Received: 1 August 2018 | Revised: 17 January 2019 | Accepted: 24 March 2019

DOI: 10.1002/qre.2482

RESEARCH ARTICLE

WILEY

Detecting the process changes for multivariate nonlinear profile data

Jeh-Nan Pan | Chung-I Li | Meng Zhe Lu

Department of Statistics, National Cheng Kung University, Tainan, Taiwan, ROC

Correspondence
Chung-I Li, Department of Statistics, National Cheng Kung University, Tainan 70101, Taiwan, ROC.
Email: cili@mail.ncku.edu.tw

Funding information
Ministry of Science and Technology of Taiwan, ROC, Grant/Award Number: MOST 107-2410-H-006-035

Abstract

In profile monitoring for a multivariate manufacturing process, the functional relationship of the multivariate profiles rarely occurs in linear form, and the real data usually do not follow a multivariate normal distribution. Thus, in this paper, the functional relationship of multivariate nonlinear profile data is described via a nonparametric regression model. We first fit the multivariate nonlinear profile data and obtain the reference profiles through support vector regression (SVR) model. The differences between the observed multivariate nonlinear profiles and the reference profiles are used to calculate the vector of metrics. Then, a nonparametric revised spatial rank exponential weighted moving average (RSREWMA) control chart is proposed in the phase II monitoring.

Moreover, a simulation study is conducted to evaluate the detecting performance of our proposed nonparametric RSREWMA control chart under various process shifts using out-of-control average run length (ARL_1). The simulation results indicate that the SREWMA control chart coupled with the metric of mean absolute deviation (MAD) can be used to monitor the multivariate nonlinear profile data when a common fixed design (CFD) is not applicable in the phase II study. Finally, a realistic multivariate nonlinear profile example is used to demonstrate the usefulness of our proposed RSREWMA control chart and its monitoring schemes.

KEYWORDS

common fixed design (CFD), mean absolute deviation (MAD), multivariate nonlinear profile data, spatial rank exponential weighted moving average (SREWMA) control chart, support vector regression (SVR)

1 | INTRODUCTION

The statistical control charts play an important role in monitoring and improving the product and process quality. Generally speaking, the use of control chart involves two phases. In phase I study, the parameters of the process are estimated based on a set of historical data and used to establish control limits for phase II monitoring. In phase II study, the data are sequentially collected over time to assess whether the parameters of the process have changed from the estimated values in phase I study. If the process triggers an alarm in phase II study, then the quality practitioners need to immediately find out and remove the assignable causes, so the process stability can be closely monitored and maintained.

With the advent of modern technology, manufacturing processes have become rather sophisticated. Two or more correlated quality characteristics are often required for monitoring the product and process quality in today's high-technology industries, such as semiconductor manufacturing, printed circuit board, aerospace and telecommunications industries, etc. If the multiple quality characteristics are assumed to be represented by a functional relationship of one or more explanatory variables, then the data generated from such a functional relationship is called multivariate profile data. The illustration of the data collection process for multivariate nonlinear profile data with four quality characteristics and *m* profiles is shown in Figure 1.

The purpose of profile monitoring is to detect whether the functional relationship of a multivariate manufacturing process has been changed in phase II study. Zou et al[1] discussed that the parametric methods used for monitoring profiles are inappropriate because the small sample distribution of parameter estimators in nonlinear regression is unobtainable. Because of the complexity of parameter estimation involved in nonlinear profile monitoring, Williams et al[2] proposed a nonparametric monitoring method and used five metrics to measure deviations from a reference profile. Zou et al[3] considered the case that the number of profiles is not large in phase I and, in fact, the number is considerably small in many practical applications. Steiner et al[4] considered an arbitrary design, where the data points collected are not necessarily the same for all the profiles. These situations were mentioned as "the situations that don't have a common fixed design (CFD) for the data set."

Generally speaking, the functional relationship of multivariate nonlinear profile data cannot be known in advance and the real data usually do not follow a multivariate normal distribution. Hence, the development of a new nonparametric monitoring method for detecting the process changes of multivariate nonlinear profile data, which does not satisfy CFD, becomes necessary. In this paper, we focus our research on phase II study to timely detect the process changes in various parameters for multivariate nonlinear profiles. However, in phase II monitoring, the reference profile is assumed to be known and needed to be determined prior to the implementation of the proposed control chart. Thus, phase I reference profile construction is discussed in section 3.2, and the training parameters for phase I simulation study are discussed in section 4.2.2.

2 | LITERATURE REVIEW

In the past decade, many researchers (such as Gupta et al[5]; Kang and Albin[6]; Mahmoud and Woodall[7]; Zou et al[1]) have proposed different methods for monitoring the profile data. Woodall et al[8] and Woodall[9] summarized a literature review of the profile data and pointed out the shortcomings of the existing methods. Generally speaking, there are two profile monitoring methods, ie, parametric or nonparametric method. On the other hand, based on the data types, statistical methods for monitoring product and process quality profiles are commonly classified into linear profile or nonlinear profile monitoring.

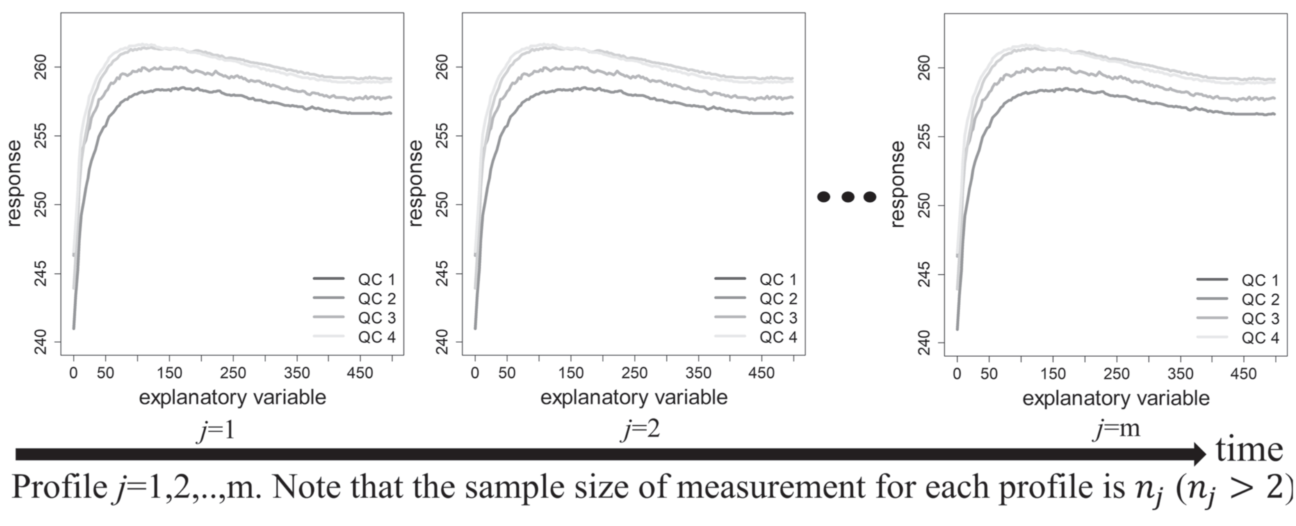

FIGURE 1 Multivariate nonlinear profile data collection process

2.1 | Linear profile

Kang and Albin[6] proposed two different approaches to monitor linear profile data in phase II study. They used the least square method to estimate the two parameters, ie, the intercept (β_0) and the slope (β_1) in their linear regression model. Their first approach was to apply T^2 multivariate control chart to monitor the intercept (β_0) and the slope (β_1). The second approach was to use EWMA/R control chart for monitoring the mean and the variance of the residuals from the samples. Kim et al[10] transformed the X values by setting up a linear regression to monitor the linear profile data. Then, they used three independent univariate EWMA control charts to monitor the intercept, the slope, and the variance of errors, respectively. The results showed that their monitoring method is better than EWMA/R control chart proposed by Kang and Albin.[6] Zou et al[11] proposed a novel multivariate exponentially weighted moving average scheme for monitoring general linear profiles. Wu et al[12] used support vector regression (SVR) to detect and classify the shifts in linear profile data. Soleimani et al[13] considered a simple linear profile and assumed that a first-order autoregressive model existed between observations in each profile. They showed that their control chart with a weaker autocorrelation has a better detecting performance than that with a stronger autocorrelation. In practical applications, the quality of industrial products or processes is often represented by two or more correlated quality characteristics. Noorossana et al[14] proposed the use of three control chart schemes for phase II monitoring of multivariate simple linear profiles. The first two methods are simply direct extensions of the methods proposed by Kang and Albin.[6] The third method is developed based on the method proposed by Kim et al.[10] Eyvazian et al[15] proposed four methods to monitor multivariate linear profiles in phase II study. Ghashghaei et al[16] extended exponentially weighted moving average semicircle and generally weighted moving average semicircle control charts to simultaneously monitor the mean vector and the covariance matrix of multivariate linear profiles in phase II study. Soleimani et al[17] investigated phase II monitoring of the multivariate linear profiles when the independence assumption of observations within profiles is violated. They pointed out that the MEWMA/χ^2 method has better detecting performance.

2.2 | Nonlinear profile

In recent years, nonlinear profile monitoring has become a hot research topic. Williams et al[2] proposed using three different T^2 statistics for phase I analysis to monitor the coefficients resulting from a parametric nonlinear regression model that was used to fit profile data. They also considered the use of nonparametric regression method and the use of metrics to measure deviations from a reference profile. Because of the fact that nonparametric regression techniques provide great flexibility in modeling the responses, Walker and Wright[18] proposed a nonparametric method (smoothing technique) to model the vertical density profiles (VDPs) of particle boards. Woodall et al[8] discussed that the methods of monitoring complicated profiles using principal component analysis, splines, and wavelets. On the basis of the functional data analysis, Moguerza et al[19] developed a new method for constructing the control limits for nonlinear profiles. Vaghefi et al[20] applied the smoothing spline method to fit nonlinear profiles and used the metrics to measure the deviations from a reference profile to monitor the process. Hung et al[21] proposed using SVR to fit-in-control profiles. Then, they employed the moving block bootstrap method to generate correlated samples for each in-control profile and obtain a simultaneous confidence region for the underlying functional relationship. The obtained confidence region was used to monitor the real AIDS data collected from hospitals in Taiwan. Lee et al[22] proposed using a smoothing spline method to construct a reference (baseline) profile. Then, a decision tree-based monitoring procedure was developed based on a feature vector with distance-based metrics/statistics. Chen et al[23] proposed a method that integrates CUSUM statistics and support vector machines (SVMs) to monitor general profiles. Li et al[24] proposed a new nonparametric monitoring method to monitor nonlinear profile data via a SVR model. They used the VDP data as an example to demonstrate the usefulness of their control chart. Guevara and Vargas[25] proposed a new method combining the process capability and principal component analysis to monitor multivariate nonlinear profiles. Steiner et al[4] proposed a nonlinear profile monitoring method for oven temperature data. They used T^2 statistic to monitor nonlinear model parameters for four locations of the oven.

