

Gaussian process modeling for engineered surfaces with applications to Si wafer production

Matthew Plumlee^{a,*}, Ran Jin^b, V. Roshan Joseph^a and Jianjun Shi^a

Received 18 June 2013; Accepted 9 July 2013

When producing engineered surfaces, the stochastic portion of the processing greatly affects the overall output quality. We propose a Gaussian process model that accounts for the impact of control variables on the stochastic elements of the produced surfaces. An optimization algorithm is outlined to find the maximum likelihood estimates of the model parameters. A case study involving the thickness surfaces of semiconductor wafers is examined that demonstrates the need for the proposed approach. Copyright © 2013 John Wiley & Sons, Ltd.

Keywords: designed experiment; functional response; quality control; statistical process control

1 Introduction

The material surface is of critical importance because of its impact on the downstream properties of products such as friction and optical behavior. We study the case where we receive measurements from surfaces created via controlled production or “engineered”. Data from engineered surfaces have become pervasive in recent years because of developments in measurement technology, information storage capacity, and processing power. A large amount of statistical research has studied monitoring processes that produce surfaces (typically termed profiles). Monitoring is usually performed by explicitly modeling a given surface (e.g., Qiu et al. (2010)). Our paper presents a different direction. Our goal is not to explicitly model a given surface but instead to model the *stochastic distribution* of produced surfaces. We study this distribution as the process inputs are altered. This problem is similar to a regression problem with functional responses (Faraway, 1997). The core difference between this work and the works found in the current literature is the focus on the *stochastic* portion of the responses.

We study the scenario where a user has the capability to alter certain variables, which we will denote *control variables* or \mathbf{u} . Surface measurements (Y) are collected corresponding to a set of explanatory variables (\mathbf{x}) that represent the spatial location of the measurement. This paper addresses the challenge of incorporating control variables into the analysis of systems with stochastic surface responses. As an example, Section 4 studies wire-sawn semiconductor wafers exhibiting undesirable waviness. Lapping, a process in which a liquid slurry containing abrasive particles is pressed onto a rotating surface, is used to reduce waviness. Here, control variables include the downward pressure and the rotation speed of the lapping pad. Table I outlines the response of a lapping process to a 2^{4-1} fractional factorial design of experiments with one center run; see Wu & Hamada (2009) for more information on designed experiments.

^aH. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30318, USA

^bGrado Department of Industrial and Systems Engineering, Virginia Institute of Technology, Blacksburg, VA 24061, USA

*Email: mplumlee@gatech.edu

Table I. Designed experiment for a lapping process.								
Run	Control variables \mathbf{u}				Profile data			
	ρ_{low}	RPM	t_{low}	t_{high}				
1	+	–	–	–	$\mathbf{Y}_{1,1}$	$\mathbf{Y}_{1,2}$...	\mathbf{Y}_{1,d_1}
2	+	+	–	+	$\mathbf{Y}_{2,1}$	$\mathbf{Y}_{2,2}$...	\mathbf{Y}_{2,d_2}
3	–	–	–	+	$\mathbf{Y}_{3,1}$	$\mathbf{Y}_{3,2}$...	\mathbf{Y}_{3,d_3}
4	–	–	+	–	$\mathbf{Y}_{4,1}$	$\mathbf{Y}_{4,2}$...	\mathbf{Y}_{4,d_4}
5	–	+	–	–	$\mathbf{Y}_{5,1}$	$\mathbf{Y}_{5,2}$...	\mathbf{Y}_{5,d_5}
6	+	–	+	+	$\mathbf{Y}_{6,1}$	$\mathbf{Y}_{6,2}$...	\mathbf{Y}_{6,d_6}
7	+	+	+	–	$\mathbf{Y}_{7,1}$	$\mathbf{Y}_{7,2}$...	\mathbf{Y}_{7,d_7}
8	–	+	+	+	$\mathbf{Y}_{8,1}$	$\mathbf{Y}_{8,2}$...	\mathbf{Y}_{8,d_8}
9	0	0	0	0	$\mathbf{Y}_{9,1}$	$\mathbf{Y}_{9,2}$...	\mathbf{Y}_{9,d_9}

The data, $\mathbf{Y}_{ij}^T = [Y_{ij}(\mathbf{x}_1), \dots, Y_{ij}(\mathbf{x}_n)]$, consist of over 150 thickness measurements on a regular grid $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ on wafer j produced at experimental index i .

