Statistical Surface Monitoring by Spatial-Structure Modeling

ANDI WANG

Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

KAIBO WANG

Tsinghua University, Beijing 100084, P. R. China

FUGEE TSUNG

Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

In some manufacturing processes, the quality characteristic is represented by a two-dimensional (2-D) surface. Surface data can generally be treated as a special profile with one response variable and two explanatory variables, for which spatial correlations are commonly observed. Existing parametric charts for profile monitoring are unable to adequately describe the spatial correlations among variables in 2-D surface data, and nonparametric charts cannot be applied to a 2-D data structure directly. In this study, we propose a new chart based on the Gaussian-Kriging model, in which the spatial correlations within the 2-D surface profile are represented by a parametric function. We construct a parametric model that considers three components of the surface—the global trend, the spatial correlations, and independent errors. Then we monitor the process by detecting changes in the estimated parameters. We utilize this method to monitor a wafer-manufacturing process and compare its performance with that of an existing profile-monitoring method through simulation.

Key Words: Gaussian-Kriging Model; Profile Monitoring; Spatial Correlation; Statistical Process Control; Surface Data.

Introduction

Due to its effectiveness in detecting the occurrence of abnormal process failures, reducing process variation, and improving product quality, statistical process control (SPC) has been widely used in diverse industries. In traditional SPC, one or multiple variables that represent the status of the process or the quality of the product are usually monitored for process changes. However, due to the increasing demand for high quality and the continuous development of data-acquisition techniques, product quality is sometimes represented by a two-dimensional (2-D) surface. Thus, the monitoring of surface data has become particularly important, and new techniques to address sophisticated surface patterns are required.

To demonstrate a typical surface-monitoring scenario and illustrate our monitoring scheme, a case study of monitoring the distribution of a wafer's thickness is considered throughout the paper. The wafer is the base material for all semiconductor devices. Processing wafers starts with slicing thin round silicon plates from silicon ingots. After the wafers are sliced, they are loaded into a lapping machine and milled under a specific pressure to improve the flat-

Mr. Wang is a Doctoral Student in the Department of Industrial Engineering and Logistic Management. His email address is awangab@ust.hk.

Dr. Wang is Associate Professor in the Department of Industrial Engineering at Tsinghua University. He is a Senior Member of ASQ. His email address is kbwang@tsinghua.edu.cn. He is the corresponding author of this paper.

Dr. Tsung is Professor in the Department of Industrial Engineering and Logistic Management. He is a Fellow of ASQ. His email address is season@ust.hk.

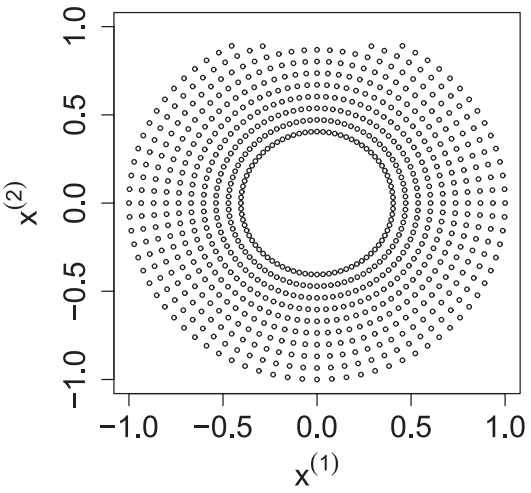

FIGURE 1. Position of Measurements Across the Round Wafer Surface. There are 3912 measurements on 10 concentric circles in total. Each of the eight inner circles has 400 measurements. The two outer circles hav smaller measurements due to the wafer being flat.

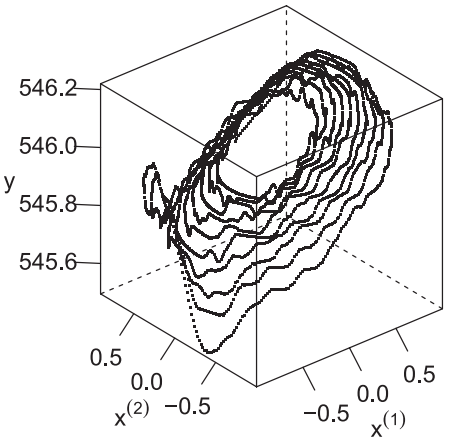

FIGURE 2. A Two-Dimensional Surface Map Showing the Distribution of a Wafer's Thickness After Lapping. $x^{(1)}$ and $x^{(2)}$ denote the coordinates of the wafer's surface and y is the measured thickness value at each site (in μm).

ness and uniformity of their surfaces. The smoothness and uniformity of the wafer are important quality metrics that depend directly on the manufacturing process. Therefore, wafers are periodically drawn from the product line and thickness values are measured at a number of fixed positions across the round plate. On the wafer, a coordinate system is usually defined based on the directions of the single-crystal silicon material. We denote the horizon and vertical axes of this coordinate system as $x^{(1)}$ and $x^{(2)}$, respectively. We further define $\boldsymbol{x}_i = (x_i^{(1)}, x_i^{(2)})$ as the location of the ith measurement point on the wafer, and Y_i represents the thickness obtained at location $\boldsymbol{x}_i$. In practice, thousands of measurements could be obtained from one wafer. Figure 1 shows the locations of the measurement points of a real dataset. The coordinates $x^{(1)}$ and $x^{(2)}$ have been rescaled to $[-1, 1]$. Figure 2 shows the thickness values obtained from one sample wafer after the lapping process, where the data are in a form of a surface map. Monitoring the surface data and detecting potential process shift with the help of the 2-D surface map is essential for product quality.

Currently, the industry metrics for measuring a wafer's quality include the average thickness value and the total thickness variation (TTV) value. The TTV is defined as the difference between the maximum thickness and the minimum thickness for all measurements on the wafer. Other metrics that are commonly utilized in the semiconductor industry to measure wafer quality include the total indicator reading (TIR) and the site TIR (STIR). These metrics reflect the variation in the thickness profile, but they do not fully utilize the rich information contained in the dataset. Consequently, even if the values fall outside the control limits, the failure patterns are not easily revealed. Properly designed monitoring schemes are called for to enhance the performance of control charts.

In most cases, surface data can be treated as a particular type of profile, which is represented by a number of response variables and their corresponding explanatory variables (Woodall et al. (2004)). However, the monitoring of surface data has numerous challenges compared with common profile-monitoring techniques. First, surface data usually exhibit spatial correlations, because adjacent points typically exhibit similar material properties and are processed under similar conditions (Colosimo (2008)). Spatial correlations, which are particularly significant in surface data, have been observed by many researchers, such as Walker and Wright (2002) and Williams et al. (2007). However, most parametric regressive models, which are commonly employed to monitor profiles, assume that the response variable Y_i and the explanatory variable $\boldsymbol{x}_i$ are related by

$$Y_i = f(\boldsymbol{x}_i, \boldsymbol{\beta}) + \varepsilon_i,$$

where $f(\boldsymbol{x}_i, \boldsymbol{\beta})$ are fixed trends and ε_i are independent error terms. Thus, the possible spatial correla-

tion is arbitrarily disregarded. For this reason, this field has called for more in-depth research on profile monitoring that makes use of spatial correlations.

Besides, rich information is contained in the surface data. The common practice of monitoring a profile using a regressive model involves monitoring the parameters depicting the shape of the deterministic part (usually denoted by $\boldsymbol{\beta}$) and the parameter depicting the intensity of the noise (σ^2). However, using the single parameter σ^2 to describe the deviation is not adequate. Process monitoring based on more complex data is expected to provide more information about process faults. For example, Reis and Saraiva (2006) examined the surface profile of paper and emphasized that the information on fine-scale roughness and large-scale waviness is embedded in the surface data. In our wafer example, the 2-D information also illustrates the thickness trend, macro-level waviness, and micro-level roughness of the wafer. Given the rich information available, more efficient process monitoring and diagnosis algorithms should be employed.

In this paper, we focus on the Phase II monitoring of the 2-D wafer surface (Woodall et al. (2004)). By considering unique 2-D structures and spatial correlations, we use a Gaussian-Kriging model to characterize variations in the wafer surface. Assuming the in-control process parameters are known, we designed two related multivariate control charts for monitoring the stability of online surface data. These charts are capable of monitoring the trend of the surface, as well as the spatial patterns, including waviness and roughness of the surface.

The remainder of this paper is organized as follows. In the next section, we review the related literature in profile-monitoring studies. After that, we introduce the model we used for charting and our monitoring scheme. Then, we illustrate the existence of spatial correlation using the real wafer's surface data and compare our method with the multiple regression method, which is modified from the existing profile chart that does not consider spatial correlation. The necessity of considering spatial correlation and the advantage of our method are discussed in the last section.

Review of Related Literature in Profile Monitoring

As mentioned above, a surface dataset can be viewed as a special type of profile, with spatial correlation, complicated structures, and usually having a moderate to large number of measurements. Some recent methods in profile-monitoring literature take spatial correlations and complicated profile structures into consideration, thus were likely to be introduced for monitoring surface data. However, several aspects of these methods are deficient. Jensen et al (2008), for example, proposed a profile-monitoring scheme that is based on the linear mixed model. Their scheme considers the autocorrelation of the errors and the fluctuation of the parameters in the profile. The correlation of the errors takes a simple form, such as compound symmetry or auto-regression. The compound symmetry form assumes that, for any two observations within a profile, the correlation coefficient of their error terms is a fixed value. This makes the compound symmetry form unsuitable for common 2-D surface data because closer observations should exhibit higher correlations. The auto-regression form, on the other hand, is not easily applied in cases where the dimension of the explanatory variable is higher than one. Although spatial correlations are integrated in the model, variations in its patterns are not specifically monitored, and the performance of the chart with respect to variations in correlations has not been evaluated. Colosimo (2008) used a spatial auto-regressive regression (SARX) model to monitor the roundness of the profile obtained by turning. Colosimo et al. (2010) extended the method to monitor a cylindrical surface. However, the SARX model requires a discretization of the measurement coordinates, to define the "neighborhood" of measurements, in order for spatial correlations to be assigned. Grimshaw et al. (2013) proposed a control chart to detect the mean shift of spatial data. In their approach, the spatial correlation structure of the data is modeled by a variogram, based on which the covariance matrix of all measurements across the profile is obtained. The authors then constructed a Hotelling T^2 chart to monitor the mean profile shifts, and their simulation studies based on the monitoring of the bottle wall shows that this method outperforms the ordinary mean chart in various out-of-control cases. However, in some applications, we are not only interested in detecting the shift in the mean profile, but also interested in detecting the change of the surface patterns like waviness and roughness. Much information about surface patterns is contained in the correlation structure of the data. Thus, increasing the performance of the mean chart by introducing a correlation structure of the data is not sufficient and the on-line

monitoring of this correlation structure is also necessary.

