A Spatial Calibration Model for Nanotube Film Quality Prediction

Xin Wang, Su Wu, Kaibo Wang, Xinwei Deng, Liang Liu, and Qi Cai

***Abstract*—A carbon nanotube (CNT) film, which is drawn from a CNT array, is a spatially distributed thin film with unique and appealing properties. Novel devices have been developed based on CNT films. The anisotropy of a CNT film, which is a spatially distributed quality index, is difficult to measure in practice due to metrology and cost constraints. As the anisotropy is highly correlated with the height of the CNT array and the height can be measured in a much easier and more cost-effective way, we propose a spatial model for predicting the anisotropy using the height. The model takes the spatially distributed two-dimensional (2-D) height as an input and provides a predicted anisotropy distribution in a 2-D space. If the anisotropy measures are obtained, the model can provide a more accurate prediction. The performance of the proposed model is verified by both a simulation study and real data samples.**

***Note to Practitioners*—Timely and accurate measurement of key product features is essential in scale-up nanomanufacturing processes. Even though a fast growth of metrology technology has been seen in recent years, some variables of nanoscale products are still hard to measure, either too costly or too time consuming, in high-speed large-scale production. However, physical mechanisms may suggest that a hard-to-measure variable may be correlated with another easy-to-measure variable. In such a case, a spatial calibration model could be constructed, based on which the prediction of the hard-to-measure variable is achievable given measures of the easy-to-measure variable. Such a calibration model provides an effective alternative to physical metrology tools in large-scale nanomanufacturing processes in which metrology technology is not fully ready yet.**

***Index Terms*—Calibration model, Gaussian process, kriging, nanomanufacturing, spatial correlation.**

Manuscript received July 01, 2014; revised December 19, 2014; accepted March 06, 2015. Date of publication April 14, 2015; date of current version April 05, 2016. This paper was recommended for publication by Associate Editor Q. Huang and Editor M. Wang upon evaluation of the reviewers' comments. The work of S. Wu was supported by the National Natural Science Foundation of China (NSFC) under Grant 71471097, the work of K. Wang was supported by NSFC under Grant 71471096 and Grant 71072012, and the work of X. Deng was supported by the National Science Foundation under Grant CMMI-1433571. *(Corresponding author: Kaibo Wang.)*

X. Wang, S. Wu, and K. Wang are with the Department of Industrial Engineering, Tsinghua University, Beijing 100084, China (e-mail: kbwang@tsinghua.edu.cn).

X. Deng is with the Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061 USA.

L. Liu and Q. Cai are with the Tianjin FuNaYuanChuang Technology Co. Ltd, Tianjin 300457, China.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASE.2015.2413833

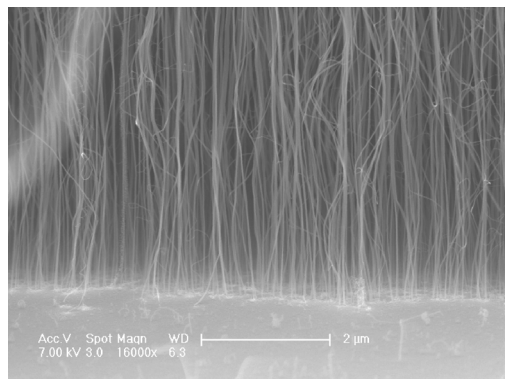


Fig. 1. SEM images of the CNT array on the silicon wafer in a side view.

I. Introduction

DUE to the fast development of nanotechnology, the requirement for high-precision and cost- effective metrology devices is extremely urgent. Certain quality characteristics are inherently difficult to measure due to their fragile physical structures. Some newly developed metrology devices are sensitive to noise, and high precision is difficult to achieve. Certain metrology procedures are destructive and cost inefficient. To tackle the physical bottleneck of metrology technology, in this paper, we seek a statistical solution to the metrology problem. It is found from engineering practice and statistical analysis that certain quality variables are highly correlated. Therefore, if an appropriate statistical model is traceable, it is potentially possible to replace the metrology procedure with the model and predict difficult-to-measure variables using easy-to-measure variables.

In this research, we focus on a nanomanufacturing process that produces carbon nanotube (CNT) films. CNTs have been considered a potential substitute for many traditional materials due to their excellent photoelectric properties, mechanical strength and flexibility [1]. Continuous and large-scale CNT films, which are drawn from a super-aligned CNT array synthesized on silicon wafers using the chemical vapor deposition (CVD) method [2], have been developed to produce novel products such as touch panels, LCDs and other optoelectronic devices [3], [4]. The CNT array (see Fig. 1), which is essentially a forest of tall and super-aligned CNTs, is synthesized by a CVD process on a wafer basis. When the CNT arrays are drawn from one side away from the wafer (see Fig. 2), a thin film is formed naturally due to the unique adhesion properties of the CNTs.

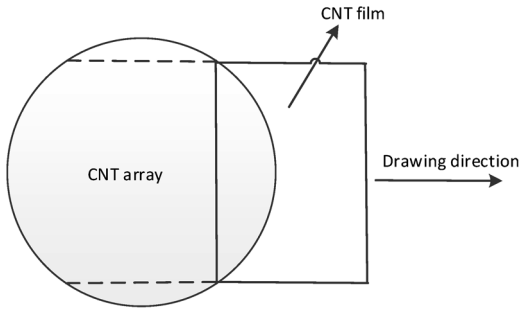


Fig. 2. Schematic drawing of the cut lines.

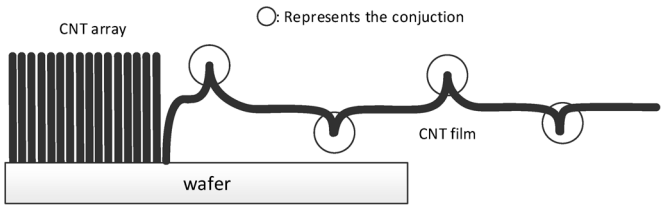


Fig. 3. Diagram of the mechanism for transforming the CNT array to film when drawing from one side of the array.

The quality of devices made from CNT films is highly dependent on the quality of the films being used. Among other factors, the anisotropy of the CNT film is one of the most important quality characteristics and is determined by the resistances of the film along two perpendicular directions: the horizontal resistance parallel to the pulling direction and the vertical resistance perpendicular to the pulling direction. The anisotropy of a CNT film directly affects the sensitivity and usability of the touch screen made from it. To measure the anisotropy, a small film sample must be torn away from the large film, that is, measuring anisotropy is destructive, and in practice, it can only be measured at very limited locations for each CNT array. In addition, measuring resistance involves connecting the small piece of CNT film in an electrical circuit. As the resistance value is relatively small, it is easily affected by the node of the electrical circuit, the positioning of the film in a fixture by different operators, the pressure applied to fix the film, and other noise factors in the measurement process.

It is first noticed from a statistical analysis of massive production data that the height and the anisotropy are linearly and positively correlated. Fig. 4 shows a scatter plot of the height and the anisotropy; the anisotropy increases as the height increases. Further discussion with the nanoscientists suggests an explanation for this phenomenon. The parallel resistance of the film is mainly driven by the resistance of the conjunctions between adjacent CNTs. See Fig. 3 for the formation of a film from an array. As the height of the CNT array increases, the number of conjunctions in the CNT film and its parallel resistance decrease. As the vertical resistance is independent of the height and the anisotropy is calculated as the ratio of the vertical and horizontal resistances, the anisotropy is therefore positively correlated with the height of the CNT array. Fortunately, the height of a CNT array can be measured easily in a nondestructive way with a very high accuracy.

In light of the above findings, in this work, we intend to develop a spatial statistical model to link the anisotropy with the

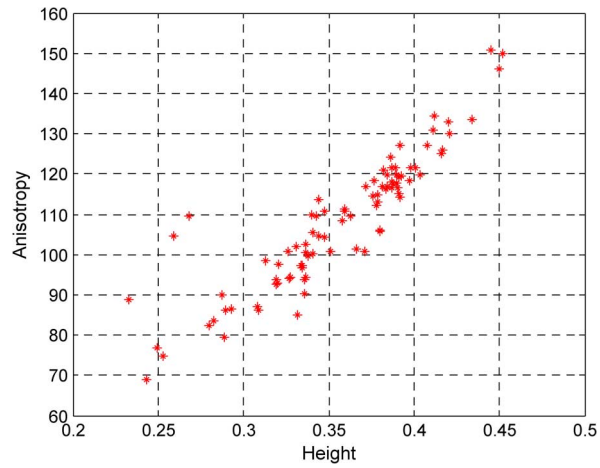


Fig. 4. Scatter plot of the height and anisotropy for the given measurement location in the CNT array.

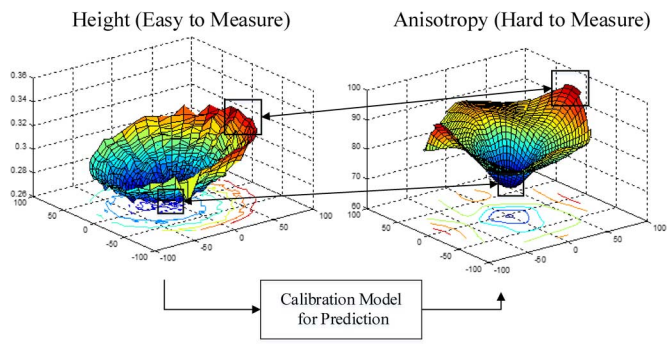


Fig. 5. Diagram of the basic idea in this research.

measured height. Based on this model, by using the height map of a CNT array, the spatially distributed anisotropy map can be predicted (see Fig. 5).

The remainder of this paper is organized as follows. Section II presents a review of the related literature. Section III proposes the spatial hierarchical model, parameter estimation, and prediction procedures. Section IV verifies the effectiveness of our method by simulation and is followed by the case study that is used to evaluate the performance of the proposed method by a comparison with two conventional methods in Section V. The last section concludes this work.

II. Literature Review

Recently, as one approach for precise uncertainty quantification and modeling, engineering knowledge-guided statistical approaches have been applied successfully in certain nanotechnology applications, such as experimental design, process modeling and control for the synthesis [5]–[8], and characterization of nanomaterials [9]–[13]. However, the spatial metrology issues in nanomanufacturing are not discussed in these works.

When two variables are correlated, and one variable is easy to measure precisely while the other requires much more effort or expense, a statistical calibration model can be developed to relate the quantities of interest [14]. Eisenhart [15] proposed two models to tackle such calibration problems. In one of the models, the author suggested treating the easily measured

quantity as a response and the other as an explanatory variable and building a model to link the two variables together. More recently, a wide spectrum of research has been described that compares and analyzes the method using different criteria in different scenarios, such as prediction interval calculation [15]–[19] and sample size determination [20]. Parker *et al.* [21] provided a general guideline for practitioners on how to choose the calibration method according to a specific application. All the works mentioned above are essentially based on univariate linear regression. However, the major challenge in this research lies in building an effective model for linking two spatial surfaces on the nanoscale, which can be considered an extension of the traditional calibration problem to spatial data and an application in nanotechnology.

