Received: 14 August 2017 | Revised: 21 March 2018 | Accepted: 23 April 2018

DOI: 10.1002/qre.2326

RESEARCH ARTICLE

WILEY

Statistical monitoring of multiple profiles simultaneously using Gaussian processes

Salman Jahani | Raed Kontar | Dharmaraj Veeramani | Shiyu Zhou

Department of Industrial and Systems Engineering, University of Wisconsin-Madison, Madison, WI, USA

Correspondence
Shiyu Zhou, Department of Industrial and Systems Engineering, University of Wisconsin-Madison, Madison, WI, USA.
Email: shiyuzhou@wisc.edu

Funding information
National Science Foundation, Grant/Award Number: 1561512

Abstract

Profile monitoring is the application of control charts to monitor the stability of a process over time when the process can be characterized by a functional relationship between a response variable and 1 or more explanatory variables. Most of the research in profile monitoring has been focused on monitoring univariate profiles, while multivariate profile data are widely observed in practice. In this paper, a monitoring approach based on a multivariate Gaussian process (MGP) model is proposed to monitor multivariate profiles simultaneously. In this regard, using a non-separable covariance function, a MGP model is fitted to represent the baseline in-control multivariate profile. Then, the stability of the process is tracked by monitoring a distance measure between the new observations of the multivariate profile and the baseline in-control model. A key advantage of this method is that it considers correlations both within profiles and between profiles. We also introduce, as a benchmark, a univariate Gaussian process-based profile monitoring scheme modified for multivariate profiles. The performance of the proposed approaches is investigated and compared through numerical studies and a real-world case study. The analysis confirms the effectiveness of the MGP-based monitoring scheme for multivariate profiles.

KEYWORDS

multivariate Gaussian process, multivariate profile monitoring, non-separable covariance, statistical process control

1 | INTRODUCTION

Advances in sensor and information technology have made measurement of many process variables easily accessible and enabled unprecedented opportunities for process condition monitoring and control. Statistical process control (SPC) has provided a variety of methods and tools to monitor and control a process. Among various SPC techniques, profile monitoring has drawn significant attention in recent years.[1-5] In this context, the process characteristic under consideration can be represented by a functional relationship between the response variable and 1 or more explanatory variables. This functional relationship is typically referred to as a "profile", and "profile monitoring" generally refers to methods used to detect changes in the profile structure due to assignable causes and to accurately identify their sources.

In many engineering processes that are monitored with multiple sensors, the process condition is characterized by 2 or more profiles, and the response variables of interest are correlated. For instance, consider the force balance calibration in wind tunnel experiments at NASA Langley Research Center.[6] In this example, 3

orthogonal force components representing the response variables and 3 orthogonal torque components as the explanatory variables are measured simultaneously. Another example is the ice making process of an ice machine. For an ice machine, it is known that the performance of the ice making process can be characterized through different temperature signals. Figure 1 shows the behavior of 4 different temperature signals in an ice machine during 3 consecutive ice making cycles. As seen in Figure 1, in each of the panels, different cycles have some similarities as well as some differences. The similarities between the cycles in each panel reflect inherent process characteristics under a specific operational condition, while the differences are due to natural random disturbances. Also, it can be seen that these 4 different signals demonstrate a correlated behavior with the same cycle mostly because all of them are collected from the same physical system. Such correlation may be due to interactions among the signals or due to the fact that all the signals are impacted by the same latent factor, such as the ambient temperature. While the signals in (A), (B), and (D) of Figure 1 exhibit the same or opposite trends and are clearly correlated, the signal in (C) is a bit different and seems to be constant in most of the range. However, by using a formal statistical hypothesis test (such as Pearson's samples correlation coefficient test), one can confirm the existence of correlation between the signal in (c) and other signals of Figure 1. For a process having multiple signal profiles, it is highly desired that these profiles be monitored simultaneously in order to quickly detect process condition changes.

SPC methods, including profile monitoring methods, typically consist of 2 phases, phase I and phase II. In phase I, a set of process data is gathered and analyzed to see if reliable control limits can be established for future observations. In this phase, any unusual patterns in the data lead to adjustment of the process and the search for some assignable causes. Once all of the assignable causes and corresponding data have been eliminated from the data set, we are left with data associated with a stable operating condition which is called the in-control (IC) data set. Using the IC data set, IC process parameters which are representative of the actual process performance are estimated. In phase II, the process is monitored to detect any change in the process using the parameters estimated in phase I. The performance of a phase II monitoring scheme is typically measured in terms of the average run length (ARL) which is the average number of samples taken from the process until the chart signals a change. For a specific monitoring procedure, the IC ARL is often set to a given level. Then, when the process is out of control (OC), the monitoring procedure with a small ARL is considered to have good performance in detecting the specific change in the process. The ARL when the process is IC is referred to as ARL_0 whereas the ARL when the process is OC is referred to as ARL_1.

The existing profile monitoring methods can be broadly classified into 2 categories: parametric methods and non-parametric methods. Many parametric approaches have been proposed in the research literature for profile monitoring using linear regression models. In most of these approaches, the stability of the linear regression model is tracked by monitoring the regression coefficients using the Hotelling's T^2 control chart. For instance, Stover and Brill[7] and Kang and Albin[8] proposed T^2 control charts based on successive vectors of the least squares estimators of the Y-intercept and slope. However, in many real-world applications, linear regression models

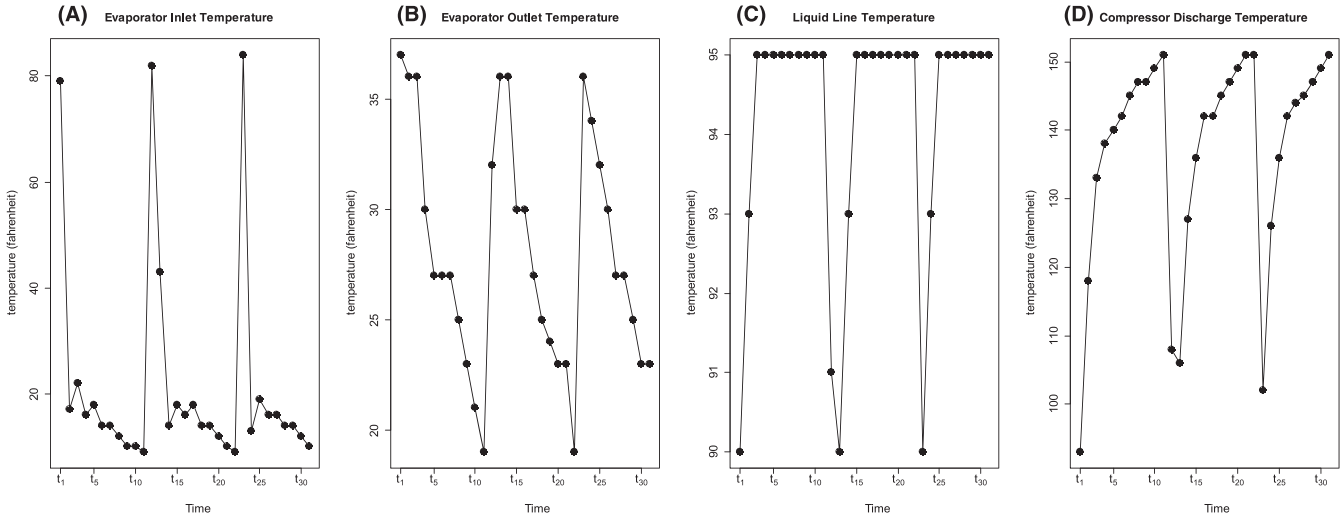

FIGURE 1 Temperature signal profiles for 3 consecutive ice making cycles

are often insufficient to represent the shape of the profiles. In such situations, a non-linear regression model may be better suited to model each profile. For a non-linear regression model, one can monitor the stability of the profile structure by monitoring the coefficients of the regression model as discussed by Williams et al.[3]

In practice, it might be difficult to assume any form of parametric relationship for the observed profiles. This issue might be due to difficulty in determining the correct parametric form of the relationship or because of the fact that the function is inherently non-parametric, ie, it does not have any finite parametric representation.[1] For instance, vertical density profiles (VDPs) data set has been extensively used in the literature.[3,9,10] Considering the VDPS data set, Chang and Yadama[10] proposed a method to identify the mean shifts and shape changes in non-parametric profiles. In their work, a discrete wavelet transformation is applied to separate the noise from the profile contours, and B-splines are used to generate critical points to define the shape of profiles. Applying principal component analysis (PCA), Shiau et al[11] used charts of principal component scores of profiles in monitoring both phase I and phase II. Chicken et al[12] proposed a threshold-based method for determining the difference between the known wavelet coefficients in phase I and the coefficients of the newly observed profiles. All the previously mentioned approaches focus on single variable profile monitoring. One challenge with these methods is how to select the right basis and the right number of basis functions in the non-parametric representation.

Gaussian process (GP) models provide an interesting alternative for non-parametric profile modeling and monitoring and offer several advantages. First, GP models are non-parametric and do not assume any underlying relationship or rigid structure between the inputs and the outputs.[13] Moreover, GP models provide a unique view on prediction errors and give a powerful tool for uncertainty quantification which can be considered as an advantage over the wavelet and PCA methods. Also, GP models can provide a robust estimate of outputs over inputs in the case where there is insufficient sampling data. These features make GP a useful tool for profile monitoring. For instance, a GP model was used by Colosimo et al[14] to model and monitor cylindrical machined surfaces generated by turning operations. Similarly, Cicorella et al[15] proposed using a GP approach to model and monitor complex machined shapes. In these methods, a template profile is used to establish a GP model, and then the newly observed profiles are compared with the model outputs. Zhang et al[4] proposed using a GP model to describe the within-profile autocorrelation (WPA) in a linear profile monitoring problem. They proposed 2 Shewart-type multivariate control charts to monitor the parameters of the GP corresponding to the linear trend and the WPA separately.

In the previously mentioned GP-based non-parametric approaches, only single variable profile is considered, and the problem of simultaneously monitoring multivariate profiles is not addressed. In a study by Zou et al,[16] a multivariate linear profile monitoring approach is developed. They proposed to apply a variable-selection-based multivariate control scheme to monitor multiple linear profiles. One restriction of their method is that it is not applicable to general nonlinear profiles. Paynabar et al[17] proposed a method for analyzing multichannel nonlinear profiles using an uncorrelated multilinear PCA in order to characterize the process variation and perform fault detection and fault diagnosis. Although the method proposed in their paper can effectively analyze the case of multichannel homogenous profile data, it cannot be used for the case of multi-sensor heterogeneous profile data where various sensors measure different variables. The case of monitoring multivariate nonlinear profiles was also examined in a study by Chou et al.[18] In their proposed approach, different profiles are first fitted by B-splines, and then the deviations of the observed profile from the fitted profile are fed into a multivariate EWMA (MEWMA) control scheme for the purpose of process monitoring. Moreover, in a study by Wu et al,[19] the performance of different population covariance matrix estimators for the MEWMA control charts has been investigated. One limitation of these approaches based on the MEWMA chart is that they also do not consider the heterogeneity in the data. Recently, Li et al[20] proposed to fit a multivariate Gaussian Process (MGP) model for multivariate profiles of the IC process in phase I and then established the IC region of the parameters of the MGP. In phase II, an MGP model for the newly observed profiles is fitted, and the fitted model parameters are monitored. The MGP model enjoys the flexibility and desirable analytical properties of the GP along with the ability to capture the correlation among multiple variable profiles. These advantages make MGP an attractive option for the multivariate profile monitoring problem. However, the approach in Li et al[20] has some limitations. First, one needs to fit a new MGP model for every incoming multi-profile sample which makes it difficult to monitor the profiles in phase II in real time, because fitting an MGP model takes a considerable amount of time due to complexity in its likelihood function. Second, it is known that the parameter fitting of GP is, in general, a non-convex optimization problem and the solution could easily get stuck at a local optimum. This means that the

distribution of the fitted parameters of the MGP is very hard to characterize, which will lead to inefficiency in monitoring the coefficients of the process for the profile monitoring problem.

In this paper, we propose a new multivariate profile monitoring chart using the MGP approach. In our approach, using kernel convolution, a non-separable covariance matrix is developed, and the MGP model is fitted to establish the baseline IC structure of the multivariate profile. In phase II, unlike Li et al's approach,[20] our MGP chart is based on monitoring the distance between the new and the baseline IC profile values. This method offers several advantages. First, it does not require fitting of a new MGP model and estimation of parameters for every incoming sample in phase II monitoring. The proposed method can not only significantly reduce the computational load in phase II, but also deal with the issues caused by the uncertainties in parameter fitting in MGP. Furthermore, the flexible non-separable covariance matrix we adopted for the MGP model can effectively address between-profile-correlation and WPA, the 2 main challenges in multivariate profile monitoring problems.