Other researchers have employed nonparametric methods to address spatial correlations. Qiu et al. (2010), who used a nonparametric mixed-effect model and local linear kernel smoothing to sketch the correlated profile, proposed a control scheme for monitoring the discrepancy between the smoothed sample profile and the estimated in-control trend. Their simulation studies, which consider heteroscedasticity and spatial correlation within the profile, demonstrated that the method is efficient in detecting common shifts in the profile's trend. However, the variation in the strength of spatial correlations (as an important feature of surface data) is not directly monitored in their approach, and the explanation of the out-of-control signal proves challenging. Reis and Saraiva (2006) proposed a multiscale SPC method for monitoring the surface profile of paper. The quality of the paper is described by two metrics: fine-scale roughness and large-scale waviness. The variations in the two metrics are generated by different mechanisms. The researchers applied wavelet transformation to decompose the profile into the two metrics and proposed monitoring their coefficients. The fact that the profile can be decomposed into two metrics makes the use of a 2-D surface map appropriate for monitoring the profile. Instead, Reis and Saraiva (2006) studied a cross-sectional profile, which is a 1-D data set and ignored the spatial information contained in the surface. In the next section, we introduce our control chart based on a Gaussian-Kriging model. By characterizing the spatial correlations within the surface using a correlation function, the Gaussian-Kriging model successfully captures the spatial correlation of the data. It also allows for the flexibility of the measurement positions and facilitates monitoring of the surface with several parameters that explicitly controls the surface's spatial patterns.

Model Building

Since its origination in the 1950s, the Gaussian-Kriging model has been studied extensively in spatial statistics (Cressie (1993)). Sacks et al. (1989) introduced this model as an interpolation method to the field of computer experiments, and Santner et al. (2003) and Fang et al. (2006) systematically introduced this area of research. This method is also implemented in machine learning, where it is referred to as Gaussian process model (Rasmussen and Williams (2006)). Zhao et al. (2011) used a Gaussian-Kriging model to illustrate the behavior of a wire-saw slicing process in wafer preparation. However, this model remains to be adopted in surface-profile monitoring.

In the Gaussian-Kriging model, the response variable Y_j, which is the thickness value measured at position $(x_j^{(1)}, x_j^{(2)})$ on the wafer, is decomposed into three parts. First, the wafer exhibits an overall thickness pattern; that is, its thickness increases roughly linearly such that it is thin at one end and thick at the other. This overall mean trend represents a variation in the macro-level thickness. Second, the wafer is smooth and the thicknesses at adjacent sites are correlated. This variation in micro-level thickness accounts for the waviness of the wafer. Last, measurement errors and other incidental errors may affect the thickness at each site independently and cause roughness in the profile. Therefore, we propose the following equation to characterize the measurements of thickness values on a sample wafer surface:

$$Y_j = \mu + \beta_1 x_j^{(1)} + \beta_2 x_j^{(2)} + Z(\boldsymbol{x}_j) + \varepsilon_j, \qquad (1)$$

where $\boldsymbol{x}_j = (x_j^{(1)}, x_j^{(2)})$ are the 2-D explanatory variable, representing the coordinates of the jth measurement point on the surface as introduced before, and Y_j is the corresponding response variable, the thickness measurement at this position. This model contains three components: the linear trend, the Gaussian field component, and the noise term. The linear trend characterizes the macro-scale variation, the Gaussian field component characterizes the micro-level spatial correlation, and the noise term represents random noise. Therefore, this model captures both the large-scale and the fine-scale variations in the wafer surface successfully. Process monitoring based on this model is expected to provide not only improved charting performance but also helpful diagnostic information. We explain this model in greater detail now.

The linear trend of the Gaussian-Kriging model is similar to the linear trend of a multiple regression model. The linear trend of a single wafer indicates that the expected thickness profile of the wafer is a planar surface. The linear trend captures the macro-level thickness pattern. In practice, this thickness pattern is produced by the lapping mechanism (see Lin and Wang (2011) for a schematic illustration of the lapping process). When wafers are mounted onto a lapping machine, the upper and lower plates of the machine rotate in opposite directions, removing the unwanted surface materials. Because the upper plate

may be tilted away from the vertical axis, more material is removed from some parts of the wafer than from other parts such that the thickness of the wafer exhibits a roughly linear trend from one end to the other.

The second term in the model $Z(\boldsymbol{x})$ depicts the spatial correlation within the surface. It is assumed to be a stationary process that satisfies the following conditions:

1. $\mathrm{E}(Z(\boldsymbol{x})) = 0$,
2. $\mathrm{Cov}(Z(\boldsymbol{x}_1), Z(\boldsymbol{x}_2)) = C(\boldsymbol{x}_1 - \boldsymbol{x}_2)$, where $C(\cdot)$ is a positive-definite function,
3. For any $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$, $(Z(\boldsymbol{x}_1), Z(\boldsymbol{x}_2), \ldots, Z(\boldsymbol{x}_n))$ follow a multivariate normal distribution.

These conditions intuitively indicate that the properties do not change across the surface and that the covariance of two measurements is only dependent on their relative positions. From condition 2, for any $\boldsymbol{x}$,

$$\mathrm{Var}(Z(\boldsymbol{x})) = C(0).$$

If we denote σ_Z^2 by $C(0)$, we obtain $\mathrm{Cov}(Z(\boldsymbol{x}_1), Z(\boldsymbol{x}_2)) = \sigma_Z^2 r(\boldsymbol{x}_1 - \boldsymbol{x}_2)$, where $r(\cdot)$ is the correlation function of the stationary Gaussian process and $r(\boldsymbol{x}_1 - \boldsymbol{x}_2)$ depicts the correlation among response variables measured at locations with distances of $\boldsymbol{x}_1 - \boldsymbol{x}_2$ apart. The variogram of the correlation function $r(\cdot)$ is defined as

$$2\gamma(\boldsymbol{h}) = \mathrm{Var}(Z(\boldsymbol{x}+\boldsymbol{h}) - Z(\boldsymbol{x})) = \sigma_Z^2(2 - 2r(\boldsymbol{h})), \quad (2)$$

which can be estimated using an empirical variogram introduced by Cressie (1993) and defined as

$$2\hat{\gamma}(\boldsymbol{h}) = \frac{1}{|N(\boldsymbol{h})|} \sum_{N(\boldsymbol{h})} (Z(\boldsymbol{s}_i) - Z(\boldsymbol{s}_j))^2, \qquad \boldsymbol{h} \in R^2, \quad (3)$$

where $N(\boldsymbol{h}) = \{(\boldsymbol{s}_i, \boldsymbol{s}_j) \mid \boldsymbol{s}_i - \boldsymbol{s}_j = \boldsymbol{h}\}$. The shape of $1 - r(\boldsymbol{h})$ may be roughly sketched by this formula. Cressie (1993) advocates selecting $r(\cdot)$ from a parametric family to ensure that the covariance function is positive-definite. In many cases, one or more parameters in the covariance model represent the strength of a spatial correlation within the profile. Some families of the univariate correlation function, such as the cubic, exponential, and Gaussian, can be found in the work of Koeler and Owen (1996). For the 2-D surface, the product correlation function

$$r(\boldsymbol{h}; \theta_1, \theta_2) = r_{\theta_1}(|h^{(1)}|) r_{\theta_2}(|h^{(2)}|)$$

is often used for mathematical convenience. In this expression, $r_\theta(\cdot)$ denotes a univariate correlation function.

Zhao et al. (2011) employed the following Gaussian correlation function to model the surface of the wafer:

$$r_\theta(d) = \exp\left\{-\frac{d^2}{2\theta^2}\right\}.$$

Thus,

$$r(\boldsymbol{x}_1 - \boldsymbol{x}_2; \theta_1, \theta_2) = \exp\left\{-\frac{(x_1^{(1)} - x_2^{(1)})^2}{2\theta_1^2} - \frac{(x_1^{(2)} - x_2^{(2)})^2}{2\theta_2^2}\right\}. \quad (4)$$

The Gaussian correlation function can provide a smooth fit of the spatial data and therefore is adopted in this study. In the simulation study section, we will demonstrate why the Gaussian correlation function is suitable for the variogram of the thickness profile through crude analysis. However, more delicate methods for selecting correlation functions can be developed in an expanded Phase I study. Our method can be easily extended to other types of correlation functions.

We now introduce some important properties of the Gaussian correlation function, many of which also apply to other correlation families. First, the correlation among the response variables is positive; it decreases as the distance between the response variables increases. This property is consistent with our intuition. In addition, θ_1 and θ_2 determine the intensity of the spatial correlation in the $x^{(1)}$ and $x^{(2)}$ direction. The value of the correlation function $r_\theta(d)$ with smaller θ will decrease more rapidly when d increases, as shown in Figure 3. Reflecting in the realization of resulted random field, larger θ_i may generate $Z(\boldsymbol{x})$ with more fluctuations in the direction of

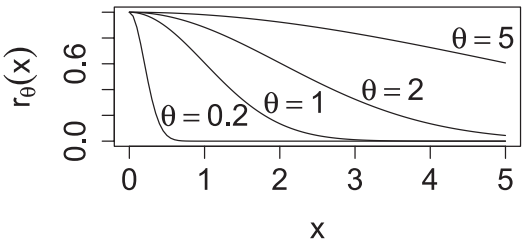

FIGURE 3. The Gaussian Correlation Function $r_\theta(x)$ with Different θ's.

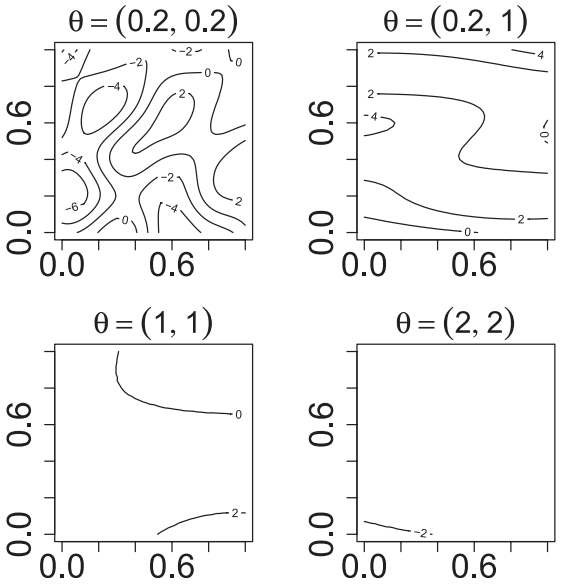

FIGURE 4. The Contours of Four Profiles of a Simulated Gaussian Process $Z(x)$ with Different θ_1's and θ_2's. $\sigma_Z^1 = 10$.

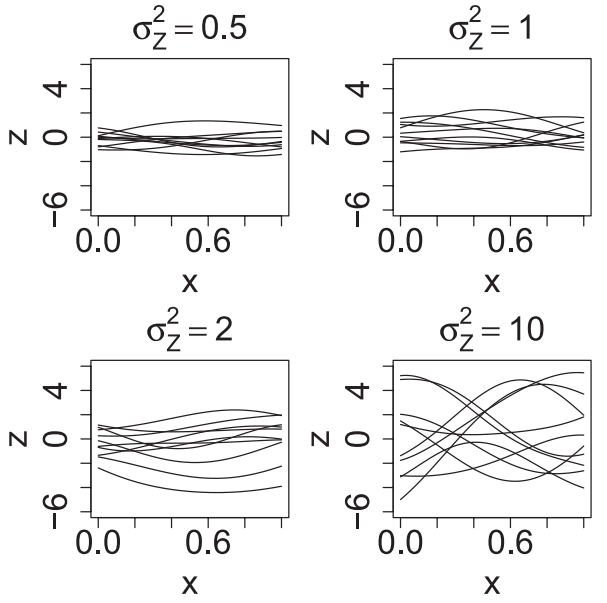

FIGURE 5. The Shape of Ten Simulated One-Dimensional Profiles of a Gaussian Process with $\theta = 0.5$ but Different σ_Z^1's.