Research has been reported to model the relationship between the control variables and the engineered surfaces by using physical or mechanical knowledge. However, these models are often deterministic and lack the ability to deal with the inherent stochasticity. The stochastic behavior of the system cannot be ignored for analysis as a multitude of studies have found that topology related to stochastic variation affects the output quality (e.g., surface friction, see Bhushan (2003)). Furthermore, there has been evidence that stochastic elements are affected by the processing used; see Pei (2002) for an example involving a grinding process applied to silicon wafers.

Toward the aim of modeling the stochastic behavior of the system, we introduce a uniquely parameterized version of a Gaussian process model. Gaussian process models have found extensive use in areas such as geostatistics, environmental science, and computer experiments; see the works of Matheron (1963), Sansó & Guenni (2000), and Santner et al. (2003) for examples. The power of this model is the ability to account for varying levels differentiability and continuity of a surface. For a more complete history of literature relating to spatial statistics, see Gelfand et al. (2010).

The core contribution of this paper is the proposed model for the interaction of control variables and stochastic behavior, outlined in Section 2. In Section 3, an iterative optimization technique for finding maximum likelihood estimates (MLEs) for the model is discussed. Section 4 outlines a case study for this model that verifies the need to model stochastic behavior as a function of the control variables. Additionally, Section 4 describes a surface model defined on polar coordinates for silicon wafers produced via a rotational mechanism. Section 5 includes conclusions and outlines possible future work.

2 Proposed modeling framework

The model discussed in this paper will emulate semi-parametric models such as those seen in Zeger & Diggle (1994), replacing the spline model with a Gaussian process. This is justified by the connections between processes generated by spline basis functions and Gaussian process models; see, for example, Wahba (1978). Each experimental index i corresponds to a set of control variables denoted \mathbf{u}_i . The model of the k^{th} point on the j^{th} surface generated at experimental index i is described as

$$Y_{ijk}(\mathbf{u}_i, \mathbf{x}_k) = \mu(\mathbf{x}_k; \boldsymbol{\beta}(\mathbf{u}_i)) + Z_j(\mathbf{x}_k; \theta(\mathbf{u}_i), \sigma^2(\mathbf{u}_i)) + \epsilon_{ijk}, \tag{1}$$

where μ , Z , and ϵ represent the underlying mean, a stochastic error due to processing, and measurement noise, respectively. Values of measurement noise are modeled as independent and normal, $\epsilon_{ijk} \sim \mathcal{N}[0, \sigma_\epsilon^2]$. The mean is modeled as an additive linear model with a set of basis functions $\mathbf{f}_x(\mathbf{x})$ that map \mathbf{x} to \mathbb{R}^p , that is, $\mu(\mathbf{x}; \boldsymbol{\beta}(\mathbf{u})) = \mathbf{f}_x^\top(\mathbf{x})\boldsymbol{\beta}(\mathbf{u})$. The proper choice of a set of basis functions for the mean depends on the particular application. In Section 4, we use a general set of spline basis functions for \mathbf{f}_x combined with a basis function that incorporates engineering knowledge. The modeling of Z_j will take the form of an independent realization of the Gaussian process,

$$Z_j(\mathbf{x}; \theta(\mathbf{u}), \sigma^2(\mathbf{u})) \sim GP[0, \sigma^2(\mathbf{u})R(\cdot, \cdot; \theta(\mathbf{u}))],$$

where $\sigma^2(\mathbf{u})$ and $\theta(\mathbf{u})$ are the variance and correlation parameters of the process.

The Gaussian process model assumes that any collection of points on a surface follows a multivariate normal distribution that depends on the variance, $\text{Var}(Z_j(\mathbf{x}_1)) = \sigma^2(\mathbf{u})$, and the correlation function, $\text{Corr}\{Z_j(\mathbf{x}_1), Z_j(\mathbf{x}_2)\} = R(\mathbf{x}_1, \mathbf{x}_2; \theta(\mathbf{u}))$. The correlation function, R , is a positive definite function that dictates the correlation between any two points on a surface. For more information on the properties and development of correlation functions, see Gelfand et al. (2010). Several correlation functions exist, including the powered exponential class, the Matérn class (Matérn, 1986), and the Cauchy class (Gneiting & Schlather, 2004). The Gaussian process model is useful in this circumstance because it has the power to account for the underlying continuity of the surface. Using the Gaussian process assumption also creates the ability to use available maximum likelihood-based methods for estimating the parameters of the model.

