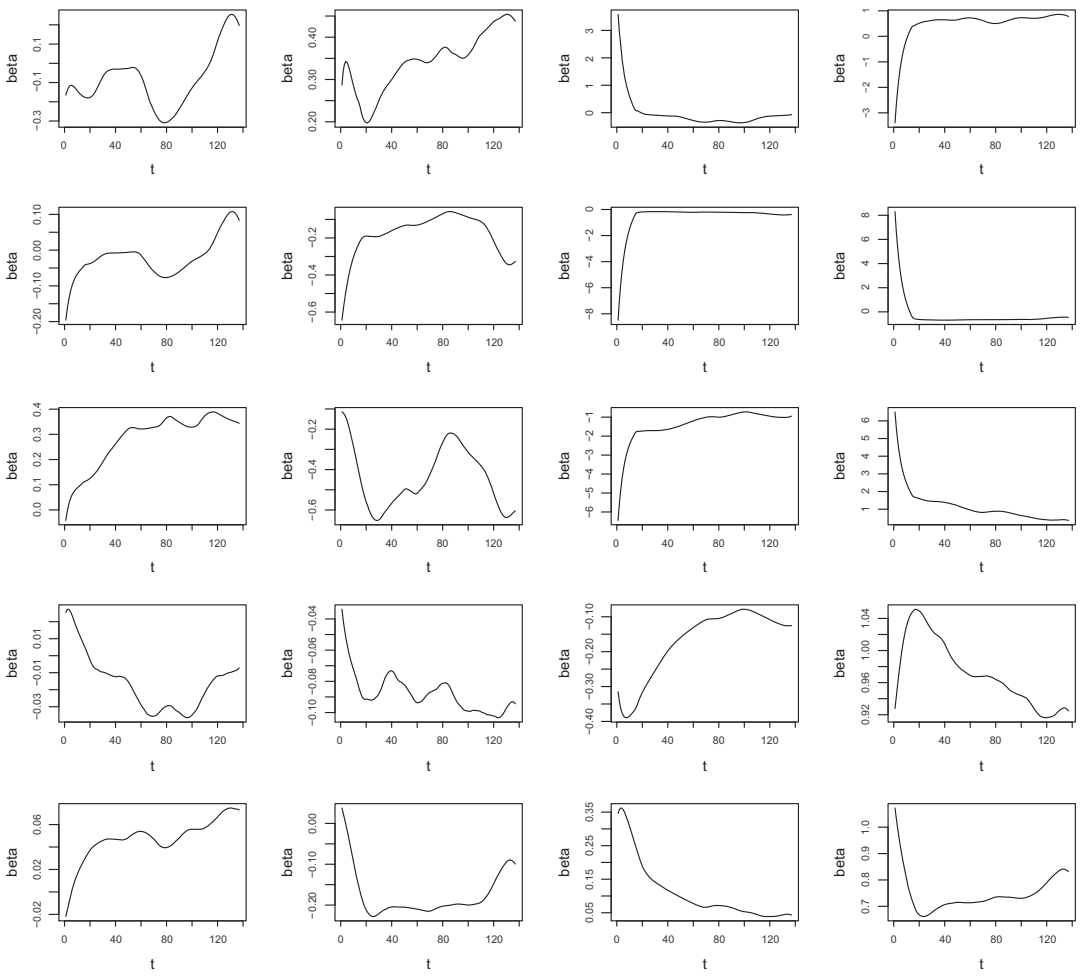

Fig. 9. The estimates of coefficient functions $\boldsymbol{\beta}(t)$.

be constant zero and many of them fluctuate a lot along with the time. This coincides with our preliminary analysis in Fig. 2 and demonstrates that profile variables have strong interrelationship. The estimates of $\boldsymbol{\alpha}_j$ for $j = 1, \ldots, 5$ are respectively (13.3, −0.413), (0.01, −0.004), (−0.013, −0.001), (0.003, −0.0005), and (−0.008, −0.003). A significance test given in Fan and Zhang (2000) is performed to determine if these values are significant in the model. The results show (not reported here) that except for the value of x_7 in $\hat{\boldsymbol{\alpha}}_4$, all the parameter estimates cannot not be rejected at the significance level 0.05.

