A Change Point Approach for Phase-I Analysis in Multivariate Profile Monitoring and Diagnosis

Kamran Paynabar, Peihua Qiu & Changliang Zou

A Change Point Approach for Phase-I Analysis in Multivariate Profile Monitoring and Diagnosis

Kamran Paynabar[1], Peihua Qiu[2] and Changliang Zou[3*]

[1]*H. Milton Stewart School of Industrial & Systems Engineering,*

Georgia Institute of Technology, USA;

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

[3]*Institute of Statistics, Nankai University, China*

Abstract

Process monitoring and fault diagnosis using profile data remains an important and challenging problem in statistical process control (SPC). Although the analysis of profile data has been extensively studied in the SPC literature, the challenges associated with monitoring and diagnosis of multichannel (multiple) nonlinear profiles are yet to be addressed. Motivated by an application in multi-operation forging processes, we propose a new modeling, monitoring and diagnosis framework for phase-I analysis of multichannel profiles. The proposed framework is developed under the assumption that different profile channels have similar structure so that we can gain strength by borrowing information from all channels. The multi-dimensional functional principal component analysis is incorporated into change-point models to construct monitoring statistics. Simulation results show that the proposed approach has good performance in identifying change-points in various situations compared with some existing methods. The codes for implementing the proposed procedure are available in the supplementary material.

Keywords: Functional data analysis; Functional principal component analysis; Multichannel profiles; Nonlinear profile; Statistical process control

*Corresponding author: nk.chlzou@gmail.com

1 Introduction

Advanced sensing technologies have facilitated real-time data collection for process monitoring and fault diagnosis in complex manufacturing systems. Profile or functional data, where the measured variable is represented as a function of some explanatory variables such as time or space, is one of the common data types collected by sensing systems. Profile monitoring has been extensively studied in statistical process control (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 and Albin 2000; Mahmoud and Woodall 2004; Zou et al. 2007; Williams et al. 2007), monitoring methods based on mixed-effect models (Jensen et al. 2008; Qiu et al. 2010; Paynabar et al. 2012), wavelet methods (Jin and Shi 1999; Chicken et al. 2009; Paynabar and Jin 2011), and nonparametric regression methods (Qiu and Zou 2010). Extensive discussion about various research problems on profile monitoring can be found in Woodall (2007), Noorossana et al. (2011) and Qiu (2014, Chap 10).

Most existing profile monitoring methods are applicable to single-stream profile data only, and they cannot be used in situations where process performance is measured by multichannel sensors. For example, as shown in Figure 1(a), in multi-operation forging processes with transfer or progressive dies, tonnage force exerted on all dies are measured by four strain sensors that are mounted on four columns of the press. In this sensing system, each sensor records the tonnage force profile at the predefined equal sampling interval of a rotational crank angle. This results in multichannel profile data shown in Figure 1(b). There is very little work in the literature focusing on the analysis of multichannel profile data. Lei et al. (2010) transformed the multichannel tonnage profiles to a single aggregated profile by adding signal magnitudes across the channels. Clearly, the aggregated profile data, cannot capture all information existing in individual profile channels. Alternatively, one may analyze each profile channel separately. This approach, however, overlooks the inter-relationship among profile channels that provides valuable information about the process (See for example Paynabar et al., 2013). In order to fully utilize the information of multichannel profiles, Paynabar et al. (2013) used multilinear dimension reduction techniques to extract informative features from multichannel profiles, while considering their inter-relationships. Specifically, they used a tensor-to-vector projection method known as Uncorrelated Multilinear Principal Component Analysis (UMPCA) for feature extraction and utilized these features for fault detection and classification in multi-operation forging processes. Their method, however, relied on

the assumption that sufficient profile data from both normal and faulty operations are available that can be used for training classifiers. Consequently, their method cannot be used for phase-I process control, in which only unlabeled profile data are available.

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. The main objective of this paper is to develop a new method for phase-I monitoring of multichannel profile data. There are several challenges associated with this task. First, the high-dimensionality of multichannel profiles prevents us from using existing multivariate statistical methods (e.g., Srivastava and Worsley 1986) due to the "curse of dimensionality." Second, it is challenging to properly model the inter-relationship of profile channels. Third, it is also difficult to check the stability of the process from which the profile data were collected. To address these challenges, in this paper, we first utilize the multi-dimensional functional principal component analysis (MFPCA) to model multichannel profiles and their inter-relationships, and to extract their important features as well. Then, we develop a change-point model that uses the extracted features for phase-I monitoring. Furthermore, a diagnostic procedure is developed to identify the set of altered profile channels causing the out-of-control alarms.

The remainder of the paper is organized as follows. Our proposed modelling, monitoring and diagnosis procedures are described in Section 2. In Section 3, the performance of the proposed method is evaluated and compared with some alternative methods using simulations. In Section 4, we revisit the forging process presented in Figure 1 and use it to demonstrate the implementation of the proposed scheme. Several remarks conclude the article in Section 5. Some technical details, including some asymptotic results and their proof, are provided in an appendix, which is available online as supplementary materials. The codes for implementing the proposed procedure are available in the supplementary material.

2 Methodology

The phase-I monitoring and diagnosis methodology developed in this paper consists of three main steps. First, in order to model multichannel profiles and their inter-relationships, the MFPCA is applied to a sample of profile data. Similar to the functional PCA, the MFPCA would enable us to

reduce the infinite-dimensional nature of the functional estimation problem to a few features that can describe the major structure of profile variation well. In the second step, based on the extracted features, a change-point model is developed for detecting possible process changes and estimating the change times. Finally, a Bayesian Information Criterion (BIC)-based variable selection procedure is used for identification of out-of-control profile channels. The details of each step are elaborated in subsequent sections.

2.1 Multi-dimensional Functional Principal Component Analysis

We first review the widely used FPCA technique as discussed in Ramsay and Silverman (2005). The basic idea of FPCA is to decompose the space of curves into principal directions of variation. Let $X(u)$, be a squared integrable random function with mean $\mu(u) = E\{X(u)\}$ and covariance function $\mathrm{Cov}\{X(u), X(u')\}$ for $u, u' \in [0, 1]$. The spectral decomposition of $\mathrm{Cov}\{X(u), X(u')\}$ is given by $\sum_{k=1}^{\infty} \lambda_k \phi_k(u) \phi_k(u')$, where $\lambda_1 \geq \lambda_2 \geq \cdots$ are ordered nonnegative eigenvalues and ϕ_k's are the corresponding orthogonal eigenfunctions with unit L_2 norms. Then, the random function $X(u)$ can be written as $X(u) = \mu(u) + \sum_{k=1}^{\infty} \xi_k \phi_k(u)$, where $\xi_k = \int \{X(u) - \mu(u)\} \phi_k(u) du$ are uncorrelated random variables, known as principal component (PC) scores or loadings, with mean zero and variance λ_k. Note that in practice, only a few eigenvalues and eigenfunctions are required to capture the important modes of variations of a sample of random functions.

We need to generalize the FPCA technique to the MFPCA that is capable of modeling multi-channel profiles by appropriately addressing the correlations between individual functional observations. Our treatment is similar to that in Di et al. (2009) which introduces the multi-level functional principal component analysis for fitting a two-way functional ANOVA model. See Dubin and Müller (2005) for a related discussion. Suppose a random sample of m profiles each with p channels is available. We require $m > p$ throughout this paper. These profiles are represented by a set of vector-valued functional data denoted by $\{X_i(u), i = 1, \ldots, m\}$, where $X_i(u) = (X_{i1}(u), \ldots, X_{ip}(u))^{\mathsf{T}}$ is the vector of profiles observations measured at u from individual channels. In some applications, like the forging process described before, u represents time. However, profiles can be functions of other independent variables, such as locations, temperature, etc. In addition, though we treat profiles as random functions for facilitating the exposition, profiles are usually observed at discrete sets of sampling points and this will be discussed later in Section 2.3.

Moreover, it is assumed that the set of multichannel profiles can be modeled as independent realizations of an underlying multivariate stochastic process,

$$X_i(u) = \mu(u) + Y_i(u), \quad i = 1, \ldots, m, \tag{1}$$

where $\mu(u)$ is the mean vector function of the stochastic process and $Y_i(u)$ is the stochastic error with $E[Y_i(u)] = 0$. Without loss of generality, we assume that $u \in \mathcal{T} = [0, 1]$. Similar to the single functional observations, the random vector-valued function $Y_i(u)$ can be represented by a set of orthonormal basis functions as follows

$$Y_i(u) = \sum_{1 \le k < \infty} \xi_{ik} \upsilon_k(u), \tag{2}$$

where the sequences $\{\xi_{ik}, i = 1, \ldots, m, \ k = 1, 2, \ldots\}$ are independently and identically distributed (i.i.d.) p-variate random variables with mean 0 and covariance matrix Σ_k, and the eigenfunctions $\upsilon_k(u)$, $k = 1, 2, \ldots$, form an orthonormal basis. The covariance matrix Σ_k is

$$\Sigma_k = E\left(\xi_{ik}\xi_{ik}^\mathsf{T}\right) = E\left\{\left(\int_0^1 Y_i(u)v_k(u)du\right)\left(\int_0^1 Y_i^\mathsf{T}(u)v_k(u)du\right)\right\}.$$

The basis functions $\upsilon_k(u)$ are not random and the randomness in the model enters via the p-variate coefficient vectors $\xi_{ik}, k = 1, 2, \ldots$. It should be emphasized that our model (2) assumes that all the p random functions share a common set of eigenfunctions and their inter-relationships are essentially described by the correlations between the components of ξ_{ik}. This assumption distinguishes from the standard multivariate FPCA (see Sec. 8.5 in Ramsay and Silverman 2005) in which the vector-valued random functions are represented by a set of multivariate eigenfunctions. Model (2) is valid in the case when the multiple profile curves exhibit similar pattern and thus we can expect that it is particularly useful for our multichannel profiles described in Section 1 (c.f. Figure 1). Model (2) allows us to integrate all the information across multiple channels, resulting in an elegant procedure as developed below.

In order to implement MFPCA, we usually need to analyze the covariance function $\Lambda(u, u') = (\Lambda_{jl}(u, u'))_{p \times p}$ with $\Lambda_{jh}(u, u') = \text{Cov}\{Y_{ij}(u), Y_{ih}(u')\}$, which is an asymmetric p by p matrix given u and u'. The analysis of $\Lambda(u, u')$ is fairly flexible and can be parametric or nonparametric. For example, if we assume that $\Lambda_{jl}(u, u') = \rho(|u - u'|; a)$ for some correlation function ρ and a coefficient a, that is the covariance depends only on the absolute difference $|u - u'|$, then the correlation structure includes the nonhomogeneous Ornstein-Uhlenbeck process and Gaussian process models. By

(2) we can see that $\Lambda_{jh}(u, u') = 0$ for any u, u' if the jth and hth profile channels are independent of each other, say $\mathrm{Cov}(\xi_{ikj}, \xi_{ikh}) = 0$ for all k, where ξ_{ikj} is the jth component of $\boldsymbol{\xi}_{ik}$.