The remainder of this paper is structured as follows. Section 2 provides the mathematical formulation of the multivariate profile monitoring problem. In Section 3, the methodology for monitoring multivariate profiles is introduced. In addition to the MGP-based method, we also introduce a method for monitoring multivariate profiles using separate GP-based univariate charts for comparison purposes. Section 4 presents a numerical study where the performance of the proposed approach is investigated under different scenarios. In Section 5, the application of the proposed methods on a real-life case study is demonstrated. We also compare the performance of the proposed method with the method proposed in Li et al.[20] Finally, our concluding remarks are presented in Section 6.

2 | PROBLEM DESCRIPTION

In this paper, we focus on monitoring general multivariate profiles. In practice, each sample of such multivariate profiles is measured at some specific time points which are referred to as design points in the GP literature. To be more specific, let us assume we have an IC data set of n independent samples of k different profiles, and denote $\boldsymbol{y}_i^{(m)}$ as the m^{th} sample of the ith profile variable, where $m = 1, 2, ..., n$ and $i = 1, 2, ..., k$. Further, we assume there are p_i different design points namely $x_{i,j}$ ($j = 1, 2, ..., p_i$) for the i^{th} profile variable. Thus, we have $\boldsymbol{y}_i^{(m)} = \left[y_i^{(m)}(x_{i,1}), y_i^{(m)}(x_{i,2}), ..., y_i^{(m)}(x_{i,p_i})\right]^T$. It is assumed that when the process is in statistical control, the underlying model for each profile i is as follows:

$$y_i^{(m)}(x_{i,j}) = g_i(x_{i,j}) + \varepsilon_i^{(m)}(x_{i,j}) \tag{1}$$

where g_i is the true underlying function of profile i and $\varepsilon_i^{(m)}$ is the intrinsic noise at the design point $x_{i,j}$. There is no restriction on the form of function g_i, and it can be either linear or non-linear. For the intrinsic noise $\varepsilon_i^{(m)}(x_{i,j})$ within each profile, we allow the possibility that $Corr\left(\varepsilon_i^{(m)}(x_{i,j}), \varepsilon_i^{(m)}(x_{i,j'})\right) > 0$.

The number of design points for each profile i (p_i) is taken to be equal and the explanatory variables $x_{i,j}$ are assumed to be fixed from sample to sample. This is typically referred to as common fixed design in the non-parametric regression context. However, this assumption is not required for the GP-based and MGP-based profile monitoring methods. It is assumed merely for convenience and for comparison with the legacy method that needs this assumption.

Here, we consider the 2 phases of a profile monitoring problem. In phase I, we assume that we have a set of IC profiles data from which the IC process parameters can be estimated. Once the IC model is established as a baseline, in phase II, we want to detect any change in the structure of the profiles. The proposed methodology to model and monitor the multivariate profiles described in Equation (1) is discussed in the next section.

3 | DESCRIPTION OF METHOD FOR MONITORING OF MULTIVARIATE PROFILES

A simple method for monitoring multiple profiles simultaneously is to monitor each profile individually and use a simple rule to determine whether the process is OC. For example, the process can be considered to be OC if one of the profiles indicates an OC situation. In order to compare with this simple method, we will introduce a monitoring scheme based on univariate GP in Section 3.1. The MGP-based monitoring scheme will be described in Section 3.2. The comparison study later shows that the MGP-based method performs better because the cross correlation between different profiles is considered in MGP.

3.1 | Univariate Gaussian process (GP) chart

First, we focus on establishing a GP-based univariate profile monitoring scheme for profile i. A GP model of profile i requires a set of training data containing 1

observation for each design point of profile i; while in phase I analysis, we usually have a set of IC samples for each profile. Because 1 GP model can be estimated from 1 observation of a profile i, we need a way to take advantage of the available multiple IC observations of profile i while estimating the GP model for profile i. In this regard, our approach is based on fitting a GP regression model for each individual IC observation of profile $\left(\boldsymbol{y}_i^{(m)}\right)$ and pooling all n different GP models into 1 representative model.

A GP is, in fact, a collection of random variables with a joint Gaussian distribution.[13] Thus, having p_i observations from sample m of profile i, we can consider them as a sample of a p_i-variate Gaussian distribution. A GP model for sample m of profile i is completely specified by its mean function $\mu_i^{(m)}\left(x_{i,j}\right)$ and covariance function $\mathcal{K}_i^{(m)}\left(x_{i,j},x_{i,j'}\right)$. Thus, the GP regression model for sample m of profile i can be expressed as follows:

$$y_i^{(m)}\left(x_{i,j}\right)=f_i^{(m)}\left(x_{i,j}\right)+\varepsilon_i^{(m)}\left(x_{i,j}\right)$$
$$f_i^{(m)}\left(x_{i,j}\right)\sim GP\left(\mu_i^{(m)}\left(x_{i,j}\right),\mathcal{K}_i^{(m)}\left(x_{i,j},x_{i,j'}\right)\right),\quad \varepsilon_i^{(m)}\sim N\left(0,\sigma^2\right) \tag{2}$$

In this model, $\varepsilon_i^{(m)}$ represents the measurement noise, the mean function $\mu_i^{(m)}\left(x_{i,j}\right)$ is usually taken to be zero, and the model is parameterized by defining a positive definite covariance function. In this study, as in most other practical applications, the covariance function is chosen to be a Gaussian kernel where $\mathcal{K}_i^{(m)}\left(x_{i,j},x_{i,j'}\right)=\rho_{i,m}^2 exp\left\{-\lambda_{i,m}\left(x_{i,j}-x_{i,j'}\right)^2\right\}$.[4,13-15] Here $\rho_{i,m}^2$ is defined as the maximum allowable covariance and $\lambda_{i,m}$ is the shape parameter for sample m of profile i. Considering the Gaussian covariance function, the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{i}}^{(\boldsymbol{m})}$ for the observations of profile i can be constructed. Moreover, having estimated the hyper-parameters of the covariance function through maximizing a log-likelihood function, the prediction for each design point j of profile i $\left(\widehat{y}_i^{(m)}\left(x_{i,j}\right)\right)$ can be calculated considering the following distribution:

$$\widehat{y}_i^{(m)}\left(x_{i,j}\right)|\boldsymbol{y}_i^{(m)}\sim\mathcal{N}\left(\boldsymbol{\eta}_{i,m}^T\left(x_{i,j}\right)\left(\boldsymbol{\Sigma}_i^{(m)}\right)^{-1}\boldsymbol{y}_i^{(m)},\Sigma_i^{(m)}\left(x_{i,j},x_{i,j}\right)-\boldsymbol{\eta}_{i,m}^T\left(x_{i,j}\right)\left(\boldsymbol{\Sigma}_i^{(m)}\right)^{-1}\boldsymbol{\eta}_{i,m}\left(x_{i,j}\right)\right) \tag{3}$$

where, $\boldsymbol{\eta}_{i,m}\left(x_{i,j}\right)=\left[\Sigma_i^{(m)}\left(x_{i,j},x_{i,1}\right),\Sigma_i^{(m)}\left(x_{i,j},x_{i,2}\right),...,\Sigma_i^{(m)}\left(x_{i,j},x_{i,p_i}\right)\right]^T$ and $\Sigma_i^{(m)}\left(x_{i,j},x_{i,j}\right)=\mathcal{K}_i^{(m)}\left(x_{i,j},x_{i,j}\right)+\sigma^2$. More details about the development of a Gaussian process regression model can be found in Rasmussen and Williams.[13]

Using the earlier method, we can fit a GP model for each observation m of profile i. As mentioned earlier, the GP model gives us the covariance matrix $\left(\boldsymbol{\Sigma}_i^{(m)}\right)$ as well as the vector of predictions $\left(\widehat{\boldsymbol{y}}_i^{(m)}\right)$ in design points for each sample m of profile i. As the next step, we need to estimate the IC profile i by pooling all fitted GP models. In this regard, for each profile i, the final vector of predicted values in its design points $\left(\widehat{\boldsymbol{y}}_i^*\right)$ can be estimated by averaging over their multiple counterparts as follows:

$$\widehat{\boldsymbol{y}}_i^*=\frac{1}{n}\sum_{m=1}^{n}\widehat{\boldsymbol{y}}_i^{(m)}=\boldsymbol{A}_i\sum_{m=1}^{n}\widehat{\boldsymbol{y}}_i^{(m)} \tag{4}$$

where $\boldsymbol{A}_i=\frac{1}{n}\boldsymbol{I}_{p_i\times p_i}$ is a linear transform from the individual samples of profile i to the vector of final predictions and $\boldsymbol{I}_{p_i\times p_i}$ is the identity matrix with dimensions $p_i\times p_i$. For this vector of predictions, the following properties can be derived:

$$\mathbb{E}\left(\widehat{\boldsymbol{y}}_i^*\right)=\mathbb{E}\left(\boldsymbol{A}_i\sum_{m=1}^{n}\widehat{\boldsymbol{y}}_i^{(m)}\right)=\boldsymbol{A}_i\,\mathbb{E}\left(\sum_{m=1}^{n}\widehat{\boldsymbol{y}}_i^{(m)}\right)=\boldsymbol{A}_i\sum_{m=1}^{n}\widehat{\boldsymbol{y}}_i^{(m)} \tag{5}$$

$$\boldsymbol{\Sigma}_i^*=\mathrm{var}\left(\widehat{\boldsymbol{y}}_i^*\right)=\boldsymbol{A}_i\,\mathrm{var}\left(\sum_{m=1}^{n}\widehat{\boldsymbol{y}}_i^{(m)}\right)\boldsymbol{A}_i^T=\boldsymbol{A}_i\left(\sum_{m=1}^{n}\boldsymbol{\Sigma}_i^{(m)}\right)\boldsymbol{A}_i^T \tag{6}$$

Regarding the above 2 properties, we can estimate the IC vector of predicted values $\left(\widehat{\boldsymbol{y}}_i^*\right)$ as well as the covariance matrix $\left(\boldsymbol{\Sigma}_i^*\right)$ of profile i. Moreover, it is guaranteed that using the above linear transform, the resulting covariance matrix is positive definite.

Having estimated the required process parameters of each profile i using the GP regression model, a monitoring scheme is required for the incoming profiles in phase II analysis. Moreover, we can only establish separate univariate charts because the GP approach can only model the univariate profiles separately. In this regard, the following charting statistic is proposed for each new sample r of profile i.

$$T_{GP_{i,r}}^2=\left(\boldsymbol{y}_i^{(r)}-\widehat{\boldsymbol{y}}_i^*\right)T\left(\boldsymbol{\Sigma}_i^*\right)-1\left(\boldsymbol{y}_i^{(r)}-\widehat{\boldsymbol{y}}_i^*\right) \tag{7}$$

where $\boldsymbol{y}_i^{(r)}=\left[y_i^{(r)}\left(x_{i,1}\right),y_i^{(r)}\left(x_{i,2}\right),...,y_i^{(r)}\left(x_{i,p_i}\right)\right]^T$ is the vector containing the new observations of profile i, $\widehat{\boldsymbol{y}}_i^*$ is the vector of estimated predicted values of profile i and $\boldsymbol{\Sigma}_i^*$ is its estimated covariance matrix. This individual chart triggers an OC signal if $T_{GP_{i,r}}^2>L_{GP_i}$, where $L_{GP_i}>0$ is the control limit chosen to satisfy

$$\Pr\left(T_{GP_{i,r}}^2>L_{GP_i}\right)=\alpha_i$$

The value α_i is called the false alarm rate. If the observations of profiles are independent from each other, then $ARL_{0_i} = \frac{1}{\alpha_i}$. Thus, by selecting an appropriate value of α_i, we can achieve a specific ARL_{0_i} value for the individual GP chart i. For a given value of α_i, the control limit (ie, L_{GP_i} here) is often determined through simulation. A general simulation procedure can be found in Qiu et al[21] and is also described in the following section where we determine the control limit for the MGP-based monitoring scheme. In the above section, we have described how to establish a univariate GP chart for a single variable profile. If there are multiple variable profiles in the process, we can establish multiple univariate GP charts accordingly. However, because we are using multiple charts as a system to monitor the process, the overall false alarm rate of the system, denoted as α, will be different from each individual false alarm rate α_i. If the profile variables are independent from each other, the relationship between α and α_i s could be simple as described in Montgomery[22]:

$$\alpha = 1 - \prod_{i=1}^{k} (1 - \alpha_i) \tag{8}$$

Here, once we select a value of α, we can use the above equation and assume all α_i s are the same to compute the α_i value. However, if the profile variables are not independent, the relationship between α and α_is could be complex and hard to identify. In this case, we can use the above equation as an approximation. This is actually 1 weakness of using multiple univariate charts to monitor multivariate profiles. In the next section, we introduce an MGP-based method for monitoring multivariate profiles simultaneously.

