Research Article

Quality and Reliability Engineering International

(wileyonlinelibrary.com) DOI: 10.1002/qre.1814

Published online in Wiley Online Library

Robust Parameter Design for Profile Quality Control

Lulu Bao,[a] Qiang Huang[b] and Kaibo Wang[a]*

In certain manufacturing processes, product quality is characterized by spatial profiles, and such profiles are expected to meet specific shape requirements. As profile shapes are affected by process conditions, properly adjusted process variables are expected to help improve profile quality. This work aims to achieve desired shapes of profiles that are sensitive to the variation of noise factors through optimizing settings of controllable factors. A hierarchical model is first built to characterize the spatial correlation of measurement points on a profile and link quality metrics with process variables. The process is then optimized using the robust parameter design technique. The performance of the proposed method is studied using a motivating example from nanomanufacturing. Copyright © 2015 John Wiley & Sons, Ltd.

Keywords: hierarchical model; Kriging; robust parameter design; spatial correlation

1. Introduction

Due to the advancement of sensing and metrology techniques, large quantities of data can be collected during production or quality inspection within a short time at a low cost. When data are collected at multiple gauge points on a product, a spatial profile is naturally formed. The variability across such a profile is crucial for those products that are desired to achieve a specific shape and must be controlled to meet quality target.

As a motivating example, we study a nanomanufacturing process that produces carbon nanotube (CNT) arrays in this paper. CNT arrays are expected to grow uniformly on wafer substrates. However, due to the physical structure the production facility has, real CNT arrays usually deviate from the ideal profile to a bowl shape, as shown in Figure 1. Such a deviation will lead to both quality and cost issues in downstream-manufacturing stages. In the current industrial practice, aggregated quality metrics are usually used to characterize product quality and provide information for process improvement. For example, the average height of nanotubes is commonly used by engineers as one objective to determine the growth time. However, aggregated metrics have obviously ignored the spatial variability a nanotube array has and are therefore ineffective if being utilized in quality control. Instead, the spatial profile obtained from product inspection conserves more information of variability[1] and could be considered for quality control.

The idea of using profiles to characterize product quality has been extensively studied in the literature on statistical process monitoring.[2–8] However, research that addresses issues of process control for improving profiles quality is still limited. As profiles are important quality metrics and usually have specific requirements, quality engineers also want to control the geometric properties of profiles by adjusting controllable process variables.

To control profile quality, the first issue is to identify process variables that potentially affect profile shapes. Some real manufacturing processes have provided such examples in which process variables have an impact on the geometric features of profiles. For instance, the cutting speed and the depth of cut in a lathe-tuning process of a titanium alloy affect profiles' roundness.[9] There is also literature analyzing the effect of process variables on shapes of curves.[2,10–12]

Once influential factors are identified in a process, the next step is to seek appropriate settings for the controllable factors to optimize process output and improve product quality. In recent years, there has been a wide spectrum of research performed on robust parameter design for functional responses, which could be seen as a special type of profiles. A natural approach in robust parameter design study for profile data is the extension of conventional location-dispersion modeling method. To use this method, statistical models should be built to link the mean, variance,[13] and even spatial dependence[14] of measurements at gauge points, to process variables; then, the optimal settings for controllable factors can be obtained. Although the location-dispersion modeling approach is easy to use, it has a drawback that interactions of controllable factors and noise factors are not reflected in the model.

[a]*Department of Industrial Engineering, Tsinghua University, Beijing, 100084, China*
[b]*Epstein Department of Industrial and Systems Engineering, University of Southern California, Los Angeles, CA, 90089, USA*
**Correspondence to: Kaibo Wang, Department of Industrial Engineering, Tsinghua University, Beijing, 100084, China.*
E-mail: kbwang@tsinghua.edu.cn

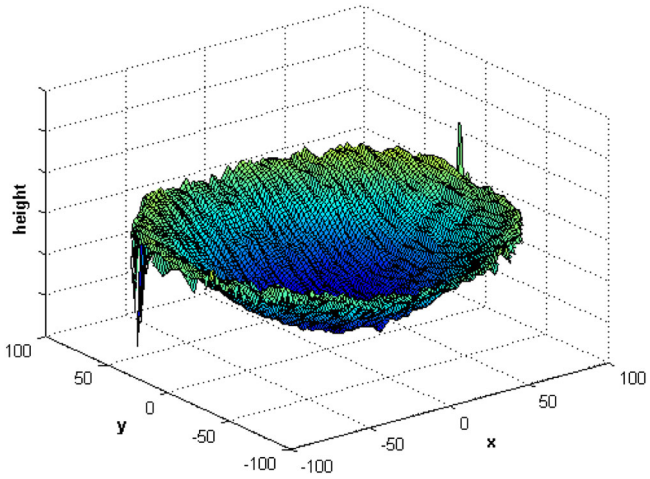

Figure 1. The three-dimensional (3D) plot of a carbon nanotube (CNT) array.

The signal-to-noise ratio approach is a mutation of the location-dispersion method, which combines mean and variance together.[15] The signal-to-noise ratio serves as an objective function and a metric of profiles' uniformity.[16] Similar to the location-dispersion modeling approach, the signal-to-noise technique uses a two-step approach to solve the objective function. Due to limitations of the signal-to-noise ratio on fitting practical situations and mixing location effects and dispersion effects, it is not recommended to use this technique in process parameter design problems.[17]

Response modeling is a commonly used approach of robust parameter design. Responses are linked to process variables via a regression model, which contains interactions of controllable factors and noise factors. Then, the controllable factors are set by solving an objective function, which is defined by combining target deviation and variance of profiles.[18] Another common approach is the Bayesian approach, proposed by Peterson,[19] which emphasizes spatial correlation of responses and uncertainty of parameter estimates. The posterior probability of achieving a desired shape is maximized when there are noise factors.

Recently, hierarchical approaches are proposed on the basis of the aforementioned methods for robust design of profiles. Nair and Taam[13] proposed a two-stage modeling approach that is a variation of the response modeling method. Del Castillo and Colosimo[20] proposed a hierarchical version of the Bayesian approach. A hierarchical approach usually contains two stages: in the first stage, quality responses of profiles are modeled as a function of locations; then, the second stage model links coefficients generated in stage 1 to process variables. Hierarchical approaches have priorities over non-hierarchical models due to their interpretability.[11] The coefficients in stage 1 usually have clear geometric meaning, which is directly related to the shapes of profiles. The models in stage 2 indicate that we can control the shape of profiles by choosing appropriate settings of controllable process variables.

For spatial profiles such as the CNT arrays studied in this paper, within-profile correlation is usually strong and cannot be neglected.[21] However, current profile modeling approaches often leave out spatial correlation. Naturally, the spatial correlation is not involved during optimization procedures of robust parameter design. A major reason is that it is a computation consumed to estimate parameters in a complex covariance matrix when the number of locations is large.[3] We will address this issue in greater details in the next section.

The aim of this work is to propose a robust parameter design method for controlling profile shapes. The proposed method will help achieve a desired shape of profiles that is tolerant to variations of noise factors by optimizing settings of controllable factors. Compared with the existing work on robust parameter design or functional response analysis, the major contribution of this work is twofolded. First, a hierarchical model, which takes the spatial correlation of measurement points into consideration using the Kriging technique, is proposed to characterize profiles and link profiles with process variables for control purpose; second, the robust parameter design methodology is integrated with the hierarchical model to improve profile quality through off-line optimization. The proposed model is effective even if the number of locations is large or the number of locations varies within profiles.

The rest of this paper is organized as follows. The hierarchical model of profiles is described in the next section, followed by the robust parameter design to obtain optimal controllable factor settings. In section 4, the motivating application from nanomanufacturing is revisited to illustrate and evaluate the performance of the proposed method. Section 5 concludes this work with suggestions for future research.

2. Hierarchical modeling for profiles

To improve CNT array quality, in this section, we first build a hierarchical model to characterize profile quality and link critical quality metrics with process variables. The hierarchical model contains two stages: stage 1 connects responses to locations of gauge points on the profiles and stage 2 links coefficients in stage 1 to process variables. Based on this hierarchical model, process variables will be optimized in the next section to improve product quality.

Assume that there are m_i points on the i-th profile, where $i = 1, \ldots, N$. The j-th point on profile i, where $j = 1, \ldots, m_i$, is recorded at location s_{ij}, and the corresponding response is y_{ij}. For $i = 1, \ldots, N$, a two-stage hierarchical model for the profiles is given by

Stage 1:

$$\mathbf{y}_i = \mathbf{S}_i\boldsymbol{\theta}_i + \mathbf{z}_i + \boldsymbol{\varepsilon}_i \tag{1}$$

Stage 2:

$$\boldsymbol{\theta}_i = \mathbf{B}\cdot\mathbf{f}(\mathbf{x}_i, \mathbf{u}_i, \mathbf{e}_i, \mathbf{n}_i) + \mathbf{v}_i \tag{2}$$

In stage 1, the responses are modeled as a function of locations. In this model, $\mathbf{y}_i = \left(y_1 \ldots, y_{m_i}\right)^T$ is a $m_i \times 1$ vector representing the responses on the i-th profile, $\boldsymbol{\theta}_i$ is a $p \times 1$ coefficients vector of stage 1 relating to the shape of profiles, $\mathbf{S}_i$ is a $m_i \times p$ regressor matrix of which the elements are functions of the locations of gage points, $\mathbf{z}_i = (z_{i1} \ldots, z_{im_i})^T$ is a vector following a Gaussian–Kriging process and characterizing spatial correlation of the gauge points, and $\boldsymbol{\varepsilon}_i$ is a white noise vector following a multivariate normal distribution. The response is decomposed into a global trend and local details that are modeled by a linear model and a Gaussian process, respectively. This is known as a universal Kriging model where global trend is not negligible because parameters in the global trend model may indicate some principal geometric features. Universal Kriging is frequently used in current literature. For example, Ba and Joseph[12] decompose responses into smooth global trend and local details and model them with two Gaussian processes separately. As another example, Plumlee and Jin[8] also use universal Kriging to build models.