Fig. 10(a)–(c) show the scatter plots of the residual data resulting from the fitted model for the x_1 against x_3, x_2 against x_4 and x_5 against x_4, respectively, at time point $t = 5$. This is an analog to Fig. 2 in which we use the classical model (1). It can be seen that the profile relationships between the residuals are insignificant, which illustrates that between-curves interrelationships have been largely accounted by the regression-adjusted procedure.

Similar benefits can also be observed in Figs. 11 and 12 which are parallel to Figs. 4 and 5 in Section 2 by using residuals from the fitted model (2) instead. The trend pattern shown in Fig. 4 has been removed in Fig. 11. All the residual curves do not have an upward or downward tend with the wafer index increases and seem to be like "real" stochastic processes. This is partially because the environmental effect has been taken consideration

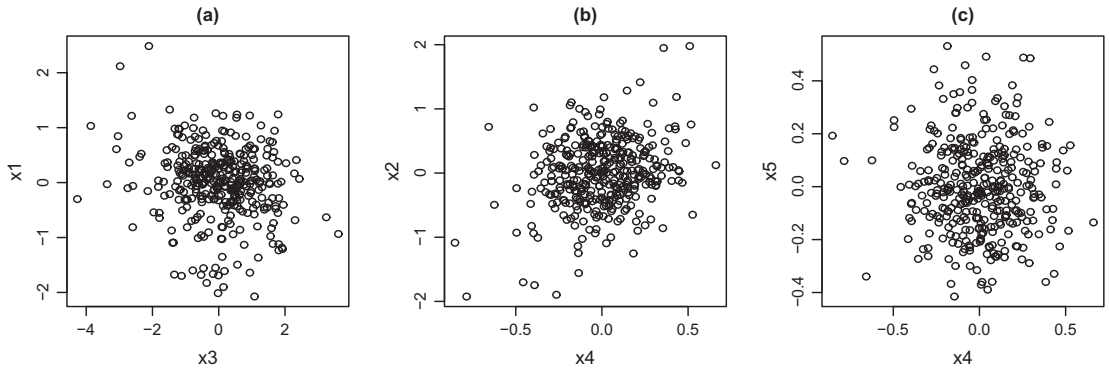

Fig. 10. The scatter plots of the residual data at $t = 5$: (a) the x_1 against x_3; (b) x_2 against x_4; and (c) x_5 against x_4.