3.2 | Multivariate Gaussian process chart

A MGP model is developed using mechanisms that can simultaneously predict different outputs considering their dependencies. The covariance matrix of a MGP model is defined by a function which considers both between-profile-correlation and WPA and can be either separable or non-separable. In a separable structure, the same set of parameters is estimated for all of the variables and the correlations between different outputs are forced. However, in a non-separable structure, different outputs have different correlations which allow the model to infer the commonalities and differences between different outputs and provide more flexible predictions. Thus, the non-separable covariance matrices, in which different outputs have shared and independent features, lead to more flexibility in defining the covariance structure over the separable ones. In this paper, a MGP model based on the non-separable covariance structure is proposed to model and monitor the multivariate profiles.

A MGP model, similar to the GP model, can be applied in the case that there is only 1 sample of each profile. However, as mentioned earlier, there are n samples of different profiles in phase I from which the IC process parameters need to be estimated. Therefore, here we need an approach to take advantage of all the IC profiles in phase I to accurately estimate 1 representative MGP model. In this regard, the proposed approach is based on fitting the MGP model for each IC sample m of different profiles $\left(\boldsymbol{y}^{(m)} = \left[\boldsymbol{y}_1^{(m)^T}, \boldsymbol{y}_2^{(m)^T}, ..., \boldsymbol{y}_I^{(m)^T}\right]^T\right)$ and then pooling all of them into 1 final model.

One common way to construct a non-separable MGP model is based on the kernel convolution[23] which is the alternative way of constructing a GP model. To be more specific, a GP model can also be constructed by convolving the GP random variables with an arbitrary kernel. Defining $s \in \mathcal{R}^D$ as the spatial region over which the design points are considered, $f_i^{(m)}(s)$ in Equation (2) can be constructed through convolving a smoothing kernel $\mathcal{K}_i^{(m)}(s)$ with a continuous Gaussian white noise process $z_i^{(m)}(s)$. The resulting white noise process $f_i^{(m)}(s)$ can be defined as $cov\left(z_i^{(m)}(s), z_i^{(m)}(s')\right) = \tau_{ss'}$ where τ is the Dirac Delta function.

$$f_i^{(m)}(s) = \int_{R^D} \mathcal{K}_i^{(m)}(s-u) z_i^{(m)}(u) du \tag{9}$$

The resulting covariance function for $f_i^{(m)}(s)$ can be derived as follows:

$$cov\left(f_i^{(m)}(s), f_i^{(m)}(s')\right) = \int_{R^D} \mathcal{K}_i^{(m)}(s-u) \mathcal{K}_i^{(m)}(s'-u) du \tag{10}$$

Therefore, considering the same procedure and a set of k outputs, the general decomposition of the MGP model for the profile i of sample m can be written as follows:

$$\begin{aligned} y_i^{(m)}(s) &= f_i^{(m)}(s) + \varepsilon_i^{(m)}(s) \qquad i = 1, 2, ..., k, \\ & m = 1, 2, ..., n \end{aligned} \tag{11}$$

where $f_i^{(m)}(s)$ is the process constructed through kernel convolution and $\varepsilon_i^{(m)}(s)$ is the noise term. The final objective of a non-separable MGP model is to derive processes $\left\{f_i^{(m)}(s)\right\}_{i=1}^{k}$ for each sample m that are dependent upon some common latent processes $\left\{z_u^{(m)}(s)\right\}_{u=1}^{U}$. Here, in the multivariate profile monitoring setting, we consider the case where each profile $y_i^{(m)}(s)$ is constructed by the sum of 3 stationary Gaussian processes. The first process is the convolution of a smoothing kernel with a unique latent function $z_i^{(m)}(s)$, the second process originates from the

convolution of the output and a common latent function $z_0^{(m)}(s)$, and the third one is the measurement noise $\varepsilon_i^{(m)}(s)$. Thus, for each profile i of sample m, we have:

$$y_i^{(m)}(s) = \mathcal{K}_{ii}^{(m)}(s)*z_i^{(m)}(s) + \mathcal{K}_{0i}^{(m)}(s)*z_0^{(m)}(s) + \varepsilon_i^{(m)}(s) \quad i = 1, 2, ..., k, \quad m = 1, 2, ..., n \tag{12}$$

where * defines the kernel convolution and $\mathcal{K}_{ui}$ is the kernel connecting the latent function $z_u^{(m)}$ $(u = 0, i)$ to the output $y_i^{(m)}$. Regarding the above formulation, we have successfully considered the independent features of each output in sample m by convolving over unique latent functions, as well as the shared features among all outputs through the common dependency on $z_0^{(m)}(s)$. Thus, this structure provides enough flexibility for the multivariate profile monitoring procedure to extract the shared information among different profiles. Figure 2 demonstrates the MGP convolution structure described earlier.

As the next step, we need to choose a kernel function, and then the MGP model would be fully parametrized through its hyper-parameters. Here, the Gaussian kernel is utilized to model the structure in the MGP. The Gaussian kernels do not require many parameters to be estimated which makes them a perfect choice for the scenario in which there is scarcity of data.[24] Also, their ability to model different spatial features and achieve reasonable performance in different applications has made them a common choice in the literature.[4,14,25] Thus, considering Equation (11), the Gaussian kernel used in this study is as follows:

$$\mathcal{K}_{ui}(s) = \frac{\rho_{ui}\sqrt{|L_{ui}|}}{\sqrt[4]{\pi^D}} exp\left(-\frac{1}{2}(s-\mu_{ui})^T L_{ui}(s-\mu_{ui})\right) \tag{13}$$

where $s, \mu \in R^D$ and L_{ui} is a $D \times D$ positive definite matrix. Following the kernel convolution discussed in Ver Hoef and Barry[26] and Kontar et al,[25] the auto and cross covariance functions considering the kernel in Equation (13) can be derived as follows:

$$\begin{aligned}
\Sigma_{ii}^{(m)}(s,s') &= cov_{ii}^{(m)}(s,s') \\
&= \rho_{ii,m}^2 exp\left(-\frac{1}{4}\times\frac{d^2}{L_{ii,m}^2}\right) + \rho_{0i,m}^2 exp\left(-\frac{1}{4}\times\frac{d^2}{L_{0i,m}^2}\right) + \sigma_{i,m}^2\delta_{ss'} \\
\Sigma_{ij}^{(m)}(s,s') &= cov_{ij}^{(m)}(s,s') \\
&= \rho_{0i,m}\rho_{0j,m}\sqrt{\frac{2|L_{0i,m}L_{0j,m}|}{L_{0i,m}^2 + L_{0j,m}^2}} exp\left(-\frac{1}{2}\times\frac{d^2}{L_{0i,m}^2 + L_{0j,m}^2}\right) \\
& j, i = 1, 2, ..., k \quad and \quad m = 1, 2, ..., n
\end{aligned} \tag{14}$$

where $d = s - s'$, $\delta(.)$ is the Kronecker delta function and $\sigma_{i,m}^2$ is the variance of $\varepsilon_i^{(m)}(s)$ corresponding to the noise component of the ith profile in sample m. The detailed derivation of Equation (14) can be found in Appendix 1. The covariance matrix of the joint Gaussian process of all k profiles in sample m can be constructed as follows:

$$\mathbf{\Sigma}^{(m)} = \begin{bmatrix} \mathbf{\Sigma}_{11}^{(m)} & \cdots & \mathbf{\Sigma}_{1k}^{(m)} \\ \vdots & \ddots & \vdots \\ \mathbf{\Sigma}_{k1}^{(m)} & \cdots & \mathbf{\Sigma}_{kk}^{(m)} \end{bmatrix} \tag{15}$$

The reliability of the MGP model depends upon the correct estimation of its hyper-parameters. Considering $\theta = \{\rho, L, \sigma\}$ as the set of hyper-parameters in the above covariance matrix, the log likelihood function can be expressed as follows:

$$l\left(\theta|\boldsymbol{y}^{(m)}\right) = -\frac{1}{2}\left(\boldsymbol{y}^{(m)}\right)^T\left(\mathbf{\Sigma}^{(m)}\right)^{-1}\boldsymbol{y}^{(m)} - \frac{1}{2}\log|\mathbf{\Sigma}^{(m)}| - \frac{\sum_{i=1}^{k} p_i}{2}\log(2\pi) \tag{16}$$

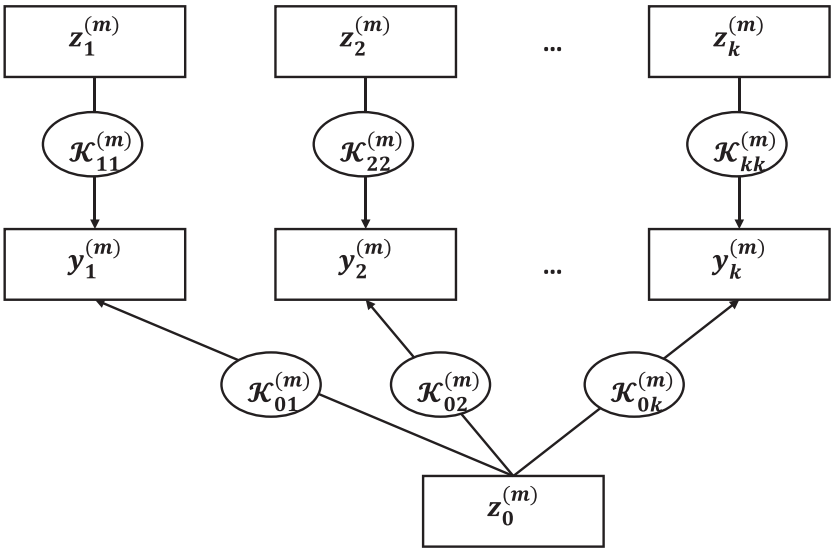

FIGURE 2 MGP convolution structure

Maximizing the above equation to get θ^*, the predictive distribution for the design point j of ith profile in sample m can be expressed as follows:

$$\widehat{y}_i^{(m)}(x_{i,j})|\boldsymbol{y}^{(m)} \sim \mathcal{N}\left(\boldsymbol{\eta}_{i,m}^T(x_{i,j})\left(\boldsymbol{\Sigma}^{(m)}\right)^{-1}\mathbf{y}^{(m)}, \Sigma_{ii}^{(m)}(x_{i,j},x_{i,j}) - \boldsymbol{\eta}_{i,m}^T(x_{i,j})\left(\boldsymbol{\Sigma}^{(m)}\right)^{-1}\boldsymbol{\eta}_{i,m}(x_{i,j})\right) \tag{17}$$

where $\boldsymbol{y}^{(m)} = \left[\boldsymbol{y}_1^{(m)T}, \boldsymbol{y}_2^{(m)T}, ..., \boldsymbol{y}_k^{(m)T}\right]^T$, $\boldsymbol{\eta}_{i,m}(x_{i,j}) = \left[\boldsymbol{\eta}_{i1}^T(x_{i,j}), ..., \boldsymbol{\eta}_{ik}^T(x_{i,j})\right]^T$ and $\boldsymbol{\eta}_{ii'}(x_{i,j}) = \left[\Sigma_{ii'}^{(m)}\left(x_{i,j}, x_{i',1}\right), \Sigma_{ii'}^{(m)}\left(x_{i,j}, x_{i',2}\right), ..., \Sigma_{ii'}^{(m)}\left(x_{i,j}, x_{i',p_{i'}}\right)\right]^T$. It should be noted that the number of parameters to be estimated from the log-likelihood function is $5k$ and it increases linearly with the number of variables. In many practical cases, the number of variables that needs to be monitored simultaneously is relatively small. Therefore, the problem we are dealing with is generally not highly complex. Also, one should note that according to Rasmussen and Williams[13] and many others, the predictions based on GP are robust to its parameters' estimate because of the Bayesian interpretation. In other words, even if the estimates of the Gaussian process parameters are sub-optimal, the prediction of the GP will be stable and relatively accurate. More details about the kernel convolution and development of a MGP model can be found at Higdon,[23] Ver Hoef and Barry,[26] and Kontar et al.[25]

The main advantage of the non-separable MGP structure as derived above is that it models the behavior of each profile considering its unique features along with the features that are shared among all the profiles. This learning structure is essential in multivariate profile monitoring where different profiles have commonalities as well as unique behaviors.