In stage 2, the coefficient vector in stage 1 is linked to process variables. In Equation (2), $\mathbf{B}$ is a $p \times (q+1)$ coefficient matrix, $\boldsymbol{f}(\mathbf{x}_i, \mathbf{u}_i, \mathbf{e}_i, \mathbf{n}_i)$ is a vector consisting of functions of the process variables, and $\mathbf{v}_i$ is a noise vector that follows a multivariate normal distribution $N_p(\mathbf{0}, \Sigma_v)$. Huang[22] uses a similar hierarchical approach that contains a term to characterize spatial correlation to model the length of nanowires. The difference is that the noise factors are not considered in the model because it is not for robust parameter design.

Similar to the method that Zhong *et al.*[23] used to optimize the process variables, this paper classifies the process variables being considered in Equation (2) into four categories: off-line setting factors, $\mathbf{x}_i$; online controllable factors, $\mathbf{u}_i$; measurable noise factors, $\mathbf{e}_i$; and unobservable noise factors, $\mathbf{n}_i$. The process variable vector $\boldsymbol{f}(\mathbf{x}_i, \mathbf{u}_i, \mathbf{e}_i, \mathbf{n}_i)$ is often defined as

$$\left(\mathbf{x}_i{}^T,\ \mathbf{u}_i{}^T,\ \mathbf{e}_i{}^T,\ \mathbf{n}_i{}^T,\ \mathbf{x}_i{}^T\otimes\mathbf{e}_i{}^T,\ \mathbf{x}_i{}^T\otimes\mathbf{n}_i{}^T, \mathbf{u}_i{}^T\otimes\mathbf{e}_i{}^T,\ \mathbf{u}_i{}^T\otimes\mathbf{n}_i{}^T\right)^T$$

where $\otimes$ *is the Kronecker product. Then, the* k-th element of $\boldsymbol{\theta}_i$ in Equation (2) can be expressed as the function of factors from the four categories:

$$\theta_{ik} = \beta_{k0} + \boldsymbol{\beta}_{k1}^T\mathbf{x} + \boldsymbol{\beta}_{k2}^T\mathbf{u} + \boldsymbol{\beta}_{k3}^T\mathbf{e} + \boldsymbol{\beta}_{k4}^T\mathbf{n} + \mathbf{x}^T\mathbf{B}_{k1}\mathbf{e} + \mathbf{u}^T\mathbf{B}_{k2}\mathbf{e} + \mathbf{x}^T\mathbf{B}_{k3}\mathbf{n} + \mathbf{u}^T\mathbf{B}_{k4}\mathbf{n} + v_{ik} \tag{3}$$

The coefficient matrix $\mathbf{B}$ is given by

$$\begin{pmatrix} \beta_{10} & \boldsymbol{\beta}_{01}^T & \boldsymbol{\beta}_{02}^T & \boldsymbol{\beta}_{03}^T & \boldsymbol{\beta}_{04}^T & (vec(\mathbf{B}_{01}))^T & (vec(\mathbf{B}_{02}))^T & (vec(\mathbf{B}_{03}))^T & (vec(\mathbf{B}_{04}))^T \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \beta_{p0} & \boldsymbol{\beta}_{p1}^T & \boldsymbol{\beta}_{p2}^T & \boldsymbol{\beta}_{p3}^T & \boldsymbol{\beta}_{p4}^T & \left(vec\left(\mathbf{B}_{p1}\right)\right)^T & \left(vec\left(\mathbf{B}_{p2}\right)\right)^T & \left(vec\left(\mathbf{B}_{p3}\right)\right)^T & \left(vec\left(\mathbf{B}_{p4}\right)\right)^T \end{pmatrix}$$

One important feature of the proposed hierarchical model is that we take spatial correlation among gauge points into account and characterize the correlation with the Kriging technique. In the literature, spatial correlation has been highlighted in many studies related to profiles such as profile modeling,[24–26] profile monitoring,[27,28] profile optimization,[29] and profile control.[1] As profile data are measured at different but adjacent locations, the data usually exhibit certain spatial correlation. Therefore, a model that considers the spatial correlation among all measurement locations usually has better performance in estimation and prediction of profile variations.[30] However, such within-profile correlation is often neglected in robust parameter design study of profiles because of the computation difficulty resulted from a large and complex covariance matrix, as we mentioned before. Some techniques can resolve this problem through simplification of covariance structures. For instance, Del Castillo and Colosimo[20] use a random effects term $\mathbf{S}_i\mathbf{v}_i$ (in our notation) to model the within-profile correlation and then select part of $f(\mathbf{x}_i, \mathbf{u}_i, \mathbf{e}_i, \mathbf{n}_i)' \otimes \mathbf{S}_i$ to replace $\mathbf{S}_i$ to simplify the covariance structure. We use Kriging technique to model the within-profile correlation, and it is more intuitive and easier to interpret than the random effects technique.

Kriging is the most widely used technique to model spatial correlation and has been extensively studied in spatial statistics[21] and computer experiments.[31] The Kriging technique has been applied to model the correlation structure of profile responses.[32,33] In the proposed hierarchical model, we use the commonly used Gaussian process structure in Kriging analysis to characterize the correlation among the multiple responses; in Equation (1), $\mathbf{z}_i = (z_{i1} \ldots, z_{im_i})^T$ is assumed to follow a multivariate Gaussian distribution with

$$\begin{aligned} E(z_i) &= 0 \\ Cov(z_i, z_j) &= \sigma_z^2 R(\delta, s_i - s_j) \end{aligned} \tag{4}$$

where σ_z^2 *is the variance of the process and* $R(\cdot)$ is a pre-specified correlation that controls the smoothing of z_i. Commonly used correlation structures include exponential, Matern, and Gaussian correlation function.[34] A proper correlation function should be

chosen based on data and specific applications. By specifying the correlation structure using Kriging technique, the parameters to be estimated in the covariance matrix is significantly reduced, and thus the model work even if there is a large number of locations.

3. Robust parameter design

In the previous section, we propose a hierarchical model for profiles to link responses to process variables. The main task of this section is to find appropriate settings for controllable factors such that the output product quality is insensitive to variations caused by noise factors. To formulate this target as an optimization problem, we must define an objective function.

Khuri and Conlon,[35] Ames *et al.*,[36] and Murphy *et al.*[37] studied different types of loss functions. Among them, one of the most commonly used loss function is a multivariate quadratic loss function, which is defined as

$$E\left[(\mathbf{y}-\boldsymbol{\tau})^T\mathbf{A}(\mathbf{y}-\boldsymbol{\tau})\right] \tag{5}$$

where A is a weight matrix and τ is a target vector of responses. The loss function in Equation (5) measures deviation between responses and the target value. There are two main advantages for the multivariate quadratic loss function: one advantage is that correlation among responses is incorporated in optimization procedures and the other advantage is that both the off-target penalty and variance are considered during the optimization.

The multivariate quadratic loss function is applicable to our problem because spatial correlation among the responses is emphasized and we aim to achieve a desired shape with small variation. Therefore, similar to the loss function in Equation (5), we define the loss function as follows

$$O(\mathbf{x},\mathbf{u}) = E_{\mathbf{e},\mathbf{n},z,v,\varepsilon}\left[\sum_{i=1}^{N}\sum_{j=1}^{m_i}\left(y_{ij}-\tau_{ij}\right)^2\right] \tag{6}$$

where τ_{ij} is the target value of the profile *i* at point s_{ij}. Differing from the loss function given in Equation (5), the weight matrix **A** is set as an identity matrix because we believe that all the gauge points are equally important and that there is no additional cost resulting from the simultaneous target deviation of pairs of responses. Thus, we can simply write the loss function as an additive combination of responses, as shown in Equation (6). By minimizing the objective function in Equation (6), the optimal settings of controllable factors are derived to make the profile remain close to the desired shape and insensitive to the variation of noise factors. The procedures of optimization will be illustrated with an example in the next section.

4. Case study and discussion

We now revisit the nanomanufacturing process for producing CNT arrays. The CNT arrays in this example were grown on silicon wafers using the chemical-vapor deposition (CVD) method.[38] In industrial practice, the nanotubes are expected to meet a specific target height and be uniform in height. It is easily learned from engineering knowledge that the height of the nanotubes is related to process variables such as the partial pressure of the precursor and the temperature of the reaction.[39] The purpose of this study is therefore seeking opportunities to improve the process output by optimizing the controllable factors.