2.3 | Small number of profiles

Because of the expensive cost incurred in the destructive testing and some industrial cases, the number of historical profiles is usually small. Zou et al[3] considered that the number of profiles is not large in phase I study. In fact, it is considerably small in many practical applications. For example, the particle boards data used by Walker and Wright[18] only contained 22 VDPs; the calibration data used by Mahmoud and Woodall[7] only contained 22 profiles; the calibration data used by Gupta et al[5] only contained 10 profiles; the dose–response data used by Williams et al[2] only contained 44 profiles, and the automotive engine testing data used by Amiri et al[26] only contained 26 profiles.

2.4 | Support vector regression

Li et al[24] proposed the use of SVR for fitting the univariate nonlinear profile data and obtained their reference profiles. The SVR is a supervised statistical learning algorithm for regression problem. In SVR, the explanatory variables are mapped onto a feature space, and a linear model in Equation (1) is constructed in the feature space.

$$g(\mathbf{X}, \mathbf{w}) = \mathbf{w}^{T}\phi(\mathbf{X}) + \mathbf{b}, \tag{1}$$

where $\mathbf{w}$ is normal vector, $\phi(.)$ is a nonlinear transformation function and $\mathbf{b}$ is bias term. The quality of estimation is measured by the loss function $L(y_i, g(\mathbf{X}_i, \mathbf{w}))$, and the SVR model is formulated into a minimization problem as listed in Equation (2).

$$\min \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\left(\xi_i + \xi_i^*\right) \tag{2}$$

subject to:

$$\begin{cases} y_i - g(\mathbf{X}_i, \mathbf{w}) \leq \epsilon + \xi_i \\ g(\mathbf{X}_i, \mathbf{w}) - y_i \leq \epsilon + \xi_i^*, \\ \xi_i, \xi_i^* \quad \geq 0 \end{cases} \tag{3}$$

where $\epsilon > 0$ is a certain threshold, constant $C > 0$ is a penalization that can be viewed as a way to control over-fitting, ξ_i and ξ_i^* are slack variables.

The loss function is given by

$$L(y_i, g(x_i, \mathbf{w})) = \begin{cases} 0 & , \quad \text{if } |y_i - g(x_i, \mathbf{w})| < \epsilon \\ |y_i - g(x_i, \mathbf{w})| - \epsilon & , \quad \text{otherwise} \end{cases} \tag{4}$$

2.5 | Spatial rank-based multivariate EWMA control chart

SREWMA control chart proposed by Zou et al[3] is a multivariate self-starting method for monitoring location parameters. The spatial rank and multivariate exponential weighted moving average methods are combined to construct the control chart.

Zou et al[3] used the following multivariate location change-point model:

$$\boldsymbol{x}_j \overset{iid}{\longrightarrow} \begin{cases} \boldsymbol{\mu}_0 + \boldsymbol{\Omega}\boldsymbol{\varepsilon}_j, & for\ j = -m_0 + 1, \ldots, 0, 1, \ldots, \tau \\ \boldsymbol{\mu}_1 + \boldsymbol{\Omega}\boldsymbol{\varepsilon}_j, & for\ j = \tau + 1, \ldots \end{cases},$$

where $\boldsymbol{x}_j \in R^p$ is the jth p-dimensional observation vector, $\boldsymbol{x}_{-m_0+1}, \ldots, \boldsymbol{x}_0$ are the independent and identically distributed historical observations, τ is the unknown change point, $\boldsymbol{\mu}_0 \neq \boldsymbol{\mu}_1$, the p vectors $\boldsymbol{\varepsilon}_j$ are independent, standardized, and centered residuals, $\boldsymbol{\Omega}$ is a full-rank $p \times p$ transformation matrix, and $\boldsymbol{\Sigma} = \boldsymbol{\Omega}\boldsymbol{\Omega}^{\boldsymbol{T}} > 0$ is a scatter matrix. The spatial sign function is defined as follows:

$$U(\boldsymbol{x}) = \begin{cases} \|\boldsymbol{x}\|^{-1}\boldsymbol{x}, & \boldsymbol{x} \neq 0 \\ 0, & \boldsymbol{x} = 0 \end{cases}; \ \|\boldsymbol{x}\| = (\boldsymbol{x}^T\boldsymbol{x})^{\frac{1}{2}}.$$

Given that the $\widehat{\boldsymbol{M}}_{j-1}$ is the triangular Cholesky inverse root of $\widehat{\boldsymbol{S}}_{j-1}$ and $\widehat{\boldsymbol{S}}_{j-1}$ is the sample covariance matrix of the historical sample at the jth time point, Zou et al[3] proposed the following spatial rank EWMA charting statistic:

$$Q_j^{R_E} = \frac{2-\lambda}{\lambda} \boldsymbol{v}_j^T \left\{ \operatorname{cov}\left[R_F\left(\boldsymbol{M}\boldsymbol{x}_j\right)\right] \right\}^{-1} \boldsymbol{v}_j;\ \mathrm{j} = 1,2, \ldots ,$$

where

$$\operatorname{cov}\left[R_F\left(\boldsymbol{M}\boldsymbol{x}_j\right)\right] = \mathrm{E}\left[\left\|R_F\left(\boldsymbol{M}\boldsymbol{x}_j\right)\right\|^2\right]\mathbf{I}_p/p$$

$$\mathrm{E}\left[\left\|R_F\left(\boldsymbol{M}\boldsymbol{x}_j\right)\right\|^2\right] \approx \frac{\left[\sum_{q=-m_0+1}^{0}\left\|\widetilde{R}_E\left(\widehat{\boldsymbol{M}}_0\boldsymbol{x}_q\right)\right\|^2 + \sum_{q=1}^{j-1}\left\|R_E\left(\widehat{\boldsymbol{M}}_{q-1}\boldsymbol{x}_q\right)\right\|^2\right]}{(m_0+j-1)}$$

$$\boldsymbol{v}_j = (1-\lambda)\boldsymbol{v}_{j-1} + \lambda R_E\left(\widehat{\boldsymbol{M}}_{j-1}\boldsymbol{x}_j\right);\ \boldsymbol{v}_0 = 0$$

$$R_E\left(\widehat{\boldsymbol{M}}_{j-1}\boldsymbol{x}_j\right) = \frac{1}{m_0+j-1}\sum_{q=-m_0+1}^{j-1} U\left(\widehat{\boldsymbol{M}}_{j-1}\left(\boldsymbol{x}_j - \boldsymbol{x}_q\right)\right).$$

Note that SREWMA control chart will trigger a signal if the $Q_j^{R_E}$ exceeds control limit L. Moreover, SREWMA control chart has the following properties:

1. It is sensitive to detect the small and moderate shifts in location parameters for skewed and heavy-tailed multivariate distributions.
2. It avoids the need for a lengthy data-gathering step because it is a self-starting scheme.

2.6 | Metrics

Williams et al[2] proposed a nonparametric method for monitoring nonlinear profile data in phase II study. They used the smoothing spline-fitting method to fit profile data and establish the reference profile for phase I in-control profile data. Then, five distance metrics were calculated and used to measure the deviations of each observed profile from the reference profile. Note that these five metrics were plotted in an individual X and moving range control chart (I-MR control chart) for monitoring the process change of the nonlinear profile data.

The five metrics proposed by Williams et al[2] are listed as follows:

1. Maximum deviation

$$M_{j1} = s_0 \cdot \max_i \left|\dot{y}_{ij} - \widetilde{y}_i\right|, \quad s_0 = \operatorname{sgn}\left(\dot{y}_{ij} - \widetilde{y}_k\right), \quad \text{where } k = argmax_i \left|\dot{y}_{ij} - \widetilde{y}_i\right|.$$

2. Sum of absolute deviation (SAD)

$$M_{j2} = \sum_{i=1}^{n} \left|\dot{y}_{ij} - \widetilde{y}_i\right|$$

3. Mean absolute deviation (MAD)

$$M_{j3} = \frac{1}{n}\sum_{i=1}^{n} \left|\dot{y}_{ij} - \widetilde{y}_i\right|$$

4. Maximum absolute deviation

$$M_{j4} = \left|M_{j1}\right|$$

5. Sum of squared differences (SSD)

$$M_{j5} = \sum_{i=1}^{n} \left(\dot{y}_{ij} - \widetilde{y}_i \right)^2,$$

where $\dot{y}_{ij}$ is the predicted value of jth profile under the corresponding value of the explanatory variable x_i, $i = 1,2, \ldots,n$.; the reference profile $\widetilde{y}_i$ is the average of estimated profiles across all m profiles.