$x^{(i)}$, $i = 1, 2$ according to experiments, as is shown in Figure 4, which displays the contours of four simulated $Z(\boldsymbol{x})$'s with the same σ_Z^2 value ($\sigma_Z^2 = 10$) but different θ_1 and θ_2 values. With smaller θ_1 and θ_2 values, the generated surface tends to be wavier in the corresponding directions.

Furthermore, σ_Z^2 may affect the amplitude of the fluctuations. For clear demonstration, we show the effect of σ_Z^2 with a 1-D x. The observations drawn from this study still hold for the above 2-D processes. Figure 5 shows the effect of σ_Z^2 on the shapes of four simulated $Z(x)$'s. The parameter θ for all four processes is 0.5, and σ_Z^2 changes from 0.5, 1, 2, to 10. If the correlation is strong (θ is large), σ_Z^2 may lead to a deviation of the entire profile from its mean value, as illustrated in Figure 6, in which the shapes of simulated $Z(x)$ with 1-D x and $\sigma_Z^2 = 1$, $\theta = 0.02, 0.2, 1,$ and 2 are presented. θ_1, θ_2, and σ_Z^2 determine the pattern of random waviness of the Gaussian term when fitting the model to the wafer data. Therefore, alarms are expected to go off when conditions that affect wafer waviness occur.

When there is strong evidence that the correlation within the profile is isotropic, we can assume $\theta_1 = \theta_2 = \theta$ in the model. However, when monitoring the wafer surfaces, we may rotate the coordinates of the wafer such that the direction of the x_1 axis is parallel to the slicing direction, and the x_2 axis is orthogonal to it. In this manner, assigning independent values for θ_1 and θ_2 has explicit meaning. Hereafter, our method assumes θ_1 and θ_2 are distinct parameters in this study.

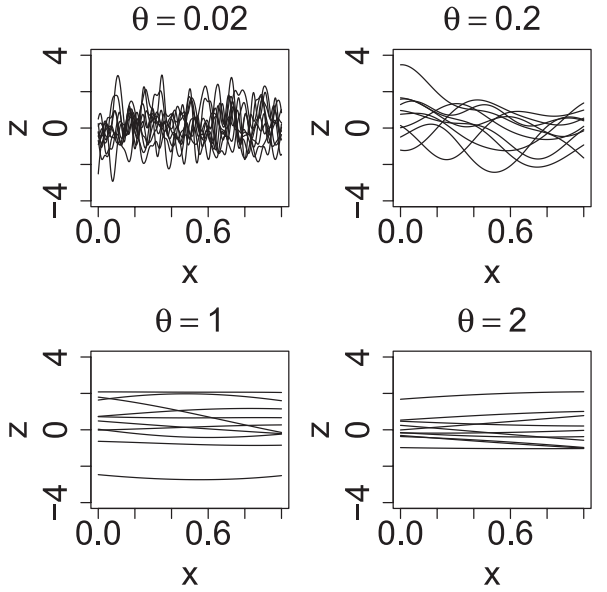

FIGURE 6.The Shape of Ten Simulated One-Dimensional Profiles of a Gaussian Process with $\sigma_Z^1 = 1$ but Different θ's.

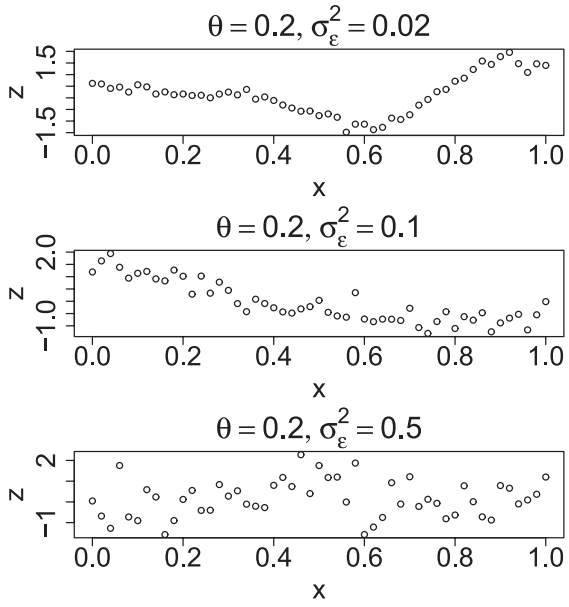

FIGURE 7. The Contours of One-Dimensional Profiles $Z(x) + \varepsilon$ Simulated from Gaussian-Kriging Model, with Different σ_ε^2's.

The third term ε_j represents the random error caused by inaccurate measurements and other unintentional events. We assume that ε_j's are independent and identically distributed (i.i.d.) random variables and follow the $N(0, \sigma_\varepsilon^2)$ distribution, and that each ε_j may affect the "roughness" of the measured surface. Figure 7 shows three simulated discrete 1-D profiles of $Z(x) + \varepsilon$ with no linear trend, the same θ and σ_Z^2 values but different σ_ε^2 values. It is evident that a larger σ_ε^2 leads to rougher surfaces.

The linear trend, the Gaussian field $Z(\boldsymbol{x})$, and the random error ε in the Gaussian-Kriging model represent the characteristics of the surface in different scope: the linear part models the major trend of the wafer, the Gaussian field models the waviness, and the noise part represents the roughness of the measured surface. There exists a total of seven parameters in this model $(\mu, \beta_1, \beta_2; \sigma_\varepsilon^2, \sigma_Z^2, \theta_1, \theta_2)$ and each one has a corresponding engineering implication.

Process Monitoring Based on the Gaussian-Kriging Model

In this section, we propose a control chart that is based on the monitoring of the parameter vectors proposed in the last section to ensure the stability of the wafer's mean thickness, inclination, pattern of waviness, and degree of roughness. The in-control values of these seven parameters are assumed to be known from a Phase I study of historical data. We monitor the stability of these parameters by comparing their estimations from the thickness measurements of each on-line wafer with their in-control values.

We denote the parameter vector by $\boldsymbol{\xi} = (\mu, \beta_1, \beta_2; \sigma_\varepsilon^2, \sigma_Z^2, \theta_1, \theta_2)'$ and assume that each surface profile has m measurements with explanatory variables (positions) as follows:

$$\boldsymbol{x}_1 = (x_1^{(1)}, x_1^{(2)}), \cdots, \boldsymbol{x}_m = (x_m^{(1)}, x_m^{(2)}).$$

The corresponding response variables (thickness) are $(Y_1, \ldots, Y_m)'$. When a Gaussian-Kriging model is used in geostatistics or computer experiments, prediction of $Y(x)$ at a given point is the objective. While in SPC, we are interested in hypothesis testing for the parameter vector $\boldsymbol{\xi}$, which determines the surface characteristics as mentioned in last section. Due to this difference in objectives, there is no well-known hypothesis-testing method for the parameter vector $H_0 : \boldsymbol{\xi} = \boldsymbol{\xi}_0$ in the Gaussian-Kriging model in computer experiments or geostatistics literature. We propose to use a control scheme based on the Wald test and multivariate control charts, to detect shifts in the parameters in the Gaussian-Kriging model. We obtain an estimator of the parametric vector and its corresponding covariance and construct a T^2 type statistic to monitor the shift in the parametric vector with the maximum-likelihood estimation of the vector and its asymptotic covariance.

Step 1: Obtain the Maximum-Likelihood Estimator (MLE) from Each Surface Sample

First, we derive the joint probability distribution of observations within a surface sample. Let

$$\boldsymbol{y} = (Y_1, \ldots, Y_m)',$$
$$\boldsymbol{z} = (Z(\boldsymbol{x}_1), \ldots, Z(\boldsymbol{x}_m))'$$

be the response variables and their corresponding Gaussian field terms. Based on the Gaussian field described in the previous section, we obtain

$$\mathrm{Cov}(Z(\boldsymbol{x}_{j_1}), Z(\boldsymbol{x}_{j_2})) = \sigma_Z^2 r(\boldsymbol{x}_{j_1} - \boldsymbol{x}_{j_2}; \theta_1, \theta_2),$$

where $r(\cdot; \theta_1, \theta_2)$ is calculated from Equation (4). Define $\boldsymbol{R}(\theta_1, \theta_2)$ as the matrix of which the (j_1, j_2)th element is $r(\boldsymbol{x}_{j_1} - \boldsymbol{x}_{j_2}; \theta_1, \theta_2)$. The covariance matrix of the Gaussian field is given by

$$\mathrm{Cov}(\boldsymbol{z}) = \sigma_Z^2 \cdot \boldsymbol{R}(\theta_1, \theta_2).$$

Furthermore,

$$\begin{aligned}\mathrm{Cov}(\boldsymbol{z}+\boldsymbol{\varepsilon}) &= \mathrm{Cov}(\boldsymbol{z},\boldsymbol{z}) + \mathrm{Cov}(\boldsymbol{\varepsilon},\boldsymbol{\varepsilon}) \\ &= \sigma_Z^2 \cdot \boldsymbol{R}(\theta_1,\theta_2) + \sigma_\varepsilon^2 \boldsymbol{I}_{m\times m}.\end{aligned}$$

Thus,

$$\mathrm{Cov}(\boldsymbol{y},\boldsymbol{y}) = \sigma_Z^2 \cdot \boldsymbol{R}(\theta_1,\theta_2) + \sigma_\varepsilon^2 \boldsymbol{I}_{m\times m}.$$

$\mathrm{Cov}(\boldsymbol{y},\boldsymbol{y})$ is determined by σ_ε^2, σ_Z^2, θ_1, and θ_2. In addition, the expectation of each element of $\boldsymbol{y}$ satisfies

$$\mathrm{E}(Y_i) = \mu + \beta_1\, x_i^{(1)} + \beta_2\, x_i^{(2)},$$

which is dominated by μ, β_1, and β_2.

This representation of the mean and covariance of $\boldsymbol{y}$ motivates us to split the seven parameters into two groups, corresponding to the mean and covariance of the profile, as follows:

$$\boldsymbol{\xi} = (\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)',$$

where

$$\boldsymbol{\xi}_1 = (\mu, \beta_1, \beta_2)',$$

and

$$\boldsymbol{\xi}_2 = (\sigma_\varepsilon^2, \sigma_Z^2, \theta_1, \theta_2)'.$$

Let

$$\boldsymbol{B} = \begin{pmatrix} 1 & x_1^{(1)} & x_1^{(2)} \\ \vdots & \vdots & \vdots \\ 1 & x_m^{(1)} & x_m^{(2)} \end{pmatrix},$$

$$\boldsymbol{\Sigma}(\boldsymbol{\xi}_2) = \sigma_Z^2 \cdot \boldsymbol{R}(\theta_1,\theta_2) + \sigma_\varepsilon^2 \boldsymbol{I}_{m\times m},$$

thus,

$$\boldsymbol{y} \sim N(\boldsymbol{B}\boldsymbol{\xi}_1, \boldsymbol{\Sigma}(\boldsymbol{\xi}_2)). \tag{5}$$

By obtaining the value of $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ that maximizes the log-likelihood function of this distribution, we can determine the MLE of $\boldsymbol{\xi}$. Fang et al. (2006) developed a general algorithm for calculating $\hat{\boldsymbol{\xi}}$. We extend it for our use (see details in the Appendix).