The above procedure can be applied on every sample m of the set of IC samples in phase I. Then, we need to pool all of the developed MGP models to estimate 1 representative MGP model of the IC process. In this regard, like the GP approach, the final vector of predicted values ($\widehat{\boldsymbol{y}}^*$) can be estimated by averaging over its multiple counterparts as follows:

$$\widehat{\boldsymbol{y}}^* = \frac{1}{n}\sum_{m=1}^{n}\widehat{\boldsymbol{y}}^{(m)} = \boldsymbol{B}\sum_{m=1}^{n}\widehat{\boldsymbol{y}}^{(m)} \tag{18}$$

where $= \frac{1}{n}\boldsymbol{I}_{P\times P}$, $P = \sum_{i=1}^{k} p_i$, and $\boldsymbol{I}_{P\times P}$ represents the identity matrix with dimension $P \times P$. Here, $\boldsymbol{B}$ is a linear transform from the individual samples to the vector of final predictions. For this vector of predictions, the following properties can be derived:

$$\mathbb{E}\left(\widehat{\boldsymbol{y}}^*\right) = \mathbb{E}\left(\boldsymbol{B}\sum_{m=1}^{n}\widehat{\boldsymbol{y}}^{(m)}\right) = \boldsymbol{B}\,\mathbb{E}\left(\sum_{m=1}^{n}\widehat{\boldsymbol{y}}^{(m)}\right) = \boldsymbol{B}\sum_{m=1}^{n}\widehat{\boldsymbol{y}}^{(m)} \tag{19}$$

$$\boldsymbol{\Sigma}^* = \mathrm{var}\left(\widehat{\boldsymbol{y}}^*\right) = \boldsymbol{B}\,\mathrm{var}\left(\sum_{m=1}^{n}\widehat{\boldsymbol{y}}^{(m)}\right)\boldsymbol{B}^T = \boldsymbol{B}\left(\sum_{m=1}^{n}\boldsymbol{\Sigma}^{(m)}\right)\boldsymbol{B}^T \tag{20}$$

Regarding the above 2 properties, we can estimate the IC vector of predicted values ($\widehat{\boldsymbol{y}}^*$) along with the positive definite covariance matrix ($\boldsymbol{\Sigma}^*$) of the MGP.

The IC process parameters in phase I can be estimated using the above procedure. Next, a procedure is required to monitor the stability of the incoming profiles in phase II. In this regard, we propose using the following charting statistic for each new sample r of the process.

$$T_{MGP_r}^2 = \left(\boldsymbol{y}^{(r)} - \widehat{\boldsymbol{y}}^*\right)^T \boldsymbol{\Sigma}^{*-1}\left(\boldsymbol{y}^{(r)} - \widehat{\boldsymbol{y}}^*\right) \tag{21}$$

where $\boldsymbol{y}^{(r)} = \left[\boldsymbol{y}_1^{(r)T}, \boldsymbol{y}_2^{(r)T}, ..., \boldsymbol{y}_k^{(r)T}\right]^T$ is the vector containing the new observations of all profiles, $\widehat{\boldsymbol{y}}^* = \left[\widehat{\boldsymbol{y}}_1^{*T}, \widehat{\boldsymbol{y}}_2^{*T}, ..., \widehat{\boldsymbol{y}}_k^{*T}\right]^T$ is the vector of estimated MGP predictions for all of the profiles and $\boldsymbol{\Sigma}^*$ is the estimation of the MGP covariance matrix. The above chart triggers an OC signal if

$$T_{MGP_r}^2 > L_{MGP}$$

where the control limit (L_{MGP}) should be chosen to achieve a specified overall ARL_0 for the whole process.

In the SPC literature, it is conventionally assumed that the IC distribution of the process measurements is known. Thus, the control limit can be searched using simulation based on this distribution. However, the IC distribution is often unknown in practice and instead there is a large IC data set. In this situation, the control limit can be searched using a resampling algorithm. In this regard, in each run of the

simulation, one can resample with replacement from the IC data set until the control chart signals an OC condition. Then, the ARL_0 value can be computed based on multiple runs of this simulation. The control limit is set iteratively to finally get a nominal value for the ARL_0. In this study, 10 000 simulation runs are considered to compute the ARL values.

The design and implementation of the MGP chart can be summarized in the following steps.

1. Fit an MGP regression model for every sample m of IC observations using the procedure discussed in this section. The MGP model gives the vector of estimated predicted values as well as the estimated covariance matrix for each sample m.
2. Estimate the IC process by averaging the vectors of estimated predicted values of IC samples obtained in step 1. Using this method, the IC vector of predicted values along with the covariance matrix of the final MGP model can be estimated as described in Equations (19) and (20).
3. Monitor the process by monitoring the $T^2_{MGP_r}$ statistic described in Equation (21) for every new sample r of the process. The incoming samples are IC if their corresponding $T^2_{MGP_r}$ statistics are less than a prespecified control limit L_{MGP}.

The MGP chart proposed here offers a key advantage. Specifically, it is established based on the large covariance matrix which captures the cross-correlations between different profiles in each sample along with the correlation within each profile. This feature is crucially important in most applications where different sensors monitor a common underlying process and the measurements are made in consecutive time intervals.

4 | NUMERICAL STUDY

In this section, the performance of the multivariate profile monitoring procedures proposed above will be investigated. First, we start with the scenario that 2 signals are monitored simultaneously where the intrinsic noise in Equation (1) is assumed to be independent and identically distributed. Then, the existence of the autocorrelation in the noise in the 2-profile case will be examined. We will also consider the situation where we are monitoring more than 2 profile variables. Moreover, the case of shift in between profile correlation will be investigated.

For the sake of completeness, in the simulation study, we will compare the performance of the proposed MGP chart with the conventional T^2 chart as shown in Mestek et al[27] for monitoring each univariate non-parametric profile. For monitoring each new sample r of the profile i, we have:

$$T^2_{i,r} = \left(\boldsymbol{y}_i^{(r)} - \bar{\boldsymbol{y}}_i\right) T \boldsymbol{S}_i^{-1} \left(\boldsymbol{y}_i^{(r)} - \bar{\boldsymbol{y}}_i\right) \tag{22}$$

where $\boldsymbol{y}_i^{(r)}$ is a vector containing the response values of rth new sample of profile i and $\bar{\boldsymbol{y}}_i = \left[\bar{y}_i(x_{i,1}), \bar{y}_i(x_{i,2}), ..., \bar{y}_i(x_{i,p_i})\right]^T$ is a vector of IC response averages over the n IC samples of profile i in its design points. Moreover, $\boldsymbol{S}_i$ is the pooled sample covariance matrix of profile i.This individual T^2 control chart triggers an OC signal for sample r if

$$T^2_{i,r} > L_i$$

where $L_i > 0$ is the control limit chosen to achieve a specific ARL_{0_i} value for the individual T^2 control chart i. Here, as with the GP control chart, the control limit of each individual T^2 chart should be chosen to achieve a specified overall ARL_0 value for the system. It is worth mentioning that this chart typically requires a large number of IC observations in order to give an accurate estimate of the covariance matrix. Moreover, it should be noted that using the conventional T^2 chart has 2 limitations. First, the number of IC samples (n) should be more than the number of design points p_i; otherwise, the sample covariance matrix of profile i would be singular. Also, it requires fixed design points to construct the covariance matrix and monitor the incoming profiles.

In Section 4.5, we will also consider an "aggregated" T^2 chart based on stacking all variables of each observation as follows:

$$T^2_r = \left(\boldsymbol{y}^{(r)} - \bar{\boldsymbol{y}}\right)^T \boldsymbol{S}^{-1} \left(\boldsymbol{y}^{(r)} - \bar{\boldsymbol{y}}\right) \tag{23}$$

where $\bar{\boldsymbol{y}} = \left[\bar{\boldsymbol{y}}_1^T, \bar{\boldsymbol{y}}_2^T, ..., \bar{\boldsymbol{y}}_k^T\right]^T$. Here, $\boldsymbol{S}$ is the pooled sample covariance matrix of all the profiles. The control limit for this chart can also be chosen to achieve a specific ARL_0 value using the procedure discussed previously.

It should be noted that the IC ARL is set to 370 in the simulation study and all the ARL values are reported based on the average of 10 000 simulations. Also, as indicated in the previous sections, both GP and MGP charts have enough flexibility to observe different design points for different profiles; however, the T^2 charts can only be used for the case of fixed design points. Thus, the design points are assumed to be fixed and same in different profiles of a sample ($p_i = p$) in order to be able to compare the performance of the different profile monitoring schemes.

TABLE 1 ARL_1 comparison of the 3 charts in the trigonometric model

		$n = 31$			$n = 50$			$n = 100$		
	Shift	**GP**	**MGP**	T^2	**GP**	**MGP**	T^2	**GP**	**MGP**	T^2
Mean	0.1	341.3	278.3	355.6	244.8	234.5	318.5	242.6	230.6	109.6
	0.2	184.1	117.1	302.0	117.6	87.2	144.2	115.5	80.1	67.9
	0.3	69.0	33.1	217.2	41.7	23.6	48.8	39.7	21.1	13.0
	0.4	20.2	8.4	138.2	13.1	6.0	15.2	15.7	6.0	5.0
	0.5	6.1	2.6	84.5	4.2	2.1	5.5	4.1	2.0	2.0
Noise	0.01	215.9	178.1	271.0	184.0	156.5	189.0	172.7	147.4	175.6
	0.02	133.0	93.5	191.7	120.0	89.5	185.9	107.1	87.2	146.9
	0.03	78.2	51.2	188.5	64.3	48.7	149.9	64.1	45.2	98.9
	0.04	49.4	31.6	157.7	42.0	30.1	112.9	41.0	30.8	67.2
	0.05	32.8	19.9	130.1	27.6	18.7	86.2	25.3	17.8	48.2
Segment	0.1	392.8	346.7	358.4	319.3	314.1	340.6	305.6	299.5	315.8
	0.2	320.9	266.4	296.7	231.6	219.3	267.6	229.7	205.0	222.3
	0.3	186.0	137.0	197.5	146.9	116.7	193.4	142.3	114.2	122.7
	0.4	121.0	80.5	214.6	85.4	58.2	84.1	84.3	54.6	61.5
	0.5	65.0	34.8	176.6	45.1	24.8	41.2	42.6	24.7	29.5

4.1 | Two-variable profile monitoring

This section investigates the performance of the proposed profile monitoring schemes for 2-variable profiles. To illustrate the performance of different charts, our simulations cover 2 models: a trigonometric model and a quadratic model. The simulation study is based on the 2 phases of SPC as mentioned earlier. In each model setting, n IC profiles are generated in phase I. Then, using the procedure for each chart, the IC process parameters are estimated, and the corresponding chart is established to monitor the in-coming profiles in phase II. For each model, 3 values of n are considered. This includes 1 sparse case with the least possible number of IC samples allowed by the T^2 chart, another case with moderate number of IC samples, and a dense case with large number of IC samples. For each model, 3 types of OC situations are considered. The first OC model investigates the performance of the proposed charts in the situation where a mean shift in the process has occurred. The second OC model deals with the shift in the noise of the process. In the third OC model, it is assumed that there is a mean shift in a segment of the process.

First, in the trigonometric model setting, we assume that the underlying process is characterized by the following 2 signals:

$$\begin{aligned} y_1^{(m)}(x_{1,j}) &= 5 + 2\cos(x_{1,j}) + \varepsilon_1^{(m)}(x_{1,j}) \qquad m = 1, 2, ..., n \\ y_2^{(m)}(x_{2,j}) &= -5 + 2\sin(x_{2,j}) + \varepsilon_2^{(m)}(x_{2,j}) \qquad m = 1, 2, ..., n \end{aligned} \tag{24}$$

where $\sigma_1 = \sigma_2 = 0.5$. In this model, it is assumed that 30 design points are equally distributed over $[0, 2\pi]$. Table 1 presents the OC ARL of 3 charts in the case of mean shift, noise shift, and segment shift in this model for different number of IC observations (n). For the segment shift, it is considered that a mean shift occurred in the middle segment of each profile, as displayed in Figure 3.