Through the statistical analysis of measurements of 60 CNT arrays collected from the manufacturing process, it is confirmed that significant interactions between an observable but uncontrollable factor and some controllable factors do exist. There are eleven factors in the process. Two factors among them are observable noise factors, and the remaining nine factors are controllable factors. We distinguish four factors that have significant effects on the height of nanotubes from these factors and name them as factors *A*, *B*, *C*, and *D* for confidentiality. Among them, factor *D* is an observable but uncontrollable factor and has interactions with the off-line controllable factors *B* and *C*. This provides an opportunity to reduce the variation transmitted from the observable factor by selecting proper settings of the controllable factors. Therefore, we can apply the proposed method to improve the quality of nanotubes. In the following, we first analyze the real dataset of nanotubes and build a hierarchical model then apply robust parameter design to acquire the optimal set values for the controllable factors. Finally, the performance of the proposed method will be studied.

4.1. Data analysis and hierarchical modeling

For quality characterization purpose, a metrology equipment scans nanotube arrays line by line, with multiple gauge points on each line, thus forming a spatial profile. As all measurement lines share a similar trend, for demonstration purpose, we choose the central measurement line, which contains the most points (i.e., the most information), as our objective profile. Figure 2 shows the central line's height plot of an example nanotube array. If all measurement lines were considered and the complete two-dimensional bowl-shape profile is defined as the objective profile, the model presented in Equation (1) was still valid except a modification to the trend term and parameter vector, but the problem formulation and all results obtained in this work still hold.

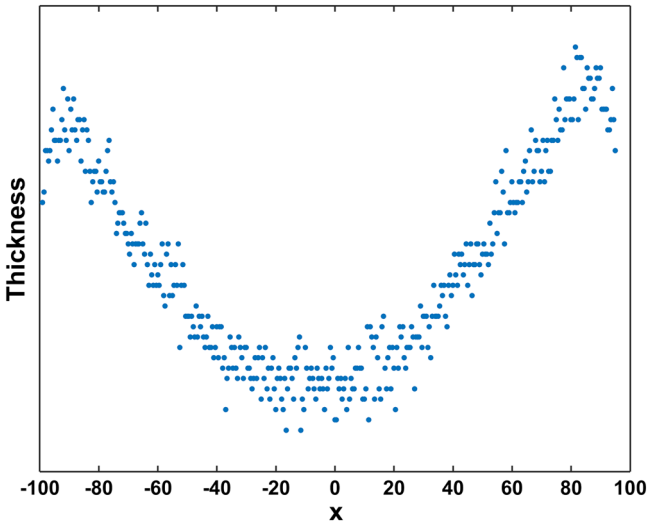

Figure 2. The height profile of nanotubes along the central line of a nanotube array.

The CNT arrays have some important features. First, the nanotube arrays usually exhibit a bowl shape with higher height close to the edges and lower height in the central area, as shown in Figures 1 and 2. Second, there are some abnormal nanotubes on the edge of the arrays resulted by non-uniform catalyst films on the edges. We eliminate these outliers before building the model, as those nanotubes are produced because of irregular catalyst distribution near the edge area and cannot be changed by adjusting factors of the CVD process. Third, the nanotube arrays are usually asymmetric because of the inhomogeneous exposure to reactant gasses in the CVD process, as Figure 2 shows.

Considering the features of the profiles, the following model was constructed:

Stage 1:

$$\begin{aligned} y_{ij} &= \mu_{ij} + z_{ij} + \varepsilon_{ij} \\ \mu_{ij} &= \begin{cases} a_{i1}s_{ij}^2 + c_i & (s_{ij} < 0) \\ a_{i2}s_{ij}^2 + c_i & (s_{ij} \geq 0) \end{cases} \end{aligned} \tag{7}$$

Stage 2:

$$\boldsymbol{\theta}_i = \begin{pmatrix} a_{i1} \\ a_{i2} \\ c_i \end{pmatrix} = \begin{pmatrix} \beta_{10} & \beta_{11} & \cdots & \beta_{16} \\ \beta_{20} & \beta_{21} & \cdots & \beta_{26} \\ \beta_{30} & \beta_{31} & \cdots & \beta_{36} \end{pmatrix} (1,\ A,\ B,\ C,\ D,\ BD,\ CD)^T + \mathbf{v}_i \tag{8}$$

where y_{ij} is the height of the nanotube at location s_{ij} on the ith profile (i.e., the response). Due to the asymmetry of profiles, models were fit to the left region and the right region separately. The four factors and their interactions deemed to have significant effects on the responses are contained in the model.

Universal Kriging is used to model the height of nanotubes. The height is the sum of a regression model and a Gaussian process. The nanotube arrays usually show a bowl shape because of their growth mechanism. Therefore, a quadratic model is used to characterize the global trend shared by CNT arrays. Nanotubes' position on wafer substrates has a significant impact on their height. Given coordinates of nanotubes, we can acquire much information on their heights. The coefficients vector $\boldsymbol{\theta}_i$ is related to locations and has specific geometric meaning, where a_{i1} and a_{i2} represent the variation scale of the nanotubes on the left and on the right, respectively, and c_i is the height of the nanotube at the center of profile i. This specific pattern will be hidden if we use ordinary Kriging because all the variation is represented using a Gaussian process. To capture this pattern, we treat the common trend as the large scale variation and use Kriging to characterize the small scale variation of each nanotube array.

Owing to the lack of theoretical and experimental support in our case, parameters of Kriging are not contained in the second stage model. The second stage model is built based on the current physical knowledge we have and the geometrical features of nanotube arrays we observe. The relationship between Kriging parameters and process factors is difficult to model. But sometimes, the small scale features may be affected by noise factors and controllable factors. If a convincing model can be built to characterize the relationships between Kriging parameters and process variables, it will be better to add Kriging parameters into the responses of second stage model. For example, Plumlee and Jin[8] model parameters of Gaussian process as a linear function of process factors and show that this kind of model better explains the model.

The vector of off-line controllable factors is $\mathbf{x} = (A, B, C)^T$, and the vector of observable factors is $\mathbf{e} = D$. To remain consistent with the symbols in Section 2, we still use $\mathbf{e}$ to represent the vector of the observable factors. Combining Equations (7) and (8), we can rewrite two separate height models for nanotubes on the left and right sides of the center point as follows.

When $s_{ij} < 0$, the height is given by

$$\begin{aligned} y_{ij} &= (\beta_{30} + \beta_{31}A + \beta_{32}B + \beta_{33}C + \beta_{34}D + \beta_{35}BD + v_{i3}) \\ &\quad + (\beta_{10} + \beta_{12}B + \beta_{13}C + \beta_{14}D + \beta_{16}CD + v_{i1}){s_{ij}}^2 + z_{ij} + \varepsilon_{ij} \\ &= \left(\beta_{30} + \boldsymbol{\beta}_{31}^T\mathbf{x} + \boldsymbol{\beta}_{33}^T\mathbf{e} + \mathbf{x}^T\mathbf{B}_{31}\mathbf{e} + v_{i3}\right) + \left(\beta_{10} + \boldsymbol{\beta}_{11}^T\mathbf{x} + \boldsymbol{\beta}_{13}^T\mathbf{e} + \mathbf{x}^T\mathbf{B}_{11}\mathbf{e} + v_{i1}\right){s_{ij}}^2 + z_{ij} + \varepsilon_{ij} \end{aligned}$$

When $s_{ij} \geq 0$, the height is governed by

$$\begin{aligned} y_{ij} &= (\beta_{30} + \beta_{31}A + \beta_{32}B + \beta_{33}C + \beta_{34}D + \beta_{35}BD + v_{i3}) \\ &\quad + (\beta_{20} + \beta_{22}B + \beta_{23}C + \beta_{24}D + \beta_{26}CD + v_{i2}){s_{ij}}^2 + z_{ij} + \varepsilon_{ij} \\ &= \left(\beta_{30} + \boldsymbol{\beta}_{32}^T\mathbf{x} + \boldsymbol{\beta}_{32}^T\mathbf{e} + \mathbf{x}^T\mathbf{B}_{32}\mathbf{e} + v_{i3}\right) + \left(\beta_{20} + \boldsymbol{\beta}_{21}^T\mathbf{x} + \boldsymbol{\beta}_{23}^T\mathbf{e} + \mathbf{x}^T\mathbf{B}_{21}\mathbf{e} + v_{i2}\right){s_{ij}}^2 + z_{ij} + \varepsilon_{ij} \end{aligned}$$

The coefficient matrix in Equation (8) is estimated, and the result is shown in Equation (9).

$$\hat{\mathbf{B}} = \begin{pmatrix} -2.12E-04 & 0 & 1.45E-04 & 6.40E-09 & 6.21E-08 & 0 & -4.19E-11 \\ -6.33E-05 & 0 & 4.61E-05 & 9.16E-10 & 1.62E-08 & 0 & -8.56E-12 \\ 12.1 & 0.0588 & -8.45 & 4.01E-05 & -0.149 & 0.106 & 0 \end{pmatrix} \tag{9}$$

4.2. Robust parameter design of nanotubes

From historical data, we found that the observable factor D follows a normal distribution with mean 130 and standard deviation 40. The target height of the nanotubes is 0.3. Due to the asymmetry of profiles, the objective function in Equation (6) is rewritten as

$$\begin{aligned} O(\mathbf{x}) &= \sum_{i=1}^{N}\sum_{j,\ s_i<0}\left\{\left[E_{e,z,v,\varepsilon}\left(y_{ij}\right) - \tau\right]^2 + Var_{e,z,v,\varepsilon}\left(y_{ij}\right)\right\} \\ &\quad + \sum_{i=1}^{N}\sum_{j,\ s_j\geq 0}\left\{\left[E_{e,z,v,\varepsilon}\left(y_{ij}\right) - \tau\right]^2 + Var_{e,z,v,\varepsilon}\left(y_{ij}\right)\right\} \end{aligned} \tag{10}$$

where τ is the target height of the nanotubes. The loss function guarantees that the nanotubes achieve the target height and remain uniform simultaneously.