Kriging (or Gaussian process) [22] is a popular technique for handling spatial data. One of the most popular models, known as universal kriging, is given as

$$Y(x) = \mu(x) + Z(x) \tag{1}$$

where $Y(x)$ is a stochastic process, $x \in \mathbb{R}^p$, $\mu(x) = \sum_{i=0}^{m} \mu_i g_i(x)$ and $Z(x)$ is a weak stationary stochastic process with mean 0 and covariance matrix $\sigma^2 R$. The $\{g_i(x), i = 1, \ldots, m\}$ is a set of pre-specified functions, and $\{\mu_i, i = 1, \ldots, m\}$ is the corresponding parameters. Under the stationary assumptions, the predictor which minimizes the mean square error can be derived as [23], [24]

$$\hat{y}(x) = \sum_{i=1}^{m} \hat{\mu}_i g_i(x) + c^T R^{-1}(Y - G\hat{\mu}) \tag{2}$$

where $c^T(x) = [r(x - x_1), r(x - x_2), \ldots, r(x - x_M)]$, G is the design matrix, $\hat{\mu} = (G^T R^{-1} G)^{-1} G^T R^{-1} Y$.

The conventional kriging methods perform poorly if the assumptions are violated [25]. Hence, some extension, such as blind kriging [25] and the composite Gaussian process [26] have been proposed to make the model more flexible and robust for different real applications. However, these kriging-based methods cannot be used to build models between two spatial variables directly because of the computational issue resulting from the high dimensionality, especially when the spatial data collection uses a very high sampling rate [27]. To address this issue, techniques such as wavelet decomposition [28], principal component analysis [29], [30], and the Kronecker products-based method [27] have been applied instead.

Shi *et al.* [31]–[33] developed a Gaussian process functional regression model, with the mean structure modeled by a functional regression and the covariance structure described by a Gaussian process, for modeling the relationship between two types of functional data. Recently, Zhang *et al.* [34] developed a bivariate spatial model to predict a surface property of a nanomaterial using another surface property, which opened a new route for the improvement of precision and the resolution of material characterization without enhanced instrumentation capabilities. The bivariate spatial model can be easily extended to the model for multiple samples that is presented in Section V, although it is developed based on only one sample. However, these models cannot deliver accurate predictions for our problem, as they do not consider the spatial correlation of the relationship between two characteristics, which is a key feature of our problem.

TABLE I
REGRESSION PARAMETERS FOR THE ANISOTROPY AND HEIGHT VALUES FROM DIFFERENT LOCATIONS

		1	**2**	**3**	**4**	**5**
Slope	Est	266.8	259.2	292.4	352.4	373.9
	Std	30.65	33.22	33.39	26.42	20.40
Intercept	Est	18.1	19.9	9.4	-1.8	-4.9
	Std	5.59	6.02	6.08	4.79	3.79
		6	**7**	**8**	**9**	**10**
Slope	Est	321.1	278.8	256.4	290.4	278.9
	Std	29.37	37.82	36.22	40.98	56.49
Intercept	Est	2.7	14.1	22.6	9.4	11.3
	Std	5.24	6.73	6.67	7.24	9.95
		11	**12**	**13**		
Slope	Est	310.3	263.9	277.7		
	Std	26.90	28.15	41.32		
Intercept	Est	5.2	15.5	10.7		
	Std	4.77	5.02	7.17		

Note: Est stands for "Estimated Value"

III. SPATIAL CALIBRATION MODEL FOR ANISOTROPY PREDICTION

A. Investigation of Spatial Correlations Equations

To achieve a better understanding of the relationship between the anisotropy and the height in a spatial domain, we first collect measurements of the two variables at 13 locations across 17 CNT array samples, treat the anisotropy as a response variable and the height as an explanatory variation, and fit a linear regression model. The details of the data collection scheme are presented in Section V. The estimated slope and intercept parameters as well as their estimated standard deviations are shown in Table I. Because the measurement process is destructive, it is difficult to get numerous CNTs array samples, thus the standard deviation of some parameters are relatively large. In kriging analysis, the semivariogram is one tool for studying spatial variations and correlations. Considering the location information of each estimated parameter, we further draw the semivariograms of the fitted slopes and intercepts based on Table I using the following equation:

$$\hat{\gamma}^k(h) = \frac{1}{2N_h} \sum_{i=1}^{N_h} \left[\beta^k(x_i + h) - \beta^k(x_i)\right]^2 \; k = 0 \; or \; 1 \tag{3}$$

where N_h is the number of location pairs with the distance of h and $\beta^k(x_i)$ is the regression parameter at location x_i. If $k = 0$, it represents the intercept; otherwise, it refers to the slope. The calculated semivariogram is further used to fit a semivariogram model of the Gaussian correlation structure as follows:

$$\gamma(h) = \begin{cases} 0, & h = 0 \\ c_0 + c_e \left\{1 - \exp\left(-\theta \cdot \|h\|^2\right)\right\}, & h \neq 0. \end{cases} \tag{4}$$

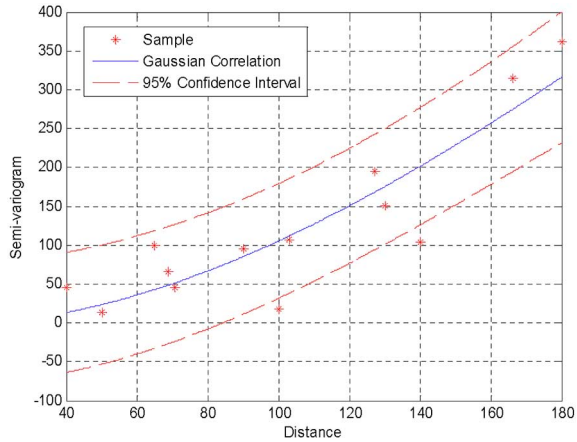


Fig. 6. Sample and fitted semivariogram for intercept.

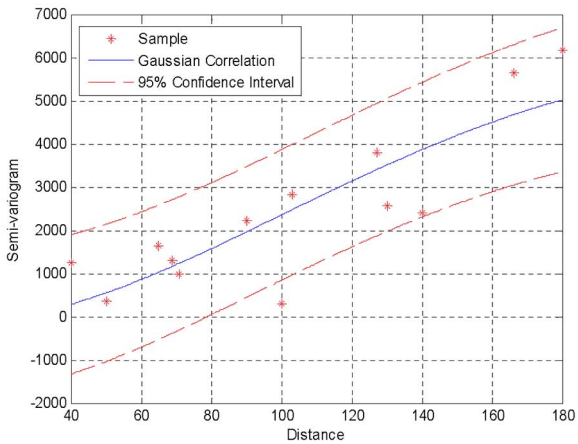


Fig. 7. Sample and fitted semivariogram for slope.

The fitted correlation models as well as their 95% confidence intervals are given in Figs. 6 and Fig. 7. It is observed that the semivariogram curves of the two regression parameters rises as the distance increases, and most of the calculated semivariogram values are located within the 95% confidence intervals. Hence, the relationship, represented by the regression parameters, possesses certain spatial correlations that can be characterized by the Gaussian correlation structure.

Therefore, a model is required to characterize not only the spatial properties of the two quantities but also the spatial correlation of the relationship between them. This point makes our research different from the methods mentioned in the previous section, such as those described by Shi *et al.* [31]–[33], which mainly focus on modeling the differences between replications instead of locations.

B. Spatial Model Development

As mentioned before, both the height and the anisotropy are considered to be spatial variables that are obtained by measuring the same quality characteristic at multiple locations of the CNT array. The measurement locations are denoted as x $= \{x_i, i = 1, \ldots, M\}$. Let $t_{x_i}, j(j = 1, \ldots, N)$ be the height at location x_i of the jth CNT array and $y_{x_i,j}(j = 1, \ldots, N)$ denote the anisotropy of location x_i at the jth CNT array. Based on the explanation and notation above, we propose using a hierarchical spatial model of two levels. The first level mainly focuses on capturing the general trend and the spatial property of the anisotropy:

$$y(t, x_i) = h(t)\beta(x_i) + Z(x_i) + \varepsilon_i, \qquad i = 1, \ldots, M \quad (5)$$

where $y(t, x_i)$ is the response (anisotropy) at location x_i given the corresponding height of t, $h(t)$ is a set of known functions. In this application, as engineering knowledge suggests a linear relationship between the height and the anisotropy at any given location, $h(t)$ is a $\{1, t\}$ vector, which has been illustrated and explained from both physical and statistical perspective in Section I.

$Z(x)$ represents the correlation structure of anisotropy at different locations (the local variation) and is also assumed to follow a Gaussian process model:

$$Z(x) \sim GP\left(0, \sigma_z^2 K(x)\right) \quad (6)$$

where $K(x)$ is the covariance structure with the (i, k)th elements being $r_z(\theta_z, x_i - x_k), i, k = 1, \ldots M, \sigma_z^2$ is the variance.

ε are the independent and normally distributed random errors that correspond to measurement errors at different locations and satisfy

$$\varepsilon \sim N(0, \tau^2 I) \quad (7)$$

where τ^2 is the variance and I is an identity matrix.

$\beta(x_i) = \{\beta^{(0)}(x_i, \beta^{(1)}(x_i)\}$ are the coefficients of $h(t)$ for a given location x_i, and its spatial property can be characterized by a second level model as follows:

$$\beta^{(0)}(x) \sim GP\left(u^{(0)}w(x), \sigma_{\beta_0} R_{\beta_0}(x)\right) \quad (8)$$

$$\beta^{(1)}(x) \sim GP\left(u^{(1)}w(x), \sigma_{\beta_1} R_{\beta_1}(x)\right) \quad (9)$$

where $u^{(0)}w(x)$ and $u^{(1)}w(x)$ are the means structures of $\beta^{(0)}(x)$ and $\beta^{(1)}(x)$ respectively, $w(x)$ are a set of prespecified functions, $u^{(0)}$ and $u^{(1)}$ are the corresponding coefficients, the covariance structures are $R_{\beta_0}(x)$ and $R_{\beta_1}(x)$ with (i, k)th elements of $r_{\beta_0}(\theta_{\beta_0}, x_i - x_k), i, k = 1, \ldots M$ and $r_{\beta_1}(\theta_{\beta_1}, x_i - x_k), i, k = 1 \ldots, M$, respectively, and σ_{β_0} and σ_{β_1} are the variances of the constant term and the slope term, respectively.

In summary, we proposed a two-level spatial model for predicting the anisotropy using the height, the hierarchical structure of which is illustrated in Fig. 8. In this model, the variability in the response can be divided into two parts, the first is the general trend, which is captured by $h(t)\beta(x_i)$, and the second is the local variation, which is characterized by $Z(x_i)$. Accordingly, the parameters are estimated separately and iteratively in the next subsection.