3 | RESEARCH METHODOLOGY

3.1 | The multivariate nonlinear profile model and its assumptions

The jth multivariate nonlinear profile is randomly collected over time. Each profile j we have the observations $(\boldsymbol{X}_j, \boldsymbol{Y}_j)$, where $\boldsymbol{Y}_j$ is $n_j \times$ p matrix of response variables and $\boldsymbol{X}_j$ is a $n_j \times 1$ vector of explanatory variable. The n_j is sample size, and it may have different values for different profiles. For each value of explanatory variable, we have p corresponding response values. Note that the mutually independence is assumed to be existed between profiles in our multivariate nonlinear profile model. When the process is in statistical control, the model can be given as follows:

$$\boldsymbol{Y}_j = \boldsymbol{F}_j + \boldsymbol{E}_{.j}; \quad j = 1,\ 2,\ \ldots, \tag{5}$$

where

$$\boldsymbol{E}_{.j} = \left[\boldsymbol{E}_{1j}\ \ \boldsymbol{E}_{2j}\ \ \ldots\ \ \boldsymbol{E}_{\mathrm{n_j}j}\right]^T,\ \boldsymbol{E}_{ij} = (\varepsilon_{ij1},\ \ldots\ , \varepsilon_{ijp})^T$$

or equivalently

$$\begin{bmatrix} y_{1j1} & \cdots & y_{1jp} \\ \vdots & \ddots & \vdots \\ y_{n_j j1} & \cdots & y_{n_j jp} \end{bmatrix}_{n_j \times p} = \begin{bmatrix} f_{1\left(x_{1j}\right)} & \cdots & f_{p\left(x_{1j}\right)} \\ \vdots & \ddots & \vdots \\ f_{1\left(x_{n_j j}\right)} & \cdots & f_{p\left(x_{n_j j}\right)} \end{bmatrix}_{n_j \times p} + \begin{bmatrix} \varepsilon_{1j1} & \cdots & \varepsilon_{1jp} \\ \vdots & \ddots & \vdots \\ \varepsilon_{n_j j1} & \cdots & \varepsilon_{n_j jp} \end{bmatrix}_{n_j \times p},$$

where $f_{k\left(x_{ij}\right)}$ is a function with certain degree of smoothness, the response variable y_{ijk} is the observed value of the explanatory variable x_{ij} with the kth quality characteristic in the jth profile, and x_{ij} is the corresponding explanatory variable such that $i = 1,2,\ \ldots\ ,n_j$ for each $j = 1,2,\ldots$. The number of interested quality characteristics is p such that $k = 1,2, \ldots,p$. The random error terms $\boldsymbol{E}_{ij}$ are generally assumed to follow some distributions, and $f_{1\left(x_{ij}\right)}, f_{2\left(x_{ij}\right)}, \ldots, f_{p\left(x_{ij}\right)}$ are the fitted values of the optimized SVR model for each quality characteristic. Their values are constant for different jth profiles.

3.2 | Constructing reference profiles in phase I study

The eps-regression machine in R package e1071 (Meyer et al[27]) is used to build the SVR model based on the training data set collected from the in-control multivariate nonlinear profile data set in phase I study. Moguerza et al[19] suggested the use of radial basis kernel function (RBF) to fit the nonlinear profile model. The kernel function RBF effectively maps the nonlinear profile data onto an infinite-dimensional feature space, and it involves some good properties including lower complexity and less execution time. In this paper, we use the grid search and fivefold cross-validation to obtain the optimal combination of parameters by using the tune function of R package e1071 (Meyer et al[27]). The tune function for selecting the optimal combination of parameters is based on the mean squared error (MSE) in the regression analysis. According to Alwee et al[28] the parameters in RBF kernel need to be estimated as below:

1. Regularization parameter (C): C is a parameter for determining the trade-off cost between minimizing training error and minimizing model complexity.
2. Kernel parameter (γ): γ represents the parameter of the RBF kernel function.
3. The tube size of ε-insensitive loss function (ε): ε is the approximation of accuracy placed on the training data points.

In phase I study, we assume that the number of in-control profiles is m_0. Then, all the m_0 in-control nonlinear profiles for each quality characteristic are combined to serve as the training data set for the kth quality characteristic, k is ranging from 1 to p. From the training data set, we use SVR to obtain the reference profiles for each of p quality characteristics. Figure 2 depicts the situation of fitting the kth reference profile graphically.

3.3 | SVR metrics

In this paper, the following three metrics with better detecting performances (Li et al[24]) are used to establish the SVR metrics by calculating the differences between the reference profiles (obtained by fitting SVR in phase I in-control profiles) and the observed profiles in phase II study, where $i = 1,2, \ldots,n_j$, $j = 1,2, \ldots$, $k = 1,2, \ldots,p$.

1. SAD

$$w_{jk1} = \sum_{i=1}^{n_j} \left| y_{ijk} - \widetilde{y}_{ik} \right| \tag{6}$$

2. MAD

$$w_{jk2} = \frac{1}{n_j} \sum_{i=1}^{n_j} \left| y_{ijk} - \widetilde{y}_{ik} \right| \tag{7}$$

3. SSD

$$w_{jk3} = \sum_{i=1}^{n_j} \left(y_{ijk} - \widetilde{y}_{ik} \right)^2 \tag{8}$$

where y_{ijk} is the observed value of the explanatory variable x_{ij} with the kth quality characteristic in the jth profile, and $\widetilde{y}_{ik}$ is the value of the reference profile at the explanatory variable x_{ij} with the kth quality characteristic.

3.4 | Developing a revised SREWMA control chart for phase II monitoring

We adopt the concept of residual control chart and employ the nonparametric SREWMA control chart proposed by Zou et al[3] to monitor the vector of metrics. In other words, we consider the vector of metrics for profile j as an

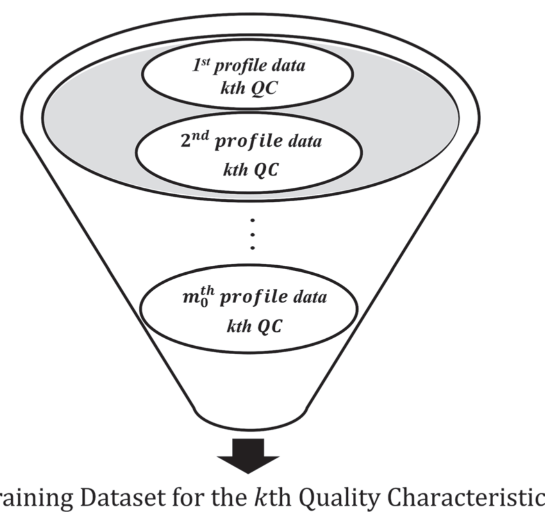

FIGURE 2 The training data set for fitting the kth reference profile in phase I

observation and propose a revised nonparametric SREWMA (RSREWMA) control chart for monitoring the multivariate location change of the vector of metrics calculated with n_j observations and p quality characteristics over j time period.

The vector of metrics is defined as follows:

$$\boldsymbol{w}_{jd} = \left(w_{j1d}, w_{j2d}, \dots, w_{jpd}\right)^T; \quad j = 1, 2, \dots, \quad d = 1, 2, 3, \tag{9}$$

where $\boldsymbol{w}_{jd}$ is a $p \times 1$ vector of metrics. Note that the symbol w_{jk1} denotes SAD for the jth profile with the kth quality characteristic; the symbol w_{jk2} denotes MAD for the jth profile with the kth quality characteristic, and the symbol w_{jk3} denotes SSD for the jth profile with the kth quality characteristic.

Let $\left\{\boldsymbol{w}^*_{1d}, \boldsymbol{w}^*_{2d}, \dots, \boldsymbol{w}^*_{m_0 d}\right\}$ be the historical sample with size m_0, $d = 1,2,3$, which are calculated from the m_0 in-control multivariate nonlinear profile data in phase I study. Let $\boldsymbol{w}_{jd}$, $j = 1,2, \dots,$ be the vector of the dth metric for the corresponding (j−1)th multivariate nonlinear profile data in phase II study. The spatial rank of $\boldsymbol{w}_{jd}$ can be calculated as follows:

$$R_E\left(\widehat{\boldsymbol{M}}_{j-1}\boldsymbol{w}_{jd}\right) = \frac{1}{m_0 + j - 1}\sum\nolimits_{q=-m_0+1}^{j-1} U\left(\widehat{\boldsymbol{M}}_{j-1}\left(\boldsymbol{w}_{jd} - \boldsymbol{w}_{qd}\right)\right); \ \boldsymbol{w}_{(-m_0+j^*)d} = \boldsymbol{w}^*_{j^*d}, \quad j^* = 1, 2, \dots, m_0, \tag{10}$$

where $\widehat{\boldsymbol{M}}_{j-1}$ is the triangular Cholesky inverse root of $\widehat{\boldsymbol{S}}_{j-1}$, and $\widehat{\boldsymbol{S}}_{j-1}$ is the sample covariance matrix of the historical sample at the jth time point.

If the observation is replaced by the spatial rank $R_E\left(\widehat{\boldsymbol{M}}_{j-1}\boldsymbol{w}_{jd}\right)$, then the control statistic of RSREWMA control chart can be written as follows:

$$Q_j^{R_E} = \frac{2-\lambda}{\lambda}\boldsymbol{v}_j^T\left\{\operatorname{cov}\left[R_F\left(\boldsymbol{M}\boldsymbol{w}_{jd}\right)\right]\right\}^{-1}\boldsymbol{v}_j \tag{11}$$

where

$$\operatorname{cov}\left[R_F\left(\boldsymbol{M}\boldsymbol{w}_{jd}\right)\right] = \mathrm{E}\left[\left\|R_F\left(\boldsymbol{M}\boldsymbol{w}_{jd}\right)\right\|^2\right]\mathbf{I}_p/p$$

$$\boldsymbol{v}_j = (1-\lambda)\boldsymbol{v}_{j-1} + \lambda R_E\left(\widehat{\boldsymbol{M}}_{j-1}\boldsymbol{w}_{jd}\right); \ \boldsymbol{v}_0 = \boldsymbol{0}$$

In order to ease the computational effect and sequentially update the estimate of $\mathrm{E}[\|R_F(\boldsymbol{M}\boldsymbol{w}_{jd})\|^2]$, we use the following approximation suggested by Zou et al[3]:

$$\mathrm{E}\left[\left\|R_F\left(\boldsymbol{M}\boldsymbol{w}_{jd}\right)\right\|^2\right] \approx \frac{\left[\sum\nolimits_{q=-m_0+1}^{0}\left\|\widetilde{R}_E\left(\widehat{\boldsymbol{M}}_0\boldsymbol{w}_{qd}\right)\right\|^2 + \sum\nolimits_{q=1}^{j-1}\left\|R_E\left(\widehat{\boldsymbol{M}}_{q-1}\boldsymbol{w}_{qd}\right)\right\|^2\right]}{(m_0 + j - 1)},$$

where

$$\widetilde{R}_E\left(\widehat{\boldsymbol{M}}_0\boldsymbol{w}_{qd}\right) = \frac{1}{m_0}\sum\nolimits_{w=-m_0+1}^{0} U\left(\widehat{\boldsymbol{M}}_0\left(\boldsymbol{w}_{qd} - \boldsymbol{w}_{wd}\right)\right); \ \boldsymbol{w}_{(-m_0+j^*)d} = \boldsymbol{w}^*_{j^*d}, \quad j^* = 1, 2, \dots, m_0.$$

If $Q_j^{R_E} > L$, then the RSREWMA control chart will trigger an alarm. Otherwise, when $Q_j^{R_E} \le L$, the $\boldsymbol{w}_{jd}$ will be included in the historical sample and the collection of new multivariate nonlinear profile data will be continued. Note that the L value needs to be determined to achieve a specified ARL_0. To facilitate the practical implementation, Zou et al[3] tabulated the values of L under various combinations of λ, p, and ARL_0 (see Table 1 in Zou et al[3]).