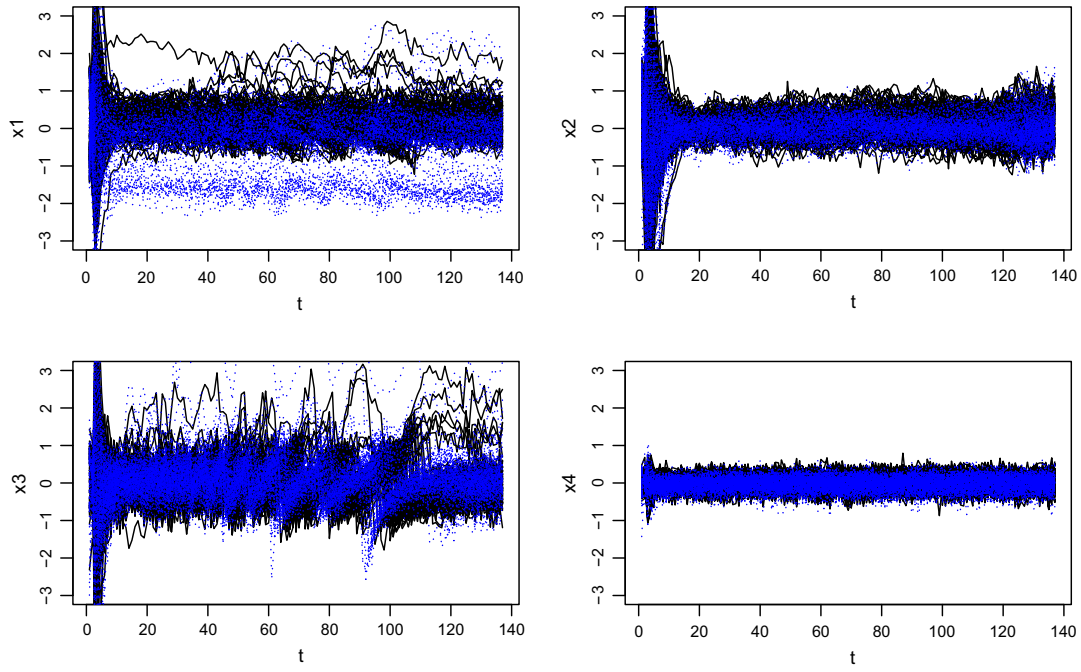

Fig. 11. The first 100 residual curves (black) and the second 100 residual curves (blue) based on the model (2) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

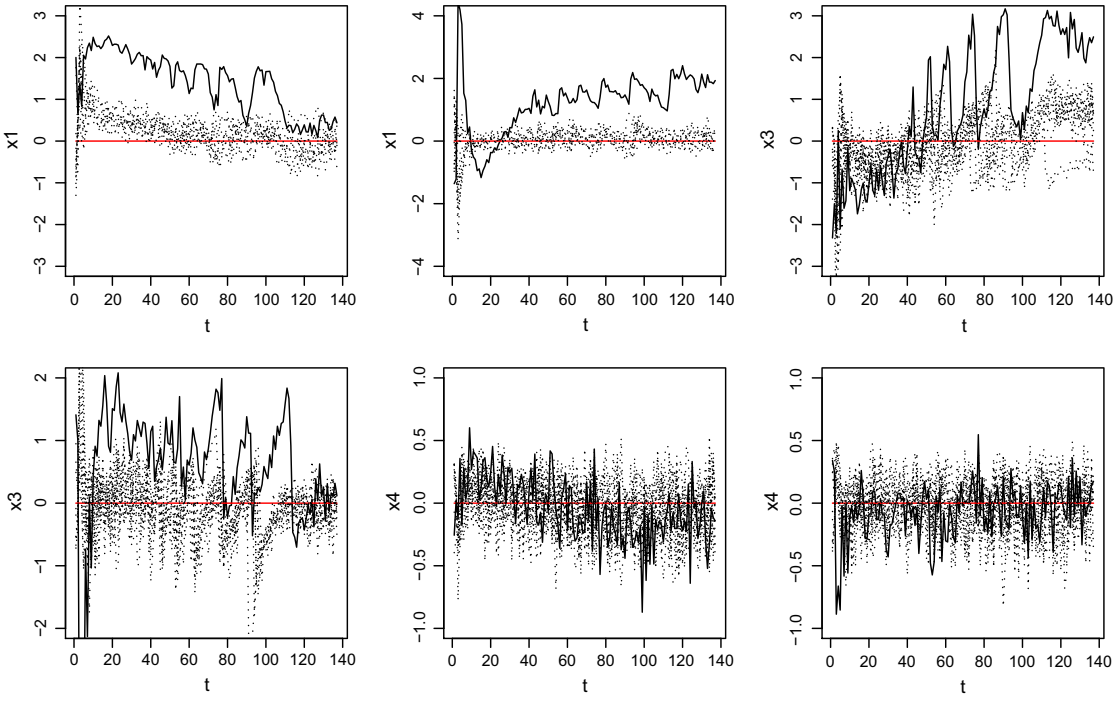

Fig. 12. The residual curves from the model (2) in the first and twelfth lots. The first wafer in each lot is represented by solid line to distinguish the other wafers (dotted line)

into the model as a parametric term. Analogously, in Fig. 12, the errors in each lot fluctuate around the zero level at each observation time, demonstrating that the intrinsic short-term pattern exhibits in Fig. 5 are removed from the present residuals. Moreover, as we expect, the "first-wafer effect" is not masked by the regression-adjusted method and still remains in these residual curves. This good feature allows us to verify the effectiveness of our proposed outlier detection procedure.