It is noted that $y(t, x_i)$ is different from the $y_{x_i,j}$ defined above. $y(t, x_i)$ is a function between the response (anisotropy) and the covariates (height and location), while $y_{x_i,j}$ is the measurement result of anisotropy at a given location x_i in the

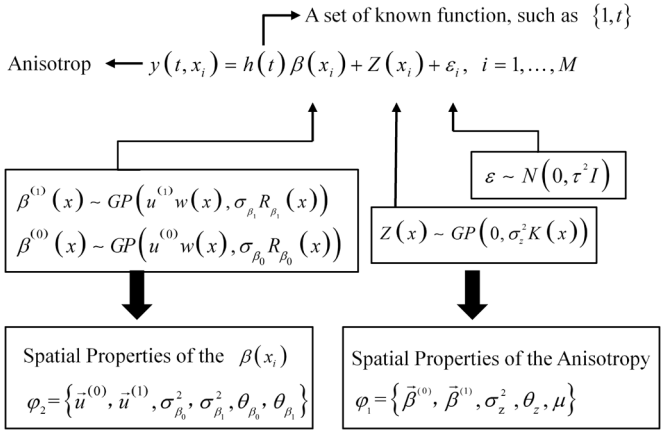


Fig. 8. Hierarchical structure of the proposed model.

jth CNTs array. For the jth replicate (or CNTs array), we denote

$$\vec{y}_j=\begin{pmatrix}y_{x_1,j}\\ \vdots\\ y_{x_M,j}\end{pmatrix}\quad \vec{t}_j=\begin{pmatrix}t_{x_1,j}\\ \vdots\\ t_{x_M,j}\end{pmatrix}$$
$$\vec{\beta}^{(0)}=\begin{pmatrix}\beta_{x_1}^{(0)}\\ \vdots\\ \beta_{x_M}^{(0)}\end{pmatrix}\quad \vec{\beta}^{(1)}=\begin{pmatrix}\beta_{x_1}^{(1)}\\ \vdots\\ \beta_{x_M}^{(1)}\end{pmatrix}.$$

It is also worth noting that the proposed model is similar to the widely used Bayesian hierarchical model in the spirit of pursuing the hierarchical structure. Compared with the usual Bayesian hierarchical model, the proposed method has the following advantages.

1) The parameters in the proposed method have better interpretability. Instead of using statistically necessary parameters and hyper-parameters, the parameters used in the proposed model have explicit engineering implications.
2) The parameters in the Bayesian hierarchical model often need to be estimated via MCMC, which can be computationally intensive and may not have closed form.

However, the Bayesian method can take advantage of the prior knowledge to improve the model, for example, Huang [7] integrated the Bayesian hierarchical modeling with physical knowledge. The combined use of Bayesian hierarchical models and engineering knowledge in solving real problems is a promising method that deserves more attention in further research.

C. Parameter Estimation

Based on the hierarchical model (5)–(9), the unknown parameters that need to be estimated include

$$\{\vec{\beta}^{(0)},\vec{\beta}^{(1)},\vec{u}^{(0)},\vec{u}^{(1)},\sigma_{\beta 0}^2,\sigma_{\beta 1}^2,\sigma_z^2,\tau^2,\theta_{\beta 0},\theta_{\beta 1},\theta_z\}. \quad (10)$$

Before estimating the parameters in (10), a training data set needs to be collected as

$$D=\{D_j,j=1,2,\ldots N\} \quad (11)$$

where $D_j=\{y_{x_i},j,t_{x_i},j,x_i,i=1,2,\ldots,M\}$.

We assume the production and measurements of each replicate (or CNT array) are independent of each other; hence

$$f(\vec{y}_1,\ldots\vec{y}_M)=\prod_{j=1}^{M}f(\vec{y}_j) \quad (12)$$

where $f(\vec{y}_1,\ldots\vec{y}_M)$ is the density function of joint distribution for $\{\vec{y}_1,\ldots\vec{y}_M\}$ and $f(\vec{y}_j)$ is the density function of the marginal distribution for $\vec{y}_j$.

For each replicate, the density function of the marginal distribution is

$$f(\vec{y}_j)=f(\vec{y}_j|\vec{\beta})f(\vec{\beta}). \quad (13)$$

Hence, based on (12) and (13), the corresponding log-likelihood function is as

$$\begin{aligned}L&=\log\left\{\prod_{j=1}^{N}f(\vec{y}_j|\vec{\beta})f(\vec{\beta})\right\}\\&=\sum_{j=1}^{N}\log f(\vec{y}_j|\vec{\beta})+N\bullet\log f(\vec{\beta})\\&=-\sum_{j=1}^{N}\left\{\log\left|\left(\sigma_z^2K+\tau^2I\right)\right|^{\frac{1}{2}}\right\}\\&\quad-\frac{1}{2}\sum_{j=1}^{N}b_j^T\left(\sigma_z^2K+\tau^2I\right)^{-1}b_j\\&\quad-N\bullet\left\{\log\left|\left(\sigma_{\beta_0}^2R_{\beta_0}\right)\right|^{\frac{1}{2}}+\log\left|\left(\sigma_{\beta_1}^2R_{\beta_1}\right)\right|^{\frac{1}{2}}+\frac{1}{2}\right\}\\&\quad-\frac{N}{2}\left(\vec{\beta}^{(1)}-u^{(1)}W\right)^T\left(\sigma_{\beta_1}^2R_{\beta_1}\right)^{-1}\left(\vec{\beta}^{(1)}-u^{(1)}W\right)\\&\quad-\frac{N}{2}\left(\vec{\beta}^{(0)}-u^{(0)}W\right)^T\left(\sigma_{\beta_0}^2R_{\beta_0}\right)^{-1}\left(\vec{\beta}^{(0)}-u^{(0)}W\right)\end{aligned} \quad (14)$$

where $b_j=(\vec{y}_j\vec{\beta}^{(0)}-\vec{t}_j\odot\vec{\beta}^{(1)})$.

After certain algebraic transformations, the log-likelihood function becomes

$$\begin{aligned}L&=-\frac{N}{2}\left\{M\log\sigma_z^2+\log|(K+\mu I)|\right\}\\&\quad-\frac{1}{2\sigma_z^2}\sum_{j=1}^{N}b_j^T(K+\mu I)^{-1}b_j\\&\quad-N\bullet\left\{\log\left|\left(\sigma_{\beta_0}^2R_{\beta_0}\right)\right|^{\frac{1}{2}}+\frac{1}{2}\left(\vec{\beta}^{(0)}-u^{(0)}W\right)^T\right.\\&\qquad\left.\times\left(\sigma_{\beta_0}^2R_{\beta_0}\right)^{-1}\left(\vec{\beta}^{(0)}-u^{(0)}W\right)\right\}\\&\quad-N\bullet\left\{\log\left|\left(\sigma_{\beta_1}^2R_{\beta_1}\right)\right|^{\frac{1}{2}}+\frac{1}{2}\left(\vec{\beta}^{(1)}-u^{(1)}W\right)^T\right.\\&\qquad\left.\times\left(\sigma_{\beta_1}^2R_{\beta_1}\right)^{-1}\left(\vec{\beta}^{(1)}-u^{(1)}W\right)\right\}\end{aligned} \quad (15)$$

where $\mu=\tau^2/\sigma_z^2$, W is the design matrix. Hence, we reorganize the parameters in (10) as $\{\varphi_1,\varphi_2\}$, where $\varphi_1=\{\vec{\beta}_{(0)},\vec{\beta}_{(1)},\sigma_z^2,\theta_z,\mu\}$, $\varphi_2\{\vec{u}^{(0)},\vec{u}^{(1)},\theta_{\beta_0}^2,\sigma_{\beta_0}^2,\theta_{\beta_0}\theta_{\beta_1}\}$. The unknown parameters can be obtained by maximizing the likelihood function. As the total number to unknown parameters to be estimated is large, a simultaneous estimation of

all unknown parameter is slow. Therefore, we here design an iterative procedure for faster parameter estimation according to the following three steps.

Step 1) We first estimate φ_1 based on the assumption that the values of φ_2 are known. For the kth iteration, when $k > 1$, φ_2 can be found directly from the previous iteration, which is denoted as $\varphi_2^{(k-1)}$. When $k = 1$, the initial values of φ_2 can be randomly selected within some pre-specified intervals or chosen based on experience.

Based on the above assumption, the log-likelihood with respect to φ_1 can be reconstructed as

$$\begin{aligned} L(\varphi_1) = &-\frac{N}{2}\left\{\log\sigma_z^2 + \log|(K+\mu I)|\right\} \\ &- \frac{1}{2\sigma_z^2}\sum_{i=1}^{N}(\vec{y}_j - T_j \bullet \vec{\beta}^*)^T \\ &\times (K+\mu I)^{-1}(\vec{y}_j - T_j \bullet \vec{\beta}^*) \\ &- N \bullet \left\{\log|R_\beta|^{\frac{1}{2}} + \frac{1}{2}(\vec{\beta}^* - X^* \bullet U^*)^T \right. \\ &\left. R_\beta^{-1}(\vec{\beta}^* - X^* \bullet U^*)\right\} \end{aligned} \tag{16}$$

where

$$\begin{aligned} \vec{\beta}^* &= \begin{pmatrix} \vec{\beta}^{(0)} \\ \vec{\beta}^{(1)} \end{pmatrix} \\ T_j &= \begin{bmatrix} 1 & & 0 & t_{x_1,j} & & 0 \\ & \ddots & & & \ddots & \\ 0 & & 1 & 0 & & t_{x_M,j} \end{bmatrix} \\ R_\beta &= \begin{bmatrix} \sigma_{\beta_0}^2 R_{\beta_0} & 0 \\ 0 & \sigma_{\beta 1}^2 R_{\beta_1} \end{bmatrix} \\ U^* &= \begin{pmatrix} \vec{u}^{(0)} \\ \vec{u}^{(1)} \end{pmatrix} \\ X^* &= \begin{bmatrix} W & 0 \\ 0 & W \end{bmatrix}. \end{aligned}$$

Given $\{\theta_z, \mu\}$, the MLE of $\vec{\beta}^*$ can be derived by taking the derivative of the log-likelihood L with respect to $\vec{\beta}^*$ as

$$\begin{aligned} \frac{dL}{d\vec{\beta}^*} = &\frac{1}{\sigma_z^2}\sum_{j=1}^{N} T_j^T (K+\mu I)^{-1}(\vec{y}_j - T_j \bullet \vec{\beta}^*) \\ &- N \bullet R_\beta^{-1}(\vec{\beta}^* - X^* \bullet U^*) \end{aligned} \tag{17}$$

Setting (17) to 0, we obtain

$$\begin{aligned} \hat{\vec{\beta}}^*(\theta_z,\mu) = &\left[\frac{1}{\sigma_z^2}\sum_{j=1}^{N} T_j^T (K+\mu I)^{-1} T_j + N \bullet R_\beta^{-1}\right]^{-1} \\ &\bullet \left[\frac{1}{\sigma_z^2}\sum_{j=1}^{T} T_j^T (K+\mu I)^{-1}\vec{y}_j \right. \\ &\left. + N \bullet R_\beta^{-1} \bullet X^* \bullet U^*\right]. \end{aligned} \tag{18}$$

Equivalently

$$\begin{aligned} \hat{\vec{\beta}}^*(\theta_z,\mu) = &\Lambda \bullet \left[\sum_{j=1}^{N} T_j^T (K+\mu I)^{-1} T_j\right]^{-1} \\ &\times \left[\sum_{j=1}^{N} T_j^T (K+\mu I)^{-1}\vec{y}_j\right] \\ &+ (I-\Lambda)(X^* \bullet U^*) \\ = &\Lambda \bullet \hat{\vec{\beta}}' + (I-\Lambda)\hat{\vec{\beta}}'' \end{aligned} \tag{19}$$

where $\hat{\vec{\beta}}'$ can be considered to be the estimation of $\hat{\vec{\beta}}$ based only on (5) and $\hat{\vec{\beta}}''$ can be considered to be the estimation based on (8) and (9).Λ is the weight matrix between $\hat{\vec{\beta}}'$ and $\hat{\vec{\beta}}''$, which is

$$\Lambda = \left[\frac{1}{\sigma_z^2}\sum_{j=1}^{N} T_j^T (K+\mu I)^{-1} T_j \bullet \left(N \bullet R_\beta^{-1}\right)\right]^{-1}. \tag{20}$$

The estimation of σ_z^2 is independent of the third term in the log-likelihood function in (16), which can be neglected. The rest can be considered to be a log-likelihood function of a universal kriging with a nugget effect of μ. Hence, the MLE of the σ_z^2 is

$$\hat{\sigma}_z^2(\theta_z,\mu) = \frac{1}{MN}\sum_{j=1}^{N} b_j^T (K+\mu I)^{-1} b_j \tag{21}$$

where $b_j = (\vec{y}_j - \vec{\beta}^{(0)} - \vec{t}_j \odot \vec{\beta}^{(1)})$.