Rather than working with $\Lambda(u, u')$, the eigenfunctions can be obtained from a scalar-valued correlation function, $c(u, u') = E\{\langle \boldsymbol{Y}_i(u), \boldsymbol{Y}_i(u')\rangle\}$ that is aggregated across channels. Here we use a straightforward definition of the inner product between multivariate functions, i.e., $\langle \boldsymbol{f}, \boldsymbol{g} \rangle = \sum_{j=1}^{p} f_j g_j$ introduced by Dubin and Müller (2005). Of course, other inner products can be used instead. $c(u, u')$ can be viewed as a quantity which measures the "overall covariance" between the two random vectors $\boldsymbol{Y}_i(u)$ and $\boldsymbol{Y}_i(u')$ but its practical interpretation is not quite clear. It is primarily a device to implement MFPCA so that we can obtain the set of eigenfunctions. Using this inner-product operator and Eq.(2), the covariance function is

$$c(u, u') = \sum_{k=1}^{\infty} \sum_{j=1}^{p} \sigma_{jk}^2 \upsilon_k(u) \upsilon_k(u'),$$

where $\sigma_{jk}^2 = E[\xi_{ikj}^2]$. We let $\lambda_k \equiv \sum_{j=1}^{p} \sigma_{jk}^2$ and $\upsilon_k(\cdot)$ denote the eigenvalues and eigenfunctions of the covariance operator $c(u, u')$, respectively. Given $c(u, u')$, λ_k and $\upsilon_k(u)$ can be calculated by solving the following system of equations

$$\int_0^1 c(u, u')\upsilon_k(u')du' = \lambda_k \upsilon_k(u), \quad u \in \mathcal{T}, \quad k = 1, 2, \ldots$$

By assuming (2), we only need to work with the scalar function $c(u, u')$ which aggregates all the covariance information across the elements of $\boldsymbol{Y}_i$ (through the inner-product operator). In contrast, in the standard MFPCA, FPCA is performed on each $\Lambda_{jh}(u, u')$ for $j \leq h$ and accordingly $p(p+1)/2$ eigen-systems need to be solved.

2.2 Estimation of MFPCA Parameters

In this section, we discuss how to use a sample of multichannel profiles $\{\boldsymbol{X}_1(u), \ldots \boldsymbol{X}_m(u)\}$ to estimate the model parameters of the MFPCA. We begin with the estimation of the covariance function $c(u, u') = E\{\langle \boldsymbol{Y}_i(u), \boldsymbol{Y}_i(u')\rangle\}$. Using the method of moments, $c(u, u')$ can be estimated by

$$\tilde{c}(u, u') = \frac{1}{m} \sum_{i=1}^{m} \langle \{\boldsymbol{X}_i(u) - \bar{\boldsymbol{X}}(u)\}, \{\boldsymbol{X}_i(u') - \bar{\boldsymbol{X}}(u')\}\rangle, \tag{3}$$

where $\bar{X}(u) = m^{-1} \sum_{i=1}^{m} X_i(u)$. Consequently, the corresponding estimators of λ_k and $\upsilon_k(\cdot)$, respectively denoted by $\tilde{\lambda}_k$ and $\tilde{\upsilon}_k(\cdot)$, are calculated by solving

$$\int_0^1 \tilde{c}(u, u') \tilde{\upsilon}_k(u') du' = \tilde{\lambda}_k \tilde{\upsilon}_k(u), \quad u \in \mathcal{T}, \quad k = 1, 2, \ldots.$$

After calculating $\tilde{\upsilon}_k(\cdot)$, the covariance matrix of $\boldsymbol{\xi}_{ik}$'s, i.e., $\boldsymbol{\Sigma}_k$, can be estimated by

$$\tilde{\boldsymbol{\Sigma}}_k = \frac{1}{m} \sum_{i=1}^{m} \int_0^1 \{X_i(u) - \bar{X}(u)\} \tilde{\upsilon}_k(u) du \int_0^1 \{X_i(u) - \bar{X}(u)\}^{\mathsf{T}} \tilde{\upsilon}_k(u) du.$$

Under some mild conditions, if the sample of profiles are collected from an in-control process, the foregoing estimators are consistent (see Proposition 3 in Appendix B).

Remark 1 Note that for a given profile sample, the eigenfunctions υ_k's can also be consistently estimated using the univariate method of moments. Particularly, one can solve the functional eigensystem with the sample covariance function $m^{-1} \sum_{i=1}^{m} \{X_{ij}(u) - \bar{X}_j(u)\}\{X_{ij}(u') - \bar{X}_j(u')\}$ for any $j \in \{1, \ldots, p\}$ to obtain consistent estimators. However, under the assumptions that model (2) is valid, the set of estimators using joint multichannel profile data would be more efficient. Some theoretical comparison can be made using similar arguments to those in Hall and Hosseini-Nasab (2006), which is beyond the scope of this paper. However, some numerical evidence is presented in Section 3. $\qquad\square$

2.3 Estimation with the discretized MFPCA

In practice, each function $X_i(u)$, is measured at a set of grid points $\{u_{it} : t = 1, \ldots, n_i\}$. If sampling points are the same across different profiles, i.e., $u_{it} = u_t$ and $n_i = n$, then the functional principal components problem can be solved by applying principal components analysis to the $n \times p$ matrix of the observed data (size m), say $\{X_i(u_t), i = 1, \ldots, m, t = 1, \ldots, n\}$, where $X_i(u_t) = (X_{i1}(u_t), \ldots, X_{ip}(u_t))^{\mathsf{T}}$. To be more specific, the covariance function $\tilde{c}(u, u')$ can be discretely evaluated at n points $u_1, \ldots, u_n$, resulting in the covariance matrix $\tilde{\mathbf{C}} = (\tilde{C}_{ts})_{n \times n}$, where

$$\tilde{C}_{ts} = \frac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{p} \{X_{ij}(u_t) - \bar{X}_j(u_t)\}\{X_{ij}(u_s) - \bar{X}_j(u_s)\}, \tag{4}$$

and $\bar{X}_j(u_t) = m^{-1} \sum_{i=1}^{m} X_{ij}(u_t)$. Then, we can solve the eigenvalue system of $\tilde{\mathbf{C}}$ and obtain the eigenvalues $\tilde{\lambda}_k$ and eigenvectors $(\tilde{v}_k(u_1), \ldots, \tilde{v}_k(u_n))^{\mathsf{T}}$. As a consequence, the (j, h)th element of $\tilde{\boldsymbol{\Sigma}}_k$, $\tilde{\sigma}_{kjh}$, can be computed by

$$\tilde{\sigma}_{kjh} = \frac{1}{m} \sum_{i=1}^{m} \left[\sum_{t=1}^{n} \{X_{ij}(u_t) - \bar{X}_j(u_t)\} \tilde{v}_k(u_t) \sum_{t=1}^{n} \{X_{ih}(u_t) - \bar{X}_h(u_t)\} \tilde{v}_k(u_t) \right]. \tag{5}$$

If the sampling grid is sparse or the sampling points are unequally spaced, one can smooth the profile data first by applying any smoothing techniques such as spline or kernel regression methods, and then use the predicted (interpolated) values on an equally-spaced grid of points. That is to say, we estimate the individual curves, sample the curves on a fine grid, and then perform FPCA on the resulting data as in (4) and (5). See Ramsay and Silverman (2005; Sec. 5) and Yao et al. (2005) for some detailed discussion. Another problem of critical importance to profile data is that we often see that variation in functional observations involves both phase and amplitude, especially when different sensors are collecting signal data. We need to transform curves by transforming their arguments u rather than the values $X(u)$ so that two functions can be compared, which we call curve registration or alignment of the data (i.e., the scale of u itself has to be distorted or transformed). Detailed discussion and general methods for curve registration can be found in Sec. 7 of Ramsay and Silverman (2005).

2.4   Change-Point Model for Phase-I Monitoring

Change-point models have been used extensively in both phase-I and phase-II SPC (e.g., Hawkins et al. 2003). Zou et al. (2006) and Mahmoud et al. (2007) employed change-point models for monitoring linear profiles. Chicken et al. (2009) and Paynabar and Jin (2011) developed change-point models based on wavelet coefficients for analyzing nonlinear profiles. Zou et al. (2009) incorporated kernel smoothing methods into change-point models for monitoring nonlinear profiles. Berkes et al. (2009) developed a methodology for detecting change-points in the mean of univariate functional data using FPCA. The foregoing change-point models, however, are only applicable to single profile data.

Assume that the mean multichannel profiles change at an unknown time point τ. Provided

that the covariance function is unchanged, model (1) can be rewritten as

$$
X_i(u) = \begin{cases} \mu_1(u) + Y_i(u), & \text{for} \quad i = 1, \ldots, \tau, \\ \mu_2(u) + Y_i(u), & \text{for} \quad i = \tau + 1, \ldots, m, \end{cases} \tag{6}
$$

where τ is the unknown change point, and $\mu_1(u)$ and $\mu_2(u)$ are the mean vectors of profiles before and after the change. If all profiles are from an identical distribution, then $\mu_1(u) = \mu_2(u)$. Thus, the following hypotheses can be used to check whether the collected profile data follow an identical distribution:

$$
\begin{cases} H_0 : \mu_1(u) = \mu_2(u), & \text{for all } 1 \leq \tau < m \text{ and } u \in \mathcal{T} \\ H_a : \mu_1(u) \neq \mu_2(u), & \text{for some } 1 \leq \tau < m \text{ in} (6). \end{cases} \tag{7}
$$

To formulate the change-point model, the classical binary segmentation procedure is adopted as follows. For each $1 \leq l < m$, define the standardized difference of profile sample means before and after a potential change-point l as

$$
\Delta_l(u) = \sqrt{\frac{l(m - l)}{m}} \left\{ \frac{1}{l} \sum_{i=1}^{l} X_i(u) - \frac{1}{m - l} \sum_{i=l+1}^{m} X_i(u) \right\}. \tag{8}
$$

If there is no shift, then this difference should be small for all $1 \leq l < m$ and $u \in \mathcal{T}$. Otherwise, it should be large when l is the true change-point.

Since the observations are in an infinite-dimensional domain, we consider to use the MFPCA to reduce the dimensionality and work with the low-dimensional projections of $\Delta_l(u)$ on the principal components of the profile data. For this purpose, the eigenfunctions of the estimated $c(u, u')$ are used for the projection. However, under the alternative hypothesis in (7), in which the sample includes both the in-control and out-of-control profiles, the set of estimators discussed in Section 2.2 would not be consistent anymore (see Claim 2 in Appendix B for more details). To address this issue, we suggest to use the moving ranges of neighboring samples to estimate the covariance function, say

$$
\hat{c}(u, u') = \frac{1}{2(m - 1)} \sum_{i=1}^{m-1} \langle \{X_{i+1}(u) - X_i(u)\} \{X_{i+1}(u') - X_i(u')\} \rangle.
$$

Similar methods for robust estimation of the covariance matrix have been studied in the literature.