The foregoing analysis demonstrates that the proposed model is capable of characterizing most features in the dataset. Now, we are ready to use the proposed FPCA method to detect outlying profiles. We use the cumulative percentage variance approach which finds that it requires $d = 1,\ 1,\ 4,\ 1$ and 1 for the five models respectively explaining 85% of the variance. We choose the overall false alarm rate as $a = 0.1$ and then use Proposition 1 to find the control limits for the five models being 9.31, 9.31, 14.5, 9.31, 9.31

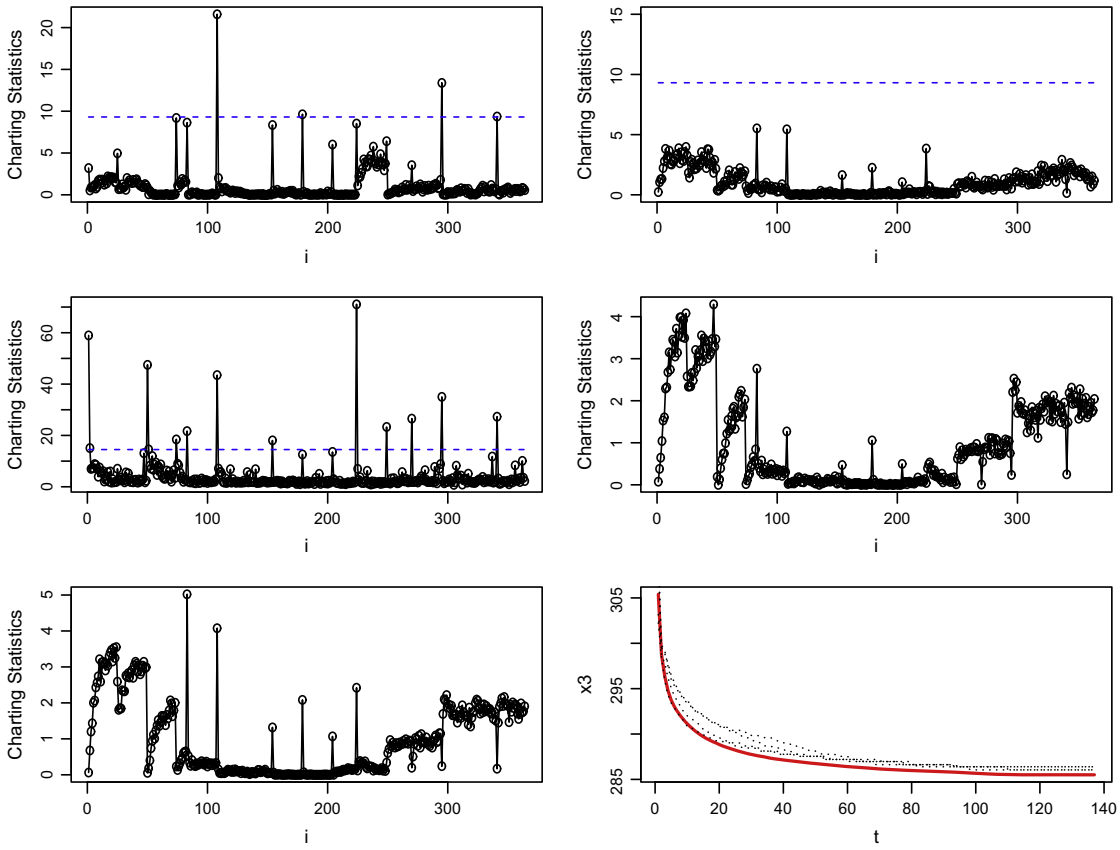

Fig. 13. The first five panels: The control charts. The last panels: The nonparametric estimate of the mean function (red solid line) and the most four significant functions (black dotted line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