Results from Table 1 show that in the sparse setting with $n = 31$ IC samples, the MGP chart outperforms both

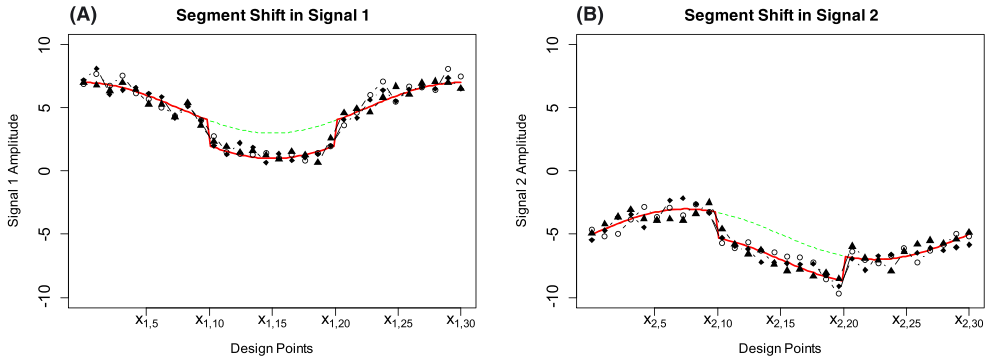

FIGURE 3 Segment shift in the trigonometric model [Colour figure can be viewed at wileyonlinelibrary.com]

the GP and T^2 charts in different shifts in the mean of the process. Moreover, considering mean shift, the T^2 chart's performance is worse than the GP chart in this setting. As mentioned earlier, the T^2 chart requires a high number of IC samples in phase I in order to accurately estimate the covariance matrix; therefore, we would expect these results. Table 1 also shows the trend in the performance of these charts in different mean shift situations with increase in the number of IC samples (n). The results show that, generally, an increase in the number of IC samples (n) leads to an improvement in the performance of the charts in terms of ARL_1 in different shifts in the mean of process. Specifically, the performance of the T^2 chart considerably improves with the increase in the number of IC samples to the extent that in the case of $n = 100$ IC samples, it performs better than the other charts in all of the mean shift situations.

From Table 1, it can be seen that in the situation where there is a shift in the noise of the process, the MGP chart consistently performs better than the GP and T^2 charts. Moreover, it can be seen that the GP chart's performance is relatively better than the T^2 chart. This is, in fact, due to the power of GP-based models in uncertainty quantification and capturing the cross-correlation between the 2 profiles. In this OC situation, as it can be seen from Table 1, increasing n improves the performance of all charts; nevertheless, the MGP chart outperforms other charts in different cases with high number of IC samples. Further, when only a segment of the profiles has shifted, the MGP chart outperforms the GP and T^2 charts in different amounts of shift. Here also, although increasing n improves the performance of the charts, the MGP chart's performance is relatively better than the others.

Next, we consider a quadratic setting in which the underlying process is characterized with 2 signals as follow.

$$\begin{aligned} y_1^{(m)}(x_{1,j}) &= 1 + 5x_{1,j} + x_{1,j}^2 + \varepsilon_1^{(m)}(x_{1,j}) \quad m = 1, 2, ..., n \\ y_2^{(m)}(x_{2,j}) &= 2x_{2,j}^2 + \varepsilon_2^{(m)}(x_{2,j}) \quad m = 1, 2, ..., n \end{aligned} \tag{25}$$

where $\sigma_1 = \sigma_2 = 1$. The measurements are made in 10 design points equally over $[0, 10]$. The OC ARL of 3 charts in the case of mean shift, noise shift, and segment shift in this model for different number of IC observations (n) are summarized in Table 2. Regarding the segment shift, here also, it is assumed that a mean shift occurred in the middle segment of each profile.

In the quadratic model, also, the MGP chart outperforms the other charts in the sparse setting when dealing with the mean shift in the process. Moreover, in this setting, the GP chart's performance is better than that of the T^2 chart. As expected, increasing the number of IC samples (n) in phase I improves the performance of the charts, to the extent that the T^2 chart outperforms the other charts in a dense setting with $n = 50$ IC samples in different mean shift situations.

Table 2, furthermore, confirms the superiority of the GP-based charts in detecting the noise shift in different settings with different number of IC samples. Especially, it can be seen that the MGP chart which considers the cross-correlation structure of 2 profiles performs relatively better than the GP chart. Regarding segment shift, the similar trend as the trigonometric model can be seen where the MGP chart outperforms the other charts.

Next, the effect of the number of design points on the performance of the 3 monitoring procedures is

TABLE 2 ARL_1 comparison of the 3 charts in the quadratic model

		$n = 11$			$n = 30$			$n = 50$		
	Shift	**GP**	**MGP**	T^2	**GP**	**MGP**	T^2	**GP**	**MGP**	T^2
Mean	0.6	122.0	81.9	339.9	119.3	80.1	208.5	120.1	80.3	51.1
	0.7	81.7	45.4	334.1	79.8	45.3	154.9	77.3	44.5	31.7
	0.8	53.9	29.9	318.9	52.7	29.6	89.3	51.6	28.0	19.7
	0.9	35.9	18.2	303.8	35.3	17.7	58.9	35.1	16.8	12.5
	1	22.7	10.6	283.0	21.9	10.3	23.3	20.9	10.2	8.1
Noise	0.02	265.4	242.6	263.6	248.5	235.0	259.1	248.7	233.5	240.6
	0.04	178.9	156.6	222.1	169.5	154.7	209.0	168.6	129.2	206.4
	0.06	128.0	109.1	179.7	122.9	107.7	162.0	121.4	104.5	133.7
	0.08	95.0	76.1	145.1	91.4	75.6	125.1	91.9	75.4	102.9
	0.1	70.7	56.4	122.7	70.3	54.8	116.8	70.7	54.8	78.4
Segment	0.6	182.3	143.0	422.9	112.9	109.3	386.9	94.6	90.6	276.9
	0.7	148.1	105.2	390.3	128.6	85.7	278.1	61.9	55.4	213.3
	0.8	114.7	77.8	344.4	64.1	34.2	222.2	39.3	32.8	170.7
	0.9	86.5	54.2	299.2	47.6	23.6	176.1	25.0	19.5	126.1
	1	67.1	38.7	261.8	35.6	16.6	132.9	16.3	11.8	90.1

TABLE 3 The effect of number of design points in the trigonometric model

Trigonometric Model with $n = 100$ IC Samples							
		30 design points			50 design points		
	Shift	GP	MGP	T^2	GP	MGP	T^2
Mean	0.1	242.6	230.6	109.6	230.7	225.1	278.5
	0.2	115.5	80.1	67.9	106.4	76.7	112
	0.3	39.7	21.1	13.0	32.1	16.8	32.1
	0.4	15.7	6.0	5.0	8.2	3.8	8.2
	0.5	4.1	2.0	2.0	2.5	1.5	2.7
Noise	0.01	172.7	147.4	171.6	165.8	146.8	238.5
	0.02	107.1	87.2	146.9	87.1	66.3	155.6
	0.03	64.1	45.2	98.9	47.0	33.4	102.4
	0.04	41.0	30.8	67.2	27.6	18.3	67.1
	0.05	25.3	17.8	48.2	17.0	10.7	47.5

investigated. In this regard, for each model, the setting with large number of IC samples is considered. For this setting, we increase the number of design points in both trigonometric and quadratic models. In order to obtain a fair comparison, we focus on the OC model with shift in the mean and noise of the process. Tables 3 and 4 show the results of this analysis for the trigonometric and the quadratic models.

The results from Tables 3 and 4 show that increasing the number of design points improves the performance of the GP and MGP charts. As mentioned earlier, the covariance matrix in the GP and MGP charts is calculated based on the radial basis kernel functions. Thus, increasing the number of design points decreases the distance between the points which leads them to learn better from each other, while large number of design points merely increases the dimensionality for the T^2 chart which results in the loss of efficiency for this chart.

4.2 | Two-variable profile monitoring with autocorrelation

In this section, we investigate the performance of the proposed profile monitoring schemes in the presence of autocorrelation. In this regard, our simulation is based on the 2 model settings introduced in the previous section. The main difference here is that the noise term in Equation (1) within each profile is not iid anymore and the successive noise terms are correlated. Specifically, we consider the case where the current noise term depends on its previous lagged noise term. Thus, the following process is assumed for the structure of the noise term:

$$\varepsilon_i^{(m)}\left(x_{i,j}\right) + \phi\varepsilon_i^{(m)}\left(x_{i,j-1}\right) \tag{26}$$

As for the trigonometric model, the 2 profiles are defined using Equation (24) with 30 equally spaced design

TABLE 4 The effect of number of design points in the quadratic model

	Quadratic Model with $n = 50$ IC Samples						
		10 design points			30 design points		
	Shift	GP	MGP	T^2	GP	MGP	T^2
Mean	0.6	120.1	80.3	51.1	63.2	30.9	31.0
	0.7	83.3	47.5	31.7	36.1	15.5	17.5
	0.8	56.6	28.0	19.7	19.7	7.9	10.8
	0.9	35.6	16.8	12.5	11.2	4.3	6.5
	1	22.9	10.2	8.1	6.2	2.5	4.2
Noise	0.02	248.7	233.5	240.6	208.3	169	262.2
	0.04	168.6	129.2	206.4	121.3	88.8	198.9
	0.06	121.4	104.5	133.7	72.4	49.2	147.2
	0.08	91.9	75.4	102.9	46.1	30.5	113.0
	0.1	70.7	54.8	78.4	30.5	18.7	88.7

TABLE 5 ARL_1 comparison of the 3 charts in the trigonometric model with autocorrelation

		$n = 31$			$n = 100$		
	Shift	GP	MGP	T^2	GP	MGP	T^2
Mean	0.1	289.9	279.7	358.1	277.0	265.4	278.8
	0.2	150.5	123.7	329.8	143.1	112.1	149.5
	0.3	63.7	38.3	274.7	32.4	30.9	70.6
	0.4	23.5	11.4	220.9	21.1	9.4	30.7
	0.5	8.6	4.0	172.1	8.1	3.2	13.2
Noise	0.01	210.7	204.0	299.4	201.3	192.3	245.9
	0.02	127.8	116.7	246.9	113.7	101.4	165.8
	0.03	81.3	64.8	202.5	81.4	60.0	109.9
	0.04	54.3	40.6	171.9	51.7	37.0	75.1
	0.05	36.6	26.1	144.5	36.7	23.4	53.8
Segment	0.1	294.8	284.3	347.1	281.3	253.0	314.4
	0.2	155.0	152.5	277.1	137.9	128.4	240.8
	0.3	86.7	82.5	190.9	81.9	71.4	162.1
	0.4	35.5	29.1	115.7	34.9	29.6	103.4
	0.5	16.7	12.2	65.2	17.7	12.6	59.4

points over $[0, 2\pi]$. Moreover, it is assumed that $\phi = 0.5$ defines the coefficient of autocorrelation in Equation (26). Again, using the same logic in the previous section, n IC samples of 2-variable profiles with autocorrelation are generated in phase I. Then, the IC parameters are estimated, and the monitoring is performed according to the procedure for each chart. Table 5 gives the ARL_1 values for the specified OC models of 3 charts in the trigonometric model under the sparse and dense settings.

Results from Table 5 show that in the sparse setting with $n = 31$ IC samples, the MGP chart performs better than the other charts in the presence of autocorrelation. Also, the GP chart's performance is better than the T^2 chart in terms of ARL_1. This, in fact, originates from the nature of the Gaussian process models that can capture the correlations within each profile. Although increasing the number of IC samples generally leads to improvement of the charts, even with $n = 100$, the MGP chart performs relatively better than the other charts. Moreover, these results show that, in the presence of autocorrelation, the GP chart performs relatively better than the T^2 chart in the dense setting with $n = 100$ IC samples.

Next, the performance of the proposed charts in the presence of autocorrelation is investigated for the quadratic model. In this regard, the functions in Equation (25) are considered, and it is assumed that the design points are equally distributed over $[0, 10]$. Here, also, $\phi = 0.5$ is considered as the IC parameter. The results of simulation for this model are given in Table 6.

The results in Table 6 show the similar trend in the performance of the 3 charts as with the trigonometric model. Here, again, in the sparse setting, the MGP chart has a superior performance in terms of ARL_1 compared with the other charts. Moreover, as expected, the GP chart performs better in the sparse setting than the T^2 chart. Also, in the dense setting with $n = 50$ IC samples, the MGP chart's performance is relatively better than that of the other 2 charts.