See, for instance, Williams et al. (2007) and the references therein. Let $\hat{\lambda}_k$ and $\hat{v}_k(\cdot)$ denote the corresponding eigenfunctions and eigenvalues of $\hat{c}(u, u')$, respectively. Then, Σ_k discussed in (2) can be estimated by

$$\hat{\Sigma}_k = \frac{1}{2(m-1)} \sum_{i=1}^{m-1} \left(\int_0^1 \{X_{i+1}(u) - X_i(u)\}\hat{v}_k(u)du \right) \left(\int_0^1 \{X_{i+1}(u) - X_i(u)\}^\mathsf{T}\hat{v}_k(u)du. \right.$$

Proposition 1 in Appendix B shows that under some mild conditions, $\hat{\lambda}_k$, $\hat{c}_k\hat{v}_k(\cdot)$ and $\hat{\Sigma}_k$ are consistent estimators of λ_k, $v_k(\cdot)$ and Σ_k, respectively, where $\hat{c}_k = \mathrm{sgn}\left\{ \int_0^1 v_k(u)\hat{v}_k(u)du \right\}$. These quantities can be discretely computed in a similar fashion to (4) and (5).

Let $\hat{\boldsymbol{\eta}}_{lk}$, for $k = 1, \ldots, d$, be the projected vectors of $\boldsymbol{\Delta}_l(u)$ corresponding to the largest d eigenvalues, which are defined by

$$\hat{\boldsymbol{\eta}}_{lk} = \int_0^1 \boldsymbol{\Delta}_l(u)\hat{v}_k(u)du, \quad l = 1, \ldots, m-1, \quad k = 1, \ldots, d$$
$$\approx \sum_{t=1}^n \boldsymbol{\Delta}_l(u_t)\hat{v}_k(u_t). \tag{9}$$

Then, $\hat{\boldsymbol{\eta}}_{lk}$'s are the low-dimensional vectors that are capable of capturing the difference between pre- and post-shift profile means at a candidate change-point l. If the profile sample is collected from an in-control process, then it can be easily checked that under the assumptions (A.1)-(A.3) in Appendix A, $\hat{\boldsymbol{\eta}}_{lk} \overset{d}{\to} N(\mathbf{0}, \Sigma_k)$. If a change-point occurs at l, then $\hat{\boldsymbol{\eta}}_{lk} \overset{d}{\to} N\left(\int_0^1 \boldsymbol{\Delta}_l(u)v_k(u)du, \Sigma_k \right)$. By a similar approach to that in Srivastava and Worsley (1986), we use a likelihood ratio test to evaluate the hypotheses in (7). To this end, because $\hat{\boldsymbol{\eta}}_{lk}$ and $\hat{\boldsymbol{\eta}}_{lk'}$ are asymptotically uncorrelated, we can use the test statistic

$$Q_m = \max_{1 \le l < m} \sum_{1 \le k \le d} \hat{\boldsymbol{\eta}}_{lk}^\mathsf{T} \hat{\Sigma}_k^{-1} \hat{\boldsymbol{\eta}}_{lk}. \tag{10}$$

A large value of Q_m exceeding a threshold (control limit) L would lead to the rejection of the null hypothesis, implying the occurrence of a change in the process. When the null hypothesis is rejected, the change-point τ can be estimated naturally by

$$\hat{\tau} = \arg \max_{1 \le l < m} \sum_{1 \le k \le d} \hat{\boldsymbol{\eta}}_{lk}^\mathsf{T} \hat{\Sigma}_k^{-1} \hat{\boldsymbol{\eta}}_{lk}. \tag{11}$$

Under certain conditions, we can establish the consistency of the change point estimator $\hat{\tau}$ under

the alternative hypothesis, i.e., $|\hat{\tau} - \tau| = O_p(1)$. See Theorem 1 in Appendix B.

With respect to the choice of d (i.e., the number of eigenfunctions υ_k used for projection), there are several approaches proposed in the literature. One approach, which is adopted in this paper, is to determine d based on the percentage of total variation explained by the extracted PC-scores. Another approach is to use the pseudo Akaike information criterion (AIC) and the cross-validation procedure (cf., Yao et al. 2005) for determining d. A thorough study, including a performance comparison of these methods, deserves much future research.

Remark 2 In traditional multivariate change-point models, in order to remove the effect of mean change on a covariance matrix estimator, the covariance matrix is often estimated from two samples (i.e., the samples before and after the change) separately, and the estimator is updated recursively at each time point (cf., Srivastava and Worsley 1986; Mahmoud et al. 2007). The updates of covariance estimates require a considerable computational effort and the computation time would grow linearly with the sample size m. Consequently, this procedure is computationally infeasible when p and m are large. To avoid this issue, we use the robust estimator of $c(u, u')$ constructed from the pooled sample in this paper, as discussed earlier. From a simulation study (available from the authors), we found that the monitoring performance of our method based on the robust estimator of the covariance matrix is similar to that with the recursively updated estimators.   □

2.5   Determining the Control Limit

In order to determine the control limit L of the proposed monitoring method, it is possible to determine the limiting distribution of the monitoring statistic Q_m under the null hypothesis (see Theorem 2 in the supplementary material). The asymptotic null distribution of Q_m is independent of the nuisance parameters $\boldsymbol{\mu}_1(u) = \boldsymbol{\mu}_2(u)$ and $c(u, u')$; thus, Q_m is asymptotically pivotal. However, in change-point problems, the rate of convergence of the test statistic distribution derived based on the binary segmentation is believed to be slow (see Sec. 1.3 of Csörgö and Horváth 1997 for some related discussions). Consequently, the asymptotic quantiles do not work well with the values of m commonly encountered in application. As an alternative approach, by the results in the following claim, we can use simulation to determine the control limit in such cases.

Claim 1: Assume that $\boldsymbol{\xi}_i$'s are normally distributed and $\upsilon_k(u)$ is known. Then, the distribution

of Q_m with $\upsilon_k(u)$'s substituting for $\hat{\upsilon}_k(u)$'s is identical to that of

$$G_m = \max_{1 \le i < m} \sum_{k=1}^{d} (\bar{\mathbf{z}}_{k,1,i} - \bar{\mathbf{z}}_{k,i+1,m})^{\mathsf{T}} \hat{\Sigma}_{zk}^{-1} (\bar{\mathbf{z}}_{k,1,i} - \bar{\mathbf{z}}_{k,i+1,m}), \tag{12}$$

where $\{\mathbf{z}_{k,i}\}$ is a set of independent standard normal multivariate observations of dimension p, $\bar{\mathbf{z}}_{k,l_1,l_2} = (l_2 - l_1 + 1)^{-1} \sum_{i=l_1}^{l_2} \mathbf{z}_{k,i}$ and $\hat{\Sigma}_{zk} = \frac{1}{2(m-1)} \sum_{i=1}^{m-1} (\mathbf{z}_{k,i+1} - \mathbf{z}_{k,i})(\mathbf{z}_{k,i+1} - \mathbf{z}_{k,i})^{\mathsf{T}}$. Consequently, in order to determine the critical value of our proposed test, denoted as $L_{m,p,d}(\alpha)$, we can first randomly generate m independent standard normal observations and then calculate the values of G_m. Then, the upper α quantile of the empirical distribution of G_m is defined as the control limit with the approximate significance level of α.

Given d and the eigenvectors $v_k(u)$, the simulation-based control limits could be sufficiently accurate regardless of which value of p by Claim 1. The choice of a sufficiently large m depends mainly on the number of d (under the Gaussian assumption) and the variance of the noise ($\boldsymbol{\xi}_{ik}$). In other words, the larger d is, the more functions $\upsilon_k(u)$ we need to estimate and consequently the larger m is needed. For example, as shown by Table 2 in Section 3, the simulation-based control limits described above are sufficiently accurate when $m = 100$ and $d = 4$. On the contrary, if eight eigenvectors/eigenvalues must be used ($d = 8$) then m should be at least equal to 200. Based on our experience, the number of m required linearly increases with d. Simulation results partly reported in the next section indicate that as long as m is not too small so that $\hat{\upsilon}_k(u) \approx \upsilon_k(u)$, this simulation-based method would yield quite accurate and reliable approximations of the control limits. Table 1 gives the $L_{m,p,d}(\alpha)$ values obtained by simulation for various combinations of m, p, and d and for three commonly used significance levels of 10%, 5% and 1%. For the other values of p, one could easily obtain the control limits by using the codes provided in the supplementary material.

It should be pointed out that these simulation-based control limits are obtained under the Gaussian assumption. Many parametric Phase I methods have statistical performance that is quite sensitive to departures from the model assumptions (Jones-Farmer et al. 2014) and, in particular, that the false alarm probability can be in real applications much larger than the nominal if the Gaussian assumption is not fulfilled. Please see Jones-Farmer et al. (2009), Capizzi and Masarotto (2013) and Bell et al. (2014) for more discussions. In practice, we may roughly check the multivariate normality of the $\hat{\boldsymbol{\eta}}_{lk}$ by some statistical tests (e.g., Mardia 1970). The independence between $\hat{\boldsymbol{\eta}}_{lk}$ and $\hat{\boldsymbol{\eta}}_{lk'}$ is more difficult to check, but we may perform a series of Pearson correlation tests for all combinations of k, k' and the channel numbers. In the cases that a parametric assumption may not

be valid, some permutation or bootstrap methods could be used instead (see Qiu et al. 2010 and Capizzi and Masarotto 2013, for some related Phase I research).

2.6 Post-Signal Diagnostics

After detecting a change in the functional mean, it is often useful to identify the set of profile channels whose functional means have changed significantly. Assuming that the estimation of the change point is sufficiently accurate, our objective in this subsection is to determine the channels that are responsible for the change. An implicit but important assumption we make here is that the probability for all process parameters to shift simultaneously is rather small. This is often a reasonable assumption because a fault is usually caused by a hidden source, which is reflected in unknown changes of one or a small set of model parameters (cf., Wang and Jiang 2009, Zou and Qiu 2009, Capizzi and Masarotto 2011).