respectively. After this, we compute the statistic S_{j,d_j} for $j = 1, \ldots, 5$ and construct the control charts along with their control limits as shown in Fig. 13. From this figure and the associated results, we can identify the indexes of the outlying wafers to be 1, 2, 50, 51, 74, 83, 108, 154, 179, 224, 249, 270, 295, 341. Among them, the wafers 1, 50, 83, 108, 179, 224, 295, 341 belong to the first-wafer category. That is to say, our testing procedure correctly identifies eight of sixteen first-wafer profiles at the level $a = 0.1$. This gives us some intuitive evidence which verifies our detection result to certain degree. The last panel in Fig. 13 shows the most significant four original curves for variable x_3 along with the nonparametric estimate of the mean function. The deviation of these curves from the mean function is quite clear.

6. Concluding remarks

Motivated by a real-data application in semiconductor industries, we develop a new Phase I modelling and monitoring framework based on the regression-adjustment technique and functional principal component analysis (FPCA) for multivariate profile data. The proposed method could effectively find the potential outlying profiles in a historical dataset as shown by numerical and real-data analysis.

There are a number of issues not addressed here that could be the topics of future research. First, note that we consider the cumulative percentage variance approach to choose d and an empirical method to determine the bandwidth h. Our simulation results show that such a criteria tailored for estimation do not produce a most powerful test. This finding is not surprising because similar conclusions have been made in the nonparametric regression testing problem. An ongoing effort of the authors is to develop a more proper adaptive selection method to make the test nearly optimal. Second, in some situations, our method does not always succeed in triggering a signal due to the so-called masking effect (Hawkins, 1980). A systematic method which is able to handle this issue to certain extent is needed. Moreover, the proposed method could readily be extended to the change-point setting by incorporating the suggested model and the classical binary segmentation technique (Berkes, Gabrys, Horváth, & Kokoszka, 2009). Finally, an efficient Phase II monitoring and diagnosis framework based on the proposed modelling method definitely deserves future studies.

Acknowledgment

The authors would like to thank the three anonymous referees, Area Editor, and Editor for their many helpful comments that have resulted in significant improvements in the article. This research was supported by the NNSF of China Grants 11431006, 11131002, 11371202, the Foundation for the Author of National Excellent Doctoral Dissertation of PR China 201232.

Appendix A. Technical conditions

(C1) The mean $\mu(t)$ is in $\mathcal{L}^2(\mathcal{T})$. The errors $Y_i(t)$ are independent and identically distributed (i.i.d.) mean zero Gaussian process. Their covariance function $c(t,s)$ is square integrable, i.e., is in $\mathcal{L}^2(\mathcal{T} \times \mathcal{T})$.

(C2) The eigenvalues λ_k satisfy, for some $d > 0$,

$$\lambda_1 > \lambda_2 > \ldots > \lambda_d > \lambda_{d+1}.$$

(C3) N, h and n satisfy: $h \to 0$, $n \to \infty$ and $nh/\log(1/h) \to \infty$ and $Nh^4 \log(1/h) = O(1)$.

(C4) The coefficient function $\boldsymbol{\beta}_j(t)$ is twice continuously differentiable.
(C5) The kernel function $K(u)$ is bounded and symmetric about 0 on $[-1,1]$.
(C6) For any j and t, $E[\mathbf{X}_{i(-j)}(t)\mathbf{X}_{i(-j)}^T(t)]$ is positive definite.

Appendix B. A sketch of the proof of Proposition 1

Lemma 1. *Suppose the conditions >(C1)–(C6) hold. Then,*

$$\sup_{t\in\mathcal{T}}||\hat{\boldsymbol{\beta}}_j(t)-\boldsymbol{\beta}_j(t)||=O_p\left(\frac{1}{\sqrt{nNh}}+h^2\right)\log^{1/2}(1/h),$$

$$\hat{\boldsymbol{\alpha}}_j=O_p((Nn)^{-1/2}).$$

By Condition (C3), this uniform convergence rate of the local linear smoother can be established similar to that of Carroll, Fan, Gijbels, and Wand (1997).