TABLE 6 ARL_1 comparison of the 3 charts in the quadratic model with autocorrelation

		$n = 11$			$n = 50$		
	Shift	**GP**	**MGP**	T^2	**GP**	**MGP**	T^2
Mean	0.6	80.4	65.2	509.7	70.8	34.9	113.0
	0.7	54.7	41.3	481.5	42.9	19.5	81.2
	0.8	36.7	25.4	462.7	27.0	10.8	57.5
	0.9	24.7	16.3	431.2	16.3	6.2	40.6
	1	16.7	10.7	389.9	9.9	3.9	28.0
Noise	0.02	240.2	224.3	303.1	209.5	188.5	257.6
	0.04	173.4	151.6	247.1	125.9	99.9	191.8
	0.06	125.3	108.2	205.9	81.4	60.8	146.3
	0.08	96.4	77.4	165.6	53.7	39.2	109.8
	0.1	72.3	55.3	140.0	36.3	25.0	82.7
Segment	0.6	160.2	158.1	406.4	95.3	88.0	161.0
	0.7	120.7	120.2	395.7	67.5	57.4	129.7
	0.8	94.7	93.8	389.9	46.9	38.0	98.5
	0.9	71.9	68.2	399.5	31.1	24.3	75.6
	1	55.5	48.6	391.1	20.8	15.7	54.8

4.3 | Monitoring profiles with more than two variables

In this section, we investigate the performance of the proposed charts in the scenario of more than 2-variable profiles. In this regard, we consider the cases where the underlying process is defined through 3 or 4 variables. First, let us consider that the following 3 functions characterize the underlying process under consideration:

$$\begin{aligned} y_1^{(m)}(x_{1,j}) &= 5 + 2\cos(x_{1,j}) + \varepsilon_1^{(m)}(x_{1,j}) \quad m = 1, 2, ..., n \\ y_2^{(m)}(x_{2,j}) &= -5 + 2\sin(x_{2,j}) + \varepsilon_2^{(m)}(x_{2,j}) \quad m = 1, 2, ..., n \\ y_3^{(m)}(x_{3,j}) &= 2e^{-0.2x_{3,j}}\cos(x_{3,j}) + \varepsilon_3^{(m)}(x_{3,j}) \quad m = 1, 2, ..., n \end{aligned} \tag{27}$$

where $\sigma_1 = \sigma_2 = \sigma_3 = 0.5$ are iid noises and the measurements are made in 30 equally spaced design points over $[0, 2\pi]$. Following the general procedure in this study, $n = 50$ IC samples of the above 3-variable profiles are generated, and the IC parameters are estimated. In phase II, the performance of the 3 proposed charts are compared in terms of ARL_1. Next, we consider a case with 4-variable profiles where the following function is added to the set of functions defined in Equation (27).

$$\begin{gathered} y_4^{(m)}(x_{4,j}) = 10 + 2e^{-0.2x_{4,j}}\sin(x_{4,j}) + \varepsilon_4^{(m)}(x_{4,j}) \\ m = 1, 2, ..., n \end{gathered} \tag{28}$$

For the 2 scenarios explained here, different OC situations including shifts in the mean, noise, and the segment containing 10 design points in the middle of each profile are considered. The results for these shifts are summarized in Table 7.

The results in Table 7 show that in the case of monitoring more than 2 variables, the performance of the MGP chart is significantly better than that of the GP and T^2 charts. First, it can be seen that when monitoring both 3 signals and 4 signals, the MGP chart consistently outperforms the other 2 charts in different OC shifts. Moreover, it can be seen that increasing the number of signals further enhances the MGP chart where with the same amount of shift in the 4 signal case, the ARL_1 for the MGP chart is less than its corresponding value for the 3 signal case. However, looking at the results for the GP and T^2 charts, it can be concluded that the inverse relationship holds and increasing the number of signals leads to a decrease in their performance.

TABLE 7 ARL_1 performance of monitoring with 3 signals and 4 signals

		Three Signals			Four Signals		
	Shift	GP	MGP	T^2	GP	MGP	T^2
Mean	0.1	243.1	224.0	269.1	277.3	206.3	301.6
	0.2	118.9	68.1	158.0	136.4	56.9	139.5
	0.3	40.3	14.2	74.4	49.1	10.2	49.2
	0.4	11.8	3.4	29.2	13.9	2.5	13.5
	0.5	3.7	1.4	11.3	4.1	1.2	4.3
Noise	0.01	175.1	143.7	238.2	190.1	132.5	276.9
	0.02	99.8	69.1	170.3	111.6	58.4	196.0
	0.03	59.8	34.5	123.2	67.8	29.0	151.4
	0.04	37.7	19.6	91.1	41.9	15.1	113.0
	0.05	24.3	12.0	69.6	27.0	9.1	87.3
Segment	0.1	286.5	285.0	307.6	333.1	287.8	363.9
	0.2	217.2	180.9	232.4	260.5	165.6	298.1
	0.3	134.9	85.9	149.9	165.4	68.8	208.0
	0.4	71.5	34.6	92.5	91.7	25.6	142.4
	0.5	35.5	13.2	52.6	44.2	8.8	87.9

4.4 | Two-variable profile monitoring with covariance shift

In this section, we investigate the performance of profile monitoring schemes in the case where there is a shift in the between-profile correlation of IC profiles. In this regard, in the trigonometric model setting, we generate the initial IC samples from a MGP with known IC parameters as stated in Equation (14) with 30 equally spaced design points over $[0, 2\pi]$. Specifically, it is assumed that the vector of IC parameters is $\{\rho_{11}, L_{11}, \rho_{22}, L_{22}, \rho_{01}, L_{01}, \rho_{02}, L_{02}, \sigma_1, \sigma_2\} = \{2,\ 1, 2, 1, 4, 1.5, 4, 1.5, 0.5, 0.5\}$. Next, using the same logic as in previous sections, n IC samples of 2-variable profile are generated in phase I. Then, using the proposed methods, the estimated IC model is built to monitor the new incoming samples. Moreover, in order to define the required OC condition, we change the corresponding cross-correlation parameters (ρ_{01}) of the covariance function. Table 8 gives the ARL_1 values of the OC model for the 3 charts.

TABLE 8 ARL_1 performance in covariance shift

	$n = 31$			$n = 50$			$n = 100$		
Shift	GP	MGP	T^2	GP	MGP	T^2	GP	MGP	T^2
0.2	390.5	288.5	360.7	379.5	279.0	333.1	313.2	268.8	324.6
0.4	232.6	190.3	344.5	229.9	186.7	278.2	192.5	179.9	270.6
0.6	112.1	104.6	318.8	113.3	101.2	233.7	106.0	88.4	209.4
0.8	57.2	45.3	273.5	60.8	45.1	164.3	55.0	40.6	134.4
1	32.3	20.5	231.6	32.4	20.8	98.7	29.9	18.8	68.3

TABLE 9 ARL_1 comparison of the MGP and T^2 charts in the trigonometric and quadratic model settings

	Trigonometric Model					Quadratic Model				
		$n = 61$		n = 100			n = 30		n = 50	
	Shift	MGP	T^2	MGP	T^2	Shift	MGP	T^2	MGP	T^2
Mean	0.1	285.9	357.6	230.6	224.0	0.6	88.1	118.6	80.3	61.7
	0.2	114.0	229.8	80.1	80.3	0.7	44.2	83.2	44.5	37.2
	0.3	31.5	198.4	21.1	25.9	0.8	29.7	57.6	28.0	23.2
	0.4	7.9	189.3	6.0	9.0	0.9	16.4	40.0	16.8	14.1
	0.5	2.5	185.0	2.0	3.8	1	11.6	27.5	10.2	8.8
Noise	0.01	284.3	381.0	147.4	248.1	0.02	245.4	284.5	233.5	273.9
	0.02	113.4	390.8	87.2	182.2	0.04	162.3	221.6	129.2	196.9
	0.03	31.5	397.5	45.2	130.8	0.06	111.5	183.2	104.5	146.9
	0.04	7.8	401.0	30.8	99.3	0.08	77.6	144.7	75.4	109.7
	0.05	2.5	379.0	17.8	76.0	0.1	55.8	114.0	54.8	83.6
Segment	0.1	380.0	366.3	299.5	330.7	0.6	90.0	344.9	90.6	173.6
	0.2	276.7	287.1	205.0	241.5	0.7	60.6	334.2	55.4	132.5
	0.3	158.6	254.5	114.2	148.9	0.8	43.2	310.9	32.8	99.5
	0.4	79.9	232.7	54.6	80.6	0.9	30.0	287.3	19.5	70.9
	0.5	35.3	215.0	24.7	41.9	1	21.6	265.9	11.8	50.2

Results from Table 8 show that MGP chart always performs better than the other charts where there is a shift in between-profile correlation. This is mostly because of the non-separable covariance function of the MGP method that explicitly considers the cross-correlation between different profiles of a sample.

4.5 | Comparison of the MGP chart with the "aggregated" T^2 chart

This section aims at comparing the performance of the MGP chart with the T^2 chart as proposed in Equation (23). In this regard, 2-variable profiles in both trigonometric and quadratic model settings are considered. The same logic as discussed in Section 4.1 is also used here to generate the IC samples. Considering the cases of moderate and large number of IC samples, the estimated IC models are built in order to monitor the new incoming samples. The OC ARL of 2 charts in the case of mean shift, noise shift, and segment shift are summarized in Table 9.

The results from Table 9 show that the MGP chart outperforms the T^2 chart described in Equation (23) in different OC conditions. The only exception is the case of mean shift when the number of IC samples is large. It should be noted that the MGP method can capture the cross-correlation among different profile variables through the shared kernel parameters. Such parameters are estimated based on the maximum likelihood method. On the other hand, the T^2 chart captures the cross-correlation by simply pooling all the observations. The T^2 often requires a large number of IC samples in phase I in order to accurately estimate the covariance matrix. Therefore, in the case of mean shift when there is a large number of IC samples, the T^2 chart performs slightly better than the MGP chart.

5 | APPLICATION TO A REAL-WORLD CASE STUDY

In this section, the application of the proposed profile monitoring procedures on a real-world case study of an ice machine is demonstrated. In an ice machine, different sensors monitor different aspects of the ice making process. Specifically, the temperature of the liquid refrigerant (evaporator inlet temperature), temperature of the vapor refrigerant (evaporator outlet temperature), the liquid line temperature, and compressor discharge temperature are considered to be the main drivers of the performance of the ice making process. These temperature signals for 3 consecutive working cycles of the ice machine are demonstrated in Figure 1.

Regarding the ice machine, data are collected every 1 minute during the machine's operation. Furthermore, the ice making process duration is not fixed from cycle to cycle and depends on some factors including the ambient temperature, the quality of water, etc. However, in order to make a fair comparison between different approaches, a fixed number of design points in different cycles are required. Thus, using an interpolation technique for each individual sample of a 4-variable profile, the design points have been fixed at 12 equally spaced time points in the process. Figure 4 shows 3 samples of the temperature signals along with the MGP estimate of the IC profile functions.

The ice machine data set contains 148 samples, each with 4 temperature variables measured in 12 equally spaced design points. The first 20 samples of data set have been considered for the phase I analysis. The IC parameters are then estimated by applying the specific procedures of the proposed charts on the first 20 samples. The resulting MGP estimated IC profiles for the temperature signals are displayed in Figure 4.

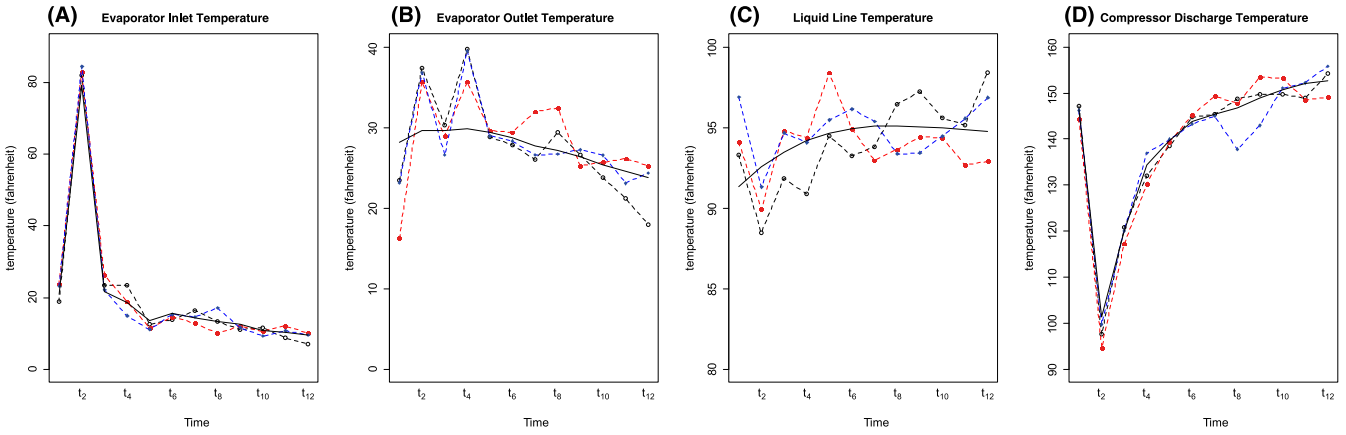

FIGURE 4 Three samples of the temperature signals (dashed lines) and the MGP estimate of the IC profile functions (solid curve) [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 10 Mean shift and segment shift analysis for case study data set

Mean Shift				Segment Shift			
Shift	**GP**	**MGP**	T^2	**Shift**	**GP**	**MGP**	T^2
0.5	161.3	152.9	209.5	0.5	187.9	184.7	210.3
1	102.1	94.3	205.0	1.0	147.1	145.1	207.8
1.5	51.3	50.5	201.5	1.5	117.0	106.1	200.7
2	29.0	22.6	195.5	2.0	83.5	70.4	197.4
2.5	16.0	10.0	186.3	2.5	59.0	44.3	193.0

TABLE 11 ARL_1 performance of monitoring the coefficients of MGP model in case study

Mean Shift		Segment Shift	
Shift	**Coefficient-based MGP chart**	**Shift**	**Coefficient-based MGP chart**
0.5	163.5	0.5	170.8
1	159.8	1	170.2
1.5	154.6	1.5	169.4
2	150.4	2	166.9
2.5	149.9	2.5	165.7

Next, we construct the proposed control charts for phase II profile monitoring by fixing the IC ARL at 200. In order to compare the monitoring procedures, we consider the scenario that a shift in the mean or the middle segment of the profile variables has occurred. The OC simulations are done based on the MGP estimate of the 4 variables and the results for 10 000 runs are summarized in Table 10.