Substituting these values into the log-likelihood function in (16), we obtain the maximum of (16) over $\vec{\beta}^*$ and σ_z^2 as

$$L_1\left(\hat{\vec{\beta}}^*, \hat{\sigma}_z^2, \theta_z, \mu\right) = -\frac{N}{2}\left\{M\log\sigma_z^2 + \log|(K+\mu I)| + M\right\} \tag{22}$$

which only depends on $\{\theta_z, \mu\}$. Hence, the MLE of $\{\theta_z, \mu\}$ can be obtained by

$$\{\hat{\theta}_z, \hat{\mu}\} = \underset{\{\theta_z,\mu\}}{\arg\min}\ M\ \log\sigma_z^2 + \log|(K+\mu I)|. \tag{23}$$

Substituting the $\{\hat{\theta}_z, \hat{\mu}\}$ back into (18) and (21), we can obtain estimations of $\vec{\beta}^*$ and σ_z^2. Then, the $\varphi_1^{(k)}$ can be found for the kth iteration.

Step 2) Now, we estimate $\varphi_2^{(k)}$ based on $\varphi_1^{(k)}$. By safely ignoring the terms in (15) that are independent of φ_2, the log-likelihood function becomes

$$\begin{aligned} L_2 = &-N \bullet \left\{\log|\left(\sigma_{\beta_0}^2 R_{\beta_0}\right)|^{\frac{1}{2}} + \frac{1}{2}\left(\vec{\beta}^{(0)} - u^{(0)}W\right)^T \right. \\ &\left. \times \left(\sigma_{\beta_0}^2 R_{\beta_0}\right)^{-1}\left(\vec{\beta}^{(0)} - u^{(0)}W\right)\right\} \\ &- N \bullet \left\{\log\left|\left(\sigma_{\beta_1}^2 R_{\beta_1}\right)\right|^{\frac{1}{2}} + \frac{1}{2}\left(\vec{\beta}^{(1)} - u^{(1)}W\right)^T \right. \\ &\left. \times \left(\sigma_{\beta_1}^2 R_{\beta_1}\right)^{-1}\left(\vec{\beta}^{(1)} - u^{(1)}W\right)\right\} \end{aligned} \tag{24}$$

Similarly, the MLEs of $\vec{u}^{(0)}$ and $\vec{u}^{(1)}$ are

$$\hat{\vec{u}}^{(0)}(\theta_{\beta_0}) = \left(X^T R_{\beta_0}^{-1} X\right)^{-1} X^T R_{\beta_0}^{-1} \vec{\beta}^{(0)} \tag{25}$$

$$\hat{\vec{u}}^{(1)}(\theta_{\beta_1}) = \left(X^T R_{\beta_1}^{-1} X\right)^{-1} X^T R_{\beta_1}^{-1} \vec{\beta}^{(1)} \tag{26}$$

and the MLEs of $\sigma_{\beta_0}^2$ and $\sigma_{\beta_1}^2$ are

$$\hat{\sigma}_{\beta_0}^2(\theta_{\beta_0}) = \frac{1}{M}\left(\vec{\beta}^{(0)} - u^{(0)}W\right)^T (R_{\beta_0})^{-1}\left(\vec{\beta}^{(0)} - u^{(0)}W\right) \tag{27}$$

$$\hat{\sigma}_{\beta_1}^2(\theta_{\beta_1}) = \frac{1}{M}\left(\vec{\beta}^{(1)} - u^{(1)}W\right)^T (R_{\beta_0})^{-1}\left(\vec{\beta}^{(0)} - u^{(1)}W\right). \tag{28}$$

Substituting these values into the log-likelihood function in (24), the log-likelihood function can be divided into two parts as follows:

$$L_{20}\left(\hat{\vec{u}}^{(0)}, \hat{\sigma}_{\beta_0}^2, \theta_{\beta_0}\right) = -\frac{N}{2}\left[M \log \sigma_{\beta_0}^2(\theta_{\beta_0}) + \log|R_{\beta_0}| + M\right] \tag{29}$$

$$L_{21}\left(\hat{\vec{u}}^{(1)}, \hat{\sigma}_{\beta_1}^2, \theta_{\beta_1}\right) = -\frac{N}{2}\left[M \log \sigma_{\beta_1}^2(\theta_{\beta_1}) + \log|R_{\beta_1}| + M\right]. \tag{30}$$

The two parts depend on θ_{β_0} and θ_{β_1} respectively. Hence, the MLEs of θ_{β_0} and θ_{β_1} can be obtained by

$$\hat{\theta}_{\beta_0} = \arg\min_{\theta_{\beta_0}} M \log \sigma_{\beta_0}^2(\theta_{\beta_0}) + \log|R_{\beta_0}| \tag{31}$$

$$\hat{\theta}_{\beta_1} = \arg\min_{\theta_{\beta_1}} M \log \sigma_{\beta_1}^2(\theta_{\beta_1}) + \log|R_{\beta_1}|. \tag{32}$$

By substituting the $\{\hat{\theta}_{\beta_0}, \hat{\theta}_{\beta_1}\}$ back into (25)–(28), we can obtain an estimate of $\{\vec{u}^{(0)}, \vec{u}^{(1)}, \sigma_{\beta_0}^2, \sigma_{\beta_1}^2\}$. Hence, the $\varphi_2^{(k)}$ can be found for the kth iteration and used to estimate $\varphi_1^{(k+1)}$ in the next iteration.

Step 3) Iterate steps 1 and 2 until a stop criterion is met.

D. Prediction

After the training data in (11) have been acquired and the parameters in (10) have been estimated, we focus on the prediction of a new sample. In this subsection, two types of predictions, which can be distinguished according to whether there is anisotropy measured (or extra information) in the new sample, are taken into account. If there is no anisotropy measured in any location of the new sample, it is called an "Unobserved sample prediction" otherwise, it is called a "Partially observed sample prediction."

Each type of prediction defined above can be further divided into two subclasses. The first one represents the case where y^* is predicted at $x^* \in \{x_i, i = 1, \ldots, M\}$ given its height t^*, while the second one refers to the case where y^* is predicted at $x^* \notin \{x_i, i = 1, \ldots, M\}$given its height t^*. Intuitively, the difference between these two subclasses is that the $\vec{\beta}^{(0)}$ and $\vec{\beta}^{(1)}$ of x^* have been available for the first subclass, while they need to be predicted for the second subclass. As a result, the entire prediction part can be grouped into four scenarios that are presented in Table II. The prediction procedures for each scenario are presented in two steps and are summarized in Table III.

TABLE II
FOUR SCENARIOS FOR THE PREDICTION

	Unobserved sample	Partially observed sample
First subclass	Scenario I	Scenario III
Second subclass	Scenario II	Scenario IV

TABLE III
SUMMARY OF THE PREDICTION PROCEDURES

Type of Prediction	Scenario	Prediction Procedures
Unobserved sample	Scenario I	Step 2 (35)
	Scenario II	Step 1—Step 2 (35)
Partially observed sample	Scenario III	Step 2 (36)
	Scenario IV	Step 1—Step 2 (36)

Step 1) For the given x^*, we predict $\vec{\beta}^{(0)}$ and $\vec{\beta}^{(1)}$ as

$$\hat{\beta}^{(0)}(x^*) = \hat{u}^T w(x^*) + c_{\beta^{(0)}}^T(x^*)\left(R_{\beta_0}\right)^{-1}\left(\vec{\beta}^{(0)} - W\hat{u}^{(0)}\right) \tag{33}$$

$$\hat{\beta}^{(1)}(x^*) = \hat{u}^T w(x^*) + c_{\beta^{(1)}}^T(x^*)\left(R_{\beta_1}\right)^{-1}\left(\vec{\beta}^{(1)} - W\hat{u}^{(1)}\right) \tag{34}$$

where $c_{\beta^{(0)}}^T(x^*) = [r_{\beta^{(0)}}(\hat{\theta}_{\beta^{(0)}}, x^* - x_1), \ldots, r_{\beta^{(0)}(\hat{\theta}_{\beta^{(0)}}, x^* - x_M)}]$ and $c_{\beta^{(1)}}^T(x^*) = [r_{\beta^{(1)}}(\hat{\theta}_{\beta^{(1)}}, x^* - x_1), \ldots, r_{\beta^{(1)}(\hat{\theta}_{\beta^{(1)}}, x^* - x_M)}]$ It is noted that for the Scenario I and Scenario III predictions, Step 1 can be skipped because $\hat{\beta}^{(0)}(x^*)$ and $\hat{\beta}^{(1)}(x^*)$, which have been estimated in the estimation procedure, do not need to be predicted.

Step 2) For the "Unobserved sample prediction," the y^* of the test point (t^*, x^*) is predicted as

$$\hat{y}(t^*, x^*) = h(t^*)\hat{\beta}(x^*) \tag{35}$$

For the "Partially observed sample prediction," there is the potential to improve the prediction of y^* by taking this extra information into consideration as

$$\hat{y}(t^*, x^*) = h(t^*)\hat{\beta}(x^*) + c_z^T(x^*)(K + \mu I)^{-1}\left(\vec{y}_0 - H_j\left[\vec{\beta}^{(0)}, \vec{\beta}^{(1)}\right]\right) \tag{36}$$

where $c_z^T(x^*) = [r_z(\hat{\theta}_z, x^* - x_1), \ldots, r_z(\hat{\theta}_z, x^* - x_M)]$ $\hat{\beta}(x^*) = [\hat{\beta}^{(0)}(x^*), \hat{\beta}^{(1)}(x^*)]^T$,

$$H_j = \begin{bmatrix} 1 & t_{x_1,j} \\ 1 & t_{x_2,j} \\ \vdots & \vdots \\ 1 & t_{x_P,j} \end{bmatrix}$$

is the design matrix, and p is the number of measured locations in the new sample.

IV. Performance Study via Simulation

Here, a simulation study is used to verify the effectiveness of the estimation algorithm and the accuracy of the prediction. The simulation study can be divided into three parts, the first part is the general simulation to verify the effectiveness of the proposed method, which are followed by the uncertainty analysis and sensitivity analysis of the algorithm in the second and third part respectively.

A. General Simulation

The size of the training sample, denoted as N, is assumed to be 20. On each CNT array (each sample), there are $M = 16$ measurement locations.