Let s denote a model in which $|s|$ profile channels have changed after the estimated change point $\hat{\tau}$. Consequently, $s \in \mathcal{S}$ where $\mathcal{S}$ is the model space under consideration with $2^p - 1$ subsets. Note that the empty set that indicates no channels have changed has been excluded from $\mathcal{S}$. After the proposed test triggers an out-of-control alarm, the change-point estimate $\hat{\tau}$ can be obtained simultaneously which divides the profile sample into two parts. Let $\boldsymbol{Z}_1 = \{\boldsymbol{X}_1(u), \ldots, \boldsymbol{X}_{\hat{\tau}}(u)\}$ and $\boldsymbol{Z}_2 = \{\boldsymbol{X}_{\hat{\tau}+1}(u), \ldots, \boldsymbol{X}_m(u)\}$ denote the two sets of independent profiles before and after $\hat{\tau}$, respectively. Then, we can generalize the best-subset searching procedure in Zou et al. (2011) to cases with functional data to determine which profile channels have changed. To this end, the following BIC criterion (Chen and Chen 2008) for functional data is calculated for each model s:

$$\text{BIC}_s = g(s) + |s|d\left(\log\frac{\hat{\tau}(m-\hat{\tau})}{m} + 2\log(pd)\right), \tag{13}$$

where $g(s) = \sum\limits_{1 \le k \le d} \hat{\boldsymbol{\eta}}_{ks}^{\mathsf{T}} \hat{\boldsymbol{\Sigma}}_k^{-1} \hat{\boldsymbol{\eta}}_{ks}$, the jth component of $\hat{\boldsymbol{\eta}}_{ks}$ is defined by

$$\hat{\boldsymbol{\eta}}_{ks}^{(j)} = \begin{cases} \sqrt{\frac{\hat{\tau}(m-\hat{\tau})}{m}} \left. \widehat{+} \right|_0^1 \left\{ \bar{\boldsymbol{X}}_{1,\tau}(u) - \bar{\boldsymbol{X}}_{\tau+1,m}(u) \right\}^{(j)} \hat{\upsilon}_k(u)du, & \text{for } j \notin s \\ 0, & \text{for } j \in s \end{cases},$$

and $\bar{\boldsymbol{X}}_{l_1,l_2}(u) = (l_2 - l_1 + 1)^{-1} \sum_{i=l_1}^{l_2} \boldsymbol{X}_i(u)$. The model $\hat{s}$ that minimizes the criterion in (13), i.e.,

$\hat{s} = \arg\min_s \text{BIC}_s$, is the model that includes the profile channels whose functional means have changed. It should be noted that the number of possible models under which the BIC criterion needs to be calculated increases exponentially with the number of channels p. This makes it hard to optimize the BIC when p is large. In such cases, we suggest using combinatorial optimization methods to find the optimal s. See more discussion about combinatorial optimization methods in Nemhauser and Wolsey (1999). The consistency of the diagnosis procedure described above is given in the supplementary material.

3   Simulation Study

In this section, we evaluate the performance of the proposed method through a simulation study. For this purpose, we first study the empirical false alarm rate of the developed change-point model, and then compare its detection power and change-point estimation performance with those of some alternatives. Finally, the performance of the proposed diagnosis procedure is examined. All the results in this section are obtained from 2,500 replicated simulations.

In order to study the false alarm rate, three different in-control scenarios were considered. In all of the following data generating models, without loss of generality (since Q_m is invariant of $\boldsymbol{\mu}_1(u)$ and $\boldsymbol{\mu}_2(u)$ under the null hypothesis), the in-control mean function $\boldsymbol{\mu}_1(u)$ ($= \boldsymbol{\mu}_2(u)$) was chosen to be 0. We fix the number of channels (profiles) as $p = 4$.

- Model (I)-IC: $\boldsymbol{Y}_i(u) = \sum_{k=1}^{4} \boldsymbol{\xi}_{ik} \upsilon_k(u)$, where $\upsilon_k(u)$'s are the first four non-constant Fourier basis functions with a period of 0.5 and $\boldsymbol{\xi}_{ik}$'s are four-dimensional multivariate normally distributed vectors with mean zero and covariance $(\Sigma_k)_{jh} = k(0.8)^{|j-h|}$;

- Model (II)-IC: $\boldsymbol{Y}_i(u) = \sum_{k=1}^{8} \boldsymbol{\xi}_{ik} \upsilon_k(u)$, where $\upsilon_k(u)$'s are the first eight non-constant Fourier bases with period 0.5 and $\boldsymbol{\xi}_{ik}$'s are four-dimensional multivariate normally distributed vectors with mean zero and covariance $(\Sigma_k)_{jh} = k(0.6)^{|j-h|}$, when $k = 1, \ldots, 4$, and $(\Sigma_k)_{jh} = k(0.4)^{|j-h|}$, when $k = 5, \ldots, 8$;

- Model (III)-IC: $\boldsymbol{Y}_i(u) = \sum_{k=1}^{4} \boldsymbol{\xi}_{ik} \upsilon_k(u)$, where $\upsilon_k(u)$'s are the first four B-spline basis functions of order three with equally spaced knots in $[0, 1]$ and $\boldsymbol{\xi}_{ik}$'s are four-dimensional multivariate normally distributed vectors with mean zero and covariance $(\Sigma_k)_{jh} = k(0.5)^{|j-h|}$.

All the three processes were realized on a grid of 50 equispaced points in $\mathcal{T} = [0, 1]$. The averages of 50 curves generated from these three models are depicted in the left column of Figure 2. In the supplementary material, three individual curves generated from these models are presented in Figure A.1. We have also conducted some other simulations with various values of p, different values of n, and the results show that the general comparison conclusions given below do not change. These additional simulation results are available from the authors upon request.

To illustrate the effectiveness of the proposed approximation method, in Table 2, we tabulate the simulated type I error rates for the three models, various values of m, and the nominal type I error rates (1%, 5% and 10%). In each replication, the number of the eigenfunctions d was chosen automatically such that the explained cumulative percentage variance is 95%. The percentage of mis-specification of the number of eigenfunctions ($d = 4, 8, 4$ for Models (I)-(III), respectively) in our simulation is rather low (usually less than 1%; See Figure 3). It can be seen that the accuracy of the approximation based on (12), increases quickly as m increases. Also, comparison of the calculated error rates from Models (I) and (II) indicates that approximation accuracy decreases as the value of d increases. Moreover, by comparing results under Models (I) and (III), it seems that our method is not affected much by the type of basis functions and the covariance matrices. Overall, $L_{m,p,d}(\alpha)$ provides a good and consistent approximation to the threshold value in most cases.

Now we turn to compare the detection power of the proposed method with some existing procedures in the literature. One natural idea to handle multichannel profiles is to stack up the profile channels and change them into a high-dimensional vector. One could then apply the FPCA to the resulting vector and extract features to construct the change-point model, as discussed earlier. This method is referred to as vectorized-FPCA (VFPCA). There are several issues in using VFPCA for multichannel profiles; see Paynabar et al. (2013) for related discussions. In particular, under our models (1) and (2), this approach breaks the correlation structure in the original data, and potentially loses the useful representations that can be obtained in the original form. Another standard PCA-based method is to use the univariate method of moments as described in Remark 1. That is, univariate FPCA is applied to each individual channel to obtain the corresponding estimates of $\upsilon_k(u)$'s. This method is referred to as individual-FPCA (IFPCA).

In addition, we also consider another two non-PCA methods. The first one takes the average

of the signals across the channels and uses the test statistic

$$\max_{1 \le l < m} \frac{1}{n} \sum_{t=1}^{n} \frac{1}{p} \sum_{j=1}^{p} \Delta_{lj}^2(u_t),$$

where $\Delta_{lj}(u)$ is the jth component of $\boldsymbol{\Delta}(u)$ in (8). The other one monitors each profile separately with a χ^2-type test statistic (Zhang and Albin 2009) and combine all the information across the channels together by

$$\max_{1 \le l < m} \max_{1 \le j \le p} \frac{1}{n} \sum_{t=1}^{n} \Delta_{lj}^2(u_t).$$

We refer these two methods as NoPCA-I and NoPCA-II, respectively. It should be emphasized that because all these benchmarks do not correctly/fully utilize the correlation information, it is difficult to obtain approximate threshold values (control limits) for them. To have a fair comparison, we have performed a size-corrected power comparison in the sense that the actual threshold values were found through simulations so that all the three tests have approximately the same type-I error rate of 0.05. Of course, such a size-correction is only for comparison purposes in the simulation setup, and it is inapplicable in practice.

The number and variety of alternative models are too large to allow a comprehensive, all-encompassing comparison. Our goal is to show the effectiveness, robustness and sensitivity of the proposed method, and thus we only chose certain representative models for illustration. Based on Models (I)-(III), we considered the following three out-of-control models.

- Model (I)-OC: $\boldsymbol{Y}_i(u)$ were generated by Model (I); Set $\boldsymbol{\mu}_1(u) = 0$, $\mu_{22}(u) = c_{\mathrm{I}} \delta \cos(4\pi u)$, $\mu_{23}(u) = c_{\mathrm{I}} \delta \sin(4\pi u)$ for $u \in [1/4, 3/4]$ and $\mu_{21}(u) = \mu_{24}(u) = 0$.

- Model (II)-OC: $\boldsymbol{Y}_i(u)$ were generated by Model (II); Set $\boldsymbol{\mu}_1(u) = 0$, $\mu_{22}(u) = c_{\mathrm{II}} \delta \cos(4\pi u)$, $\mu_{23}(u) = c_{\mathrm{II}} \delta \sin(4\pi u)$ for $u \in [1/4, 3/4]$ and $\mu_{21}(u) = \mu_{24}(u) = 0$.

- Model (III)-OC: $\boldsymbol{Y}_i(u)$ were generated by Model (III); Set $\boldsymbol{\mu}_1(u) = 0$, $\mu_{21}(u) = c_{\mathrm{III}} \delta \exp(-u)$, $\mu_{23}(u) = c_{\mathrm{III}} \delta \sin(4\pi u)$ for $u \in [0, 1]$ and $\mu_{22}(u) = \mu_{24}(u) = 0$.

All the simulations were conducted with two different change-points of $\tau = 25$ and $\tau = 50$ and one sample size of $m = 100$.

From Theorem 1 given in the supplementary material, we know that the power of the proposed test mainly depends on the quantity $\sum_{k=1}^{d} \left(+ \big|_0^1 \boldsymbol{\mu}_2^{\mathsf{T}}(u) \upsilon_k(u) du \Sigma_k^{-1} \left(+ \big|_0^1 \boldsymbol{\mu}_2(u) \upsilon_k(u) du \right.$, which measures the distance between the null and the alternative hypothesis. Therefore, it seems that is not directly comparable across the three models even with the same $\boldsymbol{\mu}_2(u)$. To make magnitudes of the shift parameter δ comparable across Models (I)-(III), we set $c_{\mathrm{I}} = 1$, $c_{\mathrm{II}} = 1.5$ and $c_{\mathrm{III}} = 0.3$, which were found through simulations. The averages of 50 curves generated from these three models ($\delta = 1$) along with the true OC functions are depicted in the right column of Figure 2. We observe that the average curves could roughly capture the pattern of true functions, though it is clearly that 50 curves cannot produce sufficiently accurate estimation.

Figure 3 presents the boxplots of the observed eigenvalues across the 2,500 replicates of the simulation under these three models with $\delta = 2$. With these results, the choice of d is clear and our method that chooses d automatically with the explained cumulative percentage variance would work well.