Step 2: Calculate the Covariance of the Estimators

In previous studies, such as those of Colosimo (2008), Williams et al. (2007), and Jensen et al. (2008), the covariance of the estimator is often obtained from a data-driven method; the variance of the estimated parameters from the historical profiles is treated as the variance of the estimators. Therefore, this estimation is highly dependent on available historical data. In this study, we consider using the theoretical value of the MLE's covariance. Mardia and Marshall (1984) proved that, when the number of data points and the domain of the design points tends to infinity, the MLE for the Gaussian process exhibits a weak consistency and $\hat{\boldsymbol{\xi}}$ asymptotically follows a multivariate normal distribution with the inverse of the Fisher information matrix as the covariance, that is,

$$\hat{\boldsymbol{\xi}} \xrightarrow{d} N(\boldsymbol{\xi}, \boldsymbol{I}^{-1}(\boldsymbol{\xi})).$$

The Fisher information matrix is defined in the Appendix. According to Mardia and Marshall (1984), $\boldsymbol{I}(\boldsymbol{\xi})$ is block-diagonal,

$$\boldsymbol{I}(\boldsymbol{\xi}) = \mathrm{diag}(\boldsymbol{I}_1(\boldsymbol{\xi}), \boldsymbol{I}_2(\boldsymbol{\xi})), \tag{6}$$

and has the representation of

$$\boldsymbol{I}_1(\boldsymbol{\xi}) = \boldsymbol{B}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\xi}_2)\boldsymbol{B}, \tag{7}$$

$$[\boldsymbol{I}_2(\boldsymbol{\xi})]_{i,j} = \frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\xi_2^{(i)}}\boldsymbol{\Sigma}^{-1}\frac{\partial\boldsymbol{\Sigma}}{\partial\xi_2^{(j)}}\right), \quad i,j = 1,\ldots,4. \tag{8}$$

Here $\xi_2^{(i)}$ denotes the ith component of $\boldsymbol{\xi}_2$.

Step 3: Construct T^2 Statistics to Monitor the Sampled Surface

Let $\boldsymbol{\xi}_{\mathrm{IC}} = (\boldsymbol{\xi}_{1,\mathrm{IC}}, \boldsymbol{\xi}_{2,\mathrm{IC}})$ be the in-control (IC) parameter vector obtained from Phase I and let $\boldsymbol{\xi} = (\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)$ be the real parameter from which the surface sample is generated. By considering the hypothesis testing problem $H_0 : \boldsymbol{\xi} = \boldsymbol{\xi}_{\mathrm{IC}}$, a multivariate T^2 control chart is designed,

$$\begin{aligned}T^2 &= (\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_{\mathrm{IC}})'\boldsymbol{I}(\boldsymbol{\xi}_{\mathrm{IC}})(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_{\mathrm{IC}}) \\ &= (\hat{\boldsymbol{\xi}}_1 - \boldsymbol{\xi}_{1,\mathrm{IC}}, \hat{\boldsymbol{\xi}}_2 - \boldsymbol{\xi}_{2,\mathrm{IC}})'\mathrm{diag}(\boldsymbol{I}_1(\boldsymbol{\xi}_1), \boldsymbol{I}_2(\boldsymbol{\xi}_2)) \\ &\quad \cdot (\hat{\boldsymbol{\xi}}_1 - \boldsymbol{\xi}_{1,\mathrm{IC}}, \hat{\boldsymbol{\xi}}_2 - \boldsymbol{\xi}_{2,\mathrm{IC}}) \\ &= (\hat{\boldsymbol{\xi}}_1 - \boldsymbol{\xi}_{1,\mathrm{IC}})'\boldsymbol{I}(\boldsymbol{\xi}_{1,\mathrm{IC}})(\hat{\boldsymbol{\xi}}_1 - \boldsymbol{\xi}_{1,\mathrm{IC}}) \\ &\quad + (\hat{\boldsymbol{\xi}}_2 - \boldsymbol{\xi}_{2,\mathrm{IC}})'\boldsymbol{I}(\boldsymbol{\xi}_{2,\mathrm{IC}})(\hat{\boldsymbol{\xi}}_2 - \boldsymbol{\xi}_{2,\mathrm{IC}}) \\ &= T_1^2 + T_2^2,\end{aligned}$$

where

$$\begin{aligned}T_1^2 &= (\hat{\boldsymbol{\xi}}_1 - \boldsymbol{\xi}_{1,\mathrm{IC}})'\boldsymbol{I}(\boldsymbol{\xi}_{1,\mathrm{IC}})(\hat{\boldsymbol{\xi}}_1 - \boldsymbol{\xi}_{1,\mathrm{IC}}), \\ T_2^2 &= (\hat{\boldsymbol{\xi}}_2 - \boldsymbol{\xi}_{2,\mathrm{IC}})'\boldsymbol{I}(\boldsymbol{\xi}_{2,\mathrm{IC}})(\hat{\boldsymbol{\xi}}_2 - \boldsymbol{\xi}_{2,\mathrm{IC}}).\end{aligned}$$

According to the asymptotic property of $\hat{\boldsymbol{\xi}}$, when the process is in control and the number of measurements is sufficiently large, this T^2 approximately follows the $\chi^2(7)$ distribution. Thus, we can calculate and monitor the T^2 statistic for each surface sample: if T^2 is larger than a specified control limit, the control chart signals an alarm. We name this chart the individual surface chart (ISC) because it monitors a single T^2 statistic.

In addition to monitoring T^2, there is another possible monitoring scheme. Because the Fisher information matrix for $\boldsymbol{\xi}$ is a block-diagonal matrix (see Appendix) and $\hat{\boldsymbol{\xi}}$ asymptotically follows a normal distribution, the MLEs of the two groups of parameters ($\hat{\boldsymbol{\xi}}_1$ and $\hat{\boldsymbol{\xi}}_2$) are approximately independent if a sufficient number of measurements exists. This notion suggests that T_1^2 and T_2^2 are approximately independent and are asymptotically distributed as $\chi^2(3)$ and $\chi^2(4)$, respectively. Therefore, we can construct two independent control charts to monitor T_1^2 and T_2^2. Given the specific control limits UCL_1 and UCL_2, if the surface sample achieves $T_1^2 > \mathrm{UCL}_1$ or $T_2^2 > \mathrm{UCL}_2$, the process signals an alarm. We name this chart the separate surface chart (SSC). The performance of the ISC and the SSC will be compared in a subsequent section.

Note that the proposed T^2 chart can be easily improved by incorporating the multivariate EWMA procedure; such an improvement should make the chart more effective in detecting small shifts in the parameters. Although similar approaches have been thoroughly discussed in the literature, we do not include this enhancement in the following simulation studies.

Diagnostic information is usually desired when a multivariate chart signals an alarm. Due to the correlations among the parameter estimates, we may not be able to attribute the alarm to a single variable. However, the traditional diagnostic methods proposed for the general T^2 chart, such as the methods proposed by Mason et al. (1995), can also be used with the proposed chart. In addition, the separate monitoring scheme that was illustrated previously can provide clues regarding the set of variables that shift, because T_1^2 consists of parameters that correspond to macro-thickness patterns, whereas T_2^2 only consists of micro-level roughness parameters.

Simulation Studies

This study is motivated by real data collected from a semiconductor manufacturing process. By constructing a model for surface characterization and monitoring, we have explicitly employed some assumptions about the data. In this section, we verify the assumptions using real data and simulation methods and then compare the performance of the proposed chart with that of a modified profile chart.

Verification of the Correlation Functions

The previous deviation is dependent on the assumption of spatial correlation within the wafer sample. In this section, we use the variogram introduced previously to show the existence of spatial correlation within the profile of the wafer's thickness and verify the validity of the Gaussian correlation function. There are 3912 measurement points on a wafer's surface. To fully utilize the data and for ease of exposition, we sketched the isotropic variogram instead of the directed variogram of Equation (3). Other modifications are needed to solve the problem that it cannot guarantee that $\gamma(h)$ is well defined for a series of equally spaced h, because the position of measurements on each wafer is not lattice and the values of h which makes $|N(h)| \neq 0$ may be irregular. On each wafer, the distance between two measurement varies between 0 (approximately the nearest distance between two measurements) and 2 (approximately the largest distance between two measurements on wafer with radius 1). We divide $[0, 2]$ into n equal intervals $N_k = (k\delta, (k+1)\delta)$, $\delta = 2/n$, and $k = 0, \ldots, n-1$, and calculate the mean of $(Z(\boldsymbol{s}_i) - Z(\boldsymbol{s}_j))^2$, $|\boldsymbol{s}_i - \boldsymbol{s}_j| \in N_k$ to sketch the isotropic empirical variogram at $k\delta$. We selected $n = 50$ because 50 intervals is big enough to let a trend in $\gamma(h)$ be observed and not so small to ensure the accuracy of each estimated $\gamma(h)$, with adequate pairs of $(\boldsymbol{s}_i, \boldsymbol{s}_j)$ fall in each interval. The isotropic empirical variogram we used is

$$2\tilde{\gamma}(k\delta) = \frac{1}{|N_k|} \sum_{|\boldsymbol{s}_i - \boldsymbol{s}_j| \in N_k} (Z(\boldsymbol{s}_i) - Z(\boldsymbol{s}_j))^2,$$

with $\delta = 2/50 = 0.04$ and $1 \leq k \leq 50$. Figure 8 sketches its trend.

Fang et al. (2006) introduced some commonly used univariate correlation functions, such as the family of linear correlation functions with

$$r(x; \theta) = \max\left\{1 - \frac{x}{\theta l}, 0\right\},$$

where l is the range of x, and the family of exponential correlation functions with

$$r(x; \theta) = \exp\left\{-\frac{x}{\theta}\right\}.$$

There is also the family of cubic correlation functions introduced by Koeler and Owen (1996), with

$$r(x; \gamma) = 1 - \frac{3(1-\rho)}{2+\gamma} x^2 + \frac{(1-\rho)(1-\gamma)}{2+\gamma} x^3,$$

where $\rho = \mathrm{Corr}(Y(0), Y(1))$ is the correlation between observations with a unit distance and $\gamma =$

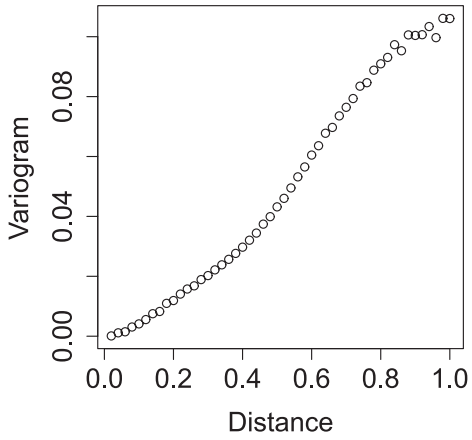

FIGURE 8. The Empirical Variogram Obtained from the Measurements on a Sample Wafer.