By minimizing the objective function in Equation (10), we can obtain the optimal settings of controllable factors. For simplification, we denote that

$$\gamma_{10}^{(ij)} = \hat{\beta}_{30} + \hat{\beta}_{10}{s_{ij}}^2,\ \boldsymbol{\gamma}_{13}^{(ij)} = \hat{\boldsymbol{\beta}}_{33} + \hat{\boldsymbol{\beta}}_{13}{s_{ij}}^2,\ \boldsymbol{\gamma}_{11}^{(ij)} = \hat{\boldsymbol{\beta}}_{31} + \hat{\boldsymbol{\beta}}_{11}{s_{ij}}^2,\ \mathbf{G}_1^{(ij)} = \hat{\mathbf{B}}_{31} + \hat{\mathbf{B}}_{11}{s_{ij}}^2$$
$$\gamma_{20}^{(ij)} = \hat{\beta}_{30} + \hat{\beta}_{20}{s_{ij}}^2,\ \boldsymbol{\gamma}_{23}^{(ij)} = \hat{\boldsymbol{\beta}}_{33} + \hat{\boldsymbol{\beta}}_{23}{s_{ij}}^2,\ \boldsymbol{\gamma}_{21}^{(ij)} = \hat{\boldsymbol{\beta}}_{31} + \hat{\boldsymbol{\beta}}_{21}{s_{ij}}^2,\ \text{and}\ \mathbf{G}_2^{(ij)} = \hat{\mathbf{B}}_{31} + \hat{\mathbf{B}}_{21}{s_{ij}}^2$$

Then, the loss function in Equation (10) can be expressed as

$$\begin{aligned} O(\mathbf{x}) &= \sum_{i=1}^{N}\sum_{j,\ s_i<0}\left\{\left[\gamma_{10}^{(ij)} + \left(\boldsymbol{\gamma}_{11}^{(ij)}\right)^T\mathbf{x} + \left(\boldsymbol{\gamma}_{13}^{(ij)}\right)^T\boldsymbol{\mu}_\mathbf{e} + \mathbf{x}^T\mathbf{G}_1^{(ij)}\boldsymbol{\mu}_\mathbf{e} - \tau\right]^2\right. \\ &\quad \left. + \left[\boldsymbol{\gamma}_{13}^{(ij)} + \left(\mathbf{G}_1^{(ij)}\right)^T\mathbf{x}\right]^T\Sigma_\mathbf{e}\left[\boldsymbol{\gamma}_{13}^{(ij)} + \left(\mathbf{G}_1^{(ij)}\right)^T\mathbf{x}\right] + \begin{pmatrix}{s_{ij}}^2 & 0 & 1\end{pmatrix}\boldsymbol{\Sigma}_{\mathbf{v}_i}\begin{pmatrix}{s_{ij}}^2 & 0 & 1\end{pmatrix}^T + {\sigma_z}^2 + {\sigma_\varepsilon}^2\right\} \\ &\quad + \sum_{i=1}^{N}\sum_{j,\ s_j\geq 0}\left\{\left[\gamma_{20}^{(ij)} + \left(\boldsymbol{\gamma}_{21}^{(ij)}\right)^T\mathbf{x} + \left(\boldsymbol{\gamma}_{23}^{(ij)}\right)^T\boldsymbol{\mu}_\mathbf{e} + \mathbf{x}^T\mathbf{G}_2^{(ij)}\boldsymbol{\mu}_\mathbf{e} - \tau\right]^2\right. \\ &\quad \left. + \left[\boldsymbol{\gamma}_{23}^{(ij)} + \left(\mathbf{G}_2^{(ij)}\right)^T\mathbf{x}\right]^T\Sigma_\mathbf{e}\left[\boldsymbol{\gamma}_{23}^{(ij)} + \left(\mathbf{G}_2^{(ij)}\right)^T\mathbf{x}\right] + \begin{pmatrix}0 & {s_{ij}}^2 & 1\end{pmatrix}\boldsymbol{\Sigma}_{\mathbf{v}_i}\begin{pmatrix}0 & {s_{ij}}^2 & 1\end{pmatrix}^T + {\sigma_z}^2 + {\sigma_\varepsilon}^2\right\} \end{aligned} \tag{11}$$

Forcing the differential of the objective function in Equation (12) to 0, we can obtain the solution to the optimization problem,

$$\mathbf{x} = -\left(\sum_{i=1}^{N}\sum_{j,\ s_i<0}\mathbf{A}_1^{(ij)} + \sum_{i=1}^{N}\sum_{j,\ s_j\geq 0}\mathbf{A}_2^{(ij)}\right)^{-1}\cdot\left(\sum_{i=1}^{N}\sum_{j,\ s_i<0}\mathbf{d}_1^{(ij)} + \sum_{i=1}^{N}\sum_{j,\ s_j\geq 0}\mathbf{d}_2^{(ij)}\right) \tag{12}$$

where

$$\mathbf{A}_k^{(ij)} = \left(\boldsymbol{\gamma}_{k1}^{(ij)} + \mathbf{G}_k^{(ij)}\boldsymbol{\mu}_\mathbf{e}\right)\left(\boldsymbol{\gamma}_{k1}^{(ij)} + \mathbf{G}_k^{(ij)}\boldsymbol{\mu}_\mathbf{e}\right)^T + \mathbf{G}_k^{(ij)}\Sigma_\mathbf{e}\left(\mathbf{G}_k^{(ij)}\right)^T \text{and}$$
$$\mathbf{d}_k^{(ij)} = \left(\boldsymbol{\gamma}_{k1}^{(ij)} + \boldsymbol{\gamma}_{k1}^{(ij)}\boldsymbol{\mu}_\mathbf{e}\right)\left(\boldsymbol{\gamma}_{k0}^{(ij)} + \left(\boldsymbol{\gamma}_{k3}^{(ij)}\right)^T\boldsymbol{\mu}_\mathbf{e} - \tau\right) + \mathbf{G}_k^{(ij)}\Sigma_\mathbf{e}\boldsymbol{\gamma}_{k3}^{(ij)}$$

Finally, using the real data collected from the CNT array process, we derived the optimal settings of controllable factors as $(A, B, C) = (0.70040, 1.4101, 1136.4)$.

It should be noted that the proposed method could be applied to general process models with both off-line controllable factors and observable noise factors. The detailed procedure for minimizing the objective function with a general process model is given in the Appendix. In the cases where dataset is small or noise is large, uncertainties of parameter estimates should not be neglected. On account of the hierarchical structure of the model, the estimation uncertainties of stage 1 parameters will transfer to stage 2. For the *i*-th profile, the best linear unbiased estimator of $\boldsymbol{\theta}_i$; also, the generalized least squares estimator is

$$\hat{\boldsymbol{\theta}}_i = \left(\mathbf{S}_i{}^T\boldsymbol{\Sigma}_i^{-1}\mathbf{S}_i\right)^{-1}\mathbf{S}_i{}^T\boldsymbol{\Sigma}_i^{-1}\mathbf{y}_i$$

where $\boldsymbol{\Sigma}_i = \sigma_{z_i}^2\mathbf{R}_i + \sigma_\varepsilon{}^2\mathbf{I}$ *and* $\boldsymbol{R}_i$ is a pre-specified correlation matrix. The variance matrix of the estimator is

$$\mathrm{var}\left(\hat{\boldsymbol{\theta}}_i\right) = \left(\mathbf{S}_i{}^T\boldsymbol{\Sigma}_i^{-1}\mathbf{S}_i\right)^{-1}$$

This estimation uncertainty of $\boldsymbol{\theta}_i$ can be addressed in the second stage model

$$\boldsymbol{\theta}_i = \mathbf{B}\cdot\mathbf{f}(\mathbf{x}_i, \mathbf{u}_i, \mathbf{e}_i, \mathbf{n}_i) + \mathbf{v}_i$$

where $\boldsymbol{v}_i$ is a noise vector that follows a multivariate normal distribution $N_p(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{v}_i})$. Therefore, if we set $\boldsymbol{\Sigma}_{\mathbf{v}_i} = \mathrm{var}\left(\hat{\boldsymbol{\theta}}_i\right)$, the uncertainty of $\hat{\boldsymbol{\theta}}_i$ is included in the model. In this way, the uncertainty of parameter estimation in stage 1 propagates to the next stage, and then variance of $\hat{\mathbf{B}}$ in stage 2 can be estimated. To take estimation uncertainties into account, we use $O\left(\mathbf{x}, \mathbf{u}\middle|\hat{\mathbf{B}}\right)$ as the new objective function. By minimizing this loss function, we can find the optimal control factors' setting that is robust to noise factors and estimation uncertainties of parameters.