The core idea of this paper is the following: *the parameters of the surfaces, especially ones that dictate stochastic behavior, ought to be modeled as a function of the control variables*. In terms of the proposed model, the mean parameter ($\boldsymbol{\beta}(\mathbf{u})$), the correlation parameter ($\theta(\mathbf{u})$), and the variance of the surface ($\sigma^2(\mathbf{u})$) are assumed to change with the control variables. This implies that the control variables affect not only the mean but also the stochastic portion of the surface model.

Because of the high cost associated with performing designed experiments in wafer manufacturing (both in terms of material and measurement costs), it is common for few design points to be present in experiments such as Table I. As a result, it is reasonable to use a linear model with higher order terms to represent how \mathbf{u} interacts with the surface parameters. The functions $\boldsymbol{\beta}(\mathbf{u})$, $\sigma^2(\mathbf{u})$, and $\theta(\mathbf{u})$ are therefore modeled as linear combinations of a set of basis functions $\mathbf{f}_u(\mathbf{u})$ that map a value of \mathbf{u} to \mathbb{R}^q . Let $\mathbf{A} \in \mathbb{R}^{q \times p}$, $\boldsymbol{\alpha}_\sigma \in \mathbb{R}^q$, and $\boldsymbol{\alpha}_\theta \in \mathbb{R}^q$; the functions are $\boldsymbol{\beta}(\mathbf{u}) = \mathbf{A}\mathbf{f}_u(\mathbf{u})$, $\log \sigma^2(\mathbf{u}) = \boldsymbol{\alpha}_\sigma^\top \mathbf{f}_u(\mathbf{u})$, and $g(\theta(\mathbf{u})) = \boldsymbol{\alpha}_\theta^\top \mathbf{f}_u(\mathbf{u})$, where g is a link function (e.g., a log-link function for a strictly positive quantity). These functions are not known explicitly, and they depend on parameters \mathbf{A} , $\boldsymbol{\alpha}_\sigma$, and $\boldsymbol{\alpha}_\theta$, respectively. We propose estimating these parameters via the algorithm in Section 3. The same basis function, $\mathbf{f}_u(\mathbf{u})$, is assumed for each parameter $\boldsymbol{\beta}(\mathbf{u})$, $\sigma^2(\mathbf{u})$, and $\theta(\mathbf{u})$. This is not a requirement, but it often simplifies notation and allows for the procedure outlined in this paper to be generally applicable.

3 Parameter estimation

The model described in the previous section has limited usability without valid parameter estimates. In this paper, parameter estimates will be determined by the maximum likelihood criterion. Given the complex structure of the model, closed-form estimates for all parameters could not be reached. Additionally, the amount of data present in experiments such as those seen in Table I makes Markov chain Monte Carlo techniques intractable. This section will detail an iterative method to find MLEs.

There is a simplified parameterization of the mean using properties of Kronecker products, \otimes . The expression for the mean vector is simplified to

$$\begin{aligned} \boldsymbol{\mu}(\mathbf{u}) &= \mathbf{F}_x \boldsymbol{\beta}(\mathbf{u}), \\ &= \mathbf{F}_x \mathbf{A} \mathbf{f}_u^\top(\mathbf{u}), \\ &= (\mathbf{F}_x \otimes \mathbf{f}_u^\top(\mathbf{u})) \text{vec}(\mathbf{A}), \\ &= (\mathbf{F}_x \otimes \mathbf{f}_u^\top(\mathbf{u})) \boldsymbol{\gamma}, \end{aligned}$$

where $\boldsymbol{\mu}^\top(\mathbf{u}) = [\mu(\mathbf{x}_1; \boldsymbol{\beta}(\mathbf{u})), \dots, \mu(\mathbf{x}_n; \boldsymbol{\beta}(\mathbf{u}))]$, \mathbf{F}_x is defined as $\mathbf{F}_x^\top = [\mathbf{f}_x(\mathbf{x}_1), \dots, \mathbf{f}_x(\mathbf{x}_n)]$, and $\boldsymbol{\gamma}$ is the new parameter for the mean of the model.