$\mathrm{Corr}(Y'(0), Y'(1))$ is the correlation between two observations of the derivative process with a unit distance. The family of Gaussian correlation functions has the representation of

$$r(x;\theta) = \exp\left\{-\frac{x^2}{2\theta^2}\right\}.$$

According to the relationship between variogram and correlation function in Equation (2), these correlation function's corresponding variograms are graphically shown in Figure 9.

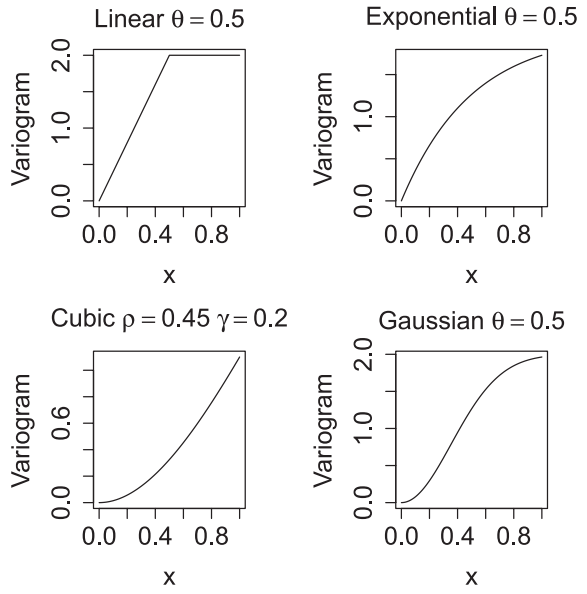

FIGURE 9. The Corresponding Variogram of Several Common Univariate Correlation Functions: Linear, Exponential, Cubic, and Gaussian.

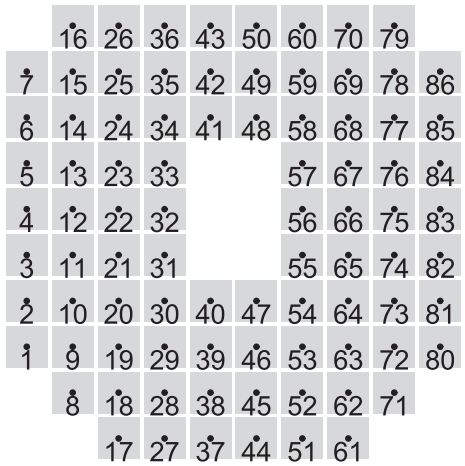

FIGURE 10. The Positions of the 86 Resampled Data Points on Each Wafer.

The ascending trend of the empirical variogram suggests that spatial correlation can be reasonably assumed. By comparing the shape of the empirical variogram and the shapes of the model variograms corresponding to various families of correlation functions, we find that Gaussian correlation functions in Figure 9 and the empirical variogram in Figure 8 have similar convexity property, and so we select a Gaussian correlation function to construct the model in this study.

Simulation Settings for the Verification of Asymptotic Distributions and Performance Comparison

We introduce our simulation settings prior to conducting simulation studies that examine the performance of the proposed scheme.

In the Gaussian-Kriging model, the computational speed of the iterative algorithm for calculating MLEs is assumed to be slow when m, the number of data points within one surface sample, is large. That is because the inversion of an $m \times m$-order covariance matrix is required in each step of the iteration. It is practically impossible to employ all 3912 data points to estimate the MLE. For this reason, we resample the 3912 data points to obtain 86 lattice points prior to constructing the Gaussian-Kriging model. In practice, which data reduction technique to use depends on the application, and special consideration should be taken to preserve as much information as possible. In this study, we select 86 data points to preserve the predominant thickness information of the wafer

and to computationally manipulate the problem. The uniform sampling scheme is adopted so that these 86 positions are arranged on a lattice (see Figure 10). There are no data points in the center area due to measurement constraints. Jin et al. (2012) can be referred to as a sequential sampling strategy based on Gaussian-Kriging model, by which the number and positions of measurement points can be determined.

Based on a conforming real wafer, the maximum-likelihood algorithm yields the following estimates:

$$\begin{aligned} \mu &= 545.91, \\ \beta_1 &= 0.167, \\ \beta_2 &= -0.0316; \\ \sigma_\varepsilon^2 &= 0.0000618, \\ \sigma_Z^2 &= 0.0120, \\ \theta_1 &= 0.276, \theta_2 = 0.464. \end{aligned} \tag{9}$$

In what follows, we treat this parameter vector as the in-control value for the performance study, with the exception of θ_1 and θ_2, which will be assigned with arbitrary values that represent either strong or weak spatial correlations. In practice, however, in-control parameter estimation should be studied more rigorously in Phase I.

According to Ranjan et al. (2011), when any pair of measurements are close in position, the covariance matrix may become either singular or not positive definite. We also found numerical problems may also occur when the spatial correlation within the profile becomes too weak: intuitively, when θ_1 and θ_2 are too small, the $Z(\boldsymbol{x}_i)$ terms of different observations will be nearly independent of each other, and they may not be distinguishable from ε_i. This indicates that the proposed Gaussian-Kriging model is not suitable for monitoring surfaces with weak spatial correlation. Some researchers (e.g., Fang et al. (2006)) have advocated the use of a penalized-likelihood approach to obtain a stable estimation. To avoid the appearance of a singular matrix in parameter estimation, Ankenman et al. (2010) suggested estimating the variance of ε first when using the Gaussian-Kriging model with ε terms, and then estimating other parameters assuming σ_ε^2 is known. In this study, it is possible to first estimate $\sigma_\varepsilon^2 = \lim_{h\to 0}\hat{\gamma}(h)$ because $\lim_{h\to 0} r(h) = 0$ and

$$\begin{aligned} \gamma(h) &= \frac{1}{2}\mathrm{Var}(Z(x+h) - Z(x)) \\ &= \sigma_e^2 + \sigma_z^2(1 - \mathrm{Corr}(Z(x), Z(x+h))) \\ &= \sigma_e^2 + \sigma_z^2(1 - r(h)). \end{aligned}$$

Then the estimation of other parameters can be obtained in a maximum-likelihood method. Ranjan et al. (2011) proposed a lower bound of $\sigma_\varepsilon^2/\sigma_z^2$ to deal with singularity problems and recommended that a small reduction from $p_k = 2$ in $r(x;\theta) = \exp(-x^{p_k}/\theta^2)$ to $p_k = 1.99$ or 1.95 may decrease the occurrence of near-singular cases. However, the methods mentioned above may not yield a simple representation of the covariances of the parameter vector. Therefore, we do not include these special measures in this study, but maintain the original control scheme instead, and limit the application of our control scheme to situations in which the spatial correlation is not excessively weak. To further lower the probability of nonconvergence, we adopt the following measures: a) restrict the value of the parameters to a reasonable region (σ_Z^2 and σ_ε^2 must be positive and θ_1 and θ_2 should not be too small); b) check the value of the likelihood function. If the likelihood function decreases, shorten the length of the step in searching for the MLE. And c) if the MLE value switches among several values and the algorithm fails to stop after 25 steps, we employ the average of the last five values as the estimator. Our simulation study confirms that the chance of nonconvergency can be disregarded with these ad-hoc measures.

Verification of the Asymptotic Distributions of the MLEs and T^2 Statistics

In the proposed chart, we assumed that the domain and number of measurements are big enough so that the MLEs of the parameters are approximately normally distributed. To verify this assumption, we computed 500 parameter estimates from 500 computer-generated surfaces (using Equation (5)) with parameters in Equation (9). The matrix plot of these estimates is shown in Figure 11, which illustrates that the joint distribution of each pair of components is approximately elliptical and indicates that the pair approximately follows a 2-D normal distribution.

Using the representation of the Fisher information matrix in Equations (6), (7), and (8), we can calculate the theoretical asymptotic covariance of the MLE, i.e., the inverse of the Fisher information matrix. Subsequently, the T^2, T_1^2, T_2^2 statistics are obtained. The empirical cumulative distribution functions (c.d.f.) of T_1^2 and T_2^2, which are shown in Figure 12, are compared with the cumulative distribution function of $\chi^2(3)$ distribution and $\chi^2(4)$ distributions, respectively. The empirical c.d.f.'s are quite

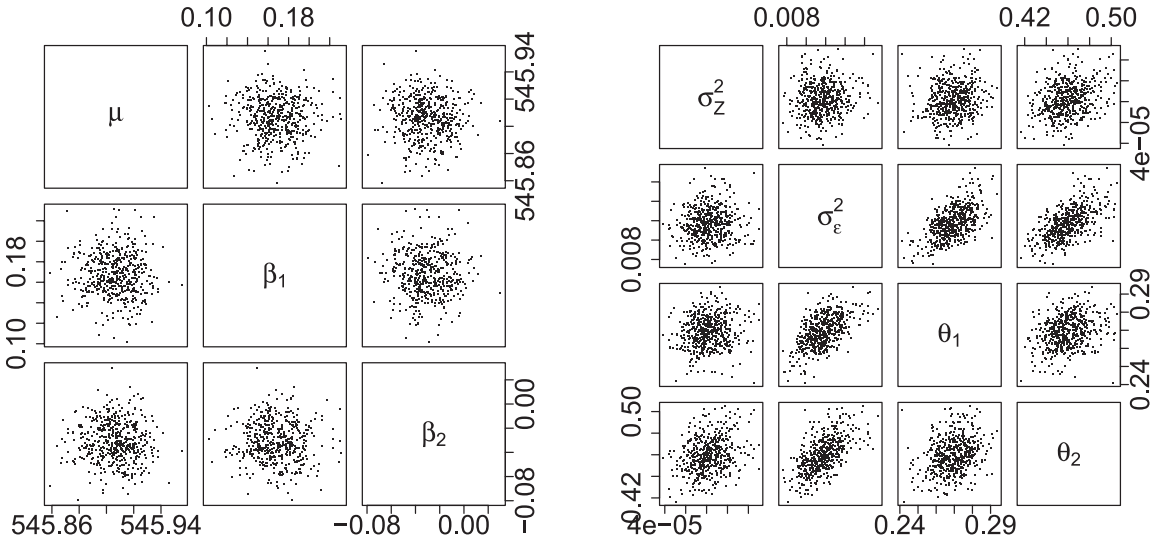

FIGURE 11. The Scatter Plots of 500 MLEs of Parameters in $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$. Each pair of components presents an elliptical-shaped distribution.

similar to the theoretical c.d.f.'s, suggesting that the assumption of an asymptotic distribution is valid.