It can be seen from Table 10 that the MGP chart outperforms the GP and T^2 charts in different mean shift and segment shift situations. This confirms the results we obtained in the simulation study that, for profile monitoring with more than 2 variables, the MGP chart performance is better than that of the GP and T^2 charts.

Next, we establish a control chart based on monitoring the coefficients of the MGP model as introduced in Li, Y. et al.[20] Using the 20 initial samples, the coefficients and the Fisher information matrix of the coefficients of MGP model are estimated. We use a T^2 statistic to monitor the coefficients of the MGP model with the Fisher information matrix as the covariance matrix. Again, the ARL_0 is fixed at 200. Table 11 demonstrates the OC simulation results in this case.

Comparing the results of Table 11 with Table 10, one can see that the profile monitoring scheme based on monitoring the coefficients of MGP model is less efficient than the MGP chart proposed in this paper. This further confirms the superiority of our proposed MGP monitoring scheme.

In addition to the above comparison, we performed an analysis based on using the first 20 and 104 samples of data to estimate the IC parameters and to find the control limit, respectively. The control limit is set using the procedure discussed in the simulation study section and considering a type I error rate of 0.04. The remaining 44 samples are also used to test the charts discussed in this paper as well as the chart based on monitoring the coefficients of the MGP model. The MGP, GP, and T^2 charts for samples 105 through 148 are shown in Figures 5 and 6. The solid line in each chart represents the control limit of that specific chart.

From Figure 5, it can be seen that the MGP chart first gives an OC signal in the 117th sample. Also, an OC condition is shown in the 119th, 121th, 124th, 128th, 130th, and 133th sample. However, regarding the GP chart, it first signals OC in sample 121 where the liquid line temperature chart is OC. Next, this chart gives OC signals in samples 126, 128, and 133. As for the T^2 chart, it gives its first OC signal in sample 122 where the compressor discharge temperature chart is OC. This chart signals OC condition in samples 126 and 130 as well.

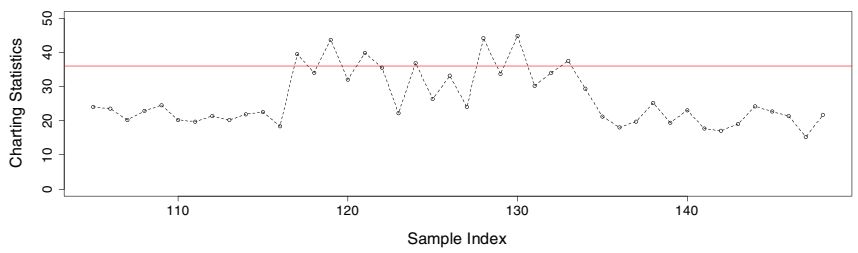

FIGURE 5 MGP chart for the ice machine data set [Colour figure can be viewed at wileyonlinelibrary.com]

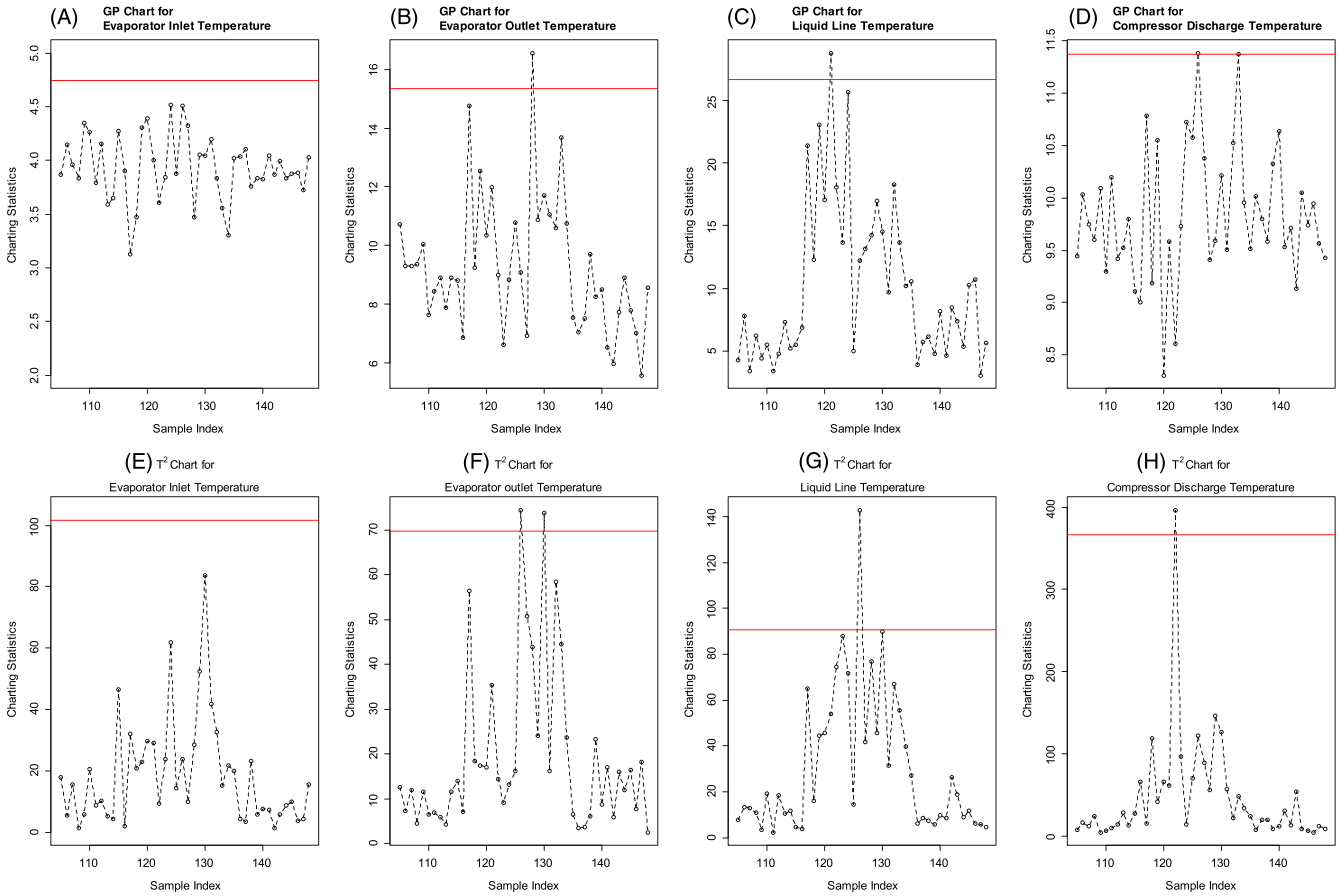

FIGURE 6 Univariate GP and T^2 charts for the ice machine data set [Colour figure can be viewed at wileyonlinelibrary.com]

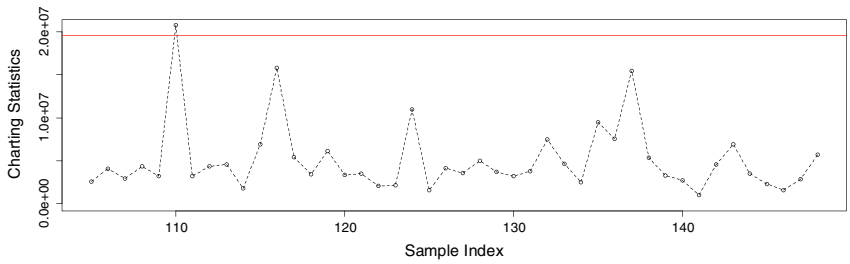

FIGURE 7 Chart of monitoring the coefficients of MGP model for the ice machine data set [Colour figure can be viewed at wileyonlinelibrary.com]

The charting statistics based on monitoring the coefficients of the MGP model for samples 105 through 148 are shown in Figure 7.

As can be seen from Figure 7, this control chart can only detect an OC condition in sample 110 and remains IC for the rest of samples. By contrast, the profile monitoring procedures introduced earlier in this paper show a change in the ice making process between the 117th and 133th samples.

The case study analysis shows that the MGP chart proposed in this paper can detect an OC condition faster and more accurately than the other monitoring schemes discussed here. It can also be concluded that scheme based on monitoring the coefficients of the MGP model is not as efficient as the MGP model proposed here, as it is unable to respond to the changes in the process between the 117th and 133th samples as detected by other charts.

6 | CONCLUSION

In this paper, modeling and monitoring of multivariate profiles using the MGP model are investigated. Because in most engineering systems different sensors are correlated, it is crucially important to consider this correlation in modeling and monitoring the resulting multivariate profiles. The MGP model enjoys the flexibility and desirable analytical properties of the GP along with the ability to capture the correlation among

multiple variables of a profile. Thus, it provides a powerful tool to model different profiles considering the correlations existing both within and between profiles. Along with the MGP-based profile monitoring scheme, methods based on modifying univariate GP and T^2 charts for the case of multivariate profiles are also presented in this paper.

The MGP chart proposed in this study is based on the non-separable covariance structure which offers high flexibility in modeling and monitoring multivariate profiles. This flexibility typically leads to better performance in term of ARL_1 which is confirmed by the results in our numerical study. The results also show that in the presence of autocorrelation in measured values of a profile, the MGP chart outperforms other charts in different out-of-control situations with different number of IC samples. An important feature of the MGP model is that it can learn from different dependent profiles. Therefore, increasing the number of profiles potentially enhances the MGP model, while this merely increases the misdetection rate for the univariate charts. This is also confirmed by the analysis in the numerical study section. Finally, the real-world case study further confirms the results from the numerical study and shows the advantages of the MGP-based chart for monitoring multivariate profiles in practice.

It should be pointed out that the efficiency of MGP models can be further improved by optimally selecting the design points in the measurement region. In this regard, one can use a space filling algorithm like Latin hypercube or sliced Latin hypercube to select the optimal design points in space. We aim to investigate along this direction in our future research. Also, it is known that the multivariate profile monitoring charts based on EWMA scheme can catch the OC faster in the case of small shifts.[18] One potential direction of research for the MGP chart proposed here could be integrating it with an EWMA scheme to enhance its performance in the case of small shifts. Finally, with the advent of stochastic kriging-based methods, it would be desirable to develop a chart using this approach and compare its performance with the MGP chart.

ACKNOWLEDGEMENT

The authors would like to thank the editor and the reviewers for their helpful comments. The financial support for this work is provided by the U.S. National Science Foundation Grant #1561512.