4.3. *Sensitivity analysis*

In the following, we treat the estimated coefficients in Equation (9) as the true parameter values of the process and apply the controllable factor settings obtained through the proposed robust design method to the process. Through the sensitivity study in the succeeding texts, we can analyze the behavior of the objective function around the optimal solutions and can evaluate the optimality of our solution intuitively.

We generate height data of nanotubes to test whether the optimal settings obtained from our algorithm are optimal compared to other controllable factor values. Because factor *A* has no interaction with factor *D* (i.e., settings of factor *A* do not impact the variation transmitted from factor *D* to the process output), we only analyze factors *B* and *C*.

Fixing factor *C* at the optimal setting that we obtained using the proposed method, we study the effect of factor *B* on the process output by moving this factor to values different from its optimal setting. At each setting of factor *B*, we generate height data of 100 profiles, which are produced under 100 different values of factor *D*. The values of factor *D* are randomly generated in accordance with a normal distribution $N(130, 40^2)$. We take the nanotube grown at the center of the profiles with location $s_{ij}=0$ as an example. Computation results of the central nanotube from 100 profiles are shown in Table I.

In Table I, the target deviation is defined as $\sum_{i=1}^{N}\left(y_{s_{ij}=0} - \tau\right)^2$, $N = 100$. The column coverage percentage of (0.25 and 0.35) shows the percentage of nanotubes with height between 0.25 and 0.35, which is defined as the target value plus/minus the standard deviation. The starred row indicates that factor *B* is set to its optimal setting.

The numerical result in Table I shows that the optimal set value of factor *B* that we obtain from the proposed method is the best among all the given settings, yielding the lowest target deviation and smallest variance. The result can be seen intuitively from Figure 3. Figure 3 shows the effect of the controllable factor *B* on the distribution of the height of the nanotube at the center of the profiles. The horizontal axis represents different settings of factor *B*, in which the setting marked with a red dotted line is the optimal value obtained by our method. Above the horizontal axis shows the distribution of the observable factor *D*. The distributions of factor *D* are the same for all the settings of factor *B*. The vertical axis of Figure 3 shows the distribution of the heights at the center point of the profile. It can be seen that the variance of responses at the optimal controllable setting is the smallest and the average height is closest to the target.

Using the same process, we study the effect of factor *C* on the process output by fixing factor *B* at the optimal setting we obtained using the proposed method. We also generate data of 100 profiles at each setting of factor *C* to compare their performance. Here, we

Table I. Study of the effect of factor *B*

Factor *B*	Average height	Standard deviation (SD)	Target deviation	Coverage percentage of (0.25 and 0.35)
1.4000	0.23358	0.04064	0.60464	32%
1.4050	0.26422	0.02103	0.60464	76%
1.4100	0.29485	0.00141	0.00285	100%
1.4101*	0.29562*	0.00092*	0.00200*	100%*
1.4150	0.32548	0.01820	0.09773	90%
1.4200	0.35612	0.03781	0.45647	47%

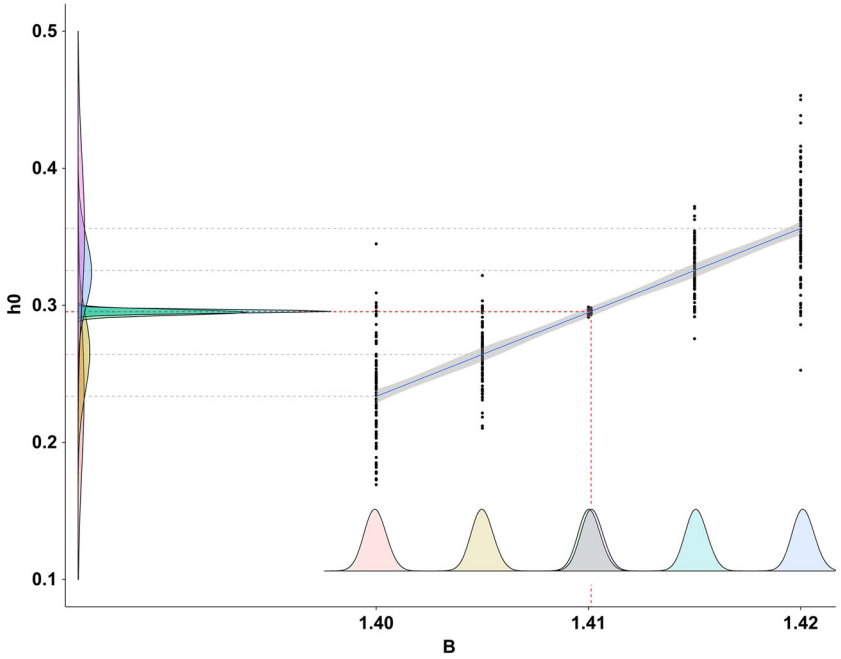

Figure 3. The effect of factor *B* on the distribution of the height at center points.

study all the nanotubes on a profile. There are $M = 91$ gage points on a profile with coordinate ranging from -90 to 90 and step length 2.

The definitions of performance metrics in Table II are similar to those in Table I. The difference is that the object of study in Table II is the average height of a nanotube profile, while that in Table I is the height of the central nanotube. Table II shows the result of different settings of factor *C*. The optimal set value of factor *C* obtained through the proposed method still has a small standard deviation and an on-target average height. Although the average height at set value 1100 is closest to target among all the set values, the corresponding variance is larger than that of the optimal setting. Similarly, the average height at set value 1200 has the smallest among all the given settings, but they are apparently off-target. The proposed method has a tradeoff between target achievement and variance reduction. The tradeoff is important in our case because uniformity of nanotubes is more desirable compared to a small target deviation. The comparison of different set values can also be seen visually in Figure 4. The vertical axis of Figure 4 shows the distribution of the average heights across the whole profile. At the optimal controllable setting, which is marked with red dotted line, the distribution of the average heights has the smallest variation, and the average height is close to the target height. Although there are some controllable settings that stay closer to the target height, the output has a larger variation.

Through the analysis previously, we can see that the proposed method produces a solution that has low deviation from the target and a small variation. The quality of nanotube profiles at the optimal controllable factor settings is robust to the variation of the observable but uncontrollable factor. Therefore, the proposed method is effective at improving the robustness of profile quality in a process where there are noise factors, and the noise factors have interactions with controllable factors.

Table II. Study of the effect of factor *C*

Factor *C*	Average height	Average standard deviation (SD)	Target deviation	Coverage percentage of (0.294 and 0.306)
900	0.29386	0.000762	0.003826	40%
1000	0.29793	0.000504	0.000456	100%
1100	0.30199	0.000245	0.000402	100%
1136.4*	0.30347*	0.000151*	0.001206*	100%*
1200	0.30605	0.000013	0.003665	0
1300	0.31012	0.000271	0.010244	0
1400	0.31418	0.000530	0.020141	0

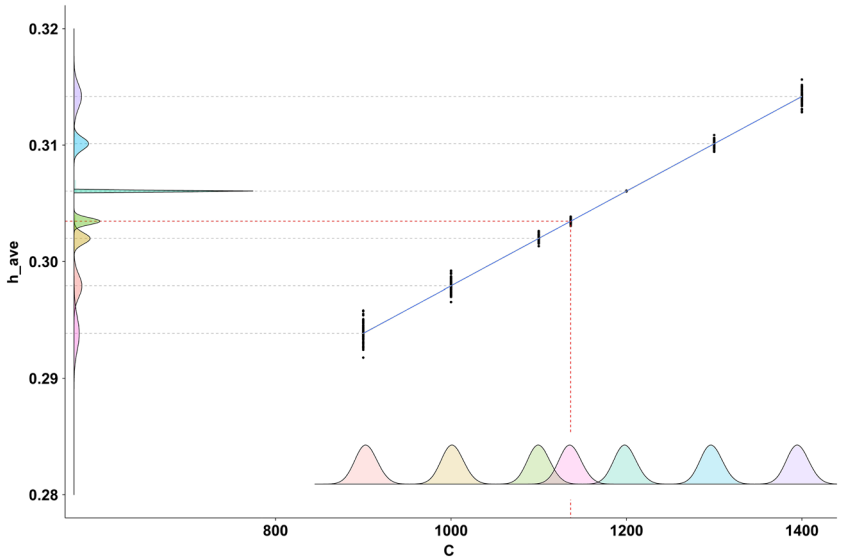

Figure 4. The effect of factor *C* on the distribution of the average height.

5. Conclusion

In this work, we propose a new method of robust parameter design for profiles. A hierarchical model is built for profiles where spatial correlation of the measurement points is characterized with Kriging technique. Then, we optimize processes by minimizing a quadratic loss function, which is defined as the sum of the squared deviations between the predicted responses and the target value. Compared with previous methods, the proposed method has the following features: first, spatial correlation is considered in the model and the optimization procedures; second, the model is built hierarchically, which makes the model meaningful and easy to interpret. The application of the method to the production of CNTs proves that the proposed method is effective on quality improvement of profiles. Through the proposed robust parameter design method, a desired shape of profiles that is insensitive to the variation caused by noise factors can be achieved.