The Comparison of Performance

In this section, we compare the performance of our proposed ISC and SSC with the traditional multiple regression chart. In SPC, the performance of control charts is usually evaluated by comparing their average run length (ARL), which is defined as the average number of steps needed for the charts to signal. The control limits of all competing charts are adjusted so that each chart has the same ARL value, ARL_{in}, when the process is in control. When unexpected process changes occur and the process goes out of control, the ARL of the control charts in such cases, ARL_{out}, should be shorter than ARL_{in}, and the chart having the shortest ARL_{out} is regarded as the one with the best charting performance.

Next we introduce the multiple regression chart, which is used as a benchmark in this simulation

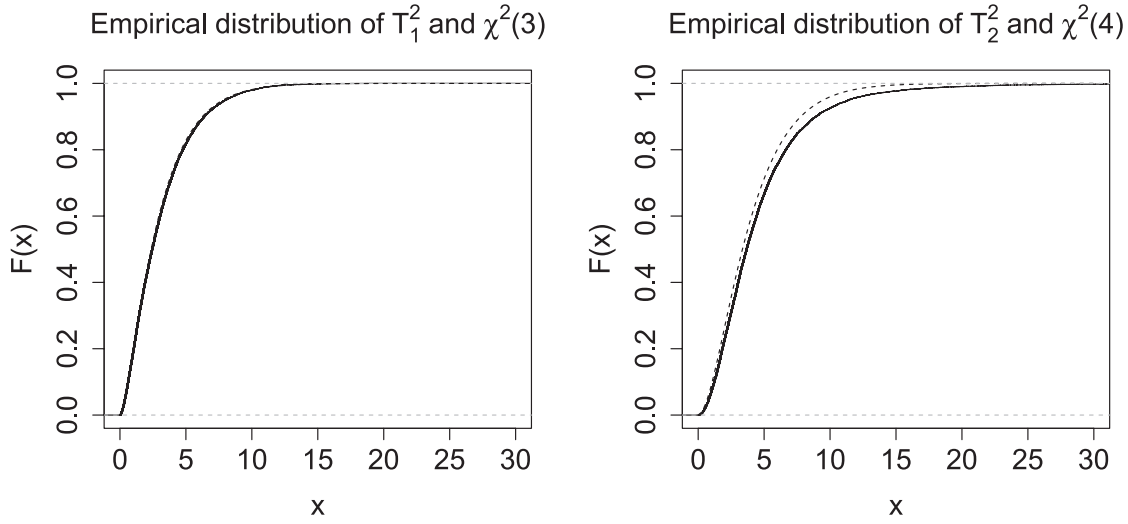

FIGURE 12. The Empirical c.d.f. of T_1^2 and T_2^2 Statistics (Indicated by Solid Curves) and the c.d.f. of $\chi^2(3)$ and $\chi^2(4)$ Distributions (Indicated by Dotted Curves). The dotted curve is hard to see in the first figure because the empirical c.d.f. of T_1^2 is very close to that of $\chi^2(3)$.

study. In the multiple regression model, the relationship between the response variable Y_j and the explanatory variables $x_j^{(1)}$ and $x_j^{(2)}$ is given by the following equation:

$$Y_j = \mu + \beta_1 x_j^{(1)} + \beta_2 x_j^{(2)} + \varepsilon_j. \tag{10}$$

In the model, the variation of the profile is simply represented by identically distributed normal random variables ε_j with mean 0 and variance σ^2. The spatial correlation within the data is disregarded. Zou et al. (2007) proposed a MEWMA chart to monitor the linear parameters ($\boldsymbol{\xi}_1 = (\mu, \beta_1, \beta_2)$) and the variance of the noise terms (σ^2) of a multiple linear profile model. This chart is modified a little to achieve a fair comparison with our proposed surface charts, ISC and SSC. The details are shown as follow.

First, Zou et al. (2007) transformed all estimated parameters of interest into a multivariate normally distributed random vector and then employed a MEWMA chart to monitor the vector. Because the MEWMA chart is more effective for small shifts, we only apply Shewhart-type or multivariate T^2 charts for a reasonable comparison. If small shifts are of interest, the scheme proposed in this study can be easily combined with an EWMA procedure to improve the performance, and then reasonably compared with the original MEWMA scheme of Zou et al. (2007).

In Zou et al. (2007), a χ^2 distributed statistic $\tilde{Z}_j(\sigma) = (n-p)\hat{\sigma}_j^2/\sigma^2$ is proposed for monitoring the variance of the noise term, (σ^2) and a multivariate-normally distributed statistic $\tilde{Z}_j(\boldsymbol{\xi}) = (\hat{\boldsymbol{\xi}}_j - \boldsymbol{\xi})/\sigma$ is proposed for monitoring the linear coefficients). In order to incorporate them in one chart, Zou et al. (2007) transformed $\tilde{Z}_j(\sigma)$ to a standard normal random variable $\tilde{Z}_j^*(\sigma)$ and combined $\tilde{Z}_j^*(\sigma)$ and $\tilde{Z}_j(\boldsymbol{\xi})$ in one multivariate-normal distributed vector. However, when spatial correlation exists, $\tilde{Z}_j(\sigma)$ is no longer χ^2 distributed, and, if we transform $\tilde{Z}_j(\sigma)$ to $\tilde{Z}_j^*(\sigma)$ using the same technique, the distribution of the resulting $\tilde{Z}_j^*(\sigma)$ will deviate dramatically from $N(0,1)$, and the performance of the charts will be critically affected. Therefore, for comparison purposes, we employ two control charts instead to monitoring $\tilde{Z}_j(\boldsymbol{\xi})$ and $\tilde{Z}_j(\sigma)$ separately, which we named the $\boldsymbol{\xi}$ chart and the σ chart. For the $\boldsymbol{\xi}$ chart, we select $\mathrm{UCL}_{\boldsymbol{\xi}}$, and when $\tilde{Z}_j(\boldsymbol{\xi})'(X'X)\tilde{Z}_j(\boldsymbol{\xi}) > \mathrm{UCL}_{\boldsymbol{\xi}}$, the $\boldsymbol{\xi}$ chart signals an alarm. For the σ chart, we select UCL_σ and LCL_σ, such that when $\tilde{Z}_j(\sigma)$ is greater than UCL_σ or smaller than LCL_σ, the σ chart signals an alarm. Assuming that the monitoring of $\boldsymbol{\xi}$ and σ^2 has similar importance, we select these control limits so that the $\boldsymbol{\xi}$ chart and σ chart have equivalent in-control ARL, thus equivalent in-control alarm probability as follows:

$$\begin{aligned} &P(\tilde{Z}_j(\sigma) > \mathrm{UCL}_\sigma \text{ or } \tilde{Z}_j(\sigma) < \mathrm{LCL}_\sigma) \\ &= P(\tilde{Z}_j(\boldsymbol{\xi})'(X'X)\tilde{Z}_j(\boldsymbol{\xi}) > \mathrm{UCL}_{\boldsymbol{\xi}}). \end{aligned}$$

Assuming the σ chart can detect an increase and a decrease in σ with similar efficiencies, let

$$P(\tilde{Z}_j(\sigma) > \mathrm{UCL}_\sigma) = P(\tilde{Z}_j(\sigma) < \mathrm{LCL}_\sigma).$$

The $\boldsymbol{\xi}$ chart and σ chart are used collectively and the control limits are selected to obtain a predetermined joint in-control ARL. Because we do not employ nominal control limits (the quantile of the χ^2 distribution) from the theoretical distributions, it is obvious that, once the in-control ARL is selected, the performance of these charts is not dependent on the selection of the model's in-control parameter σ; we may substitute $\tilde{Z}_j(\boldsymbol{\xi})'(X'X)\tilde{Z}_j(\boldsymbol{\xi})$ with $(\hat{\boldsymbol{\xi}}_j - \boldsymbol{\xi})'(X'X)(\hat{\boldsymbol{\xi}}_j - \boldsymbol{\xi})$ in the $\boldsymbol{\xi}$ chart and substitute $\tilde{Z}(\sigma)$ with $\hat{\sigma}^2$ in the σ chart.

In conclusion, to provide a reasonable comparison with our ISC and SSC, the chart for monitoring linear profiles is modified and implemented as follows: (a) Similar to the Gaussian- Kriging model, adopt the in-control value of $\boldsymbol{\xi} = (\mu, \beta_1, \beta_2)'$. (b) For each sample profile, calculate the least-square estimator of $\boldsymbol{\xi}$, which is denoted by $\hat{\boldsymbol{\xi}}$, and the unbiased estimator of σ^2, which is denoted by $\hat{\sigma}^2$. (c) Calculate the charting statistics, $Z(\boldsymbol{\xi}) = (\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})'(\boldsymbol{x}'\boldsymbol{x})^{-1}(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})$ and $Z(\sigma) = \hat{\sigma}^2$. (d) If $Z(\boldsymbol{\xi}) > \mathrm{UCL}_{\boldsymbol{\xi}}$, $Z(\sigma) > \mathrm{UCL}_\sigma$ or $Z(\sigma) < \mathrm{LCL}_\sigma$, the process signals an alarm. $\mathrm{UCL}_{\boldsymbol{\xi}}$, UCL_σ and LCL_σ are chosen such that

$$P(Z(\sigma) > \mathrm{UCL}_\sigma) = P(Z(\sigma) < \mathrm{LCL}_\sigma)$$

and

$$\begin{aligned} &P(Z(\sigma) > \mathrm{UCL}_\sigma \text{ or } Z(\sigma) < \mathrm{LCL}_\sigma) \\ &= P(Z(\boldsymbol{\xi}) > \mathrm{UCL}_{\boldsymbol{\xi}}). \end{aligned}$$

Because this method was originally proposed for linear profile monitoring, we hereafter name this chart the LPC.

Next we compare the LPC with the ISC and SSC proposed in this work. In our simulation studies, all random surface samples are generated from the Gaussian-Kriging model. The in-control parameters of the model, except for θ_1 and θ_2, are set to be the same as the estimates given in Equation (9). A preliminary study of historical wafer samples shows that θ_1 and θ_2 usually fall within 0.22 and 0.61. Therefore,

in the following, we study the performance of the control charts under $\theta_1 = \theta_2 = 0.4$ and $\theta_1 = \theta_2 = 0.6$, which represent medium and strong spatial correlations of the wafer measurements. Surfaces with even weaker spatial correlations (generated by θ_1 or θ_2 values less than 0.2) were not examined because the Gaussian-Kriging model may not be capable of handling surfaces with very weak spatial correlations properly. In summary, the in-control parameters for our two groups of simulation study are

$$\begin{aligned}\mu &= 545.91;\\ \beta_1 &= 0.167;\\ \beta_2 &= -0.0316;\\ \sigma_\varepsilon^2 &= 0.0000618;\\ \sigma_Z^2 &= 0.0120;\\ \theta_1 &= 0.400;\\ \theta_2 &= 0.400,\end{aligned}$$

and

$$\begin{aligned}\mu &= 545.91;\\ \beta_1 &= 0.167;\\ \beta_2 &= -0.0316;\\ \sigma_\varepsilon^2 &= 0.0000618;\\ \sigma_Z^2 &= 0.0120;\\ \theta_1 &= 0.600;\\ \theta_2 &= 0.600.\end{aligned}$$

We consider the out-of-control conditions where each parameter is shifted by $1sd$ and $2sd$, where sd denotes the standard deviation of each parameter's MLE. The MLEs are derived from the diagonal elements of the inverse of the Fisher information matrix. Because the engineering implications of β_1, β_2 and θ_1, θ_2 are equivalent, we modify β_1 and θ_1 but maintain the original β_2 and θ_2. In addition, the shifts in μ or β_1 of the same magnitude but in different directions should yield the same ARL performance, because the variance of $\hat{\mu}$ and $\hat{\beta}_1$ is not related to their real values. Thus, we do not include the cases in which μ or β_1 has a negative shift. A total of 16 out-of-control combinations and their corresponding in-control ARLs are provided in Table 1. Each ARL is calculated from 5000 replicates. The findings from this ARL comparison study are summarized below.