4 | THE SIMULATION STUDY

4.1 | Performance evaluation for the proposed control chart

There are two kinds of average run lengths (ARLs). One is called the in-control ARL (ARL_0), the other is called the out-of-control ARL (ARL_1). In statistical process control, the out-of-control ARL ($\text{ARL}_1 = \frac{1}{1-\beta}$, where β is the probability of a type II error) is commonly used to evaluate and compare the detecting performance among different control charts when the in-control ARL ($\text{ARL}_0 = 1/$, where α is the probability of a type I error or a false alarm) is fixed at 200 or 370. Thus, the detecting performance of our proposed RSREWMA control chart will be evaluated by ARL_1 when the process change of multivariate nonlinear profile data occurs. Based on Zou et al[3] we set the smoothing constant (λ) equals 0.05 and the number of simulations equals 5000 times in performing the Monte Carlo simulation. The ARL values can be calculated after 5000 replications. By fixing $\text{ARL}_0 = 200$ (ie, the probability of a type I error is .005), the ARL_1 is used to evaluate the detecting performance of our proposed RSREWMA control chart. Then, an appropriate judgment can be drawn based on the simulation results under different conditions.

4.2 | Simulation settings

4.2.1 | Model assumption

Suppose that the jth random profile sample data are collected over time and we have the observations $\{(x_{ij}, y_{ijk});\ i = 1,2, \ldots, n_j, j = 1,2, \ldots,; k = 1,2,3,4\}$, where y_{ijk} is the observed value of the explanatory variable x_{ij} with the kth quality characteristic in the jth profile, x_{ij} is the ith explanatory variable at time j, n_j is the sample size for the jth profile. Then, we consider the multivariate nonlinear in-control profile model proposed by Steiner et al[4]

$$\begin{cases} y_{ij1} = \theta_{11}*\left(1-\theta_{12}*e^{-0.036*x_{ij}}\right) + \dfrac{(258.556-\theta_{11})}{1+e^{0.023*\left(x_{ij}-313.350\right)}} + \varepsilon_{ij1} \\ y_{ij2} = 257.930*(1-0.052*e^{-0.048*x_{ij}}) + \dfrac{(260.361-257.930)}{1+e^{0.021*\left(x_{ij}-294.988\right)}} + \varepsilon_{ij2} \\ y_{ij3} = \theta_{31}*(1-0.060*e^{-0.046*x_{ij}}) + \dfrac{(261.192-\theta_{31})}{1+e^{0.022*\left(x_{ij}-296.595\right)}} + \varepsilon_{ij3} \\ y_{ij4} = 258.932*(1-0.054*e^{-0.050*x_{ij}}) + \dfrac{(261.697-258.932)}{1+e^{0.019*\left(x_{ij}-280.978\right)}} + \varepsilon_{ij4} \end{cases}, \tag{12}$$

where $\theta_{11} = 256.748$, $\theta_{12} = 0.058$, $\theta_{31} = 259.074$.

For the vector of error terms $(\varepsilon_{ij1}, \varepsilon_{ij2}, \varepsilon_{ij3}, \varepsilon_{ij4})$, the following two scenarios are considered:

1. Multivariate normal distribution: $\text{N}_4(\mathbf{0}, \mathbf{\Sigma})$
2. Multivariate exponential distribution: $MVE_4(\mathbf{1}, \boldsymbol{\rho})$

TABLE 1 The range values for the parameters (C, γ, ε)

Parameters	Range Value
C	2^{-1} to 2^{7}
γ	2^{-4} to 2^{2}
ε	0.01-0.05

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \rho\sigma_1\sigma_3 & \rho\sigma_1\sigma_4 \\ \rho\sigma_2\sigma_1 & \sigma_2^2 & \rho\sigma_2\sigma_3 & \rho\sigma_2\sigma_4 \\ \rho\sigma_3\sigma_1 & \rho\sigma_3\sigma_2 & \sigma_3^2 & \rho\sigma_3\sigma_4 \\ \rho\sigma_4\sigma_1 & \rho\sigma_4\sigma_2 & \rho\sigma_4\sigma_3 & \sigma_4^2 \end{bmatrix}, \quad \boldsymbol{\rho} = \begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix}. \tag{13}$$

In the simulation study, we assume $\sigma_1 = 1, \sigma_2 = 1, \sigma_3 = 1, \sigma_4 = 1$, and let $\rho = 0.1$, $\rho = 0.5$, $\rho = 0.9$ stand for low, medium, and high correlation cases, respectively.

4.2.2 | Training parameters for simulation study in phase I

If the studied multivariate nonlinear profile data do not meet the requirement of CFD, we then use the MAD to simulate the subsequent result. Let $x_{.j^*} = \left(x_{1j^*}, x_{2j^*}, \ldots, x_{n_{j^*}j^*}\right), j^* = 1,2, \ldots, m_0$ be a random sample taken from the sequence (0, 3, 6, 9, ..., 500) and the sample size of the studied multivariate nonlinear profiles is 161, 162, or 163, respectively. Note that the studied multivariate nonlinear profiles have different sample sizes and sampling points.

We first use SVR to fit the multivariate nonlinear profile data. Then, the reference profiles for four different quality characteristics can be obtained. In simulation study, the RBF kernel function is adopted in the SVR. The parameters need to be determined are the regularization parameter C, kernel parameter γ, and the tube size of ε. We use the grid search and K-fold cross-validation to select optimal combination of parameters. Because of the lengthy computer computing time on training SVR, the K-fold cross-validation is set to 5. Following Urbanek's[29] parallel machine in R package, we also use CPU multicore operations to decrease computing time. The optimal combination of parameters (C,γ,ε) is set to (128, 2, 0.05) for fitting the multivariate nonlinear dataset using SVR model in phase I study.

It is worthy to note that the residuals (ie, the differences between the observed multivariate profiles and the reference profiles) are used to calculate the vector of metrics. Then, the vector of the second metric $\boldsymbol{w}_{j^*2}^* = \left(w_{j^*12}, w_{j^*22}, \; w_{j^*32}, \; w_{j^*42}\right)$, where $j^* = 1,2, \ldots, m_0$ can be calculated. Let $\left\{ \boldsymbol{w}_{12}^*, \boldsymbol{w}_{22}^*, \ldots, \boldsymbol{w}_{m_02}^* \right\}$ be a historical data with sample size $= m_0$ for the second metric, which can be obtained from the in-control multivariate nonlinear profile data in phase I study, and the number of profiles (m_0) is equal to 20.

4.2.3 | Simulation results without considering within-profile correlation

In phase II study, we let $x_{.\mathrm{j}} = \left(x_{1j}, x_{2j}, \ldots, x_{n_jj}\right)$, $j = 1,2, \ldots$, be a random sample taken from the sequence (0, 1, 2, 3, ..., 500). The sample sizes of our studied multivariate nonlinear profiles are 161, 162, or 163, respectively.

We assume that various shifts in the parameters of the multivariate nonlinear profile model listed in Equation (12) are occurred and their amounts of shifts are based on the units of σ listed in Equation (13).

Considering the following types of process shifts occurred in phase II study, the simulation results of ARLs are summarized from Tables 2 to 12.

TABLE 2 The ARL_0 and ARL_1 values when θ_{11} shift to $\delta_{11}\sigma_1 + \theta_{11}$ for $\rho = 0.9$

$m_0 = 20, \lambda = 0.05, L = 10.986$		
δ_{11}	$N_4(0,\Sigma)$	$MVE_4(1,\rho)$
0.0	201.092 (188.381)	201.890 (190.633)
0.4	126.518 (163.354)	31.025 (57.769)
0.6	21.922 (28.448)	9.768 (2.017)
0.8	10.551 (2.758)	7.671 (0.773)
1.0	8.347 (1.244)	7.204 (0.517)
1.5	7.170 (0.579)	6.974 (0.377)

Note. Standard deviations are listed in parentheses.

Abbreviations: ARL_0, in-control average run length; ARL_1, out-of-control average run length.

1. Shifts for θ_{11}: $\theta_{11} \rightarrow \delta_{11}\sigma_1 + \theta_{11}$; $\delta_{11} = \pm 0.4, \pm 0.6, \pm 0.8, \pm 1.0, \pm 1.5$
2. Shifts for θ_{12}: $\theta_{12} \rightarrow \delta_{12}\sigma_1 + \theta_{12}$; $\delta_{12} = \pm 0.004, \pm 0.008, \pm 0.015, \pm 0.02$
3. Shifts for θ_{31}: $\theta_{31} \rightarrow \delta_{31}\sigma_3 + \theta_{31}$; $\delta_{31} = \pm 0.4, \pm 0.6, \pm 0.8, \pm 1.0, \pm 1.5$
4. Shifts for σ_1: $\sigma_1 \rightarrow \gamma_1\sigma_1$; $\gamma_1 = 1.04, 1.06, 1.08, 1.10$
5. Shifts for σ_3: $\sigma_3 \rightarrow \gamma_3\sigma_3$; $\gamma_3 = 1.04, 1.06, 1.08, 1.10$

The ARL values for various shifts in the θ_{11} when the random error terms follow different distributions are summarized in Table 2. The ARL values for various shifts in the θ_{12} when the random error terms follow different distributions are summarized in Table 3. Tables 2 and 3 indicate that the RSREWMA control chart can effectively detect the parameter shifts when random error terms follow either multivariate normal distribution or multivariate exponential distribution. Note that the control limit will not be affected when random errors follow different distributions and ARL_1 will be gradually decreased when the parameter shifts get bigger.