With Lemma 1, it is straightforward to verify that Lemma 2 in Yu et al. (2012) is valid in the present setting as well. Accordingly, the assertion in Proposition 1 is a direct application of Theorem 1 in Yu et al. (2012).

References

Akima, H. (1970). A new method of interpolation and smooth curve fitting based on local procedures. *Journal of the Association for Computing Machinery, 17*, 589–602.
Berkes, I., Gabrys, R., Horváth, L., & Kokoszka, P. (2009). Detecting changes in the mean of functional observations. *Journal of Royal Statistical Society: Series B, 71*, 927–946.
Bosq, D. (2000). *Linear processes in function spaces*. New York: Springer.
Carroll, R. J., Fan, J., Gijbels, I., & Wand, M. P. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association, 92*, 477–489.
Chicken, E., Pignatiello, J., & Simpson, J. (2009). Statistical process monitoring of nonlinear profiles using wavelets. *Journal of Quality Technology, 41*, 198–212.
Chou, S. H., Chang, S. I., & Tsai, T. R. (2014). On monitoring of multiple nonlinear profiles. *International Journal of Production Research, 52*, 3209–3224.
Chuang, S. C., Hung, Y. C., & Yang, S. F. (2013). A framework for nonparametric profile monitoring. *Computers & Industrial Engineering, 64*, 482–491.
Colosimo, B. M., Cicorella, P., Pacella, M., & Blaco, M. (2014). From profile to surface monitoring SPC for cylindrical surfaces via gaussian processes. *Journal of Quality Technology, 46*, 95–113.
Colosimo, B. M., & Pacella, M. (2007). On the use of principle component analysis to identify systematic patterns in roundness profiles. *Quality and Reliability Engineering International, 23*, 707–725.
Ding, Y., Zeng, L., & Zhou, S. (2006). Phase I analysis for monitoring nonlinear profiles in manufacturing processes. *Journal of Quality Technology, 38*, 199–216.
Fan, J., & Gijbels, I. (1996). *Local polynomial modeling and its applications*. London: Chapman and Hall.
Fan, J., & Zhang, J.-T. (2000). Two-step estimation of functional linear models with applications to longitudinal data. *Journal of Royal Statistical Society B, 63*, 303–322.
Fan, J., & Zhang, W. (2008). Statistical methods with varying coefficient models. *Statistics and its Interface, 1*, 179–195.
Hastie, T. J., & Tibshirani, R. J. (1993). Varying-coefficient models (with discussion). *Journal of the Royal Statistical Society, Series B, 55*, 757–796.
Hawkins, D. M. (1980). *Identification of outliers*. New York: Chapman and Hall.
Hawkins, D. M. (1991). Multivariate quality control based on regression-adjusted variables. *Technometrics, 33*, 61–75.
Hawkins, D. M. (1993). Regression adjustment for variables in multivariate quality control. *Journal of Quality Technology, 25*, 170–182.
Hung, Y. C., Tsai, W. C., Yang, S. F., Chuang, S. C., & Tseng, Y. K. (2012). Nonparametric profile monitoring in multi-dimensional data spaces. *Journal of Process Control, 22*, 397–403.
Jensen, W. A., Birch, J. B., & Woodall, W. H. (2008). Monitoring correlation within linear profiles using mixed models. *Journal of Quality Technology, 40*, 167–183.
Kang, L., & Albin, S. L. (2000). On-line monitoring when the process yields a linear profile. *Journal of Quality Technology, 32*, 418–426.
Khedmati, M., & Niaki, S. T. A. (2015). Phase II monitoring of general linear profiles in the presence of between-profile autocorrelation. *Quality and Reliability Engineering International*. http://dx.doi.org/10.1002/qre.1762 (in press).
Lada, E. K., Lu, J.-C., & Wilson, J. R. (2002). A wavelet-based procedure for process fault detection. *IEEE Transactions on Semiconductor Manufacturing, 15*, 79–90.