Figures 4–6 respectively illustrate the detection power curves of MFPCA, VFPCA, IFPCA, NoPCA-I and NoPCA-II against different shift magnitudes (δ) for OC Models (I)-(III). All plots indicate that the proposed MFPCA method has superior efficiency, as expected. Also, the IFPCA outperforms the VFPCA because the latter breaks the structure of the original data by reshaping them into vectors but the former only incurs partial loss of the information. The two non-PCA methods are not as efficient as the PCA-based methods in the considered settings which are in favor of dimension-reduction-based methods. Of course, the poor performance of PCA-based methods can be expected when dimension reduction by PCA is not successful under some other model assumptions such as with-profiles correlations are ignorable. Moreover, we consider the cases that measurement errors exist, say $\boldsymbol{Y}_i(u_t) = \tilde{\boldsymbol{Y}}_i(u_t) + \sigma \boldsymbol{\epsilon}_i(u_t)$, where $\tilde{\boldsymbol{Y}}_i(u_t)$ is generated through OC Models (I)-(III) and $\boldsymbol{\epsilon}_i(u)$'s are i.i.d p-dimensional standard normal random vectors. The results with $\sigma = 0.25$ are given in Figures A.2-A.4 in the supplementary material. All the other settings are the same as those in Figures 4-6 in the revision. From those figures, we can see that the advantage of the proposed MFPCA-based method is still clear in all the cases.

Next, we evaluate the performance of the proposed approach in estimating the change-point after the null hypothesis is rejected. Again, the IFPCA and VFPCA methods were included as the benchmarks and OC Models (I)-(III) were considered in simulations. To compare these methods, various accuracy and precision criteria including the bias and standard deviation (Sd) of the

change-point estimates, the relative square root of MSE of estimates between benchmarks and our proposed method (RSM), and $\Pr(|\hat{\tau}-\tau| \leq 1)$ and $\Pr(|\hat{\tau}-\tau| \leq 3)$ (denoted as P_1 and P_3, respectively) were reported in Table 3.

As can be seen from Table 3, for a relatively small shift $\delta = 1$, all three estimators appear to be considerably biased in estimating the true change-point in the case of $\tau = 25$. However, our proposed estimator $\hat{\tau}$ outperforms the other two estimators in terms of the five accuracy and precision criteria reported in the table in most such cases. Other simulation results in the table for various other parameter combinations also demonstrate that $\hat{\tau}$ appears to be the best in estimating τ.

Finally, to evaluate the effectiveness of the proposed diagnostic procedure, the probabilities of detecting the correct set of changed profile channels, $\Pr(\hat{s} = s_T)$, were calculated under OC Models (I)-(III) with $m = 100$, where s_T represents the true fault isolation model that contains all the indices of the variables that have changed their means at τ. These values, reported in Table 4, indicate that the estimate $\hat{s}$ is fairly accurate and the diagnosis accuracy quickly increases as the shift magnitude gets larger.

4 Case Study: Phase-I Monitoring and Diagnosis of Multichannel Tonnage Profiles

In this section, the proposed methodology is used for phase-I monitoring of multichannel tonnage profile data collected in a multi-operation forging process. In this process, a forging machine (shown in Figure 1a) is comprised of multiple dies, each assigned to perform one operation during a stroke. Figure 7a shows intermediate workpieces after each operation of the forging process. In order to monitor the performance of the process, four strain gauge sensors, each mounted on one column of the forging machine, measure the exerted tonnage force of the press uprights. This results in a four-channel tonnage profile in each cycle of operation, in which the length of each profile channel is 1200 (See Figure 1b). A sample of 496 multichannel profiles was collected under different experimental settings. This sample includes 151 in-control profiles collected under the normal production condition, and 5 groups of 69 out-of-control profiles. Each group of out-of-

control (faulty) profiles corresponds to a faulty operation condition in which one part is missing in an operation station shown in Figure 7a. For example, the Fault-4 group contains profiles measured when a part is missing in Station 4 (i.e., piercing). Each faulty condition may have different effect on the individual profile channels, depending on the proximity of the station to the sensors measuring profiles. For the illustration purpose, the average profiles of these six groups are shown in Figure 7b.

The tonnage profile dataset was used in previous research by Lei et al. (2010) and Paynabar et al. (2013). However, as discussed earlier, both of these studies used a subset of labeled profiles to train, and another subset to test their models. Differently, in this case study, we focus on constructing a phase-I monitoring method and thus, we do not use the information of profile classes (i.e., normal, fault 1, etc.) in estimating our model. Using this dataset, we define five subsets, each of which includes the in-control profiles followed by one of the five out-of-control profile groups. Thus, each subset includes 151 in-control and 69 out-of-control profiles. To reduce the measurement noise of profile channels, we first use the non-overlapping moving average function with the window size of 6, which reduces the length of each profile to 200. Then, we apply the MFPCA to each subset and calculate the first 15 eigen-functions whose corresponding eigenvalues account for more than 85% of the profiles' total variation. The reason for choosing 15 eigen-functions is that the absolute differences of consecutive eigenvalues are infinitesimal after the 15th eigenvalue. Next, the eigen-functions are used to construct the change-point model for phase-I monitoring. The control limit for type-I error rate of 0.05 is computed to be 203.32 by the procedure discussed in Section 2.4. It should be noted that the control limit may slightly be underestimated as the number of profiles in each subset ($m = 220$) is not large enough for $d = 15$, which requires m to be around 400.

Our analysis shows that in all subsets, the proposed MFPCA-based change-point model is able to detect the change and accurately estimate the time of change (i.e.,$\hat{\tau} = 151$). Furthermore, using Subset 4 (normal data followed by Fault 4 data), we compare our method with two benchmarks commonly used for analyzing the dataset. The reason for the use of Subset 4 is that as can be seen in Figure 7b, profile samples corresponding to Fault 4 and the normal operation are very similar, thus difficult to separate. In Method 1, FPCA is applied on aggregated profiles averaged over 4 channels, then the eigen-functions are utilized for process monitoring using the discussed change-point model. In Method 2, each profile channel is monitored separately. FPCA is applied on individual profile channels (IFPCA) and, then four change-point models are developed based

on the estimated eigen-functions of the signal channels. The resulting phase I control charts are plotted in Figure 8. As can be seen from the figure, MFPCA correctly detects the change and the monitoring statistic is maximized at $\tau = 151$, indicating the time when the process change occurred. The aggregated method cannot detect the process change. This is because in-control channels can mask the change in out-of-control channels when they are averaged. Although, the IFPCA for channels 3 and 4 indicate an out-of-control condition, the estimated time of change ($\hat{\tau} = 89$ and 36, respectively) by neither of charts is correct. We also perform a similar analysis for other subsets. Our analysis results indicate similar conclusions for Subset 5. For other subsets, however, all methods can correctly detect the change as the difference between the normal and faulty profiles in those subsets is sufficiently large (See Figure 7b).

Furthermore, we use the Mardia test (Mardia 1970) to check the normality assumption of $\hat{\boldsymbol{\eta}}_{lk}$; $k = 1, 2, ..., 15$. After correcting the significance level of multiple tests using the Bonferroni correction, the resulting p-values for the Mardia kurtosis and skewness tests indicate no deviation from the normality assumption except for one case in which the skewness test is rejected. Also, in order to check the independence assumption between $\hat{\boldsymbol{\eta}}_{lk}$ and $\hat{\boldsymbol{\eta}}_{lk'}$, we perform a series of Pearson correlation test for all combinations of k, k' and the channel numbers. The test results, after correcting the significance level, indicate that in only 8% of the combinations the independence assumption is rejected. In short, the normality and independence assumptions of $\hat{\boldsymbol{\eta}}_{lk}$ seem to be reasonable for the tonnage data.

After detecting the change, the fault diagnosis procedure discussed in Section 2.6 was performed to identify out-of-control profile channels. Figure A.5 in the supplementary material reports the BIC values of different channel sets for the fault group 4. As can be seen from the figure, two sets {4} and {1, 4} have minimum BIC values (around 265), meaning that channels 1 and 4 contributed to the out-of-control alarm. From Figure 1(a), we can see that both sensors 1 and 4 are mounted on the front side of the forging machine where the die for station 4 is located. This is the reason that the effect of missing parts in station 4 on the tonnage force is mostly picked up by sensors 1 and 4. Additionally, since station 4 is closer to sensor 4, its corresponding BIC value is smaller than that of sensor 1.

5   Concluding Remarks and Future Work

Phase-I analysis of multichannel (multiple) profile monitoring is a challenging problem and it has not been thoroughly investigated in the literature. In this paper, we proposed an approach that combines the classical binary segmentation procedure with the MFPCA. The proposed framework is developed under the assumption that different profile channels have similar structure, i.e., (2). This assumption allows us to gain strength by borrowing information from all channels. Our proposed approach effectively incorporates the correlation information of multichannel profiles for testing their stability and estimating the change-point. The simulation results on empirical type-I error probabilities indicated that the approximated threshold is reasonably accurate especially for large sample sizes. Through simulation, we also showed that the proposed monitoring scheme based on MFPCA outperforms conventional approaches in terms of the change-detection power and the estimation accuracy of the change-point.

Checking stability with no certainty on the underlying process distribution is quite difficult in Phase I, especially in the multivariate framework and in presence of complex and autocorrelated data (Capizzi and Masarotto 2013). An extension of the proposed MFPCA-based method to the non-normal or autocorrelated case could be a worthwhile and necessary future contribution to the existing multivariate Phase I methods (Capizzi and Masarotto 2008; Bell et al. 2014).

The model (2) assumes that all the variation in the observed curves $X_i(u)$ is functional variation, and that "white noise" is negligible. The presence of non-negligible white noise would adversely affect the model estimation and may consequently result in certain size distortion for the proposed test. It should also be pointed out our proposal handles only the case of a "single" and "sustained" change-point in the process mean. Many historical samples may actually contain more than one change-point in practice. Patterned, oscillatory mean changes (but also in the variance) can also easily happen and thus it needs to investigate in Phase I the presence not only of multiple change-points but also of several other types of changes. Extension of the proposed scheme to multiple change-points detection (c.f., Hawkins 2001) for profile monitoring warrants further research. The proposed modelling method and the MFPCA techniques can be readily extended for outlier detection (Yu et al. 2012; Zou et al. 2014) and for phase-II profile monitoring as well.

Acknowledgement

The authors thank the editor, associate editor, and two anonymous referees for their many helpful comments that have resulted in significant improvements in the article.

Supplementary Materials

Technical Details and Codes: This file provides all the asymptotic results mentioned in the paper and their proofs and the Fortran code used in the simulation study. The tonnage data and R codes for its analysis are also included.

References

Bell, R. C., Jones-Farmer, L. A., and Billor, N. (2014), "A Distribution-Free Multivariate Phase I Location Control Chart for Subgrouped Data from Elliptical Distributions," *Technometrics*, in press.

Berkes, I., Gabrys, R., Horváth, L. and Kokoszka, P. (2009), "Detecting Changes in the Mean of Functional Observations," *Journal of the Royal Statistical Society: Series B*, 71, 927–946.