For the parameter settings, the $w(x)$ are assumed to be $\{1, x_1, x_2, x_1^2, x_2^2\}$ and $\vec{u}^{(0)} = \{100, 0, 0, 10, 10\}^T$, $\vec{u}^{(1)} = \{10, 0, 0, 10, 10\}^T$, $\theta_{\beta_0} = \{0.5, 0.6\}$, $\theta_{\beta_1} = \{0.5, 0.6\}$, $\theta_z = \{0.2, 0.3\}$, $\sigma_{\beta_0}^2 = 5$, $\sigma_z^2 = 10$, $\tau^2 = 8$, which means that $\mu = 0.8$. It is noted that we assume the covariance structure is heteroscedastic. In addition, it is observed that CNTs near the center area are usually lower than that near the edge area, which forms an approximated quadratic surface with the vertex at the center of the wafer. Therefore, in this simulation study, the height function is assumed as a quadratic surface with the vertex at the center of the surface such that the interaction term is neglected in the $w(x)$.

The following Gaussian correlation function is used to characterize the correlation structure:

$$r(\theta, x_i, x_j) = \exp\left(-\theta^{(1)}(x_{i,1} - x_{j,1})^2 - \theta^{(2)}(x_{i,2} - x_{j,2})^2\right). \tag{37}$$

Based on the parameter settings, the training data are generated using the following steps:

Step 1) Generate the height value at each measurement location for the N CNT arrays as

$$t_{x_i,j} = a_0 + a_1 x_{i,1} + a_2 x_{i,2} + a_3 x_{i,1}^2 + a_4 x_{i,2}^2 + \xi_t \tag{38}$$

where, in this simulation, $a_0 = 20 + 2 \times i$, $a_1 = a_2 = 0$, $a_3 = a_4 = 1$, and $\xi_t \sim N(0, 1)$.

Step 2) Generate $\vec{\beta}^{(0)}$ and $\vec{\beta}^{(1)}$ for each measurement location according to (8) and (9) as well as the related parameter settings given above.

Step 3) Generate the anisotropy data $\{y_{x_i}, j, i = 1, \ldots, M, j = 1, \ldots, N\}$ according to (5).

Now, we have a complete training data set D as defined in (11). The simulation proceeds according to the algorithm in Section III. The initial values of $\{\sigma_{\beta_0}^2, \sigma_{\beta_0}^2, \theta_{\beta_0}, \theta_{\beta_1}\}$ are selected randomly under the constraints of $\{\sigma_{\beta_0}^2, \sigma_{\beta_0}^2\} \in (0, 100]$ and $\{\theta_{\beta_0}, \theta_{\beta_1}\} \in (0, 1)$ The initial values of $\{\vec{u}^{(0)}, \vec{u}^{(1)}\}$ are set to 80% of their true values. The changes in the log-likelihood value with increasing iterations are shown in Fig. 9. There are two types of stopping rules in the proposed algorithm. The first is that the average reduction of the likelihood function value within M steps is lower than ε, which is formulated as

$$\left| \frac{L^{(k+M)} - L^{(k)}}{M \bullet L^{(k)}} \right| \leq \varepsilon \tag{39}$$

where $L^{(k)}$ is the Log-Likelihood function value in the kth iteration, M is a pre-specified number of steps and assumed to be ten in our simulation study.ε is the pre-specified value and assumed to be 0.01 in our simulation study.

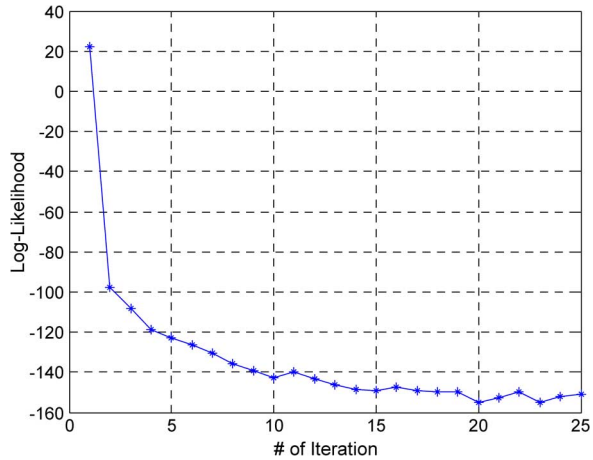


Fig. 9. Changes of the log-likelihood versus the number of iterations.

TABLE IV
Estimation Results of β_0, β_1

#	β_0		β_1	
	True	Estimated	True	Estimated
1	118.3	120.14	31.58	31.37
2	110.9	111.28	21.92	21.95
3	113.9	111.06	18.55	19.23
4	122.6	119.61	26.11	26.5
5	107.8	111.41	23.56	23.01
6	100.4	102.6	14.08	13.95
7	103.9	102.4	11.07	11.62
8	113.1	110.91	17.81	18.22
9	107.69	111.75	24.37	23.95
10	101.31	103.09	16.02	15.86
11	104.22	103.01	14.84	15.19
12	112.58	111.49	21.09	21.34
13	117.76	120.83	33.98	33.67
14	111.62	112.17	25.86	25.73
15	112.59	112.13	25.55	25.7
16	120.1	120.65	31.9	32.03

The second stopping rule is that the iteration reach a pre-specified number of steps N^I, in our simulation the N^I is assumed as 500.

The estimated parameters are given in Tables IV–VI. It can be seen that most of the parameters can be estimated accurately while the estimation errors of σ_{β_0} and σ_{β_1} are relatively larger. This happens mainly due to the inconsistence of the variance parameter estimation in a GP model [35].

The parameters estimated above are used to predict the characteristics of the new sample. The prediction comprises four parts that correspond to the four scenarios defined in Section III-D). Four new samples are generated in the same way as the 20 training samples for the four scenarios, respectively. The measurement locations in the predicted samples are identical to those

TABLE V
ESTIMATION RESULTS OF $\vec{u}$

$\vec{u}^{(0)}$	1	2	3	4	5
True value	100	0	0	10	10
Estimated	100.5	0.32	-0.15	10.05	9.71
$\vec{u}^{(1)}$	1	2	3	4	5
True value	10	0	0	10	10
Estimated	11.86	1.68	-1.49	10.38	8.64

TABLE VI
ESTIMATION RESULTS OF σ AND θ

	σ_{β_0}	σ_{β_1}	σ_z	θ_{β_0}	
True value	5	5	10	0.5	0.6
Estimated	2.0	0.11	7.28	0.69	0.59
	θ_{β_1}		θ_z		μ
True value	0.5	0.6	0.2	0.3	0.8
Estimated	0.79	0.70	0.28	0.39	0.73

TABLE VII
MSE OF THE PREDICTION RESULTS

Type of prediction	Scenario	# of Measured Points	MSE
Unobserved sample	Scenario I	0	64.98
	Scenario II	0	225.78
Partially observed sample	Scenario III	2	40.43
		4	37.45
		8	10.76
		12	0.36
	Scenario IV	2	157.65
		4	153.22
		8	111.35
		12	63.45

of the training samples for Scenario I and Scenario III, while they are different for Scenario II and Scenario IV. In addition, Scenario III and Scenario IV, which are for the "Partially observed sample prediction," are further grouped into four situations according to the number of locations that have been measured in the predicted sample, which are 2, 4, 8, and 12, respectively. It is noted that the measured locations of the situations with relatively fewer points are a subset of those in situations with more points. For each situation, the characterization of interest of every location except for the measured locations is predicted using the procedure illustrated in Table III. The mean square error (MSE) is calculated and is presented inTable VII. As can be seen, on one hand, the "Partially observed sample prediction" can provide better results than the "Unobserved sample prediction", and its accuracy can be improved by increasing the number of measured locations in the new sample. On the other hand, the predictions for Scenario I and Scenario III are more accurate than those for Scenario II and Scenario IV, respectively, which agrees with our intuitive understanding.

TABLE VIII
ESTIMATION RESULTS OF β_0, β_1

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
MSE of β_0	6.08	5.52	6.01	6.35	6.69	6.6	5.53	4.43
MSE of β_1	0.14	0.18	0.15	0.11	0.2	0.3	0.19	0.11
	9	**10**	**11**	**12**	**13**	**14**	**15**	**16**
MSE of β_0	6.47	6.39	4.67	3.15	4.99	4.2	3.36	2.58
MSE of β_1	0.2	0.27	0.18	0.1	0.11	0.12	0.12	0.07

TABLE IX
ESTIMATION RESULTS OF $\vec{u}$

$\vec{u}^{(0)}$	**1**	**2**	**3**	**4**	**5**
TRUE	100	0	0	10	10
MEAN	99.88	-0.04	-0.01	10	10.03
MSE	0.48	0.06	0.05	0.05	0.09
Std	0.69	0.23	0.22	0.22	0.3
$\vec{u}^{(1)}$	**1**	**2**	**3**	**4**	**5**
TRUE	10	0	0	10	10
MEAN	10.22	-0.24	0.13	9.89	9.76
MSE	2.52	0.87	1.1	0.55	1.03
Std	1.59	0.91	1.05	0.74	1

B. *Uncertainty Analysis*

Here, we study the uncertainty of the estimates via more extensive simulations; the parameters are set exactly the same as that in Section IV-A, and the training samples are also generated according to the procedure described in Section IV-A. The estimation algorithm is repeated 50 times with different training samples such that the uncertainty of the estimates can be calculated; the corresponding results are presented in Tables VIII–X. Since the training samples used for each run are different, the true values of β_0 and β_1 are also different. Therefore, only the MSEs of estimated β_0 andβ_1 are calculated and listed in Table VIII.

From the simulation results, it is can be seen that most of the parameters can be estimated accurately except the σ_{β_0} and σ_{β_1}. The inaccurate estimation of σ_{β_0} and σ_{β_1}, which is caused by the limited sample points, has little influence on the prediction in practice.