When output quality is characterized by profiles, robust parameter design is helpful for off-line optimization of process variables. While for a process that is in a running status, online optimization of process variables becomes important. Therefore, online control and the integration of robust parameter design with online control for profile quality improvement are important topics that deserve more future research efforts.

Acknowledgements

The authors are grateful to the editor, Professor D. C. Montgomery, and two anonymous referees for their valuable suggestions, which have helped improve this work greatly. Professor Wang's work was supported by the National Natural Science Foundation of China under grants 71471096 and 71072012. Professor Huang's work was supported by National Science Foundation with CAREER grant number NSF CMM-1055394.

References

1. Bao L, Wang K, Wu T. A run-to-run controller for product surface quality improvement. *International Journal of Production Research* 2014; **52**:4469–4487.
2. Jin J, Shi J. Feature-preserving data compression of stamping tonnage information using wavelets. *Technometrics* 1999; **41**:327–339.
3. Kim K, Mahmoud MA, Woodall WH. On the monitoring of linear profiles. *Journal of Quality Technology* 2003; **35**:317–328.
4. Woodall WH, Spitzner DJ, Montgomery DC, Gupta S. Using control charts to monitor process and product quality profiles. *Journal of Quality Technology* 2004; **36**:309–320.
5. Wang K, Tsung F. Using profile monitoring techniques for a data-rich environment with huge sample size. *Quality and Reliability Engineering International* 2005; **21**:677–688.
6. Jensen WA, Birch JB. Profile monitoring via nonlinear mixed models. *Journal of Quality Technology* 2009; **41**:18–34.
7. Wang K, Jiang W. High-dimensional process monitoring and fault isolation via variable selection. *Journal of Quality Technology* 2009; **41**:247–258.
8. Plumlee M, Jin R, Roshan Joseph V, Shi J. Gaussian process modeling for engineered surfaces with applications to Si wafer production. *Stat* 2013; **2**:159–170.
9. Del Castillo E, Colosimo BM. Statistical shape analysis of experiments for manufacturing processes. *Technometrics* 2011; **53**:1–15.
10. Colosimo BM, Pacella M. Analyzing the effect of process parameters on the shape of 3D profiles. *Journal of Quality Technology* 2011; **43**:169–195.
11. Govaerts B, Noel J. Analysing the results of a designed experiment when the response is a curve: methodology and application in metal injection moulding. *Quality and Reliability Engineering International* 2005; **21**:509–520.

12. Ba S, Joseph VR. Composite Gaussian process models for emulating expensive functions. *The Annals of Applied Statistics* 2012; **6**:1838–1860.
13. Nair VN, Taam W, Ye KQ. Analysis of functional responses from robust design studies. *Journal of Quality Technology* 2002; **34**:355–370.
14. Taam W, Hamada M. Detecting spatial effects from factorial experiments: an application from integrated-circuit manufacturing. *Technometrics* 1993; **35**:149–160.
15. Wu CFJ, Hamada MS. Experiments: Planning, Analysis, and Optimization (2nd edn). John Wiley & Sons: Hoboken, 2009; 511.
16. Guo RS, Sachs E. Modeling, optimization and control of spatial uniformity in manufacturing processes. *IEEE Transactions on Semiconductor Manufacturing* 1993; **6**:41–57.
17. Montgomery DC. Design and Analysis of Experiments (8th edn). John Wiley & Sons: Hoboken, 2012; 558.
18. Hughes-Oliver JM, Lu J-C, Davis JC, Gyurcsik RS. Achieving uniformity in a semiconductor fabrication process using spatial modeling. *Journal of the American Statistical Association* 1998; **93**:36–45.
19. Peterson JJ. A posterior predictive approach to multiple response surface optimization. *Journal of Quality Technology* 2004; **36**:139–153.
20. Del Castillo E, Colosimo BM, Alshraideh H. Bayesian modeling and optimization of functional responses affected by noise factors. *Journal of Quality Technology* 2012; **44**:117–135.
21. Cressie NA. Statistics for Spatial Data (Revised edn). John Wiley & Sons: New York, 1993.
22. Huang Q. Physics-driven Bayesian hierarchical modeling of the nanowire growth process at each scale. *IIE Transactions* 2010; **43**:1–11.
23. Zhong J, Shi J, Wu JC. Design of DOE-based automatic process controller with consideration of model and observation uncertainties. *IEEE Transactions on Automation Science and Engineering* 2010; **7**:266–273.
24. Cressie N. The origins of Kriging. *Mathematical Geology* 1990; **22**:239–252.
25. Rue H, Held L. Gaussian Markov Random Fields: Theory and Applications. Chapman & Hall/CRC, Boca Raton, 2005.
26. Bao L, Wang K, Jin R. A hierarchical model for characterising spatial wafer variations. *International Journal of Production Research* 2014; **52**:1827–1842.
27. Jensen WA, Birch JB, Woodall WH. Monitoring correlation within linear profiles using mixed models. *Journal of Quality Technology* 2008; **40**:167–183.
28. Qiu P, Zou C, Wang Z. Nonparametric profile monitoring by mixed effects modeling. *Technometrics* 2010; **52**:265–277.
29. Miro-Quesada G, Del Castillo E, Peterson JJ. A Bayesian approach for multiple response surface optimization in the presence of noise variables. *Journal of Applied Statistics* 2004; **31**:251–270.
30. Haran M. Gaussian Random Field Models for Spatial Data. In Handbook of Markov Chain Monte Carlo, Brooks S, Gelman A, Jones G, Meng X-L (eds). Chapman & Hall/CRC, Boca Raton, 2011; 449–478.
31. Fang KT, Li R, Sudjianto A. Design and Modeling for Computer Experiments, Vol. **304**. Chapman & Hall/CRC: Boca Raton, 2010.
32. Alshraideh H, Del Castillo E. Gaussian process modeling and optimization of profile response experiments. *Quality and Reliability Engineering International* 2014; **30**:449–462.
33. Hung Y, Joseph VR, Melkote SN. Analysis of computer experiments with functional response. *Technometrics* 2013; **57**:35–44.
34. Koehler J, Owen A. Computer Experiments. In Handbook of Statistics 13: Design and Analysis of Experiments, Maddala GS, Rao CR (eds). Elsevier Science B.V., Amsterdan, 1996; 261–308.
35. Khuri A, Conlon M. Simultaneous optimization of multiple responses represented by polynomial regression functions. *Technometrics* 1981; **23**:363–375.
36. Ames AE, Mattucci N, Macdonald S, Szonyi G, Hawkins DM. Quality loss functions for optimization across multiple response surfaces. *Journal of Quality Technology* 1997; **29**:339–346.
37. Murphy TE, Tsui K-L, Allen JK. A review of robust design methods for multiple responses. *Research in Engineering Design* 2005; **15**:201–215.
38. Zhang X, Jiang K, Feng C, *et al.* Spinning and processing continuous yarns from 4-inch wafer scale super-aligned carbon nanotube arrays. *Advanced Materials* 2006; **18**:1505–1510.
39. Jiang K, Feng C, Liu K, Fan S. A vapor-liquid-solid model for chemical vapor deposition growth of carbon nanotubes. *Journal of Nanoscience and Nanotechnology* 2007; **7**:1494–1504.

Authors' biographies

Lulu Bao is a PhD candidate in the Department of Industrial Engineering, Tsinghua University, Beijing, China. She received her bachelor-s degree in Industrial Engineering in 2011 from Tsinghua University, Beijing, China. Her research focuses on statistical modeling, monitoring, and control of industrial process.

Qiang Huang received his PhD degree in Industrial and Operations Engineering in 2003 from the University of Michigan-Ann Arbor. He is currently an associate professor and Gordon S. Marshall Early Career Chair in Engineering in the Daniel J. Epstein Department of Industrial and Systems Engineering, University of Southern California, Los Angeles. His research focuses on modeling and analysis of complex systems for quality and productivity improvement, with special interest in integrated nanomanufacturing and nanoinformatics, and additive manufacturing. Professor Huang is a member of the Institute of Electrical and Electronics Engineers (IEEE), Institute of Industrial Engineers, INFORMS, and American Society of Mechanical Engineers. He received 2013 IEEE Transactions on Automation Science and Engineering Best Paper Award. He has been an associate editor of the IEEE Transactions on Automation Science and Engineering since 2012 and has been a member of the scientific committee (editorial board) for the North American Manufacturing Research Institution of SME, 2009–2011 and 2013–2015.

Kaibo Wang is an associate professor in the Department of Industrial Engineering, Tsinghua University, Beijing, China. He received his BS and MS degrees in Mechatronics from Xi'an Jiaotong University, Xi'an, China, and his PhD in Industrial Engineering and Engineering Management from the Hong Kong University of Science and Technology, Hong Kong. He is in the editorial board of Journal of Quality Technology. He has published papers in journals such as Journal of Quality Technology, IIE Transactions, Quality and Reliability Engineering International, International Journal of Production Research, and others. Dr Wang's research focuses on statistical quality control and data-driven complex system modeling, monitoring, diagnosis, and control, with a special emphasis on the integration of engineering knowledge and statistical theories for solving problems from real industries. He is a member of INFORMS, Institute of Industrial Engineers, and a senior member of American Society for Quality.