First, one parameter in linear coefficients $\boldsymbol{\xi}_1$ shifts. When μ increases, the ISC outperforms the LPC. The SSC outperforms the ISC in both cases, when $\theta_1 = \theta_2 = 0.4$ and when $\theta_1 = \theta_2 = 0.6$. When β_1 increases, the ISC becomes the worst chart, whereas the SSC remains the best chart.

Second, for the shifts in $\boldsymbol{\xi}_2$, we studied the cases in which the parameters shifted either upward or downward. Unlike situations in which the parameter in $\boldsymbol{\xi}_1$ shifts, the ISC is more sensitive than the SSC in most

TABLE 1. The ARLs with Zero or One-Parameter Shift

Shift		$\boldsymbol{\theta}_{\text{IC}} = (0.4, 0.4)$			$\boldsymbol{\theta}_{\text{IC}} = (0.6, 0.6)$		
		ISC	SSC	LPC	ISC	SSC	LPC
IC		201.26	208.36	202.27	201.45	199.83	197.93
μ	+1 sd.	150.92	76.03	192.07	175.72	77.81	189.85
	+2 sd.	52.95	12.96	160.19	107.31	13.60	157.67
β_1	+1 sd.	150.11	74.84	98.53	178.50	76.69	113.99
	+2 sd.	52.96	12.66	24.09	108.01	13.45	34.33
σ_ε^2	−2 sd.	218.06	263.82	195.63	232.89	278.09	192.05
	−1 sd.	262.16	245.33	205.05	235.20	238.22	191.58
	+1 sd.	65.54	95.77	205.58	118.07	154.75	202.77
	+2 sd.	15.59	24.13	204.25	35.49	73.23	206.01
σ_Z^2	−2 sd.	12.52	14.73	4.09	10.73	14.40	1.74
	−1 sd.	67.44	97.01	71.38	66.75	108.73	58.01
	+1 sd.	52.40	47.29	41.89	76.53	45.98	34.06
	+2 sd.	8.93	11.13	13.03	15.13	13.30	10.77
θ_1	−2 sd.	46.04	98.23	336.42	92.71	111.75	214.21
	−1 sd.	185.20	233.82	257.47	268.55	257.38	217.82
	+1 sd.	57.98	78.43	156.05	69.97	99.25	160.98
	+2 sd.	13.00	19.68	124.09	20.01	33.79	122.07

TABLE 2. The ARLs with σ_Z^2 Shift and Five Profiles in Each Subgroup

Shift		$\boldsymbol{\theta}_{\mathrm{IC}} = (0.4, 0.4)$			$\boldsymbol{\theta}_{\mathrm{IC}} = (0.6, 0.6)$		
		ISC	SSC	LPC	ISC	SSC	LPC
IC		196.27	203.22	201.15	211.81	199.51	196.42
σ_Z^2	-2 sd.	1.00	1.00	1.01	1.00	1.00	1.00
	-1 sd.	4.15	3.66	6.38	4.75	4.42	6.54
	$+1$ sd.	3.31	3.68	13.31	3.51	4.36	15.40
	$+2$ sd.	1.07	1.07	2.63	1.10	1.14	3.53

cases when the parameter in $\boldsymbol{\xi}_2$ shifts. Specific findings are summarized as follows:

- Shifts in σ_ε^2. The ARLs of the ISC and SSC change as σ_ε^2 shifts. In particular, the ARLs of the ISC or SSC decrease when σ_ε^2 increases. However, when σ_ε^2 decreases, the ARLs increase monotonically for the SSC; it first increases and then decreases for the ISC. Calculation of the Fisher information matrix indicates that, when σ_ε^2 decreases, the variance of its estimator $\mathrm{Var}(\hat{\sigma}_\varepsilon^2)$ also decreases. When small downward shifts occur, the effect of the parameter's variance decreases (which extends the ARL) and is more significant than the effect of the parameter's mean shift (which shortens the ARL). Thus, the overall result is that the ARL increases.

  Regarding the performance of the LPC, we observed that, when σ_ε^2 shifts, the ARL only slightly changes. This can be explained as follows. σ_ε^2 is very small compared with σ_Z^2, and the effect of the term ε may be seen as being similar to that of adding a small random error to each measurement. In the LPC, the waviness of the surface contributes to most of the sum of square errors, thus, the chart monitoring the variance cannot detect a shift in σ_ε^2. The ε_i values are too small to be able to influence modify the regression trend, thus the chart monitoring the linear terms cannot detect shifts in σ_ε^2. In practice, we are more interested in detecting increases in σ_ε^2. Unlike the LPC, our proposed ISC and SSC can efficiently detect increases in σ_ε^2.

- Shifts in σ_Z^2. The ISC, SSC, and LPC can all detect shifts in σ_Z^2. However, our proposed spatial charts are not as efficient as the conventional LPC in many cases, especially when $\boldsymbol{\theta}_{\mathrm{IC}} = (0.6, 0.6)$. We found that, when the profile exhibits strong spatial correlations and the area of $(x^{(1)}, x^{(2)})$ is bounded, an increment in σ_Z^2 may lead to a random shift of the entire profile (recall Figure 6). In this case, the accuracy of the estimates of μ and σ_Z^2 will be affected, because the shift in μ and increase in σ_Z^2 may either generate this random shift. However, if each time we sample more than one surface as a subgroup, this situation can be improved a lot. In such cases, the shift in μ, which results in a synchronic shift of all surfaces across the subgroup, can be distinguished from the enlargement of σ_Z^2, which causes the trends of the surfaces within the subgroup shift independently. Consequently, much better estimates of μ and σ_Z^2 can be obtained. Table 2 illustrates the ARLs of the three competing charts when each subgroup contains five wafer profiles and when σ_Z^2 shifts. The performance of the spatial charts is greatly enhanced and exceeds that of the LPC in both cases of $\theta_{1,2} = 0.4$ and $\theta_{1,2} = 0.6$. If the measurement points of all profiles are the same, the computational burden will not significantly increase because the covariance matrices of all profiles ($\boldsymbol{\Sigma}(\boldsymbol{\xi}_2)$'s) are the same. In that case, the computational burden is mainly derived from the inversion of $\boldsymbol{\Sigma}(\boldsymbol{\xi}_2)$, which is calculated only once in each iteration step.

- Shifts in θ_1. The ISC and SSC can efficiently detect increases in this parameter and the ARLs are not the same when shifts in θ_1 are positive or negative with the same magnitude, as in the case where σ_ε^2 shifts as previously discussed. When θ_1 increases, the ARLs decrease; but when θ_1 decreases, if the magnitude of the decrease is minor, the ARLs of these charts may increase. The reason for this phenomenon

is similar to the case in which σ_ε^2 shifts. The shift in θ_1 may not only affect the mean value of $\hat{\theta}_1$, but may also affect its variance: when θ_1 decreases, the variance of $\hat{\theta}_1$ decreases. Regarding the LPC, for $\theta_1 = 0.4$, a larger θ_1 results in a smaller ARL; for $\theta_1 = 0.6$, a larger θ_1 results in a larger ARL. These results indicate that the performance of the LPC may vary depending on the in-control parameter values and that the LPC is not suitable for monitoring shifts in the strengths of the spatial correlations.

Overall, the SSC exhibits the best performance for monitoring shifts in linear coefficients μ and β_1. When each subgroup contains one single surface, the surface charts (ISC and SSC) can detect shifts in σ_Z^2, but may not as efficiently as the LPC can when spatial correlation is strong. However, when each subgroup contains multiple surfaces, the surface charts(SSC and ISC) can detect shifts in σ_Z^2 and θ_1 more efficiently than the LPC. Between the ISC and SSC, the SSC is more sensitive to shifts in the trend component, whereas the ISC is more sensitive to shifts in the variation component.

Conclusion and Remarks

In this paper, we developed an innovative surface control chart by proposing a method to detect shifts in parameters in the Gaussian-Kriging model. Using the Gaussian-Kriging model in surface monitoring has the following merits: the spatial correlation among data points on the surface is considered, parameters in the model have specific engineering implications, and various types of out-of-control conditions can be detected. An investigation of real surface data demonstrates that the profiles of wafer thickness are indeed spatially correlated. Although our monitoring scheme is not uniformly better than the conventional multiple linear profile chart, it can detect changes in the profile's spatial patterns more efficiently, as long as the profile is adequately represented by a Gaussian-Kriging model.

Two-dimensional data are widely used in modern manufacturing processes. This study is one of the few to address the statistical monitoring of such data. We believe that considerable work remains to be undertaken. First, a more rigorous Phase I analysis should be considered. Specific questions worth exploring include whether the data are suitable for our monitoring scheme, what constitutes a suitable correlation function, and which observations in the historical data are outliers. Second, additional comparison studies should be performed. For instance, our method can be further compared with those of Colosimo (2008), Jensen et al. (2008), or Qiu et al. (2010), especially when the simulated profiles are generated from different spatial models or real data are employed. Third, to increase the efficiency of the control chart and simultaneously decrease the computational burden for wafer inspection, the determination of the number of data points and the measurement sites on the wafer surface in on-line monitoring deserve detailed investigation.

Acknowledgments

The authors would like to thank the Editor and anonymous Referees whose suggestions have resulted in a clearer presentation and significant improvement of our article. Dr. Fugee Tsung's work was supported by the RGC Competitive Earmarked Research Grants 620010 and 619612. Dr. Kaibo Wang's work was supported by the National Natural Science Foundation of China under grant No. 71072012 and Tsinghua University Initiative Scientific Research Program.

Appendix: Details on the Algorithm for Calculating MLEs

We now illustrate the algorithm for solving the MLE of $\boldsymbol{\xi}$. The algorithm combines the general procedure introduced by Fang et al. (2006) and some modifications recommended by Mardia and Marshall (1984).