C. *Sensitivity Analysis*

Here, we analyze the sensitivity of the initial value and sample size respectively. For the first part, the initial value of $\{\vec{u}^{(0)}\vec{u}^{(1)}\}$ are given a bias, which are from 20% to 180%, and the simulation results are shown in Tables XI and XII. In order to make the results of different parameters comparable, we used the relative estimation error as the criterion which is defined as

$$\left|\frac{\hat{\varphi}-\varphi_{\text{true}}}{\varphi_{\text{true}}}\right|. \tag{40}$$

TABLE X
ESTIMATION RESULT OF θ AND σ

	θ_{β_0}		θ_{β_1}		θ_z	
TRUE	0.5	0.6	0.5	0.6	0.2	0.3
Mean	0.51	0.67	0.51	0.66	0.20	0.33
MSE	0.04	0.04	0.03	0.05	0.01	0.02
Std	0.19	0.19	0.16	0.21	0.10	0.14
	μ	σ_z	σ_{β_0}	σ_{β_1}		
TRUE	0.8	10	5	5		
Mean	0.75	10.03	0.15	3.02		
MSE	0.03	4.49	23.55	10.33		
Std	0.17	2.14	0.21	2.55		

TABLE XI
SENSITIVE ANALYSIS OF THE INITIAL VALUE (PART ONE, FOR σ, β AND θ)

	0.2	**0.4**	**0.6**	**0.8**	**1.0**	**1.2**	**1.4**	**1.6**	**1.8**
σ	0.74	0.80	0.57	0.58	0.64	0.60	0.86	0.86	0.86
β	0.59	0.42	0.25	0.08	0.01	0.06	0.24	0.42	0.57
θ	1.94	1.22	0.51	0.11	0.16	0.20	0.60	1.04	1.89

TABLE XII
SENSITIVE ANALYSIS OF THE INITIAL VALUE (PART TWO, FOR u)

u	**0.2**	**0.4**	**0.6**	**0.8**	**1.0**	**1.2**	**1.4**	**1.6**	**1.8**
1	12.64	8.73	4.83	0.94	0.05	0.87	3.81	7.73	10.67
2	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.02	0.03
3	0.03	0.02	0.02	0.02	0.02	0.02	0.03	0.03	0.03
4	0.70	0.52	0.35	0.15	0.02	0.18	0.36	0.54	0.71
5	0.72	0.54	0.37	0.18	0.01	0.17	0.35	0.53	0.70
6	1.46	1.11	0.76	0.42	0.09	0.26	0.62	0.97	1.31
7	0.14	0.14	0.15	0.15	0.15	0.16	0.13	0.13	0.13
8	0.20	0.21	0.21	0.21	0.22	0.22	0.20	0.20	0.20
9	0.12	0.10	0.08	0.06	0.04	0.01	0.00	0.03	0.05
10	0.18	0.16	0.14	0.12	0.10	0.08	0.06	0.04	0.02

It is noted that some of true value of u are 0, hence, (40) cannot be used. In this case, we used $|\hat{\varphi} - \varphi_{true}|$ as the criterion for u. As can be seen from Tables XI and XII, when the bias is between 80% and 120%, the estimation result is relatively acceptable (the estimation error are within 20%). The σ are still difficult to estimate very accurately as mentioned before.

The prediction results for different initial values are shown in Table XIII, in which the MSE are used as criterion. From the result in Table XIII, the initial value has great influence on the prediction accuracy. However, the algorithm can deliver acceptable when the bias locates between 80% and 120%, which is similar with results of parameter estimation.

For the second part, we analyzed the sensitivity of the sample size from two factors, the number of the CNTs arrays and the number of points in each CNTs array. The number of the CNTs arrays are categorized into five levels (10, 20, 30, 40, and 50), while the number of points has three levels (9, 16. and 25). We

TABLE XIII
PREDICTION RESULTS FOR DIFFERENT INITIAL VALUE (CRITERION: MSE)

MSE	**0.20**	**0.40**	**0.60**	**0.80**	**1.00**
Scenario I	798.63	290.70	353.06	97.54	57.42
Scenario II	1924.58	480.49	483.12	238.61	195.72
Scenario III	107.95	65.12	74.65	49.42	25.06
Scenario IV	315.06	221.41	77.26	58.56	47.89
	1.20	**1.40**	**1.60**	**1.80**	
Scenario I	253.76	402.22	356.45	795.80	
Scenario II	513.48	769.51	763.97	1500.01	
Scenario III	45.29	68.89	57.76	68.56	
Scenario IV	55.06	70.22	106.09	152.52	

TABLE XIV
FULL FACTORIAL DESIGN FOR THE SENSITIVITY ANALYSIS OF SAMPLE SIZE

Order	**# of CNTs arrays**	**# of Points**	**Order**	**# of CNTs arrays**	**# of Points**
1	10	9	**9**	30	25
2	10	16	**10**	40	9
3	10	25	**11**	40	16
4	20	9	**12**	40	25
5	20	16	**13**	50	9
6	20	25	**14**	50	16
7	30	9	**15**	50	25
8	30	16			

TABLE XV
SENSITIVE ANALYSIS OF THE SAMPLE SIZE (PART ONE, FOR σ, β AND θ)

# of CNTs arrays		**10**	**20**	**30**	**40**	**50**
# of Points						
	σ	0.70	0.84	0.61	0.66	0.62
9	β	0.03	0.01	0.02	0.01	0.02
	θ	1.72	1.38	1.12	0.99	0.91
	σ	0.64	0.63	0.56	0.55	0.62
16	β	0.02	0.02	0.01	0.01	0.02
	θ	1.19	0.35	0.28	0.33	0.32
	σ	0.78	0.62	0.68	0.53	0.50
25	β	0.02	0.04	0.02	0.01	0.01
	θ	1.09	0.37	0.27	0.32	0.30

investigated the effect of these two factors by full factorial design shown in Table XIV.

The results of parameter estimation are shown in Tables XV and XVI (the order is corresponding to the combination of factors shown in Table XIV), the criterion is the same as the sensitivity analysis for initial values. It can be seen that the accuracy of the parameter estimation increases with the increase of the number of CNTs array, however, the accuracy changes slightly when the number of CNTs array is larger than 20. Compared with the number of CNTs array, the number of points has less influence on the accuracy of parameter estimation.

TABLE XVI
SENSITIVE ANALYSIS OF THE SAMPLE SIZE (PART TWO, FOR $\boldsymbol{u}$)

Order / u	1	2	3	4	5	6	7	8
1	1.96	2.35	1.14	1.18	1.88	1.36	1.37	1.29
2	0.07	0.05	0.03	0.05	0.05	0.03	0.05	0.07
3	0.09	0.05	0.05	0.06	0.05	0.06	0.04	0.02
4	0.19	0.17	0.12	0.15	0.15	0.12	0.16	0.16
5	0.17	0.18	0.12	0.16	0.17	0.13	0.18	0.13
6	0.47	0.31	0.33	0.61	0.50	0.33	0.47	0.41
7	0.28	0.27	0.21	0.29	0.29	0.16	0.30	0.24
8	0.30	0.27	0.37	0.28	0.27	0.25	0.27	0.24
9	0.28	0.25	0.10	0.25	0.23	0.07	0.20	0.30
10	0.31	0.17	0.16	0.30	0.24	0.12	0.25	0.26

Order / u	9	10	11	12	13	14	15
1	0.60	1.37	1.38	1.38	0.92	0.90	0.83
2	0.04	0.04	0.04	0.04	0.01	0.01	0.01
3	0.06	0.05	0.04	0.04	0.02	0.01	0.02
4	0.13	0.17	0.16	0.16	0.13	0.11	0.11
5	0.12	0.17	0.16	0.16	0.12	0.12	0.11
6	0.40	0.52	0.56	0.56	0.14	0.11	0.13
7	0.31	0.26	0.23	0.23	0.10	0.04	0.05
8	0.41	0.25	0.24	0.24	0.12	0.07	0.13
9	0.08	0.14	0.11	0.11	0.09	0.07	0.05
10	0.09	0.27	0.16	0.16	0.08	0.06	0.05

TABLE XVII
PREDICTION RESULTS FOR DIFFERENT SAMPLE SIZE (CRITERION: MSE)

# of Points / # of CNTs arrays		10	20	30	40	50
9	Scenario I	274.73	290.85	219.73	191.59	125.25
	Scenario II	311.82	293.12	345.34	297.64	240.63
	Scenario III	111.01	129.54	146.73	126.54	118.54
	Scenario IV	159.45	158.01	191.41	171.78	140.14
16	Scenario I	161.26	178.86	141.52	193.32	135.54
	Scenario II	200.08	248.86	127.96	107.96	78.41
	Scenario III	84.52	100.67	112.50	119.00	93.78
	Scenario IV	80.72	65.89	59.53	66.50	48.77
25	Scenario I	76.76	70.41	62.66	63.89	57.14
	Scenario II	95.29	97.48	97.45	69.94	78.63
	Scenario III	36.15	27.84	22.17	26.19	22.71
	Scenario IV	63.12	54.49	36.80	32.90	31.14

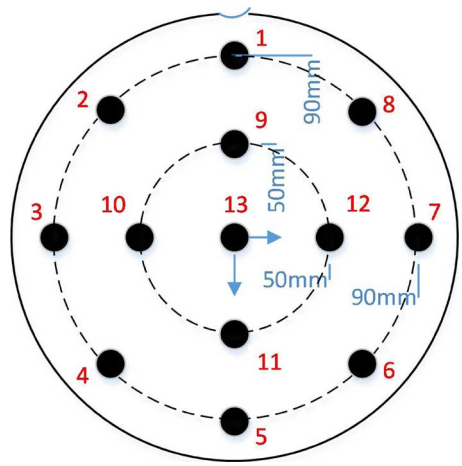


Fig. 10. Diagram of the 13 measurement locations.

The prediction results for different sample size are shown in Table XVII, in which the MSE are used as criterion. Clearly, both factors of sample size has great influence on the prediction accuracy, the prediction accuracy increases with the rise of two factors as expected.

V. APPLICATION TO A REAL EXAMPLE

In this section, we use the real data collected from 17 CNT arrays to evaluate the performance of the proposed method. On each CNT array, the anisotropy and the height are measured at 13 pre-specified locations, as shown in Fig. 10. The contours of the height and anisotropy of the CNT array are presented in Figs. 11 and 12, respectively. As can be seen, the height near the edge is usually higher than the height near the center, and the anisotropy surface basically resembles the corresponding height surface.

To determine the function form of the $w(x)$ in (8) and (9), a linear regression function is fitted using data from each measurement location. Then, we obtain the coefficients of the linear models representing the relationship between the height and the anisotropy at different locations, which are used to estimate the coefficients at other locations by interpolating. The contours of the intercept term (see Fig. 13) and the slope term (see Fig. 14) show that a quadratic function seems appropriate for describing the mean parts of (8) and (9). In addition, the result of the linear regression is used as the initial value of the proposed method in this case study.

The first part of this case study mainly focuses on verification of the proposed model using real data. We randomly select one sample from the 17 CNT arrays to be a testing sample and use the rest as training samples. The situations are still categorized according to the number of measured locations in the predicted sample, which are 0, 2, 4, 8, and 12. It is noted that "0" corresponds to the "Unobserved sample prediction" and the others constitute the "Partially observed sample prediction" in this case study. The prediction results are presented in Fig. 15 for Scenario I and Scenario II and in Fig. 16 for Scenario III and Scenario IV. As can be seen, although the situation with more measured locations can provide better results, the predicted values for the "Unobserved sample" (Scenario I and Scenario III) are already very close to the true value, which verifies the effectiveness of the proposed method for two types of prediction. In addition to the graphical illustration, we calculate the MSE of the predicted result, which decreases from 60.2 to 3.3 for the Scenario II prediction and from 102.3 to 5.4 for the Scenario IV prediction as the number of measured locations increases from 2 to 12.

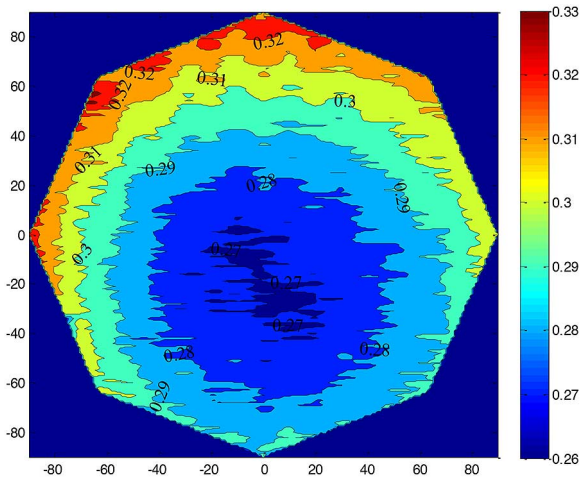


Fig. 11. Contour of the surface height for the CNT array, taking the first CNT array as an example.