Appendix

The appendix gives details of the optimization process of the objective function in Equation (10). Here, we give a general case where there are controllable factors and observable factors in a process. The vector of observable factors follows a distribution with mean $\boldsymbol{\mu}_{\mathbf{e}}$ and covariance matrix $\Sigma_{\mathbf{e}}$.

For simplification, we denote that

$$\gamma_{10}^{(ij)} = \hat{\beta}_{30} + \hat{\beta}_{10}s_{ij}^2,\ \boldsymbol{\gamma}_{13}^{(ij)} = \hat{\boldsymbol{\beta}}_{33} + \hat{\boldsymbol{\beta}}_{13}s_{ij}^2,\ \boldsymbol{\gamma}_{11}^{(ij)} = \hat{\boldsymbol{\beta}}_{31} + \hat{\boldsymbol{\beta}}_{11}s_{ij}^2,\ \mathbf{G}_1^{(ij)} = \hat{\mathbf{B}}_{31} + \hat{\mathbf{B}}_{11}s_{ij}^2$$
$$\gamma_{20}^{(ij)} = \hat{\beta}_{30} + \hat{\beta}_{20}s_{ij}^2,\ \boldsymbol{\gamma}_{23}^{(ij)} = \hat{\boldsymbol{\beta}}_{33} + \hat{\boldsymbol{\beta}}_{23}s_{ij}^2,\ \boldsymbol{\gamma}_{21}^{(ij)} = \hat{\boldsymbol{\beta}}_{31} + \hat{\boldsymbol{\beta}}_{21}s_{ij}^2,\ \text{and}\ \mathbf{G}_2^{(ij)} = \hat{\mathbf{B}}_{31} + \hat{\mathbf{B}}_{21}s_{ij}^2$$

When $s_{ij} < 0$, the expected response in Equation (10) is

$$\begin{aligned} &E_{\mathbf{e},z,v,\varepsilon}\left(y_{ij}\right) \\ &\quad = E_{\mathbf{e},z,v,\varepsilon}\left[\left(\beta_{30} + \boldsymbol{\beta}_{31}^T\mathbf{x} + \boldsymbol{\beta}_{33}^T\mathbf{e} + \mathbf{x}^T\mathbf{B}_{31}\mathbf{e} + v_{i3}\right) + \left(\beta_{10} + \boldsymbol{\beta}_{11}^T\mathbf{x} + \boldsymbol{\beta}_{13}^T\mathbf{e} + \mathbf{x}^T\mathbf{B}_{11}\mathbf{e} + v_{i1}\right)s_{ij}^2 + z_{ij} + \varepsilon_{ij}\right] \\ &\quad = \left(\hat{\beta}_{30} + \hat{\boldsymbol{\beta}}_{31}^T\mathbf{x} + \boldsymbol{\beta}_{33}^T\boldsymbol{\mu}_{\mathbf{e}} + \mathbf{x}^T\hat{\mathbf{B}}_{31}\boldsymbol{\mu}_{\mathbf{e}}\right) + \left(\hat{\beta}_{10} + \hat{\boldsymbol{\beta}}_{11}^T\mathbf{x} + \hat{\boldsymbol{\beta}}_{13}^T\boldsymbol{\mu}_{\mathbf{e}} + \mathbf{x}^T\hat{\mathbf{B}}_{11}\boldsymbol{\mu}_{\mathbf{e}}\right)s_{ij}^2 \\ &\quad = \gamma_{10}^{(ij)} + \left(\boldsymbol{\gamma}_{11}^{(ij)}\right)^T\mathbf{x} + \left(\boldsymbol{\gamma}_{13}^{(ij)}\right)^T\boldsymbol{\mu}_{\mathbf{e}} + \mathbf{x}^T\mathbf{G}_1^{(ij)}\boldsymbol{\mu}_{\mathbf{e}} \end{aligned} \tag{A1}$$

The variance of response in Equation (10) can be expressed as the sum of two parts, shown in Equation (A3).

$$Var_{\mathbf{e},z,v,\varepsilon}\left(y_{ij}\right) = E_{\mathbf{e}}\left[Var_{z,v,\varepsilon}\left(y_{ij}\right)\right] + Var_{\mathbf{e}}\left[E_{z,v,\varepsilon}\left(y_{ij}\right)\right] \tag{A2}$$

Then, we calculate the two parts separately, shown in Equations (A3) and (A4), respectively.

$$\begin{aligned} &Var_{\mathbf{e}}\left[E_{z,v,\varepsilon}\left(y_{ij}\right)\right] \\ &\quad = Var_{\mathbf{e}}\left\{E_{z,v,\varepsilon}\left[\left(\beta_{30} + \boldsymbol{\beta}_{31}^T\mathbf{x} + \boldsymbol{\beta}_{33}^T\mathbf{e} + \mathbf{x}^T\mathbf{B}_{31}\mathbf{e} + v_{i3}\right) + \left(\beta_{10} + \boldsymbol{\beta}_{11}^T\mathbf{x} + \boldsymbol{\beta}_{13}^T\mathbf{e} + \mathbf{x}^T\mathbf{B}_{11}\mathbf{e} + v_{i1}\right)s_{ij}^2 + z_{ij} + \varepsilon_{ij}\right]\right\} \\ &\quad = Var_{\mathbf{e}}\left[\left(\hat{\beta}_{30} + \hat{\boldsymbol{\beta}}_{31}^T\mathbf{x} + \hat{\boldsymbol{\beta}}_{33}^T\mathbf{e} + \mathbf{x}^T\hat{\mathbf{B}}_{31}\mathbf{e}\right) + \left(\hat{\beta}_{10} + \hat{\boldsymbol{\beta}}_{11}^T\mathbf{x} + \hat{\boldsymbol{\beta}}_{13}^T\mathbf{e} + \mathbf{x}^T\hat{\mathbf{B}}_{11}\mathbf{e}\right)s_{ij}^2\right] \\ &\quad = Var_{\mathbf{e}}\left[\left(\hat{\boldsymbol{\beta}}_{33}^T\mathbf{e} + \mathbf{x}^T\hat{\mathbf{B}}_{31}\mathbf{e}\right) + \left(\hat{\boldsymbol{\beta}}_{13}^T\mathbf{e} + \mathbf{x}^T\hat{\mathbf{B}}_{11}\mathbf{e}\right)s_{ij}^2\right] \\ &\quad = \left[\left(\hat{\boldsymbol{\beta}}_{33} + \hat{\mathbf{B}}_{31}{}^T\mathbf{x}\right) + \left(\hat{\boldsymbol{\beta}}_{13} + \hat{\mathbf{B}}_{11}{}^T\mathbf{x}\right)s_{ij}^2\right]^T\Sigma_{\mathbf{e}}\left[\left(\hat{\boldsymbol{\beta}}_{33} + \hat{\mathbf{B}}_{31}{}^T\mathbf{x}\right) + \left(\hat{\boldsymbol{\beta}}_{13} + \hat{\mathbf{B}}_{11}{}^T\mathbf{x}\right)s_{ij}^2\right] \\ &\quad = \left[\boldsymbol{\gamma}_{13}^{(ij)} + \left(\mathbf{G}_1^{(ij)}\right)^T\mathbf{x}\right]^T\Sigma_{\mathbf{e}}\left[\boldsymbol{\gamma}_{13}^{(ij)} + \left(\mathbf{G}_1^{(ij)}\right)^T\mathbf{x}\right] \end{aligned} \tag{A3}$$

$$\begin{aligned} &E_{\mathbf{e}}\left[Var_{z,v,\varepsilon}\left(y_{ij}\right)\right] \\ &\quad = E_{\mathbf{e}}\left\{Var_{z,v,\varepsilon}\left[\left(\beta_{30} + \boldsymbol{\beta}_{31}^T\mathbf{x} + \boldsymbol{\beta}_{33}^T\mathbf{e} + \mathbf{x}^T\mathbf{B}_{31}\mathbf{e} + v_{i3}\right) + \left(\beta_{10} + \boldsymbol{\beta}_{11}^T\mathbf{x} + \boldsymbol{\beta}_{13}^T\mathbf{e} + \mathbf{x}^T\mathbf{B}_{11}\mathbf{e} + v_{i1}\right)s_{ij}^2 + z_{ij} + \varepsilon_{ij}\right]\right\} \\ &\quad = E_e\left[Var_{z,v,\varepsilon}\left(v_{i3} + v_{i1}s_{ij}^2 + z_{ij} + \varepsilon_{ij}\right)\right] \\ &\quad = \left(s_{ij}^2 \quad 0 \quad 1\right)\boldsymbol{\Sigma}_{\mathbf{v}_i}\left(s_{ij}^2 \quad 0 \quad 1\right)^T + \sigma_z{}^2 + \sigma_\varepsilon{}^2 \end{aligned} \tag{A4}$$

Using the same methods, we can obtain the expectation and variation of responses when $s_{ij} \geq 0$.