We assume that the random error terms shown in Table 4 follow a multivariate normal distribution; then, the ARL values for various shifts in σ_1 are summarized in Table 4. Similarly, the ARL_1 values for simultaneous shifts in both parameter θ_{11} and θ_{12} are summarized in Table 5. Moreover, the ARL_1 values for simultaneous shifts in both parameter θ_{11} and θ_{31} are summarized in Table 6, and the ARL_1 values for simultaneous shifts in both parameter σ_1 and σ_3 are summarized in Table 7. Tables 4, 5, 6, and 7 indicate that the RSREWMA control chart can effectively detect various shifts in different parameters since the ARL_1 values become smaller when the parameter shifts get bigger. It is worthy to note that the above findings and conclusions drawn from Tables 2 to 7 are based on the simulation results for $\rho = 0.9$. Similar findings and conclusions can be drawn for the cases of $\rho = 0.1$ and $\rho = 0.5$. Also notice that the ARL_1 values become smaller when the parameter shifts get bigger regardless of the directions of simultaneous shifts in the parameters (see Tables 5 and 6 for details).

To explore the effect of correlation coefficients on the detecting performance of RSREWMA, we consider the correlation coefficients with $\rho = 0.1$, $\rho = 0.5$, and $\rho = 0.9$. The ARL values for various shifts in the θ_{11} with different

TABLE 3 The ARL_0 and ARL_1 values when θ_{12} shift to $\delta_{12}\sigma_1 + \theta_{12}$ for $\rho = 0.9$

$m_0 = 20, \lambda = 0.05, L = 10.986$		
δ_{12}	$N_4(0, \Sigma)$	$MVE_4(1, \rho)$
0.000	201.092 (188.381)	201.890 (190.633)
0.004	198.867 (193.226)	151.390 (167.841)
0.006	118.082 (151.633)	63.850 (102.230)
0.008	36.798 (67.527)	21.113 (28.102)
0.010	14.314 (11.052)	12.398 (4.918)
0.015	8.513 (1.331)	8.537 (1.370)
0.020	7.568 (0.833)	7.645 (0.866)

Note. Standard deviations are listed in parentheses.

Abbreviations: ARL_0, in-control average run length; ARL_1, out-of-control average run length.

TABLE 4 The ARL_0 and ARL_1 values when σ_1 shift to $\gamma_1\sigma_1$ for $\rho = 0.9$

$m_0 = 20, \lambda = 0.05, L = 10.986$	
γ_1	$N_4(0, \Sigma)$
1.00	201.092 (188.381)
1.04	25.850 (42.882)
1.06	12.742 (5.655)
1.08	9.841 (2.010)
1.10	8.706 (1.274)

Note. Standard deviations are listed in parentheses.

Abbreviations: ARL_0, in-control average run length; ARL_1, out-of-control average run length.

TABLE 5 The ARL_1 values under simultaneous shifts from θ_{11} to $\delta_{11}\sigma_1 + \theta_{11}$ and θ_{12} to $\delta_{12}\sigma_1 + \theta_{12}$ for $\rho = 0.9$

	δ_{11}				
δ_{12}	**−0.4**	**−0.6**	**−0.8**	**−1.0**	**−1.5**
−0.020	7.213 (0.659)	7.179 (0.608)	7.104 (0.555)	7.034 (0.541)	6.928 (0.472)
−0.015	7.598 (0.856)	7.369 (0.730)	7.199 (0.632)	7.128 (0.533)	6.966 (0.462)
−0.008	9.756 (2.038)	8.399 (1.216)	7.731 (0.883)	7.375 (0.693)	7.031 (0.500)
−0.004	19.812 (21.383)	10.797 (2.849)	8.499 (1.245)	7.711 (0.907)	7.055 (0.506)
0.004	56.982 (88.644)	13.867 (6.432)	9.403 (1.972)	7.990 (1.046)	7.094 (0.574)
0.008	13.550 (7.427)	9.977 (2.337)	8.254 (1.195)	7.607 (0.799)	7.093 (0.541)
0.015	7.964 (0.994)	7.625 (0.815)	7.400 (0.707)	7.231 (0.632)	6.971 (0.478)
0.020	7.401(0.726)	7.272 (0.670)	7.168 (0.617)	7.099 (0.578)	6.979 (0.482)
	δ_{11}				
δ_{12}	**0.4**	**0.6**	**0.8**	**1.0**	**1.5**
−0.020	7.250 (0.645)	7.218 (0.591)	7.115 (0.603)	7.056 (0.530)	6.968 (0.483)
−0.015	7.646 (0.835)	7.450 (0.740)	7.283 (0.655)	7.164 (0.562)	6.974 (0.481)
−0.008	10.285 (2.619)	8.928 (1.548)	7.948 (0.981)	7.489 (0.730)	7.035 (0.504)
−0.004	28.320 (41.837)	12.071 (4.110)	9.081 (1.632)	7.973 (1.065)	7.116 (0.539)
0.004	80.197 (126.443)	17.603 (20.628)	10.006 (2.324)	8.223 (1.152)	7.141 (0.571)
0.008	14.944 (13.540)	10.321 (2.645)	8.514 (1.367)	7.730 (0.895)	7.101(0.537)
0.015	8.081 (1.120)	7.702 (0.889)	7.416 (0.731)	7.241 (0.634)	7.004 (0.496)
0.020	7.441 (0.746)	7.312 (0.701)	7.199 (0.614)	7.101 (0.551)	6.964 (0.476)

Note. Standard deviations are listed in parentheses.

Abbreviation: ARL_1, out-of-control average run length.

correlation coefficients are summarized in Table 8. Table 8 indicates that the detecting performance of our proposed control chart becomes better when the correlation coefficient increases.

To explore the effect of historical sample size on the detecting performance of RSREWMA, we consider m_0=10, m_0=20, and m_0=40. The ARL values for various shifts in the θ_{11} with different historical sample sizes are summarized in Table 9. Table 9 indicates that the detecting performance of our proposed control chart becomes better when the historical sample size increases. Note that the control chart with a small historical sample size also has a good detecting performance.

To find a better selection of metrics when the multivariate nonlinear profiles meet the requirement of CFD, we consider the sample size equal to 162 and the sampling points equal {0,3,6,9, ... ,483} for phases I and II profiles. The ARL values for various shifts in the θ_{11} with different metrics are summarized in Table 10. The ARL values for various shifts in the θ_{12} with different metrics are summarized in Table 11. Tables 10 and 11 indicate that the control chart with the metric of SSD has a better detecting performance than the one with other metrics, ie, SAD and MAD. According to the testing result of simulation study, the charting statistic of the metric for SSD is bigger than those of other metrics with the same amount of parameter shifts. Thus, the ARL_1 values for SSD are smaller than those of other metrics.

4.2.4 | Simulation results considering within-profile correlation

Because of the spatial autocorrelation or time collapse, the assumption of uncorrelated observations within each profile may be invalid. Thus, if the within-profile correlation exists in each multivariate nonlinear profile, then the impact of autocorrelation on the detecting performance of RSREWMA control chart need to be further explored.

Assume that the dependent relationship exists within each multivariate nonlinear profile, then the multivariate nonlinear profile model in the presence of with-profile autocorrelation for vector autoregressive model of order one, denoted as VAR (1), is given as follows:

TABLE 6 The ARL_1 values under simultaneous shifts from θ_{11} to $\delta_{11}\sigma_1+\theta_{11}$ and θ_{31} to $\delta_{31}\sigma_3+\theta_{31}$ for $\rho=0.9$

	δ_{11}				
δ_{31}	**−0.4**	**−0.6**	**−0.8**	**−1.0**	**−1.5**
−1.5	7.574 (0.695)	7.557 (0.673)	7.590 (0.712)	7.621 (0.740)	7.618 (0.723)
−1.0	8.288 (1.088)	8.362 (1.154)	8.304 (1.101)	8.105 (1.007)	7.645 (0.730)
−0.8	9.613 (1.751)	9.659 (1.984)	9.161 (1.638)	8.458 (1.255)	7.552 (0.680)
−0.4	51.174 (80.723)	16.517 (9.560)	9.958 (2.266)	8.130 (1.026)	7.189 (0.549)
0.4	65.240 (100.445)	16.246 (8.431)	10.092 (2.238)	8.220 (1.098)	7.230 (0.605)
0.8	11.141 (2.788)	10.555 (2.335)	9.507 (1.448)	8.678 (1.179)	7.792 (0.867)
1.0	8.939 (1.364)	8.959 (1.331)	8.887 (1.174)	8.618 (1.083)	8.069 (0.918)
1.5	7.714 (0.806)	7.875 (0.838)	8.038 (0.849)	8.117 (0.834)	8.348 (0.959)
	δ_{11}				
δ_{31}	**0.4**	**0.6**	**0.8**	**1.0**	**1.5**
−1.5	7.607 (0.749)	7.802 (0.781)	7.897 (0.784)	8.054 (0.837)	8.382 (0.959)
−1.0	8.489 (1.158)	8.690 (1.135)	8.651 (1.061)	8.554 (0.996)	8.234 (0.984)
−0.8	9.791 (1.907)	9.815 (1.703)	9.363 (1.396)	8.739 (1.165)	7.918 (0.935)
−0.4	62.441 (104.018)	19.342 (16.196)	11.009 (2.865)	8.729 (1.414)	7.345 (0.689)
0.4	100.189 (138.175)	25.661 (38.955)	10.786 (2.845)	8.442 (1.226)	7.200 (0.562)
0.8	11.237 (3.105)	11.032 (3.004)	9.995 (2.156)	8.318 (1.102)	7.155 (0.545)
1.0	8.973 (1.401)	8.842 (1.381)	8.728 (1.329)	8.297 (1.197)	7.107 (0.543)
1.5	7.631 (0.728)	7.565 (0.659)	7.321 (0.635)	7.259 (0.599)	7.100 (0.537)

Note. Standard deviations are listed in parentheses.