Capizzi, G., and Masarotto, G. (2008), "Practical Design of Generalized Likelihood Ratio Control Charts for Autocorrelated Data," *Technometrics*, 50, 357–370.

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

Capizzi, G., and Masarotto, G. (2013), "Phase I Distribution-Free Analysis of Univariate Data," *Journal of Quality Technology*, 45, 273–284.

Chen, J., and Chen, Z. (2008), "Extended Bayesian Information Criterion for Model Selection with Large Model Spaces," *Biometrika*, 95, 759–771.

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

Csörgö, M. & Horváth, L. (1997). *Limit theorems in change-point analysis*, John Wiley, New York.

Di, C. Z., Crainiceanu, C. M., Caffo, B. S., and Punjabi, N. M. (2009), "Multilevel Functional Principal Component Analysis", *Annals of Applied Statistics*, 3, 458–488.

Dubin, J. A., and Müller, H.-G, (2005), "Dynamical Correlation for Multivariate Longitudinal Data", *Journal of the American Statistical Association*, 100, 872–881.

Hawkins, D. M. (2001), "Fitting Multiple Change-Point Models to Data," *Computational Statistics and Data Analysis*, 37, 323–341.

Hawkins, D. M., Qiu, P., and Kang, C.W. (2003), "The changepoint model for statistical process control," *Journal of Quality Technology*, **35**, 355–366.

Jensen, W. A., Birch, J. B., and Woodall, W. H. (2008), "Monitoring Correlation Within Linear Profiles Using Mixed Models," *Journal of Quality Technology*, 40, 167–183.

Jin, J., and Shi, J. (1999), "Feature-Preserving Data Compression of Stamping Tonnage Information Using Wavelet," *Technometrics*, 41, 327–339.

Jones-Farmer, L. A., Jordan, V., and Champ, C. W. (2009), "Distribution-Free Phase I Control Charts for Subgroup Location," *Journal of Quality Technology*, 41, 304–316.

Jones-Farmer, L. A., Woodall, W. H., Steiner, S. H., and Champ, C. W. (2014). "An Overview of Phase I Analysis for Process Improvement and Monitoring," *Journal of Quality Technology*, 46, 265–280.

Kang, L., and Albin, S. L. (2000), "On-Line Monitoring When the Process Yields a Linear Profile," *Journal of Quality Technology*, 32, 418–426.

Lei, Y., Zhang Z. and Jin, J. (2010), "Automatic Tonnage Monitoring For Missing Part Detection in Multi-Operation Forging Processes," *ASME Transactions, Journal of Manufacturing Science and Engineering*, 132, 051010.1-10.

Mahmoud, M. A., Parker, P. A., Woodall, W. H., and Hawkins, D. M. (2007), "A Change Point Method for Linear Profile Data," *Quality and Reliability Engineering International*, 23, 247–268.

Mahmoud, M. A., and Woodall, W. H. (2004), "Phase I Analysis of Linear Profiles with Calibration Applications," *Technometrics*, 46, 380–391.

Mardia, K. V. (1970), "Measures of Multivariate Skewness and Kurtosis With Applications," *Biometrika*, 57, 519–530.

Nemhauser, G. L., and Wolsey, L. A. (1999), *Integer and Combinatorial Optimization*, John Wiley & Sons, New York.

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

Paynabar, K., and Jin, J. (2011), "Characterization of Non-Linear Profiles Variations Using Mixed-Effect Models and Wavelets," *IIE Transactions*, 43, 275–290.

Paynabar, K., Jin, J., Agapiou, J., and Deeds, P. (2012), "Robust Leak Tests for Transmission Systems Using Nonlinear Mixed-Effect Models," *Journal of Quality Technology*, 44, 265–278.

Paynabar, K., Jin, J., and Pacella, M. (2013), "Analysis of Multichannel Nonlinear Profiles Using Uncorrelated Multilinear Principal Component Analysis with Applications in Fault Detection and Diagnosis," *IIE Transactions*, 45, 1235–1247.

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

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

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

Ramsay, J. O., and Silverman, B. W. (2005), *Functional Data Analysis*, Springer, New York.

Srivastava, M. S., and Worsley, K. J. (1986), "Likelihood Ratio Tests for a Change in the Multivariate Normal Mean," *Journal of the American Statistical Association*, 81, 199–204.

Wang, K., and Jiang, W. (2009), "High-Dimensional Process Monitoring and Fault Isolation via Variable Selection," *Journal of Quality Technology*, 41, 247–258.

Williams, J. D., Birch, J. B., Woodall. W. H., and Ferry, M. (2007), "Statistical Monitoring of Heteroscedastic Dose-Response Profiles from High-Throughput Screening," *Journal of Agricultural, Biological, and Environmental Statistics*, 12, 216–235.

Woodall, W. H. (2007), "Current Research on Profile Monitoring," *Producão*, 17, 420–425.

Yao, F., Müller, H.-G. and Wang, J.-L. (2005), "Functional Linear Regression Analysis for Longitudinal Data," *The Annals of Statistics*, 33, 2873–2903.

Yu, G., Zou, C., and Wang, Z, (2012), "Outlier Detection in the Functional Observations with Applications to Profile Monitoring," *Technometrics*, 54, 308–318.

Zhang, H., and Albin, S. (2009), "Detecting Outliers in Complex Profiles Using a χ^2 Control Chart Method," *IIE Transactions*, 41, 335–345.

Zou, C. Jiang, W., and Tsung, F. (2011), "A LASSO-Based Diagnostic Framework for Multivariate Statistical Process Control", *Technometrics*, 53, 297–309.

Zou, C., and Qiu, P. (2009), "Multivariate Statistical Process Control Using LASSO," *Journal of the American Statistical Association*, 104, 1586–1596.

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

Zou, C., Tseng, S-T., and Wang, Z. (2014), "Outlier Detection in General Profiles Using Penalized Regression Method" *IIE Transactions*, 46, 106–117.

Zou, C., Tsung, F., and Wang, Z. (2007), "Monitoring General Linear Profiles Using Multivariate EWMA Schemes," *Technometrics*, 49, 395–408.

Zou, C., Zhang, Y., and Wang, Z. (2006), "Control Chart Based on Change-Point Model for Monitoring Linear Profiles," *IIE Transactions*, 38, 1093–1103.

Table 1: Threshold values of Q_m for various m, p, d and α obtained based on (12) under the Gaussian assumption.

m	p	$d = 1$			$d = 2$			$d = 3$			$d = 4$		
		1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
50	2	19.0	13.8	11.5	24.2	18.9	16.6	29.1	23.1	20.7	33.8	27.3	24.4
	3	22.5	17.2	14.7	31.0	24.6	21.6	37.8	31.1	27.7	43.0	36.3	33.0
	4	27.1	20.6	17.9	37.6	29.5	26.4	45.1	37.6	34.1	53.6	45.4	41.6
	5	32.2	24.4	21.3	44.0	35.4	32.0	55.4	45.9	41.8	63.9	54.4	50.1
100	2	18.0	13.4	11.4	24.0	18.6	16.4	27.8	22.6	20.4	31.8	26.5	24.1
	3	21.7	16.4	14.4	28.5	23.1	20.6	34.5	28.9	26.3	40.5	34.3	31.5
	4	24.8	19.5	17.2	33.8	27.6	24.9	41.0	35.1	32.4	49.6	42.3	38.8
	5	27.9	22.4	19.6	38.3	32.3	29.2	48.2	40.8	37.6	57.2	49.9	46.2
200	2	17.3	13.3	11.5	22.9	18.3	16.2	27.3	22.3	20.2	31.3	26.4	24.1
	3	20.3	16.1	14.1	27.7	23.0	20.5	33.5	28.3	25.8	39.6	33.9	31.4
	4	23.0	18.6	16.5	32.4	27.0	24.8	39.6	34.1	31.5	46.6	40.8	37.9
	5	25.6	21.1	19.0	36.6	31.0	28.6	45.7	39.6	36.6	53.9	47.4	44.4
400	2	17.4	13.6	11.8	22.4	18.5	16.5	27.1	22.4	20.4	31.5	26.6	24.3
	3	20.3	16.4	14.5	27.4	22.9	20.8	33.4	28.3	26.1	38.6	33.4	31.1
	4	22.6	18.8	16.7	31.1	26.6	24.4	38.9	33.8	31.2	45.8	40.2	37.5
	5	25.4	20.8	18.9	36.2	30.4	28.1	44.2	38.6	36.1	52.8	46.9	44.0

Table 2: Empirical sizes (%) of the test using the threshold values given in Table 1

m	Model (I)			Model (II)			Model (III)		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
50	0.019	0.071	0.130	0.025	0.096	0.171	0.016	0.066	0.115
100	0.016	0.064	0.115	0.024	0.075	0.134	0.011	0.051	0.101
200	0.012	0.054	0.107	0.015	0.072	0.127	0.009	0.053	0.114
400	0.010	0.051	0.099	0.011	0.054	0.110	0.010	0.050	0.106

Table 3: Comparison of the three change-point methods in terms of location estimation under OC Models (I)-(III) with $m = 100$. Sd: the the standard deviations of estimators; RSM: the relative square root of MSE between the other two estimators and our proposed method. P_1: $\Pr(|\hat{\tau} - \tau| \leq 1)$; P_3: $\Pr(|\hat{\tau} - \tau| \leq 3)$.

Model	δ	Method	$\tau = 25$					$\tau = 50$				
			Bias	Sd	RSM	P_1	P_3	Bias	Sd	RSM	P_1	P_3
		IFPCA	3.62	16.2	1.49	0.361	0.550	0.54	12.5	1.88	0.404	0.617
	1.0	VFPCA	18.8	33.7	3.45	0.058	0.113	0.83	31.9	4.79	0.062	0.117
		MFPCA	2.00	10.9	–	0.480	0.686	0.03	6.65	–	0.529	0.741
(I)												
		IFPCA	0.04	1.41	1.60	0.864	0.974	0.02	1.27	1.60	0.879	0.972
	2.0	VFPCA	2.65	14.0	16.2	0.564	0.741	0.16	4.12	5.18	0.832	0.924
		MFPCA	0.02	0.88	–	0.914	0.994	0.01	0.80	–	0.933	0.995
		IFPCA	7.38	22.1	1.40	0.242	0.386	0.31	15.2	1.54	0.313	0.494
	1.0	VFPCA	21.9	34.5	2.45	0.047	0.097	0.78	31.6	3.20	0.053	0.112
		MFPCA	3.61	16.2	–	0.381	0.559	0.29	9.86	–	0.467	0.658
(II)												
		IFPCA	0.05	1.61	1.58	0.857	0.965	0.03	1.46	1.73	0.853	0.965
	2.0	VFPCA	5.22	18.6	18.9	0.372	0.541	0.24	9.58	11.4	0.516	0.716
		MFPCA	0.02	1.02	–	0.909	0.981	0.01	0.84	–	0.935	0.991
		IFPCA	11.7	26.4	1.24	0.167	0.289	0.08	21.0	1.09	0.202	0.329
	1.0	VFPCA	17.2	31.9	1.95	0.087	0.148	0.02	29.2	2.09	0.106	0.183
		MFPCA	9.29	24.2	–	0.200	0.336	0.59	20.1	–	0.233	0.367
(III)												
		IFPCA	0.18	3.26	1.17	0.678	0.872	0.06	2.61	1.22	0.686	0.888
	2.0	VFPCA	2.12	11.7	15.7	0.506	0.697	0.30	6.45	7.48	0.557	0.765
		MFPCA	0.17	3.00	–	0.700	0.888	0.00	2.36	–	0.716	0.893

Table 4: Diagnosis accuracy of the proposed procedure, $\Pr(\hat{s} = s_T)$, under OC Models (I)-(III) with $m = 100$, where s_T represents the true fault isolation model that contains all the indices of the variables that have changed their means at τ and $\hat{s}$ is the estimated one.