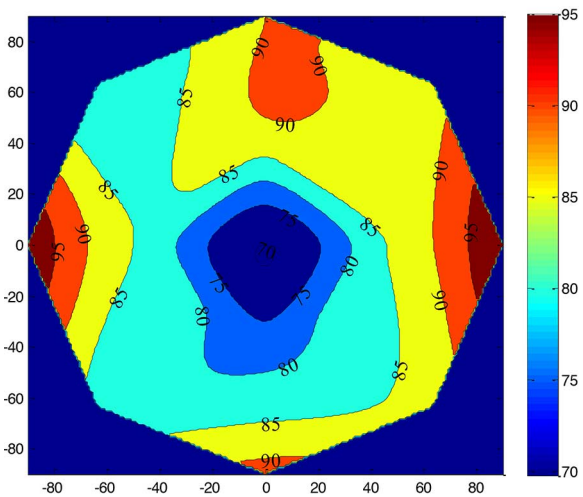


Fig. 12. Contour of the anisotropy surface for the CNT array, taking the first CNT array as an example.

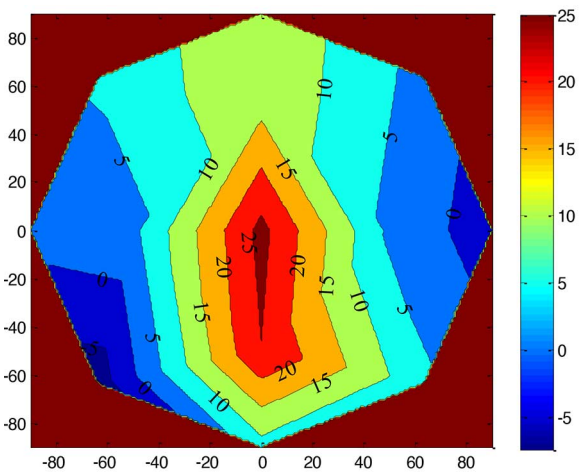


Fig. 13. Contour of the intercept term.

The second part of this section aims to illustrate the superiority of the proposed method for the application of nanomaterial surface characterization. For this purpose, two conventional

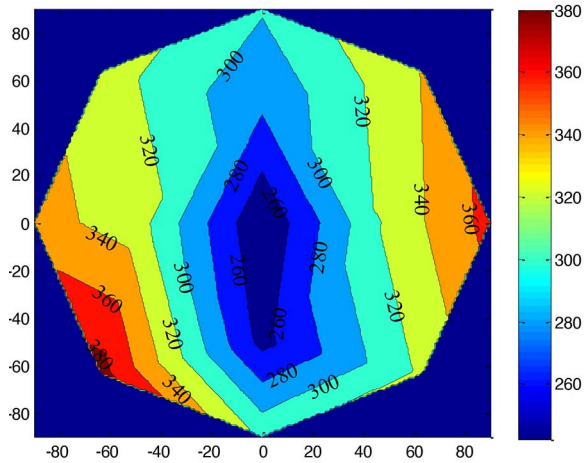


Fig. 14. Contour of the slope term.

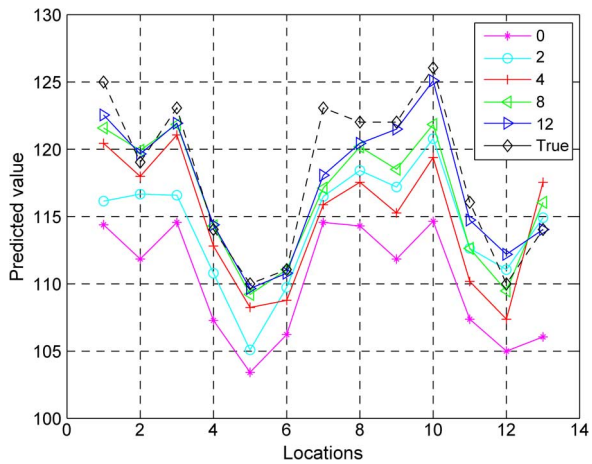


Fig. 15. Scenarios I and II prediction results, where "0" corresponds to the Scenario I prediction and the others correspond to the Scenario II prediction.

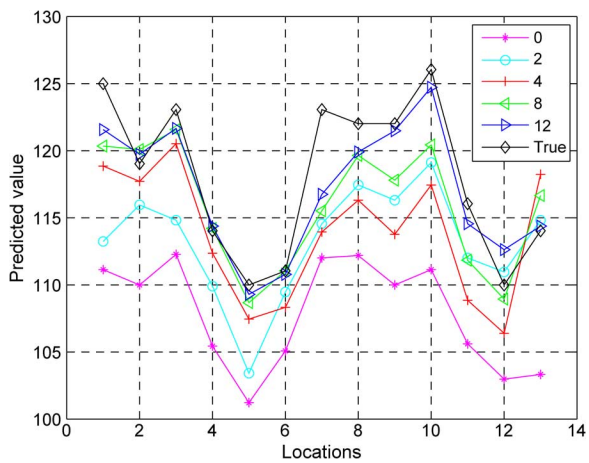


Fig. 16. Scenarios III and IV prediction results, where "0" corresponds to the Scenario III prediction and the others correspond to the Scenario IV prediction.

methods are chosen as the benchmarks to compare with the proposed method. The first method involves applying the linear regression directly for each measurement location as

$$y(t, x_i) = h(t)\beta_i + \varepsilon_i, \qquad i = 1, \ldots, M \tag{41}$$

where the coefficients are estimated as

$$\hat{\beta}_i = \left(F_i^T F_i\right)^{-1} F_i^T \vec{y}_i, \qquad i = 1, \ldots, M \tag{42}$$

where F_i is the design matrix for the ith measurement location, and the anisotropy at x_i in the new CNT array can be predicted as

$$\hat{y}_0(x_i, t^*) = \hat{\vec{\beta}}_i \bullet [1 \quad t^*]. \tag{43}$$

The second benchmark is the extension of the bivariate spatial model proposed by Zhang *et al.* [34] to multiple replicates, of which the model should be

$$y(t, x) = h(t)\beta + Z(x) + \varepsilon. \tag{44}$$

Its log-likelihood is

$$\begin{aligned} L &= \log\left\{\prod_{j=1}^{N} f(\vec{y}_j)\right\} \\ &= -\frac{N}{2}\left\{M \log \sigma_z^2 + \log\left|(K + \mu I)\right|\right\} \\ &\quad - \frac{1}{2\sigma_z^2}\sum_{j=1}^{N} b_j^T (K + \mu I)^{-1} b_j \end{aligned} \tag{45}$$

where $b_j = (\vec{y}_j - \vec{\beta}^{(0)} - \vec{t}_j \odot \vec{\beta}^{(1)})$.

By maximizing the log-likelihood function, the MLE of parameters can be easily obtained, and the anisotropy can be predicted as

$$\hat{y}(t^*, x^*) = h(t^*)\hat{\beta} + c_z^T(x^*)(K + \mu I)^{-1}(\vec{y}_0 - H_j\hat{\beta}). \tag{46}$$

It is unrealistic to measure the anisotropy at more than two locations for each CNT array in real manufacturing, and, as a result, only two locations are selected to be measured for anisotropy in most of the produced CNT arrays. Therefore, we compared the proposed method with two traditional methods under the "Partially observed sample prediction" with two measured locations, which correspond with Scenarios III and IV (defined in Section III-D). For the 17 CNT arrays, one sample at a time is selected as the test sample, and the rest are training samples. Then, three methods are applied such that the predicted value is obtained for each location for each of the different methods. We iterate this process until every sample has been used as the predicted sample. After that, the MSE of prediction for each sample is calculated as shown in Table XVIII. It is noted that, for the Scenario IV prediction, the predicted location should be deleted from the training sample. Therefore, the method based on linear regression cannot be used for this scenario, and only two methods are applied for the Scenario IV prediction.

From the results in Table XVIII, the proposed method can deliver more accurate predictions than the other two traditional methods for both scenarios. The reason for this phenomenon can be explained as follows. The LR method assumes that the anisotropy at different locations is independent, while the BSM method neglects the relationship differences between two characteristics at each location. However, the proposed method not only considers the spatial correlation of the characteristics but also uses a Gaussian process to model the coefficients; such a representation presents the relationship between two characteristics at each location and forms a hierarchical model that has the potential to deliver better performance.

TABLE XVIII
COMPARISON OF THE RESULTS OF THREE METHODS USING AN MSE CRITERION

Samples	Scenario III			Scenario IV	
	HSM	LR	BSM	HSM	BSM
1	47.15	85.74	77.83	47.37	74.3
2	18.78	16.66	24.37	18.81	26.02
3	7.42	20.53	42.67	7.4	52.38
4	28.51	31.37	42.67	28.63	40.21
5	33.41	71.93	82.33	32.53	91.95
6	13.05	30.98	21.25	12.23	19.62
7	28.76	20.79	9.9	28.01	9.94
8	30.83	39.68	31.16	30.24	31.55
9	17.08	19.97	16.4	16.81	16.18
10	136.15	61.67	406.5	144.76	397.17
11	193.38	235.33	106.57	194.17	105.86
12	60.01	29.86	33.15	64.55	39.02
13	25.17	29.03	12.85	30.27	14.03
14	19.17	43.77	10.79	22.25	11.97
15	53.26	427.84	9.45	58.55	11.56
16	342.25	110.43	621.98	357.02	587.5
17	285.7	172.65	898.02	299.1	857.91
Average	**78.82**	**85.19**	**143.99**	**81.92**	**140.42**

Note: "HSM" Refers to the Proposed Hierarchical Spatial Model, "LR" Refers to the Method Based on Linear Regression, and "BSM" Refers to Simple Extension of the Bivariate Spatial Model

Practitioners should be fully aware that, compared with LR and BSM, HSM estimates parameter iteratively, and thus is more computational demanding. In most of our simulated scenarios, the HSM method can finish estimation within several minutes on a normal personal computer, which guarantees the practical usefulness of the proposed method.

VI. CONCLUSION

As one of the critical quality indices of CNT film, anisotropy has a great influence on the performance of devices made from CNT film. However, the anisotropy cannot be measured accurately at a reasonable cost because the measurement process is destructive and extremely sensitive to environmental disturbance. At the same time, the anisotropy has been found to be closely correlated with the height of the CNT array, which is another spatial quality characteristic that can be obtained accurately and inexpensively. This research proposes a hierarchical spatial model consisting of two levels. The first level mainly focuses on capturing the general trend and the spatial property of the anisotropy, while the second level is used to characterize the spatial property of the relationship between the two quality indices. After that, an iterative MLE-based algorithm is developed to estimate the related parameters. A prediction procedure is also derived for different prediction scenarios, and its effectiveness is verified by a simulation study. Finally, a case study is utilized to evaluate the performance of the proposed method,

which illustrates that the proposed method has the ability to deliver better results for the prediction of anisotropy compared with two conventional methods.