Abbreviation: ARL_1, out-of-control average run length.

TABLE 7 The ARL_1 values under simultaneous shifts from σ_1 to $\gamma_1\sigma_1$ and σ_3 to $\gamma_3\sigma_3$ for $\rho=0.9$

	γ_1			
γ_3	**0.96**	**0.94**	**0.92**	**0.90**
1.04	19.106 (24.730)	12.692 (4.777)	10.279 (2.207)	9.086 (1.479)
1.06	13.178 (5.346)	11.306 (2.920)	9.918 (1.942)	9.067 (1.430)
1.08	10.752 (2.571)	10.087 (1.921)	9.474 (1.599)	8.881 (1.262)
1.10	9.451 (1.573)	9.247 (1.459)	8.943 (1.277)	8.608 (1.095)

Note. Standard deviations are listed in parentheses.

Abbreviation: ARL_1, out-of-control average run length.

$$\boldsymbol{Y}_j=\boldsymbol{F}_j+\boldsymbol{E}_{.j};\ j=1,\ 2,\ \dots,\ ;\ \boldsymbol{E}_{.j}=\begin{bmatrix}\boldsymbol{E}_{1j} & \boldsymbol{E}_{2j} & \cdots & \boldsymbol{E}_{n_j j}\end{bmatrix}^T$$

$$\boldsymbol{E}_{ij}=\boldsymbol{\phi}\boldsymbol{E}_{(i-1)j}+\boldsymbol{A}_{ij},\quad \boldsymbol{A}_{ij}\overset{iid}{\rightarrow}N_4(\boldsymbol{0},\boldsymbol{\Sigma});\ \mathrm{i}=1,\ 2,\ \dots,\ n_j,\quad \boldsymbol{\phi}=\begin{bmatrix}\phi & 0 & 0 & 0\\ 0 & \phi & 0 & 0\\ 0 & 0 & \phi & 0\\ 0 & 0 & 0 & \phi\end{bmatrix},$$

where $\boldsymbol{E}_{ij}$ are correlated error terms and $\boldsymbol{A}_{ij}$ are independently and identically distributed multivariate normal random variables with zero mean vector and covariance matrix $\boldsymbol{\Sigma}$. The simulation setting of explanatory variables is the same as the one listed in sections 4.2.2 and 4.2.3.

TABLE 8 The ARL_0 and ARL_1 values when θ_{11} shift to $\delta_{11}\sigma_1 + \theta_{11}$ with different correlation coefficients

$\boldsymbol{m_0 = 20, \lambda = 0.05, L = 10.986}$			
δ_{11}	$\rho = 0.1$	$\rho = 0.5$	$\rho = 0.9$
0.0	201.903 (191.792)	200.650 (185.312)	201.092 (188.381)
0.4	131.319 (157.576)	129.873 (154.141)	126.518 (163.354)
0.6	42.312 (73.277)	38.651 (71.194)	21.922 (28.448)
0.8	13.782 (6.423)	13.210 (7.655)	10.551 (2.758)
1.0	9.724 (1.932)	9.533 (1.824)	8.347 (1.244)
1.5	7.465 (0.650)	7.418 (0.640)	7.170 (0.579)

Note. Standard deviations are listed in parentheses.

Abbreviations: ARL_0, in-control average run length; ARL_1, out-of-control average run length.

TABLE 9 The ARL_0 and ARL_1 values when θ_{11} shift to $\delta_{11}\sigma_1 + \theta_{11}$ for $\rho = 0.9$ under various historical sample sizes

$\boldsymbol{\lambda = 0.05, ARL_0 = 200}$			
δ_{11}	$\boldsymbol{m_0 = 10, L = 10.811}$	$\boldsymbol{m_0 = 20, L = 10.986}$	$\boldsymbol{m_0 = 40, L = 11.055}$
0.0	202.848 (186.326)	201.092 (188.381)	202.374 (192.751)
0.4	167.277 (177.177)	126.518 (163.354)	71.577 (104.542)
0.6	67.266 (108.437)	21.922 (28.448)	14.662 (6.646)
0.8	19.919 (37.350)	10.551 (2.758)	8.957 (1.729)
1.0	11.644 (10.338)	8.347 (1.244)	7.279 (0.795)
1.5	9.172 (2.068)	7.170 (0.579)	6.217 (0.416)

Note. Standard deviations are listed in parentheses.

Abbreviations: ARL_0, in-control average run length; ARL_1, out-of-control average run length.

TABLE 10 The ARL_0 and ARL_1 values when θ_{11} shift to $\delta_{11}\sigma_1+\theta_{11}$ with different metrics

$\boldsymbol{m_0 = 20, \lambda = 0.05, L = 10.986}$			
δ_{11}	$\sum_{i=1}^{n} \lvert e_{ijk} \rvert$	$\frac{1}{n}\sum_{i=1}^{n} \lvert e_{ijk} \rvert$	$\sum_{i=1}^{n} (e_{ijk})^2$
0.0	201.060 (188.030)	202.548 (187.528)	200.641 (188.302)
0.4	104.145 (150.267)	102.145 (141.188)	82.124 (123.001)
0.6	19.642 (26.078)	19.416 (25.282)	15.251 (21.939)
0.8	10.254 (2.522)	10.266 (2.528)	9.291 (1.888)
1.0	8.299 (1.183)	8.283 (1.165)	7.878 (0.995)
1.5	7.153 (0.551)	7.159 (0.552)	7.051 (0.524)

Note. Standard deviations are listed in parentheses.

Abbreviations: ARL_0, in-control average run length; ARL_1, out-of-control average run length.

Considering the autocorrelation coefficients with $\phi = 0.1$, $\phi = 0.5$, and $\phi = 0.9$, the simulation results of ARL values are summarized in Table 12. Table 12 indicates that the detecting performance of our proposed control chart gets better when the autocorrelation coefficient decreases, and the in-control ARLs will not be affected when the within-profile correlation exists in each multivariate nonlinear profile, ie, the detecting performance for our proposed control chart with a weaker autocorrelation outperforms that with a stronger autocorrelation. These simulation results are similar to those shown in Soleimani et al.[13]

TABLE 11 The ARL_0 and ARL_1 values when θ_{12} shift to $\delta_{12}\sigma_1 + \theta_{12}$ with different metrics

$m_0 = 20, \lambda = 0.05, L = 10.986$			
δ_{12}	$\sum_{i=1}^{n}\lvert e_{ijk}\rvert$	$\frac{1}{n}\sum_{i=1}^{n}\lvert e_{ijk}\rvert$	$\sum_{i=1}^{n}(e_{ijk})^2$
0.000	201.060 (188.030)	202.548 (187.528)	200.641 (188.302)
0.004	166.305 (183.797)	170.560 (184.985)	155.761 (177.004)
0.006	67.188 (109.795)	68.756 (110.619)	41.800 (71.010)
0.008	17.775 (19.768)	18.485 (27.853)	12.868 (5.391)
0.010	10.875 (2.950)	10.938 (4.016)	9.105 (1.748)
0.015	7.782 (0.863)	7.771 (0.842)	7.286 (0.681)
0.020	7.200 (0.519)	7.202 (0.530)	6.989 (0.498)

Note. Standard deviations are listed in parentheses.

Abbreviations: ARL_0, in-control average run length; ARL_1, out-of-control average run length.

TABLE 12 The ARL_0 and ARL_1 values when θ_{11} shift to $\delta_{11}\sigma_1 + \theta_{11}$ with different autocorrelation coefficients

$m_0 = 20, \lambda = 0.05, L = 10.986$			
δ_{11}	$\phi = 0.1$	$\phi = 0.5$	$\phi = 0.9$
0.0	201.512 (188.787)	200.333 (187.406)	201.254 (186.371)
0.4	96.841 (138.535)	121.823 (158.828)	192.135 (183.726)
0.6	17.297 (14.358)	34.727 (57.287)	174.413 (180.144)
0.8	9.932 (2.240)	13.881 (7.788)	142.238 (172.353)
1.0	8.193 (1.157)	10.129 (2.611)	111.964 (150.338)
1.5	7.152 (0.591)	7.685 (0.941)	42.620 (76.422)

Note. Standard deviations are listed in parentheses.

Abbreviations: ARL_0, in-control average run length; ARL_1, out-of-control average run length.

5 | A NUMERICAL EXAMPLE

In this paper, an oven temperature data set given by Steiner et al[4] is used as an example to demonstrate the application of our proposed RSREWMA control chart in detecting the process changes for multivariate nonlinear profile data.

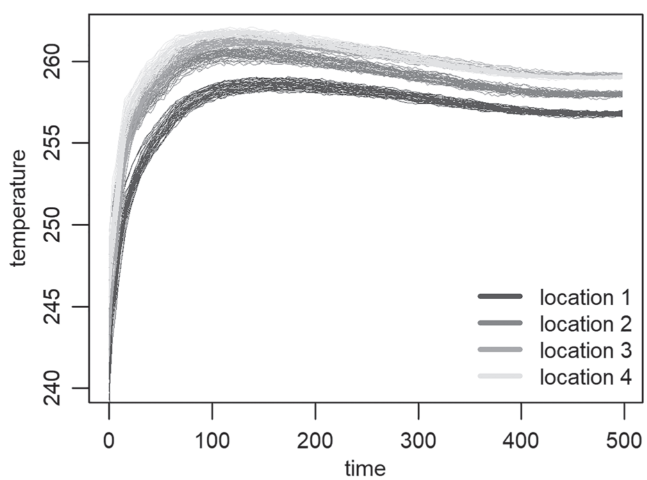

FIGURE 3 In-control profiles for the oven temperature

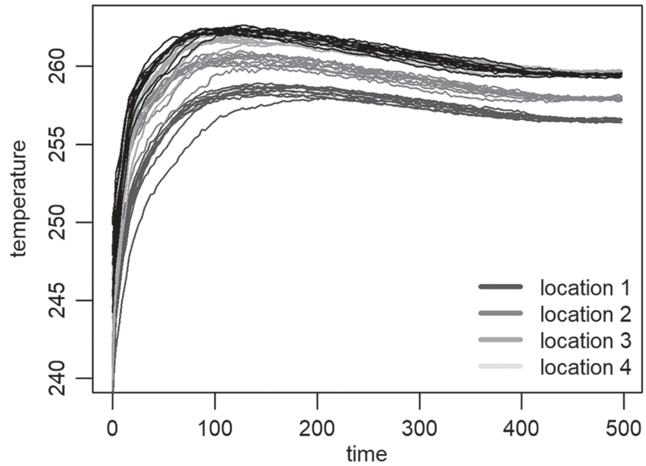

FIGURE 4 Out-of-control profiles for the oven temperature

It is worthy to note that we assume the mutual independence exists between multivariate nonlinear profiles in our multivariate nonlinear profile model. Thus, the between-profile correlations will not be considered in the numerical example.