Consider maximizing the log-likelihood function of $\boldsymbol{y}$,

$$l(\boldsymbol{y} \mid \boldsymbol{\xi}) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log(\det(\boldsymbol{\Sigma}(\boldsymbol{\xi}_2))) - \frac{1}{2}(\boldsymbol{y} - \boldsymbol{B}\boldsymbol{\xi}_1)'\boldsymbol{\Sigma}(\boldsymbol{\xi}_2)^{-1}(\boldsymbol{y} - \boldsymbol{B}\boldsymbol{\xi}_1).$$

The Fisher information matrix of $\boldsymbol{\xi}$ is defined as

$$\boldsymbol{I}(\boldsymbol{\xi}) = \mathrm{E}\left[\nabla_{\boldsymbol{\xi}} l(\boldsymbol{y}|\boldsymbol{\xi})\right]^2.$$

Mardia and Marshall (1984) proved the Fisher information of the model in Equation (5) is blockwise diagonal and has the representation

$$\boldsymbol{I}(\boldsymbol{\xi}) = \mathrm{diag}(\boldsymbol{I}_1(\boldsymbol{\xi}), \boldsymbol{I}_2(\boldsymbol{\xi})),$$

where $\boldsymbol{I}_1$ and $\boldsymbol{I}_2$ are the Fisher information matrix with respect to $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ individually. They have

following expressions:

$$\boldsymbol{I}_1(\boldsymbol{\xi}) = \boldsymbol{y}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\xi}_2)\boldsymbol{y}$$

and the (i,j)th element of $\boldsymbol{I}_2(\boldsymbol{\xi})$ is

$$[\boldsymbol{I}_2(\boldsymbol{\xi})]_{i,j} = \frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}(\boldsymbol{\xi}_2)\frac{\partial}{\partial \boldsymbol{\xi}_2^{(i)}}\boldsymbol{\Sigma}(\boldsymbol{\xi}_2)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\xi}_2)\frac{\partial}{\partial \boldsymbol{\xi}_2^{(j)}}\boldsymbol{\Sigma}(\boldsymbol{\xi}_2)\right),$$

where $\hat{\boldsymbol{\xi}}_i^{(k)}$ is the kth parameter in $\hat{\boldsymbol{\xi}}_i$, $i = 1, 2$.

Next we show the algorithm that iteratively updates the values of $\hat{\boldsymbol{\xi}}_1$ and $\hat{\boldsymbol{\xi}}_2$ until they converge.

Step 1: Select the Initial Value of $\hat{\boldsymbol{\xi}}_2$

Fang et al. (2006) suggest setting the initial value of $\hat{\boldsymbol{\xi}}_1$ to the value of its least-square estimator: $(\boldsymbol{B}'\boldsymbol{B})^{-1}\boldsymbol{B}'\boldsymbol{y}$. However, this may reduce the convergence speed in iteration and numerical problems may arise frequently. Our experiments show that, if we initially set $\hat{\boldsymbol{\xi}}_2$ as $\hat{\boldsymbol{\xi}}_2$'s in-control value, significantly fewer iterations will be needed.

Step 2: Update $\hat{\boldsymbol{\xi}}_1$ Using $\hat{\boldsymbol{\xi}}_2$

The profile MLE of $\boldsymbol{\xi}_1$ given $\boldsymbol{\xi}_2 = \hat{\boldsymbol{\xi}}_2$ is its generalized least-square estimator,

$$\hat{\boldsymbol{\xi}}_1 = (\boldsymbol{B}'\boldsymbol{\Sigma}(\hat{\boldsymbol{\xi}}_2)^{-1}\boldsymbol{B})^{-1}\boldsymbol{B}'\boldsymbol{\Sigma}(\hat{\boldsymbol{\xi}}_2)^{-1}\boldsymbol{y}.$$

Step 3: Update $\hat{\boldsymbol{\xi}}_2$ Using $\hat{\boldsymbol{\xi}}_1$

Mardia and Marshall (1984) suggested using the one-step Fisher scoring algorithm to update the value of $\boldsymbol{\xi}_2$. Denote $\hat{\boldsymbol{\xi}}_2$ as the original value of $\hat{\boldsymbol{\xi}}_2$ and $\hat{\boldsymbol{\xi}}_2'$ as the new value. We use

$$\hat{\boldsymbol{\xi}}_2' = \hat{\boldsymbol{\xi}}_2 + \boldsymbol{I}_2^{-1}(\hat{\boldsymbol{\xi}})\boldsymbol{V}_2(\boldsymbol{y};\hat{\boldsymbol{\xi}})$$

to update $\hat{\boldsymbol{\xi}}_2$. Here, $\boldsymbol{V}_2(\boldsymbol{y};\hat{\boldsymbol{\xi}})$ is the score function with respect to $\boldsymbol{\xi}_2$,

$$\begin{aligned}\boldsymbol{V}_2(\boldsymbol{y};\hat{\boldsymbol{\xi}}) &= \nabla_{\boldsymbol{\xi}_2} l(\boldsymbol{y}|\hat{\boldsymbol{\xi}}) \\ &= \nabla_{\boldsymbol{\xi}_2}\left[-\frac{1}{2}\log\det(\boldsymbol{\Sigma}(\hat{\boldsymbol{\xi}}_2)) \right. \\ &\qquad \left. -\frac{1}{2}(\boldsymbol{y}-\boldsymbol{B}\hat{\boldsymbol{\xi}}_1)'\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\xi}}_2)(\boldsymbol{y}-\boldsymbol{B}\hat{\boldsymbol{\xi}}_1)\right] \\ &= -\frac{1}{2}\nabla_{\boldsymbol{\xi}_2}\left[\log(\det(\boldsymbol{\Sigma}(\hat{\boldsymbol{\xi}}_2)))\right] \\ &\quad -\frac{1}{2}(\boldsymbol{y}-\boldsymbol{B}\hat{\boldsymbol{\xi}}_1)'\nabla_{\boldsymbol{\xi}_2}\left[\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\xi}}_2)\right](\boldsymbol{y}-\boldsymbol{B}\hat{\boldsymbol{\xi}}_1),\end{aligned} \tag{12}$$

which can be calculated using the following two matrix differentiation formulae:

$$\begin{aligned}&\frac{\partial}{\partial \hat{\boldsymbol{\xi}}_2^{(k)}}\log(\det(\boldsymbol{\Sigma}(\hat{\boldsymbol{\xi}}_2))) \\ &\quad = \sum_{j=1}^{m}\sum_{i=1}^{m}(\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\xi}}_2))_{i,j} \times \left(\frac{\partial}{\partial \hat{\boldsymbol{\xi}}_2^{(k)}}\boldsymbol{\Sigma}(\hat{\boldsymbol{\xi}}_2)\right)_{i,j}, \\ &\frac{\partial}{\partial \hat{\boldsymbol{\xi}}_2^{(k)}}\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\xi}}_2) \\ &\quad = \boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\xi}}_2)\left(\frac{\partial}{\partial \hat{\boldsymbol{\xi}}_2^{(k)}}\boldsymbol{\Sigma}(\hat{\boldsymbol{\xi}}_2)\right)\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\xi}}_2), \\ &\qquad\qquad k = 1, 2, 3, 4.\end{aligned}$$

Step 4

Run Step 2 and Step 3 iteratively until $\hat{\boldsymbol{\xi}}_1$ and $\hat{\boldsymbol{\xi}}_2$ converge.

References

Ankenman, B.; Nelson, B. L.; and Staum, J. (2010). "Stochastic Kriging for Simulation Metamodeling". *Operations Research* 58(2), pp. 371–382.

Colosimo, B. M. (2008). "Statistical Process Control for Geometric Specifications: On the Monitoring of Roundness Profiles". *Journal of Quality Technology* 40(1), pp. 1–18.

Colosimo, B. M.; Mammarella, F.; and Petro, S. (2010). "Quality Control for the Manufactured Surfaces". *Frontiers in Statistical Quality Control* 9.

Cressie, N. A. C. (1993). *Statistics for Spatial Data*. Wiley series in probability and mathematical statistics applied probability and statistics, revised edition. New York, NY: J. Wiley.

Fang, K.; Li, R.-Z.; and Sudjianto, A. (2006). *Design and Modeling for Computer Experiments*. Computer science and data analysis series. London. UK: Chapman and Hall/CRC.

Grimshaw, S. D.; Blades, N. J.; and Miles, M. P. (2013). "Spatial Control Charts for the Mean". *Journal of Quality Technology* 45(2), pp. 130–148.

Jensen, W. A.; Birch, J. B.; and Woodall, W. H. (2008). "Monitoring Correlation Within Linear Profiles Using Mixed Models". *Journal of Quality Technology* 40(2).

Jin, R.; Chang, C.-J.; and Shi, J. (2012). "Sequential Measurement Strategy for Wafer Geometric Profile Estimation". *IIE Transactions* 44(1), pp. 1–12. Amsterdam: Elsevier Science B.V.

Koehler, J. R. and Owen, A. B. (1996). "Computer Experiments". *Handbook of Statistics* 13, pp. 261–308.

Lin, J. and Wang, K. (2011). "Online Parameter Estimation and Run-to-Run Process Adjustment Using Categorical Observations". *International Journal of Production Research* 49(3), pp. 4103–4117.

Mardia, K. V. and Marshall, R. (1984). "Maximum Likelihood Estimation of Models for Residual Covariance in Spatial Regression". *Biometrika* 71(1), pp. 135–146.

Mason, R. L.; Tracy, N. D.; and Young, J. C. (1995). "Decomposition of T^2 for Multivariate Control Chart Interpretation". *Journal of Quality Technology* 27(2).

Qiu, P.; Zou, C.; and Wang, Z. (2010). "Nonparametric Profile Monitoring by Mixed Effects Modeling". *Technometrics* 52(3), pp. 265–277.

Ranjan, P.; Haynes, R.; and Karsten, R. (2011). "A Computationally Stable Approach to Gaussian Process Interpolation of Deterministic Computer Simulation Data". *Technometrics* 53(4), pp. 366–378.

Rasmussen, C. E. and Williams, C. (2006). "Gaussian Processes for Machine Learning". Cambridge, MA: The MIT Press, 38, pp. 715–719.

Reis, M. and Saraiva, P. M. (2006). "Multiscale Statistical Process Control of Paper Surface Profiles". *Quality Technology and Quantitative Management* 3(3), pp. 263–282.

Sacks, J.; Welch, W. J.; Mitchell, T. J.; and Wynn, H. P. (1989). "Design and Analysis of Computer Experiments". *Statistical science* 4(4), pp. 409–423.

Santner, T. J.; Williams, B. J.; and Notz, W. (2003). *The Design and Analysis of Computer Experiments*. New York, NY: Springer.

Walker, E. and Wright, S. P. (2002). "Comparing Curves Using Additive Models". *Journal of Quality Technology* 34(1), pp. 118–129.

Williams, J. D.; Woodall, W. H.; and Birch, J. B. (2007). "Statistical Monitoring of Nonlinear Product and Process Quality Profiles". *Quality and Reliability Engineering International* 23, pp. 925.

Woodall, W. H.; Spitzner, D. J.; Montgomery, D. C.; and Cupta, S. (2004). "Using Control Charts to Monitor Process and Product Quality Profiles". *Journal of Quality Technology* 36(3), pp. 309.

Zhao, H.; Jin, R.; Wu, S.; and Shi, J. (2011). "PDE-Constrained Gaussian Process Model on Material Removal Rate of Wire Saw Slicing Process". *ASME Transactions, Journal of Manufacturing Science and Engineering*, to appear.

Zou, C.; Tsung, F.; and Wang, Z. (2007). "Monitoring General Linear Profiles Using Multivariate Exponential Weighted Moving Average Schemes". *Technometrics* 49(4), pp. 395.