δ	Model (I)		Model (II)		Model (III)	
	$\tau = 25$	$\tau = 50$	$\tau = 25$	$\tau = 50$	$\tau = 25$	$\tau = 50$
1	0.56	0.91	0.52	0.89	0.55	0.90
2	0.78	0.98	0.75	0.96	0.76	0.99
3	0.98	1	0.96	1	0.98	1
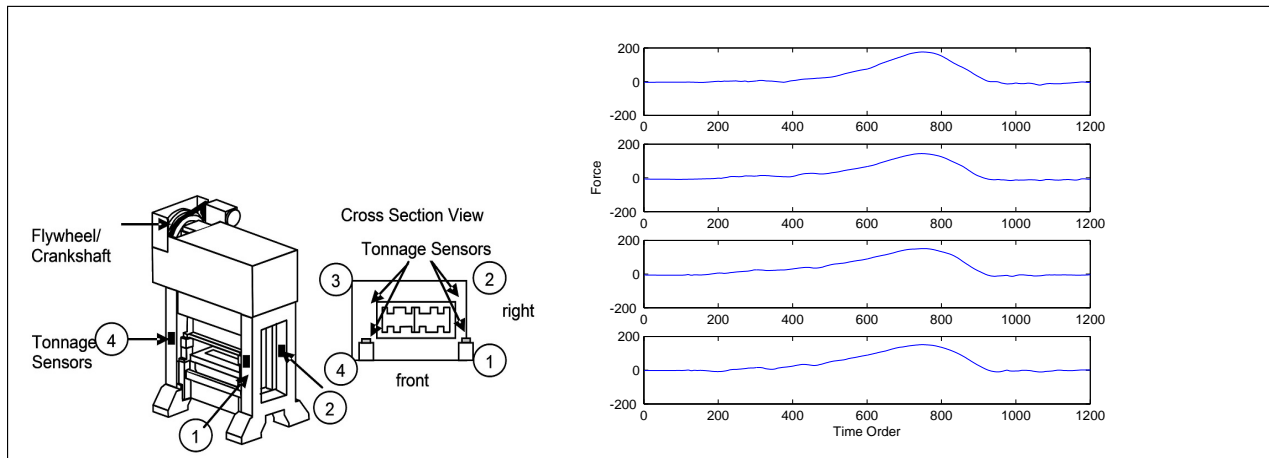
Figure 1: (a)(left panel): A forging machine with four strain gages. (b)(right panel): A sample of four-channel profiles.
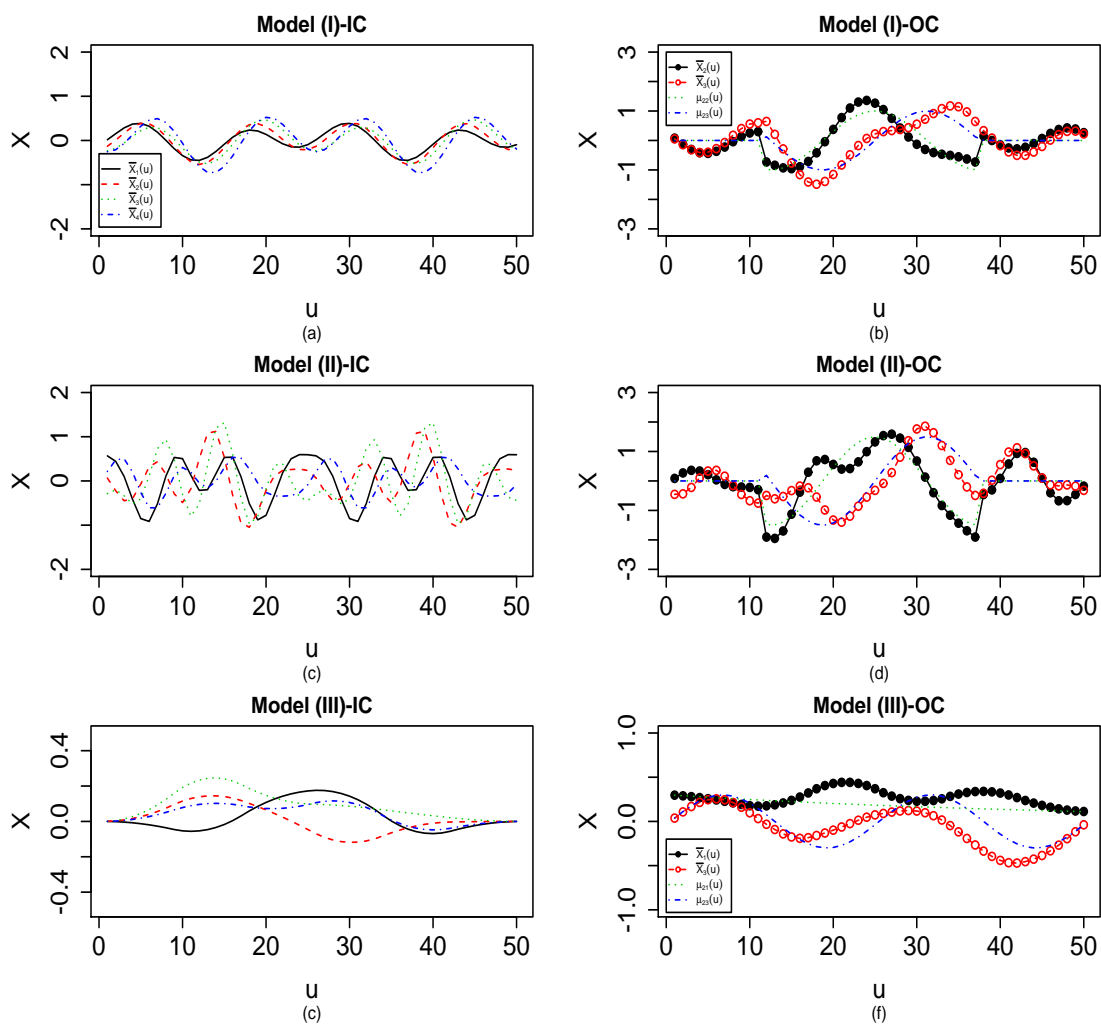
Figure 2: Left column: The averages of 50 curves of all the four channels generated from IC Models (I)-(III); Right column: The averages of 50 curves of the changed channels generated from OC Models (I)-(III) along with the true OC functions when $\delta = 1$.
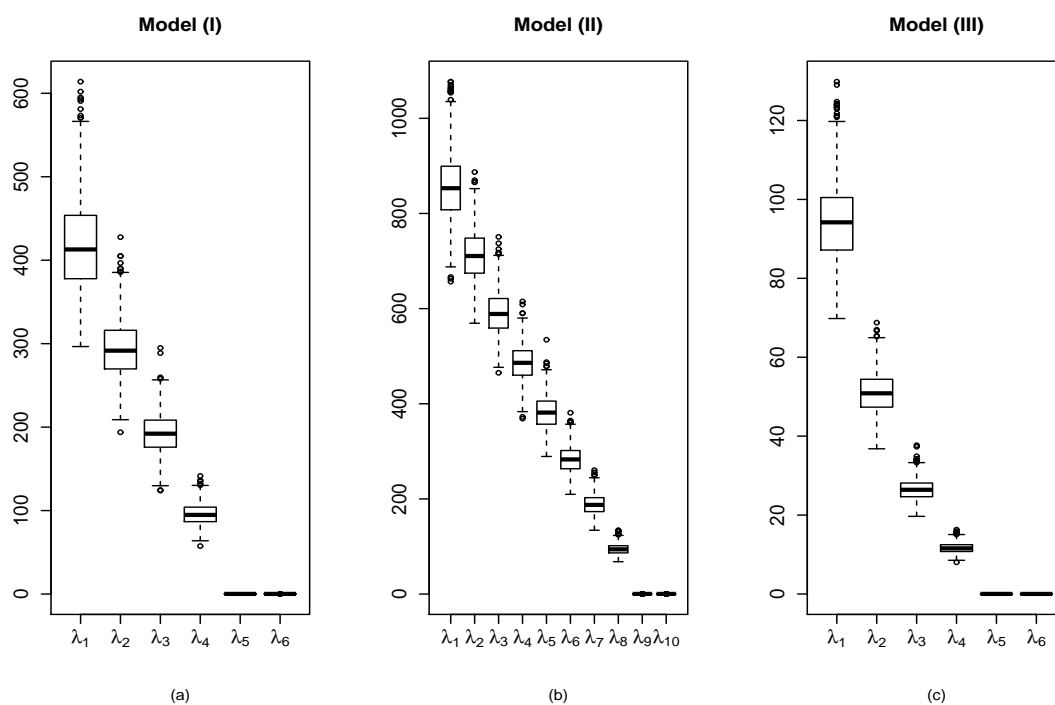
Figure 3: Boxplots of the observed eigenvalues across the 2,500 replicates of the simulation under OC Models (I)-(III) with $\delta = 2$ and $\tau = 50$
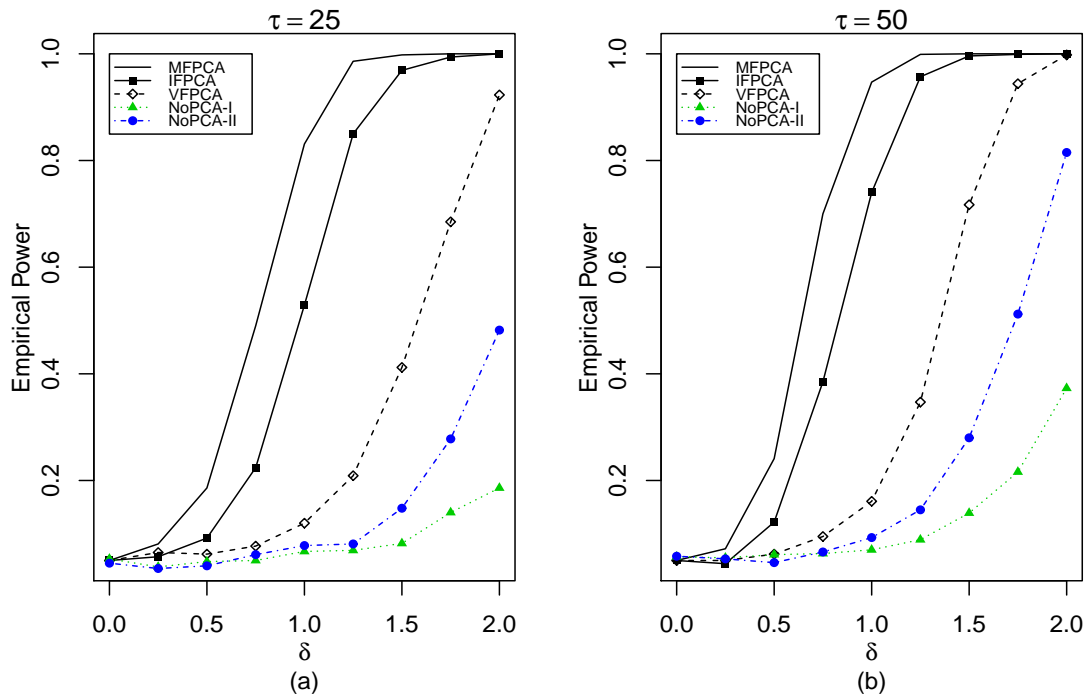
Figure 4: Size-corrected power comparison among the procedures MFPCA, VFPCA, IFPCA, NoPCA-I and NoPCA-II under Model (I)-OC: (a) $\tau = 25$; (b) $\tau = 50$
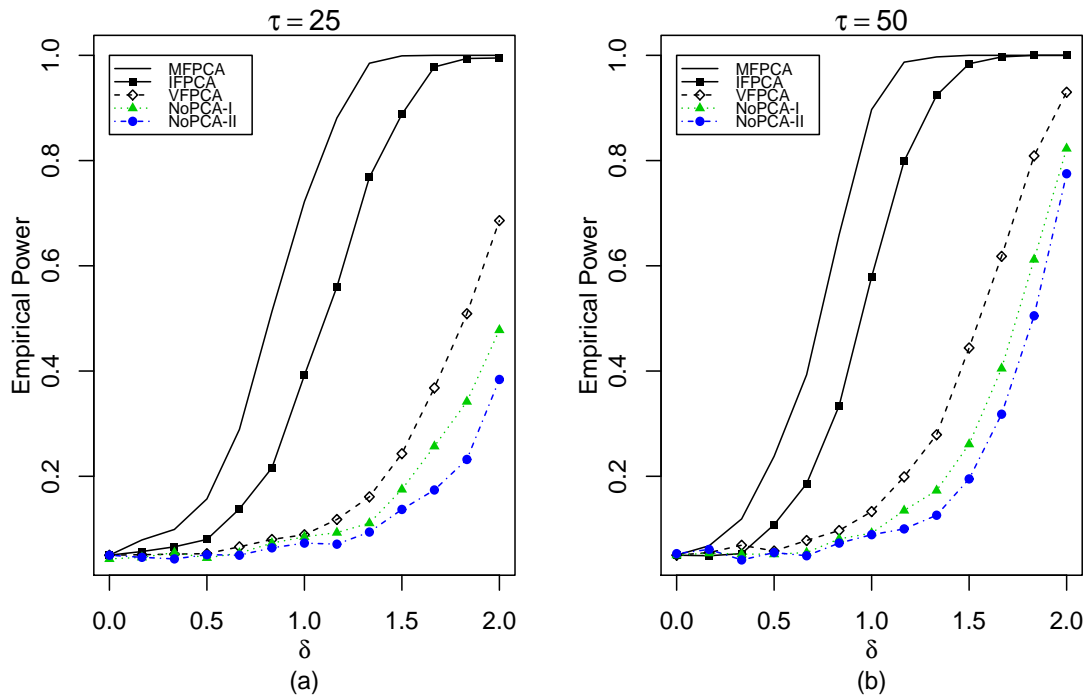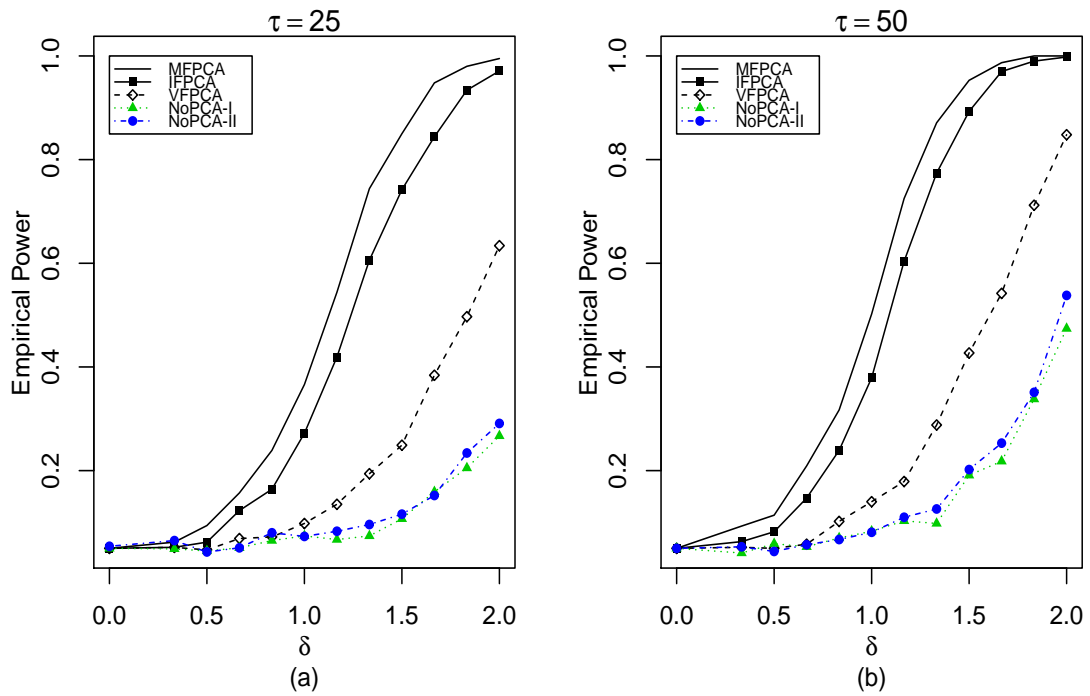
Figure 5: Size-corrected power comparison among the procedures MFPCA, VFPCA, IFPCA, NoPCA-I and NoPCA-II under Model (II)-OC: (a) $\tau = 25$; (b) $\tau = 50$