For each run (profile), the temperature is measured at four different locations. Since the temperatures of four locations are important, they are the key quality characteristics. The sensors used in the oven allow us to gather a temperature value roughly every 3 seconds throughout the oven run, which results in approximately 160 data points in a 500-second data-collection period. Although the actual oven run can be much longer, the temperature becomes stable when it reaches a plateau beyond 500 seconds. In the numerical example, we only concern about the initial start-up period in each run; thus, these data points do not necessarily occur at the same time interval.

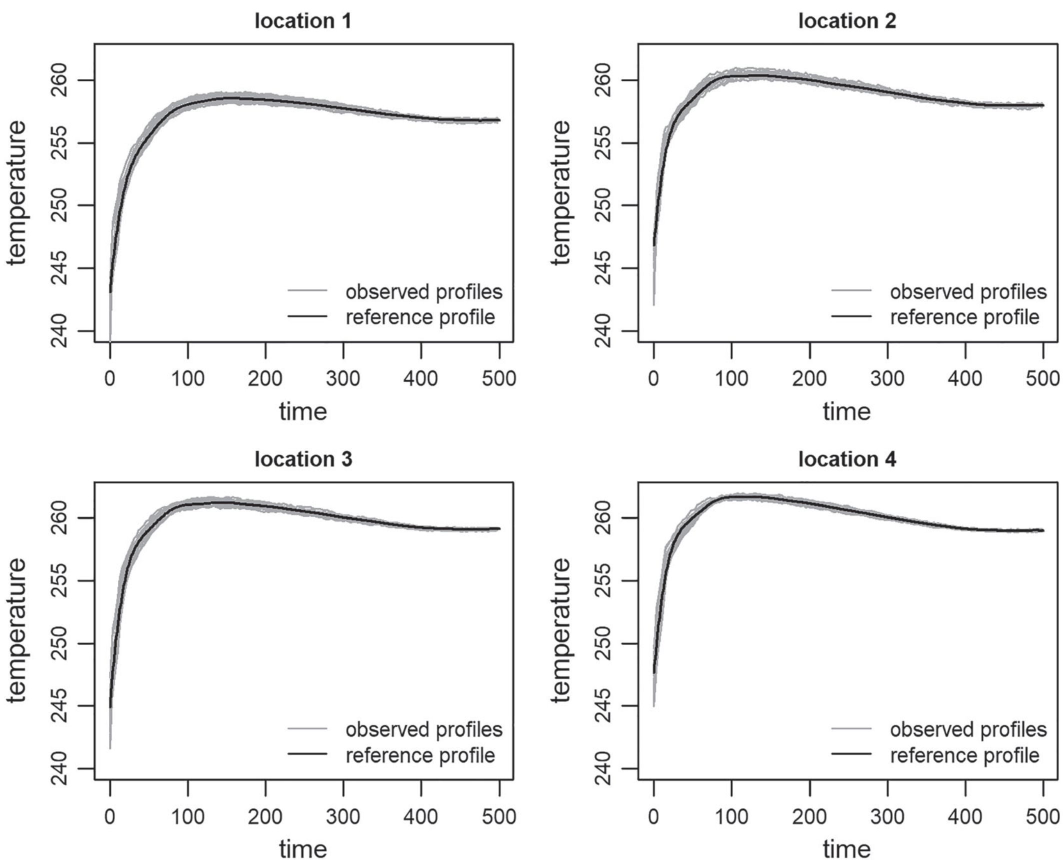

FIGURE 5 The observed and reference profiles for four different locations

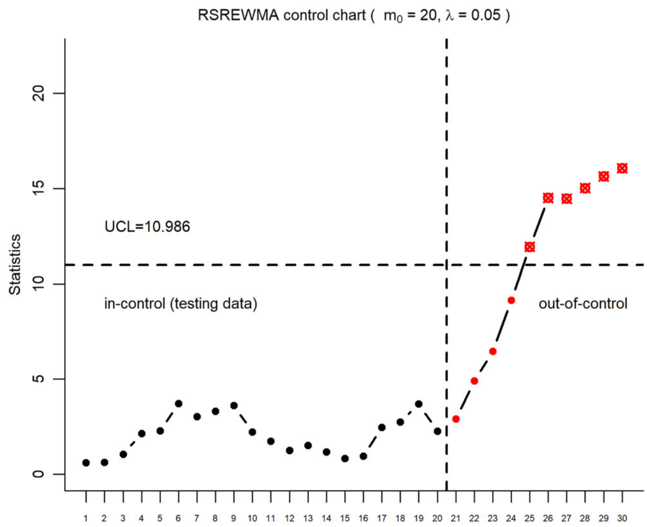

FIGURE 6 The revised spatial rank exponential weighted moving average (RSREWMA) control chart for monitoring the oven temperature data [Colour figure can be viewed at wileyonlinelibrary.com]

According to the parameter estimates by Steiner et al[4] 40 runs of similar parameter values are used to plot our in-control multivariate nonlinear profile curves, and 10 runs of different parameter values are used to plot out-of-control multivariate nonlinear profile curves. These multivariate nonlinear profiles are shown in Figures 3 and 4, respectively. Note that the modified multivariate portmanteau test was proposed by Hosking.[30] The testing results for the existence of within-profile correlation in the multivariate nonlinear profile are listed in Table A.1. These testing results confirm that the VAR (1) structure between observations in each profile is valid. Note that the 40 in-control profile data are divided into two parts; the first 20 profile data are considered as the historical data, and the rest of 20 profile data are considered as the testing data. The testing data and out-of-control data are listed in Table B.1. In phase I study, we first fit the historical in-control profile data using SVR model and obtain the reference profiles. The optimal combination of parameters (C,γ,ε) is set to (128, 2, 0.05). Then, both the observed profiles and the reference profiles are shown in Figure 5.

The RSREWMA control chart with the smoothing constant $\lambda = 0.05$ and the historical sample size m_0= 20 is plotted in Figure 6. Figure 6 shows that the testing in-control data are within the control limit (phase I), but the fifth data for the out-of-control profile exceeds the control limit. Thus, it indicates that our proposed control chart can effectively detect the process changes for multivariate nonlinear profile data in phase II monitoring.

6 | CONCLUSIONS AND FUTURE RESEARCH AREAS

With the advent of modern technology, manufacturing processes have become rather sophisticated. Two or more correlated quality characteristics are often required for monitoring the product and process quality in today's high-technology industries.

In many practical applications, we often do not know the distribution of the studied quality characteristics. Thus, in this paper, a nonparametric method is proposed for monitoring the multivariate nonlinear profile data. We first employ SVR model to fit the multivariate nonlinear profile data and obtain the reference profiles. Second, three distance metrics with better detecting performances, as shown in Li et al[24] are used to measure the deviations of each observed profile from the reference profile. Then, a nonparametric RSREWMA control chart is proposed to monitor multivariate nonlinear profile data in phase II study.Finally, both the simulation results and numerical example demonstrate the usefulness of our proposed RSREWMA control chart and its monitoring schemes. The significant findings and contributions of this paper can be summarized as below:

1. Even when multivariate nonlinear profile data does not meet the requirement of CFD, the proposed control chart still can detect the process changes effectively.
2. Even when the number of multivariate profiles is not large in phase I study, the proposed control chart still can detect the process changes effectively.

3. Even when the within-profile correlation exists in each multivariate nonlinear profile, the proposed control chart still can detect the process changes effectively.
4. The simulation results show that the metric of MAD has a better detecting performance for various process shifts when the multivariate nonlinear profile data does not satisfy CFD. But, when the multivariate nonlinear profile data satisfies CFD, the metric of SSD has a better detecting performance than other metrics.

Hence, our proposed RSREWMA control chart has a better performance for monitoring multivariate nonlinear profile data in phase II study. The quality practitioners may use our proposed RSREWMA control chart to timely detect the process changes in various parameters of multivariate nonlinear profiles for their manufacturing processes.

It is also worthy to note that we assume that the mutual independence exists between the multivariate nonlinear profiles and only the correlations within profiles are considered in this research. However, the manufacturing processes are often affected by time or space. For example, both the correlations within profiles and between profiles for multivariate profile data based on a multivariate Gaussian process model were considered by Jahani et al.[31] Thus, the correlations between multivariate nonlinear profiles for $p < m_0$ need to be further studied in future research. Moreover, when the number of historical data set (m_0) is smaller than the number of key quality characteristics (p), the correlation problem for $p \geq m_0$ needs to be further studied too. Although SVR is used to fit reference profile for each quality characteristic, the correlations among various quality characteristics were not considered. Future researchers may adopt the multidimensional support vector regression (MSVR) method to detect the process changes for nonlinear profile data with correlated quality characteristics.

ACKNOWLEDGEMENTS

We are grateful to the two anonymous reviewers for their helpful comments. The first author would like to gratefully acknowledge financial support (MOST 107-2410-H-006-035) from the Ministry of Science and Technology of Taiwan, ROC.