Then, the objective function shown in Equation (10) is written as

$$\begin{aligned} O(\mathbf{x}) &= \sum_{i=1}^N\sum_{j,\ s_i<0}\left\{\left[E_{e,z,v,\varepsilon}\left(y_{ij}\right) - \tau\right]^2 + Var_{e,z,v,\varepsilon}\left(y_{ij}\right)\right\} \\ &\quad + \sum_{i=1}^N\sum_{j,\ s_j\geq 0}\left\{\left[E_{e,z,v,\varepsilon}\left(y_{ij}\right) - \tau\right]^2 + Var_{e,z,v,\varepsilon}\left(y_{ij}\right)\right\} \\ &= \sum_{i=1}^N\sum_{j,\ s_i<0}\left\{\left[\gamma_{10}^{(ij)} + \left(\boldsymbol{\gamma}_{11}^{(ij)}\right)^T\mathbf{x} + \left(\boldsymbol{\gamma}_{13}^{(ij)}\right)^T\boldsymbol{\mu}_{\mathbf{e}} + \mathbf{x}^T\mathbf{G}_1^{(ij)}\boldsymbol{\mu}_{\mathbf{e}} - \tau\right]^2\right. \\ &\quad \left. + \left[\boldsymbol{\gamma}_{13}^{(ij)} + \left(\mathbf{G}_1^{(ij)}\right)^T\mathbf{x}\right]^T\Sigma_{\mathbf{e}}\left[\boldsymbol{\gamma}_{13}^{(ij)} + \left(\mathbf{G}_1^{(ij)}\right)^T\mathbf{x}\right] + \left(s_{ij}^2 \quad 0 \quad 1\right)\boldsymbol{\Sigma}_{\mathbf{v}_i}\left(s_{ij}^2 \quad 0 \quad 1\right)^T + \sigma_z{}^2 + \sigma_\varepsilon{}^2\right\} \\ &\quad + \sum_{i=1}^N\sum_{j,\ s_j\geq 0}\left\{\left[\gamma_{20}^{(ij)} + \left(\boldsymbol{\gamma}_{21}^{(ij)}\right)^T\mathbf{x} + \left(\boldsymbol{\gamma}_{23}^{(ij)}\right)^T\boldsymbol{\mu}_{\mathbf{e}} + \mathbf{x}^T\mathbf{G}_2^{(ij)}\boldsymbol{\mu}_{\mathbf{e}} - \tau\right]^2\right. \\ &\quad \left. + \left[\boldsymbol{\gamma}_{23}^{(ij)} + \left(\mathbf{G}_2^{(ij)}\right)^T\mathbf{x}\right]^T\Sigma_{\mathbf{e}}\left[\boldsymbol{\gamma}_{23}^{(ij)} + \left(\mathbf{G}_2^{(ij)}\right)^T\mathbf{x}\right] + \left(0 \quad s_{ij}^2 \quad 1\right)\boldsymbol{\Sigma}_{\mathbf{v}_i}\left(0 \quad s_{ij}^2 \quad 1\right)^T + \sigma_z{}^2 + \sigma_\varepsilon{}^2\right\} \end{aligned} \tag{A5}$$

Then, the differential of the objective function is given as in Equation (A6).

$$\begin{aligned}\frac{d\,O(\mathbf{x})}{d\,\mathbf{x}} &= \sum_{i=1}^{N}\sum_{j,\ s_i<0} 2\left(\boldsymbol{\gamma}_{11}^{(ij)} + \mathbf{G}_1^{(ij)}\boldsymbol{\mu}_{\mathbf{e}}\right)\left[\gamma_{10}^{(ij)} + \left(\boldsymbol{\gamma}_{13}^{(ij)}\right)^T\boldsymbol{\mu}_{\mathbf{e}} + \left(\boldsymbol{\gamma}_{11}^{(ij)} + \mathbf{G}_1^{(ij)}\boldsymbol{\mu}_{\mathbf{e}}\right)^T\mathbf{x} - \tau\right] \\ &+ \sum_{i=1}^{N}\sum_{j,\ s_i<0} 2\mathbf{G}_1^{(ij)}\Sigma_{\mathbf{e}}\left[\boldsymbol{\gamma}_{13}^{(ij)} + \left(\mathbf{G}_1^{(ij)}\right)^T\mathbf{x}\right] \\ &+ \sum_{i=1}^{N}\sum_{j,\ s_j\geq 0} 2\left(\boldsymbol{\gamma}_{21}^{(ij)} + \mathbf{G}_2^{(ij)}\boldsymbol{\mu}_{\mathbf{e}}\right)\left[\gamma_{20}^{(ij)} + \left(\boldsymbol{\gamma}_{23}^{(ij)}\right)^T\boldsymbol{\mu}_{\mathbf{e}} + \left(\boldsymbol{\gamma}_{21}^{(ij)} + \mathbf{G}_2^{(ij)}\boldsymbol{\mu}_{\mathbf{e}}\right)^T\mathbf{x} - \tau\right] \\ &+ \sum_{i=1}^{N}\sum_{j,\ s_j\geq 0} 2\mathbf{G}_2^{(ij)}\Sigma_{\mathbf{e}}\left[\boldsymbol{\gamma}_{23}^{(ij)} + \left(\mathbf{G}_2^{(ij)}\right)^T\mathbf{x}\right]\end{aligned} \tag{A6}$$

By forcing the differential to 0, we obtain the optimal solution of the controllable factors as follows

$$\begin{aligned}\mathbf{x} = -&\left\{\sum_{i=1}^{N}\sum_{j,\ s_i<0}\left[\left(\boldsymbol{\gamma}_{11}^{(ij)} + \mathbf{G}_1^{(ij)}\boldsymbol{\mu}_{\mathbf{e}}\right)\left(\boldsymbol{\gamma}_{11}^{(ij)} + \mathbf{G}_1^{(ij)}\boldsymbol{\mu}_{\mathbf{e}}\right)^T + \mathbf{G}_1^{(ij)}\Sigma_{\mathbf{e}}\left(\mathbf{G}_1^{(ij)}\right)^T\right]\right. \\ &\left.+\sum_{i=1}^{N}\sum_{j,\ s_j\geq 0}\left[\left(\boldsymbol{\gamma}_{21}^{(ij)} + \mathbf{G}_2^{(ij)}\boldsymbol{\mu}_{\mathbf{e}}\right)\left(\boldsymbol{\gamma}_{21}^{(ij)} + \mathbf{G}_2^{(ij)}\boldsymbol{\mu}_{\mathbf{e}}\right)^T + \mathbf{G}_2^{(ij)}\Sigma_{\mathbf{e}}\left(\mathbf{G}_2^{(ij)}\right)^T\right]\right\}^{-1} \\ &\cdot\left\{\sum_{i=1}^{N}\sum_{j,\ s_i<0}\left(\boldsymbol{\gamma}_{11}^{(ij)} + \mathbf{G}_1^{(ij)}\boldsymbol{\mu}_{\mathbf{e}}\right)\left[\gamma_{10}^{(ij)} + \left(\boldsymbol{\gamma}_{13}^{(ij)}\right)^T\boldsymbol{\mu}_{\mathbf{e}} - \tau\right] + \mathbf{G}_1^{(ij)}\Sigma_{\mathbf{e}}\boldsymbol{\gamma}_{13}^{(ij)}\right. \\ &\left.+\sum_{i=1}^{N}\sum_{j,\ s_j\geq 0}\left(\boldsymbol{\gamma}_{21}^{(ij)} + \mathbf{G}_2^{(ij)}\boldsymbol{\mu}_{\mathbf{e}}\right)\left[\gamma_{20}^{(ij)} + \left(\boldsymbol{\gamma}_{23}^{(ij)}\right)^T\boldsymbol{\mu}_{\mathbf{e}} - \tau\right] + \mathbf{G}_2^{(ij)}\Sigma_{\mathbf{e}}\boldsymbol{\gamma}_{23}^{(ij)}\right\}\end{aligned} \tag{A7}$$

Denoting that

$$\mathbf{A}_k^{(ij)} = \left(\boldsymbol{\gamma}_{k1}^{(ij)} + \mathbf{G}_k^{(ij)}\boldsymbol{\mu}_{\mathbf{e}}\right)\left(\boldsymbol{\gamma}_{k1}^{(ij)} + \mathbf{G}_k^{(ij)}\boldsymbol{\mu}_{\mathbf{e}}\right)^T + \mathbf{G}_k^{(ij)}\Sigma_{\mathbf{e}}\left(\mathbf{G}_k^{(ij)}\right)^T \textit{ and}$$

$$\underset{\mathbf{d}}{^{k(ij)}} = \left(\boldsymbol{\gamma}_{k1}^{(ij)} + \boldsymbol{\gamma}_{k1}^{(ij)}\boldsymbol{\mu}_{\mathbf{e}}\right)\left(\gamma_{k0}^{(ij)} + \left(\boldsymbol{\gamma}_{k3}^{(ij)}\right)^T\boldsymbol{\mu}_{\mathbf{e}} - \tau\right) + \mathbf{G}_k^{(ij)}\Sigma_{\mathbf{e}}\boldsymbol{\gamma}_{k3}^{(ij)}$$

we can get

$$\mathbf{x} = -\left(\sum_{i=1}^{N}\sum_{j,\ s_i<0}\mathbf{A}_1^{(ij)} + \sum_{i=1}^{N}\sum_{j,\ s_j\geq 0}\mathbf{A}_2^{(ij)}\right)^{-1}\cdot\left(\sum_{i=1}^{N}\sum_{j,\ s_i<0}\mathbf{d}_1^{(ij)} + \sum_{i=1}^{N}\sum_{j,\ s_j\geq 0}\mathbf{d}_2^{(ij)}\right)$$