Lowry, C. A., Woodall, W. H., Champ, C. W., & Rigdon, S. E. (1992). Multivariate exponentially weighted moving average control chart. *Technometrics, 34*, 46–53.
Mahmoud, M. A., & Woodall, W. H. (2004). Phase I analysis of linear profiles with calibration applications. *Technometrics, 46*, 380–391.
Noorossana, R., Eyvazian, M., Amiri, A., & Mahmoud, M. A. (2010). Statistical monitoring of multivariate multiple linear regression profiles in Phase I with calibration application. *Quality and Reliability Engineering International*, 291–303.
Noorossana, R., Saghaei, A., & Amiri, A. (2011). *Statistical analysis of profile monitoring*. Wiley Series in Probability and Statistics.
Paynabar, K., & Jin, J. (2011). Characterization of nonlinear profiles variations using mixed-effect models and wavelets. *IIE Transactions, 43*, 275–290.
Paynabar, K., Jin, J., Agapiou, J., & Deeds, P. (2012). Robust leak tests for transmission systems using nonlinear mixed-effect models. *Journal of Quality Technology, 44*, 265–278.
Qiu, P. (2014). *Introduction to statistical process control*. Boca Raton, FL: Chapman & Hall.
Qiu, P., Zou, C., & Wang, Z. (2010). Nonparametric profile monitoring by mixed effect modeling (with discussions). *Technometrics, 52*, 265–277.
Ramsay, J. O., & Silverman, B. W. (2005). *Functional data analysis*. New York: Springer.
Soleimani, P., & Noorossana, R. (2014). Monitoring multivariate simple linear profiles in the presence of between profile autocorrelation. *Communications in Statistics-Theory and Methods, 43*, 530–546.
Soleimani, P., Noorossana, R., & Amiri, A. (2009). Simple linear profiles monitoring in the presence of within profile autocorrelation. *Computers & Industrial Engineering, 57*, 1015–1021.
Soleimani, P., Noorossana, R., & Niaki, S. T. A. (2013). Monitoring autocorrelated multivariate simple linear profiles. *International Journal of Advanced Manufacturing Technology, 67*, 1857–1865.
Viveros-Aguilera, R., Steiner, S. H., & Mackay, R. J. (2014). Monitoring product size and edging from bivariate profile data. *Journal of Quality Technology, 46*, 199–215.
Wang, F. K., & Tamirat, Y. (2014). Process yield analysis for autocorrelation between linear profiles. *Computers & Industrial Engineering, 71*, 50–56.
Williams, J. D., Woodall, W. H., & Birch, J. B. (2007). Statistical monitoring of nonlinear product and process quality profiles. *Quality and Reliability Engineering International, 23*, 925–941.
Woodall, W. H. (2007). Current research on profile monitoring. *Producão, 17*, 420–425.
Woodall, W. H., Spitzner, D. J., Montgomery, D. C., & Gupta, S. (2004). Using control charts to monitor process and product quality profiles. *Journal of Quality Technology, 36*, 309–320.
Yao, F., Müller, H.-G., & Wang, J.-L. (2005). Functional linear regression analysis for longitudinal data. *Annals of Statistics, 33*, 2873–2903.
Yu, G., Zou, C., & Wang, Z. (2012). Outlier detection in the functional observations with applications to profile monitoring. *Technometrics, 54*, 308–318.
Zhang, H., & Albin, S. (2009). Detecting outliers in complex profiles using a χ^2 control chart method. *IIE Transactions, 41*, 335–345.
Zou, C., Ning, X., & Tsung, F. (2012). LASSO-based multivariate linear profile monitoring. *Annals of Operation Research, 192*, 3–19.
Zou, C., Tsung, F., & Wang, Z. (2007). Monitoring general linear profiles using multivariate EWMA schemes. *Technometrics, 49*, 395–408.
Zou, C., Tsung, F., & Wang, Z. (2008). Monitoring profiles based on nonparametric regression methods. *Technometrics, 50*, 512–526.