REFERENCES

1. Noorossana R, Saghaei A, Amiri A. *Statistical Analysis of Profile Monitoring*. 865 John Wiley & Sons; 2011.
2. Jensen WA, Birch JB. Profile monitoring via nonlinear mixed models. *J Qual Technol*. 2009;41(1):18-34.
3. Williams JD, Woodall WH, Birch JB. Statistical monitoring of nonlinear product and process quality profiles. *Qual Reliab Eng Int*. 2007;23(8):925-941.
4. Zhang Y, He Z, Zhang C, Woodall WH. Control charts for monitoring linear profiles with within-profile correlation using Gaussian process models. *Qual Reliab Eng Int*. 2014;30(4):487-501.
5. Woodall WH. Current research on profile monitoring. *Production*. 2007;17(3):420-425.
6. Parker P, Morton M, Draper N, Line W. A single-vector force calibration method featuring the modern design of experiments. Paper presented at: 39th Aerospace sciences meeting and exhibit; 2001.
7. Stover FS, Brill RV. Statistical quality control applied to ion chromatography calibrations. *J Chromatogr A*. 1998;804(1-2):37-43.
8. Kang L, Albin SL. On-line monitoring when the process yields a linear profile. *J Qual Technol*. 2000;32(4):418-426.
9. Walker E, Wright SP. Comparing curves using additive models. *J Qual Technol*. 2002;34(1):118-129.
10. Chang SI, Yadama S. Statistical process control for monitoring non-linear profiles using wavelet filtering and B-spline approximation. *Int J Prod Res*. 2010;48(4):1049-1068.
11. Shiau J-JH, Huang H-L, Lin S-H, Tsai M-Y. Monitoring nonlinear profiles with random effects by nonparametric regression. *Commun Stat Theory Methods*. 2009;38(10):1664-1679.
12. Chicken E, Pignatiello Jr JJ, Simpson JR. Statistical process monitoring of nonlinear profiles using wavelets. *J Qual Technol*. 2009;41(2):198-212.
13. Rasmussen CE, Williams CKI. *Gaussian Processes for Machine Learning*. the MIT Press; 2006.
14. Colosimo BM, Cicorella P, Pacella M, Blaco M. From profile to surface monitoring: SPC for cylindrical surfaces via Gaussian processes. *J Qual Technol*. 2014;46(2):95-113.
15. Cicorella P, Colosimo BM, Pacella M. Statistical process monitoring of Complex Shapes via Gaussian Process modeling. 2013.
16. Zou C, Ning X, Tsung F. LASSO-based multivariate linear profile monitoring. *Ann Oper Res*. 2012;192(1):3-19.
17. Paynabar K, Jin J, Pacella M. Monitoring and diagnosis of multichannel nonlinear profile variations using uncorrelated multilinear principal component analysis. *IIE Trans*. 2013;45(11):1235-1247.
18. Chou S-H, Chang SI, Tsai T-R. On monitoring of multiple nonlinear profiles. *Int J Prod Res*. 2014;52(11):3209-3224.
19. Wu X, Miao R, Li Z, et al. Process monitoring research with various estimator-based MEWMA control charts. *Int J Prod Res*. 2015;53(14):4337-4350.
20. Li Y, Zhou Q, Huang X, Zeng L. Pairwise estimation of multivariate Gaussian process models with replicated observations: application to multivariate profile monitoring. *Technometrics*. 2017.
21. Qiu P, Zou C, Wang Z. Nonparametric profile monitoring by mixed effects modeling. *Technometrics*. 2010;52:265-277.
22. Montgomery DC. *Statistical Quality Control*. 7 New York: Wiley; 2009.
23. Higdon D. *Space and Space-Time Modeling Using Process Convolutions. Quantitative Methods for Current Environmental Issues*. Springer; 2002:37-56.

24. Xia H, Ding Y, Wang J. Gaussian process method for form error assessment using coordinate measurements. *IIE Trans*. 2008;40:931-946.
25. Kontar R, Zhou S, Horst J. Estimation and monitoring of key performance indicators of manufacturing systems using the multi-output Gaussian process. *Int J Prod Res*. 2017;55(8):2304-2319.
26. Ver Hoef JM, Barry RP. Constructing and fitting models for cokriging and multivariable spatial prediction. *J Stat Plan Inference*. 1998;69(2):275-294.
27. Mestek O, Pavlík J, Suchánek M. Multivariate control charts: control charts for calibration curves. *Fresenius J Anal Chem*. 1994;350(6):344-351.

Salman Jahani received the BS degree in industrial engineering from University of Tehran, Tehran, Iran, in 2013. He also received an MS degree in industrial engineering from Sharif University of Technology, Tehran, Iran, in 2015. He is currently working towards the PhD degree in industrial and systems engineering and the MS degree in statistics from the University of Wisconsin-Madison, Madison, WI, USA.

Raed Kontar is an Assistant Professor in the Industrial and Operations Engineering department at the University of Michigan, Ann Arbor. He received his PhD in Industrial and Systems Engineering in 2018 and MS in statistics in 2017 from the University of Wisconsin Madison. He also received his BS in Civil and Environmental Engineering with a minor in Mathematics from the American University of Beirut (AUB) in 2014.

Dharmaraj Veeramani is the Robert Ratner Chair Professor in the Department of Industrial and Systems Engineering at the University of Wisconsin-Madison. He received his PhD and MS degrees in Industrial Engineering from Purdue University and his BS degree in Mechanical Engineering from the Indian Institute of Technology, Madras. Dr. Veeramani's research focuses on emerging frontiers of digital business, IoT technologies and applications, smart and connected systems, and supply chain management. He has received numerous research grants from federal agencies and industry. In recognition of his scholarly contributions, Dr. Veeramani has received multiple honors and awards from organizations such as the National Science Foundation, SME, SAE International, and ASEE.

Shiyu Zhou is a Professor in the Department of Industrial and Systems Engineering at the University of Wisconsin-Madison. He received his BS degree from the University of Science and Technology of China in 1993, and his master's degree and PhD from the University of Michigan in 2000. His research focuses on industrial analytics and system informatics methodologies for quality and productivity improvement and operation optimization. He has received numerous research awards and grants from various federal agencies. He is currently a Fellow of IISE and ASME.

How to cite this article: Jahani S, Kontar R, Veeramani D, Zhou S. Statistical monitoring of multiple profiles simultaneously using Gaussian processes. *Qual Reliab Engng Int*. 2018;1–20. https://doi.org/10.1002/qre.2326

APPENDIX 1

Considering the general decomposition of profile i in Equation (12) and the Gaussian Kernel in Equation (13), the $cov_{ij^{(m)}}(s,s')$ can be written as follows:

$$\begin{aligned} cov_{ij^{(m)}}(s,s') = & \int_{-\infty}^{+\infty} \frac{\rho_{ii,m}\sqrt{|\boldsymbol{L}_{ii,m}|}}{\sqrt[4]{\pi^D}} exp\left\{-\frac{1}{2}(\boldsymbol{u}+\boldsymbol{d}-\boldsymbol{\mu}_{ii,m})^T \boldsymbol{L}_{ii,m}(\boldsymbol{u}+\boldsymbol{d}-\boldsymbol{\mu}_{ii},m)\right\} \\ & \times \frac{\rho_{ij,m}\sqrt{|\boldsymbol{L}_{ij,m}|}}{\sqrt[4]{\pi^D}} exp\left\{-\frac{1}{2}\left(\boldsymbol{u}-\boldsymbol{\mu}_{ij,m}\right)^T \boldsymbol{L}_{ij,m}\left(\boldsymbol{u}-\boldsymbol{\mu}_{ij,m}\right)\right\} d^D\boldsymbol{u} \\ & + \int_{-\infty}^{+\infty} \frac{\rho_{0i,m}\sqrt{|\boldsymbol{L}_{0i,m}|}}{\sqrt[4]{\pi^D}} exp\left\{-\frac{1}{2}(\boldsymbol{u}+\boldsymbol{d}-\boldsymbol{\mu}_{0i,m})^T \boldsymbol{L}_{0i,m}(\boldsymbol{u}+\boldsymbol{d}-\boldsymbol{\mu}_{0i,m})\right\} \\ & \times \frac{\rho_{0j,m}\sqrt{|\boldsymbol{L}_{0j,m}|}}{\sqrt[4]{\pi^D}} exp\left\{-\frac{1}{2}\left(\boldsymbol{u}-\boldsymbol{\mu}_{0j,m}\right)^T \boldsymbol{L}_{0j,m}\left(\boldsymbol{u}-\boldsymbol{\mu}_{0j,m}\right)\right\} d^D\boldsymbol{u} \end{aligned} \quad \text{(a.1)}$$

Thus, for each integral term, we have that

$$\int_{-\infty}^{+\infty} exp\left\{-\frac{1}{2}(\boldsymbol{u}+2\boldsymbol{d}-\boldsymbol{\mu}_{ui,m})^T\boldsymbol{L}_{ui,m}(\boldsymbol{u}+\boldsymbol{d}-\boldsymbol{\mu}_{ui,m})\right\}exp\left\{-\frac{1}{2}\left(\boldsymbol{u}-\boldsymbol{\mu}_{uj,m}\right)^T\boldsymbol{L}_{uj,m}\left(\boldsymbol{u}-\boldsymbol{\mu}_{uj,m}\right)\right\}d^D\boldsymbol{u}$$
$$= exp\left(-\frac{1}{2}\varphi\right)\exp\left\{-\frac{1}{2}(\boldsymbol{u}-\boldsymbol{\phi})^T\boldsymbol{\Sigma}(\boldsymbol{u}-\boldsymbol{\phi})\right\}d^D\boldsymbol{u} \tag{a.2}$$

Considering $\boldsymbol{\Psi} = \boldsymbol{L}_{ui,m}^{-1} + \boldsymbol{L}_{uj,m}^{-1}$, we have that $\boldsymbol{\Sigma} = \boldsymbol{L}_{ui,\ m} + \boldsymbol{L}_{uj,\ m}$, $\boldsymbol{\phi} = \boldsymbol{\Sigma}^{-1}\{\boldsymbol{L}_{uj,\ m}\boldsymbol{\mu}_{uj,\ m} + \boldsymbol{L}_{ui,\ m}(\boldsymbol{d} - \boldsymbol{\mu}_{ui,\ m})\}$ and $\varphi = \psi^T\boldsymbol{\Psi}\psi$ where $\psi = \boldsymbol{d} - \{\boldsymbol{\mu}_{ui,\ m} - \boldsymbol{\mu}_{uj,\ m}\}$. We can derive the following:

$$\frac{(2\pi)^{\frac{D}{2}}}{\sqrt{\left|\boldsymbol{L}_{ui,m}+\boldsymbol{L}_{uj,m}\right|}}exp\left\{-\frac{1}{2}\left(\boldsymbol{d}-\left\{\boldsymbol{\mu}_{ui,m}-\boldsymbol{\mu}_{uj,m}\right\}\right)^T\boldsymbol{\Psi}^{-1}\left(\boldsymbol{d}-\left\{\boldsymbol{\mu}_{ui,m}-\boldsymbol{\mu}_{uj,m}\right\}\right)\right\} \tag{a.3}$$

And we have:

$$\frac{\left|\boldsymbol{L}_{ui,m}\right|^{\frac{1}{2}}\left|\boldsymbol{L}_{uj,m}\right|^{\frac{1}{2}}}{\left|\boldsymbol{L}_{ui,m}+\boldsymbol{L}_{uj,m}\right|^{\frac{1}{2}}} = \frac{\left|\boldsymbol{L}_{uj,m}^{-1}\right|^{\frac{1}{2}}\left|\boldsymbol{L}_{uj,m}\right|^{\frac{1}{2}}}{\left|\boldsymbol{L}_{ui,m}^{-1}+\boldsymbol{L}_{uj,m}^{-1}\right|^{\frac{1}{2}}} = \frac{1}{\left|\boldsymbol{L}_{ui,m}^{-1}+\boldsymbol{L}_{uj,m}^{-1}\right|^{\frac{1}{2}}} \tag{a.4}$$

Finally, considering the $L\ \epsilon\ R$ and adjusting (a.1) for the measurement noise the covariance functions in Equation (14) can be derived.

APPENDIX 2

NOMENCLATURE

$\boldsymbol{y}_i^{(m)}$	mth sample vector of the ith profile variable
$x_{i,\ j}$	design point j of profile variable i
p_i	number of design points in each profile i
$\varepsilon_i^{(m)}(x_{i,j})$	intrinsic noise of profile i in sample m at design point $x_{i,\ j}$
$\mu_i^{(m)}$	mean function of profile i in sample m of GP model
$\mathcal{K}_i^{(m)}$	Gaussian kernel of profile i in sample m of GP model
$\boldsymbol{\Sigma}_{\boldsymbol{i}}^{(\boldsymbol{m})}$	covariance matrix of profile i in sample m of GP model
$\widehat{\boldsymbol{y}}_i^{(m)}$	vector of GP prediction for each profile i of sample m
$\widehat{\boldsymbol{y}}_i^{*}$	GP chart estimate of profile i
$\boldsymbol{\Sigma}_i^{*}$	GP chart covariance estimate of profile i
$\boldsymbol{y}^{(m)}$	vector of in-control observations of all profiles in each sample m
$z_i^{(m)}$	Gaussian white noise process of profile i in sample m
τ	Dirac Delta function
$\widehat{\boldsymbol{y}}^{(m)}$	vector of MGP prediction for each sample m
$\boldsymbol{\Sigma}^{(m)}$	MGP estimate of covariance matrix of all k profiles in sample m
$\widehat{\boldsymbol{y}}^{*}$	MGP chart estimate of all profiles
$\boldsymbol{\Sigma}^{*}$	MGP chart covariance estimate of all profiles
$\boldsymbol{S}_i$	Pooled T^2 sample covariance matrix of profile i
$\boldsymbol{S}$	Pooled sample covariance matrix of aggregated T^2
$T^2_{GP_{i,r}}$	GP charting statistic of each new sample r of profile i
$T^2_{MGP_r}$	MGP charting statistic of each new sample r
$T^2_{i,r}$	T^2 charting statistic of each new sample r of profile i
T^2_r	Aggregated T^2 charting statistic of each new sample r
ϕ	coefficient of autocorrelation