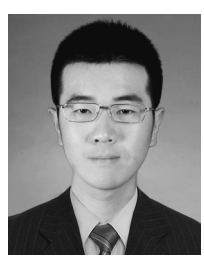
Considering the many challenges facing nanotechnology research and complexity of different novel processes, the proposed method can be extended to other situations with similar metrology needs. Although the change of real applications may lead to a change of the model format and potentially incurs new technical difficulties, we believe the adoption of statistical methods in scaling-up nanotechnology research and production is a promising area for future research.

ACKNOWLEDGMENT

The authors would like to thank the associate editor and two anonymous referees for their valuable comments, which have helped improve this work greatly.

REFERENCES

[1] M. Kumar and Y. Ando, "Chemical vapor deposition of carbon nanotubes: A review on growth mechanism and mass production," *J. Nanosci. Nanotechnol.*, vol. 10, pp. 3739–3758, 2010.
[2] K. L. Jiang, Q. Q. Li, and S. S. Fan, "Nanotechnology: Spinning continuous carbon nanotube yarns—Carbon nanotubes weave their way into a range of imaginative macroscopic applications," *Nature*, vol. 419, pp. 801–801, 2002.
[3] C. Feng, K. Liu, J.-S. Wu, L. Liu, J.-S. Cheng, and Y. Zhang *et al.*, "Flexible, stretchable, transparent conducting films made from superaligned carbon nanotubes," *Adv. Functional Mater.*, vol. 20, pp. 885–891, 2010.
[4] K. Jiang, J. Wang, Q. Li, L. Liu, C. Liu, and S. Fan, "Superaligned carbon nanotube arrays, films, yarns: A road to applications," *Adv. Mater.*, vol. 23, pp. 1154–1161, 2011.
[5] X. Wang, S. Wu, and K. Wang, "A run-to-run profile control algorithm for improving the flatness of nano-scale products," *IEEE Trans. Autom. Sci. Eng.* vol. 12, no. 1, pp. 192–203, Feb. 2015.
[6] Q. Huang, L. Wang, T. Dasgupta, L. Zhu, P. K. Sekhar, and S. Bhansali *et al.*, "Statistical weight kinetics modeling and estimation for silica nanowire growth catalyzed by pd thin film," *IEEE Trans. Autom. Sci. Eng.*, vol. 8, no. 2, pp. 303–310, Apr. 2011.
[7] Q. Huang, "Physics-driven Bayesian hierarchical modeling of the nanowire growth process at each scale," *IIE Trans.*, vol. 43, pp. 1–11, 2011.
[8] T. Dasgupta, C. Ma, V. R. Joseph, Z. L. Wang, and C. F. J. Wu, "Statistical modeling and analysis for robust synthesis of nanostructures," *J. Amer. Stat. Assoc.*, vol. 103, pp. 594–603, 2008.
[9] L. J. H. Xu and Q. , "EM estimation of nanostructure interactions with incomplete feature measurement and its tailored space filling designs," *IEEE Trans. Autom. Sci. Eng.*, vol. 10, no. 3, pp. 579–587, Jun. 2013.
[10] L. J. Xu and Q. Huang, "Modeling the interactions among neighboring nanostructures for local feature characterization and defect detection," *IEEE Trans. Autom. Sci. Eng.*, vol. 9, no. 5, pp. 745–754, Oct. 2012.
[11] F. Wang, Y. Hwang, P. Z. G. Qian, and X. Wang, "A statistics-guided approach to precise characterization of nanowire morphology," *ACS Nano*, vol. 4, pp. 855–862, 2010.
[12] W. Mai and X. Deng, "The applications of statistical quantification techniques in nanomechanics and nanoelectronics," *Nanotechnology*, vol. 21, 2010.
[13] X. Deng, V. R. Joseph, W. Mai, Z. L. Wang, and C. F. J. Wu, "Statistical approach to quantifying the elastic deformation of nanomaterials," *Proc. Nat. Acad. Sci. USA*, vol. 106, pp. 11845–11850, 2009.
[14] H. Scheffe, "Statistical-theory of calibration," *Ann. Statistics*, vol. 1, pp. 1–37, 1973.
[15] C. Eisenhart, "the interpretation of certain regression methods and their use in biological and industrial research," *Ann. Math. Statist*, vol. 10, p. 162, 1939.
[16] V. Centner, D. L. Massart, and S. de Jong, "Inverse calibration predicts better than classical calibration," *Fresenius J. Anal. Chemistry*, vol. 361, pp. 2–9, 1998.
[17] K. R. Eberhardt and R. W. Mee, "Constant-width calibration intervals for linear-regression," *J. Quality Technol.*, vol. 26, pp. 21–29, 1994.
[18] R. G. Krutchko, "Classical and inverse regression methods of calibration in extrapolation," *Technometrics*, vol. 11, p. 605-&, 1969.
[19] R. W. Mee, K. R. Eberhardt, and C. P. Reeve, "Calibration and simultaneous tolerance intervals for regression," *Technometrics*, vol. 33, pp. 211–219, 1991.
[20] J. Tellinghuisen, "Inverse vs. classical calibration for small data sets," *Fresenius J. Anal. Chemistry*, vol. 368, pp. 585–588, 2000.
[21] P. A. Parker, G. G. Vining, S. R. Wilson, J. L. Szarka, and N. G. Johnson, "The prediction properties of classical and inverse regression for the simple linear calibration problem," *J. Quality Technol.*, vol. 42, pp. 332–347, 2010.
[22] C. N, *Statistics for Spatial Data*. New York, NY, USA: Wiley, 1993.
[23] T. J. Santer, B. J. Williams, and W. I. Notz, *The Design and Analysis of Computer Experiments*. New York, NY, USA: Springer, 2003.
[24] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT, 2006.
[25] V. R. Joseph, Y. Hung, and A. Sudjianto, "Blind kriging: A new method for developing metamodels," *J. Mech. Design*, vol. 130, p. 031102, Mar. 2008.
[26] S. Ba and V. R. Joseph, "Composite Gaussian process models for emulating expensive functions," *Ann. Appl. Statistics*, vol. 6, pp. 1838–1860, Dec. 2012.
[27] Y. Hung, V. R. Joseph, and M. S. N. , "Analysis of computer experiements with functional response," *Technometrics*, vol. 57, pp. 35–44, 2015.
[28] M. J. Bayarri, J. O. Berger, R. Paulo, J. Sacks, J. A. Cafeo, and J. Cavendish *et al.*, "A framework for validation of computer models," *Technometrics*, vol. 49, pp. 138–154, May 2007.
[29] K. T. Fang, R. Li, and A. Sudjianto, *Design and Modeling for Computer Experiments*. Boca Raton, FL, USA: CRC, 2006.
[30] J. O. Ramsay and B. W. Silverman, *Functional Data Analysis*, 2nd ed. New York, NY, USA: Springer, 2005.
[31] J. Q. Shi and B. Wang, "Curve prediction and clustering with mixtures of Gaussian process functional regression models," *Statistics and Computing*, vol. 18, pp. 267–283, Sep. 2008.
[32] J. Q. Shi, B. Wang, R. Murray-Smith, and D. M. Titterington, "Gaussian process functional regression Modeling for batch data," *Biometrics*, vol. 63, pp. 714–723, Sep. 2007.
[33] J. Q. Shi, B. Wang, E. J. Will, and R. M. West, "Mixed-effects Gaussian process functional regression models with application to dose-response curve prediction," *Statistics in Medicine*, vol. 31, pp. 3165–3177, Nov. 20, 2012.
[34] Q. Zhang, X. Deng, P. Z. G. Qian, and X. Wang, "Spatial modeling for refining and predicting surface potential mapping with enhanced resolution," *Nanoscale*, vol. 5, pp. 921–926, 2013.
[35] H. Zhang, "Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics," *J. Amer. Stat. Assoc.*, vol. 99, pp. 250–261, 2004.



Xin Wang received the B.S. degree from Xi Dian University, Xi'an, China, in 2009. He is currently working toward the Ph.D. degree at Tsinghua University, Beijing, China.

His general research interests include statistical modeling, process control and experiment design for nano manufacturing.

Mr. Wang is a member of the Institute for Operations Research and the Management Sciences (INFORMS).

Su Wu received the Ph.D. degree in mechanical engineering from Tsinghua University, Beijing, China, in 1994.

He is currently a Professor with the Department of Industrial Engineering, Tsinghua University, Beijing, China. He is a member of the Editorial Board for *International Journal of Production Economics*. His research focus on manufacturing and service quality control, manufacturing technology, reliability, and maintenance.

Prof. Wu is an associate member of China High Technology Industrialization.

Kaibo Wang received the B.S. and M.S. degrees in mechatronics from Xi'an Jiaotong University, Xi'an, China, and the Ph.D. degree in industrial engineering and engineering management from the Hong Kong University of Science and Technology, Hong Kong.

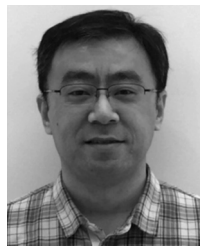
He is an Associate Professor with the Department of Industrial Engineering, Tsinghua University, Beijing, China. He has published papers in journals such as *Journal of Quality Technology*, *IIE Transactions*, *Quality and Reliability Engineering International*, *International Journal of Production Research*, and others. His 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.

Dr. Wang is a member of INFORMS, IIE, and a senior member of ASQ.

Xinwei Deng received the B.S. degree in mathematics from Nanjing University, Nanjing, China, in 2003, and the Ph.D. degree in industrial and systems engineering from Georgia Institute of Technology, Atlanta, GA, USA, in 2009.

From 2009 to 2011, he was a Visiting Assistant Professor with the Department of Statistics, University of Wisconsin-Madison. Since 2011, he has been an Assistant Professor with the Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA. His research interest includes interface between design of experiments and machine learning, model and analysis of high-dimensional data, and statistical approaches to nanotechnology.

Mr. Deng is a member of the American Statistical Association, the Institute for Operations Research, and the Management Sciences (INFORMS).



Liang Liu received the B.S. degree in physics, M.S. degree in acoustics, and Ph.D. degree in condensed physics from Tsinghua University, Beijing, China, in 1996, 1999, and 2005, respectively.

From 2005 to 2009, he was a Postdoctoral Research Assistant with the Department of Chemistry, Tsinghua University, Beijing, China. From 2009 to 2012, he was the General Manager of the Beijing Funate Innovation Technology Co. Since 2012, he has been the R&D Director of the Tianjin Funa Yuanchuang technology. Co. Ltd., Tianjin, China. He is the author of more than 40 articles and the inventor of more than 90 patents. His research interests and working focus include mass production and quality control of the super-aligned carbon nanotube arrays, films and yarns, facility and automation development for carbon nanotube production quality inspection, and applications of carbon nanotubes in touch panels, batteries, field emission devices, and other area.

Qi Cai received the B.S., M.S., and Ph.D. degrees in materials science and engineering from Anhui University, Hefei, Anhui, China, in 2001, 2004, and 2007, respectively.

From 2008 to 2011, he was a Research Assistant with the Department of Physics, Tsinghua University, Beijing, China. Since 2012, he has been a Senior R&D Engineer with the R&D Department, Tianjin Funayuanchuang Technology Co., Ltd., Tianjin, China. His research interests include carbon nanotube paste and large scale production of super-aligned arrays of carbon nanotubes.