REFERENCES

1. Zou C, Tsung F, Wang Z. Monitoring profiles based on nonparametric regression methods. *Dent Tech*. 2008;50(4):512-526.
2. Williams JD, Woodall WH, Birch JB. Statistical monitoring of nonlinear product and process quality profiles. *Qual Reliab Eng Int*. 2007;23(8):925-941.
3. Zou C, Wang Z, Tsung F. A spatial rank-based multivariate EWMA control chart. *Nav Res Logist*. 2012;59(2):91-110.
4. Steiner S, Jensen WA, Grimshaw SD, Espen B. Nonlinear profile monitoring for oven-temperature data. *J Qual Technol*. 2016;48(1):84-97.
5. Gupta S, Montgomery D, Woodall W. Performance evaluation of two methods for online monitoring of linear calibration profiles. *Int J Prod Res*. 2006;44(10):1927-1942.
6. Kang L, Albin SL. On-line monitoring when the process yields a linear profile. *J Qual Technol*. 2000;32(4):418-426.
7. Mahmoud MA, Woodall WH. Phase I analysis of linear profiles with calibration applications. *Dent Tech*. 2004;46(4):380-391.
8. Woodall WH, Spitzner DJ, Montgomery DC, Gupta S. Using control charts to monitor process and product quality profiles. *J Qual Technol*. 2004;36(3):309-320.
9. Woodall WH. Current research on profile monitoring. *Production*. 2007;17(3):420-425.
10. Kim K, Mahmoud MA, Woodall WH. On the monitoring of linear profiles. *J Qual Technol*. 2003;35(3):317-328.
11. Zou C, Tsung F, Wang Z. Monitoring general linear profiles using multivariate exponentially weighted moving average schemes. *Dent Tech*. 2007;49(4):395-408.
12. Wu IL, Lee HT, Cheng CS. Application of support vector regression in monitoring process and product profiles. *J Qual*. 2014;21(3):189-203.
13. Soleimani P, Noorossana R, Amiri A. Simple linear profiles monitoring in the presence of within profile autocorrelation. *Comput Ind Eng*. 2009;57(3):1015-1021.
14. Noorossana R, Eyvazian M, Vaghefi A. Phase II monitoring of multivariate simple linear profiles. *Comput Ind Eng*. 2010;58(4):563-570.
15. Eyvazian M, Noorossana R, Saghaei A, Amiri A. Phase II monitoring of multivariate multiple linear regression profiles. *Qual Reliab Eng Int*. 2011;27(3):281-296.
16. Ghashghaei R, Amiri A, Khosravi P. New control charts for simultaneous monitoring of the mean vector and covariance matrix of multivariate multiple linear profiles. *Commun Stat Simul Comput*. 2018;1-24.
17. Soleimani P, Noorossana R, Niaki STA. Monitoring autocorrelated multivariate simple linear profiles. *Int J Adv Manuf Technol*. 2013;67(5–8):1857-1865.

18. Walker E, Wright SP. Comparing curves using additive models. *J Qual Technol*. 2002;34(1):118-129.
19. Moguerza JM, Munoz A, Psarakis S. Monitoring nonlinear profiles using support vector machines. *Lect Notes Comput Sci*. 2007;4756:574-583.
20. Vaghefi A, Tajbakhsh SD, Noorossana R. Phase II monitoring of nonlinear profiles. *Commun Stat Simul Comput*. 2009;38(11):1834-1851.
21. Hung YC, Tsai WC, Yang SF, Chuang SC, Tseng YK. Nonparametric profile monitoring in multi-dimensional data spaces. *J Process Control*. 2012;22(2):397-403.
22. Lee HT, Huang JY, Cheng CS. The monitoring of nonlinear profiles using smoothing spline and decision tree model. *J Qual*. 2015;22(2):77-87.
23. Chen PW, Peng WS, Cheng CS, Wang YT. Monitoring of nonlinear profile variations using support vector machines. *J Qual*. 2017;24(5):313-323.
24. Li CI, Pan JN, Liao CH. Monitoring nonlinear profile data using support vector regression method. *Qual Reliab Eng Int*. 2019;35(1):127-135.
25. Guevara RD, Vargas JA. Evaluation of process capability in multivariate nonlinear profiles. *J Stat Comput Simul*. 2016;86(12):2411-2428.
26. Amiri A, Jensen WA, Kazemzadeh RB. A case study on monitoring polynomial profiles in the automotive industry. *Qual Reliab Eng Int*. 2010;26(5):509-520.
27. Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F. e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. *R package;* 2015.
28. Alwee R, Shamsuddin H, Mariyam S, Sallehuddin R. Hybrid support vector regression and autoregressive integrated moving average models improved by particle swarm optimization for property crime rates forecasting with economic indicators. *Scientific World Journal*. 2013;2013:1-11.
29. Urbanek S. Multicore: Parallel processing of R code on machines with multiple cores or CPUs. *R package*; 2011.
30. Hosking JR. The multivariate portmanteau statistic. *J Am Stat Assoc*. 1980;75(371):602-608.
31. Jahani S, Kontar R, Veeramani D, Zhou S. Statistical monitoring of multiple profiles simultaneously using Gaussian processes. *Qual Reliab Eng Int*. 2018;34(8):1510-1529. https://doi.org/10.1002/qre.2326

AUTHOR BIOGRAPHIES

Dr Jeh-Nan Pan is currently a Professor of Statistics at College of Management, National Cheng Kung University (NCKU). Professor Pan's expertise is in the fields of quality management and reliability engineering, industrial statistics, and statistical consulting. He has founded the first Statistical Consulting Center in Taiwan and has served as a Chairman of Statistics Department from 2001 to 2004. Then, he has served as a Vice Dean for more than 9 years. He was previously on the staff of USC and California State University in USA and has more than 10 years of industrial experience, started with a Senior Quality Assurance Engineer and then promoted to Chief Statistician at Leach Corporation and Statistical Process Control Program Manager at Teledyne Inc. in Los Angeles, USA. He has authored two textbooks: *Quality Management: Principles and Practice* and *Proactive Quality Assurance* (in Chinese) and coauthored (with Dr William Kolarik) a textbook *Creating Quality* (McGraw Hill Inc.) He has published 78 papers in both domestic and renowned SCI/SSCI international journals such as: *IEEE Transaction on Reliability*, *European Journal of Operational Research*, *Omega-The International Journal of Management Science, Computers and Industrial Engineering, Quality and Reliability Engineering International, Quality Engineering*, *International Journal of Production Research*, etc.

Currently, he serves on the editorial boards of *Journal of Quality, Management and Systems*, *Asia Pacific Management Review Journal*, etc. He holds MS and PhD degrees in Industrial Engineering from Texas Tech University, Lubbock, Texas. He is ASQ certified Quality and Reliability Engineer. He has received numerous distinguished awards including three Research and Teaching Excellence awards from NCKU, four best paper awards from Chinese Society for Quality (CSQ), and Asian Network for Quality (ANQ). He has also been elected as a CSQ fellow in 2016.

Dr Chung-I Li is an assistant professor in the Department of Statistics, National Cheng Kung University. He received his PhD degree in statistics from National Cheng Kung University. His current research interest is in the fields of statistical quality control and applied statistics.

Mr Meng Zhe Lu is a graduate student at National Cheng Kung University, Tainan, Taiwan. His chosen field is statistics.

How to cite this article: Pan J-N, Li C-I, Lu MZ. Detecting the process changes for multivariate nonlinear profile data. *Qual Reliab Engng Int*. 2019;1–21. https://doi.org/10.1002/qre.2482

APPENDIX A

TABLE A.1 The testing results for the existence of within correlation in the multivariate nonlinear profile

Historical Sample	The *j*th Profile in the Study of Steiner et al[4]	[a]ADF Test Result of Location 1	ADF Test Result of Location 2	ADF Test Result of Location 3	ADF Test Result of Location 4	[b]Modified Multivariate Portmanteau Test Result
1	160	Stationary	Stationary	Stationary	Stationary	Exist
2	948	Stationary	Stationary	Stationary	Stationary	Exist
3	457	Stationary	Stationary	Stationary	Stationary	Exist
4	105	Stationary	Stationary	Stationary	Stationary	Exist
5	542	Stationary	Stationary	Stationary	Stationary	Exist
6	850	Stationary	Stationary	Stationary	Stationary	Exist
7	153	Stationary	Stationary	Stationary	Stationary	Exist
8	135	Stationary	Stationary	Stationary	Stationary	Exist
9	451	Stationary	Stationary	Stationary	Stationary	Exist
10	554	Stationary	Stationary	Stationary	Stationary	Exist
11	973	Stationary	Stationary	Stationary	Stationary	Exist
12	540	Stationary	Stationary	Stationary	Stationary	Exist
13	552	Stationary	Stationary	Stationary	Stationary	Exist
14	550	Stationary	Stationary	Stationary	Stationary	Exist
15	757	Stationary	Stationary	Stationary	Stationary	Exist
16	495	Stationary	Stationary	Stationary	Stationary	Exist
17	568	Stationary	Stationary	Stationary	Stationary	Exist
18	689	Stationary	Stationary	Stationary	Stationary	Exist
19	1003	Stationary	Stationary	Stationary	Stationary	Exist
20	197	Stationary	Stationary	Stationary	Stationary	Exist

Abbreviation: ADF, augmented Dickey–Fuller.

[a]ADF test is performed whether the time series data are stationary.

[b]Modified multivariate portmanteau test is performed whether the serial correlation exists in the time series.

APPENDIX B

TABLE B.1 The 20 testing data and 10 out-of-control data for the oven temperature

Testing Data	The *j*th Profile in the Steiner et al[4]	Out-of-control Data	The *j*th Profile in the Steiner et al[4]
1	750	1	352
2	667	2	290
3	8	3	417

(Continues)

TABLE B.1 (Continued)

Testing Data	The *j*th Profile in the Steiner et al[4]	Out-of-control Data	The *j*th Profile in the Steiner et al[4]
4	558	4	346
5	987	5	390
6	455	6	398
7	165	7	275
8	849	8	381
9	151	9	400
10	732	10	269
11	690		
12	547		
13	616		
14	488		
15	532		
16	553		
17	498		
18	955		
19	485		
20	472		