Figure 6: Size-corrected power comparison among the procedures MFPCA, VFPCA, IFPCA, NoPCA-I and NoPCA-II under Model (III)-OC: (a) $\tau = 25$; (b) $\tau = 50$
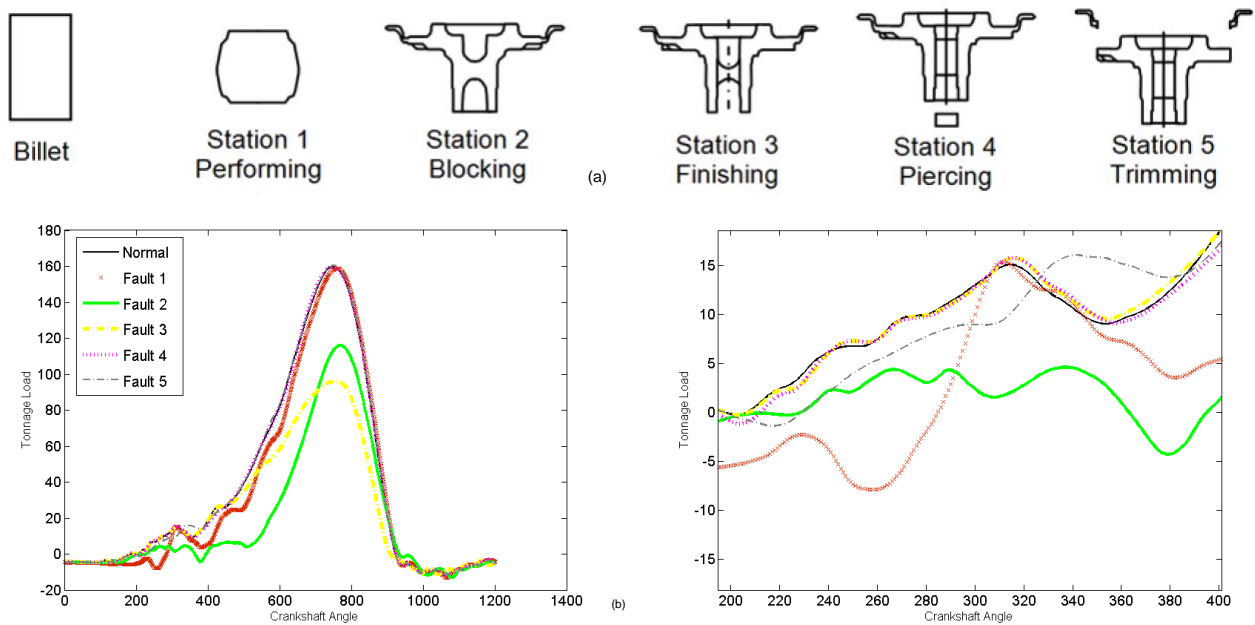
Figure 7: Top panel (a): shape of workpieces at each operation. Bottom left panel (b): Average profiles of aggregated tonnage profiles for normal and faulty operations. Bottom right panel (b): Zoomed-in average profiles for angles 200 to 400.
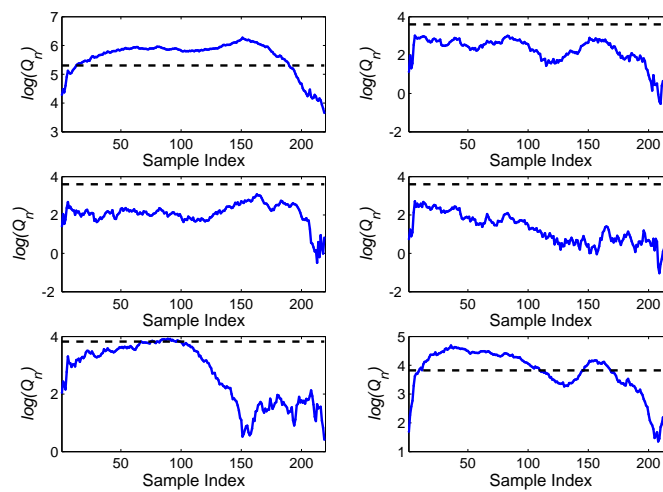
Figure 8: Phase-I control chart for tonnage profiles including 151 in-control profiles and 69 out-of-control profiles of Fault 4. True change-point is $\tau = 151$.