By the model in (1), $\mathbf{Y}_{ij} \sim \mathcal{N}_n[\boldsymbol{\mu}(\mathbf{u}_i), \sigma_i^2(\mathbf{R}(\theta_i) + \sigma_\epsilon^2 \mathbf{I})]$, where $\theta_i = \theta(\mathbf{u}_i; \boldsymbol{\alpha}_\theta)$ and $\sigma_i^2 = \sigma^2(\mathbf{u}_i; \boldsymbol{\alpha}_\sigma)$. The matrix $\mathbf{R}(\theta_i)$ is the surface correlation matrix composed of elements $\{\mathbf{R}(\theta_i)\}_{jk} = R(\mathbf{x}_j, \mathbf{x}_k; \theta_i)$. The total negative log-likelihood is therefore proportional to (up to a constant)

$$l = \sum_{i=1}^m \sum_{j=1}^{d_i} \left\{ n \log(\sigma_i^2) + \log(|\mathbf{R}_i^*|) + \frac{\mathbf{r}_{ij}^\top \mathbf{R}_i^{*-1} \mathbf{r}_{ij}}{\sigma_i^2} \right\}, \tag{2}$$

where m is the number of runs in the experiment, $\mathbf{r}_{ij}^\top = (\mathbf{Y}_{ij} - (\mathbf{F}_x \otimes \mathbf{f}_u^\top(\mathbf{u}_i)) \boldsymbol{\gamma})^\top$, and $\mathbf{R}_i^* = \mathbf{R}(\theta_i) + \frac{\sigma_\epsilon^2}{\sigma_i^2} \mathbf{I}$. Minimizing (2) is equivalent to maximizing the likelihood. The maximum likelihood estimate of the mean parameter, $\boldsymbol{\gamma}$, has a closed-form solution, given the parameters $\boldsymbol{\alpha}_\theta$ and $\boldsymbol{\alpha}_\sigma$,

$$\hat{\boldsymbol{\gamma}} = \left[\sum_{i=1}^m \sum_{j=1}^{d_i} \frac{1}{\sigma_i^2} (\mathbf{F}_x \otimes \mathbf{f}_u^\top(\mathbf{u}_i))^\top \mathbf{R}_i^{*-1} (\mathbf{F}_x \otimes \mathbf{f}_u^\top(\mathbf{u}_i)) \right]^{-1} \left[\sum_{i=1}^m \sum_{j=1}^{d_i} \frac{1}{\sigma_i^2} (\mathbf{F}_x \otimes \mathbf{f}_u^\top(\mathbf{u}_i))^\top \mathbf{R}_i^{*-1} \mathbf{Y}_{ij} \right]. \tag{3}$$

The MLE estimate of the parameter $\boldsymbol{\alpha}_\sigma$ also exhibits simplifications using the following proposition:

Proposition 3.1

Assume \mathbf{Y}_i , $i = 1, \dots, m$, follows a multivariate normal distribution with mean $\boldsymbol{\mu}(\mathbf{u}_i)$ and covariance $\sigma^2(\mathbf{u}_i) \mathbf{R}(\mathbf{u}_i)$. Assume that $\boldsymbol{\mu}(\mathbf{u}_i)$ and $\mathbf{R}(\mathbf{u}_i)$ are given and $\sigma^2(\mathbf{u}_i) = \boldsymbol{\alpha}_\sigma^\top \mathbf{f}_u(\mathbf{u}_i)$. The generalized linear model estimate (Nelder & Wedderburn, 1972) of $\boldsymbol{\alpha}_\sigma$, with gamma distribution, input vectors $\mathbf{f}_u(\mathbf{u}_1), \dots, \mathbf{f}_u(\mathbf{u}_m)$ and a log-link function, to the response $\sigma_1^2, \dots, \sigma_m^2$, where $\sigma_i^2 = \frac{1}{n} (\mathbf{Y}_i - \boldsymbol{\mu}(\mathbf{u}_i))^\top \mathbf{R}^{-1}(\mathbf{u}_i) (\mathbf{Y}_i - \boldsymbol{\mu}(\mathbf{u}_i))$, is equal to the maximum likelihood estimate of $\boldsymbol{\alpha}_\sigma$.

The preceding result is easily derived from the definition of generalized linear models. This demonstrates that we can use popular software methods to find numerical estimates for $\boldsymbol{\alpha}_\sigma$.

Using the preceding results, we construct an optimization algorithm that finds the MLEs of a group of parameters given all other parameters. This idea is similar in spirit to the classic EM algorithm (Dempster et al., 1977) viewed as an iterative maximization problem (Neal & Hinton, 1998). The parameters are grouped as $\{\boldsymbol{\alpha}_\theta, \sigma_\epsilon\}$, $\boldsymbol{\alpha}_\sigma$, and $\boldsymbol{\gamma}$. We denote each iterate with the superscript $\{k\}$ and suppress the two superscripts in the variance terms for clarity. The algorithm is as follows:

1. Initialize $\sigma_i^{\{0\}}$, $\theta_i^{\{0\}}$, and $\sigma_{\epsilon,i}^{\{0\}}$ for each experimental index i . Set $k = 0$.
2. Use $\sigma_i^{\{0\}}$ and $\theta_i^{\{0\}}$ to initialize $\boldsymbol{\alpha}_\theta^{\{0\}}$, $\boldsymbol{\alpha}_\sigma^{\{0\}}$ via least squares regression and initialize $\sigma_\epsilon^{\{0\}} = \frac{1}{m} \sum_i \sigma_{\epsilon,i}^{\{0\}}$.
3. Set $\sigma_i^{\{k\}} = \sigma^2(\mathbf{u}_i; \boldsymbol{\alpha}_\sigma^{\{k\}})$, $\theta_i^{\{k\}} = \theta(\mathbf{u}_i; \boldsymbol{\alpha}_\theta^{\{k\}})$, and $\mathbf{R}_i^* = \mathbf{R}(\theta_i^{\{k\}}) + \frac{\sigma_{\epsilon,i}^{\{k\}}}{\sigma_i^{\{k\}}} \mathbf{I}$ for all $i = 1, \dots, m$.
4. Estimate $\boldsymbol{\gamma}^{\{k+1\}}$ via (3) using \mathbf{R}_i^* and $\sigma_i^{\{k\}}$.

5. Estimate the variance term for each surface via

$$\sigma_{ij}^2 = \frac{\left(\mathbf{Y}_{ij} - \left(\mathbf{F}_x \otimes \mathbf{f}_u^\top(\mathbf{u}_i) \right) \boldsymbol{\gamma}^{\{k+1\}} \right)^\top \mathbf{R}_i^{*-1} \left(\mathbf{Y}_{ij} - \left(\mathbf{F}_x \otimes \mathbf{f}_u^\top(\mathbf{u}_i) \right) \boldsymbol{\gamma}^{\{k+1\}} \right)}{n}.$$

Fit a GLM with a log link with response σ_{ij}^2 , an input vector $\mathbf{f}_u(\mathbf{u}_i)$, and a gamma distribution to yield $\boldsymbol{\alpha}_\sigma^{\{k+1\}}$.

Set $\sigma_i^{\{k+1\}} = \sigma^2 \left(\mathbf{u}_i; \boldsymbol{\alpha}_\sigma^{\{k+1\}} \right)$ and $\mathbf{R}_i^* = \mathbf{R}(\theta_i^{\{k\}}) + \frac{\sigma_\epsilon^{\{k+1\}}}{\sigma_i^{\{k\}}} \mathbf{I}$. Repeat to convergence of $\boldsymbol{\alpha}_\sigma^{\{k+1\}}$.

6. Use a generic optimization algorithm to maximize the likelihood given in (2) using $\boldsymbol{\gamma} = \boldsymbol{\gamma}^{\{k+1\}}$ and $\sigma_i = \sigma_i^{\{k+1\}}$, $i = 1, \dots, m$, to choose $\boldsymbol{\alpha}_\theta^{\{k+1\}}$ and $\sigma_\epsilon^{\{k+1\}}$. Set $\theta_i^{\{k+1\}} = \theta \left(\mathbf{u}_i; \boldsymbol{\alpha}_\theta^{\{k+1\}} \right)$, and $\mathbf{R}_i^* = \mathbf{R} \left(\theta_i^{\{k+1\}} \right) + \frac{\sigma_\epsilon^{\{k+1\}}}{\sigma_i^{\{k+1\}}} \mathbf{I}$.

7. Repeat 4 to 6 until convergence.

This iterative formulation reduces the computational burden for the likelihood by exploiting known partial maximization in steps 4 and 5. One might question the convergence of the preceding algorithm. The estimates will converge to the true MLEs if the negative log-likelihood is convex with respect to all parameters because we are continually improving. However, there is no guarantee of convexity. But, with adequate initialization, good estimates can be obtained. Initial estimates of $\sigma_i^{\{0\}}$, $\theta_i^{\{0\}}$, and $\sigma_\epsilon^{\{0\}}$ can be found via several statistical packages by fitting a Gaussian process model to each surface $\mathbf{Y}_{i1}, \dots, \mathbf{Y}_{id_i}$ and averaging the result.

4 Case study: Si wafer thickness variation after mechanical lapping

The geometric uniformity of the silicon wafers determines the incoming variation for the downstream processing to create products such as integrated circuits and solar cells. Thus, this uniformity is a major control objective in wafer manufacturing. This section presents analysis of a designed experiment to better understand the effect of a set of control variables on a wafer flattening procedure.

4.1. Background and experimental description

Wire sawing is the most common approach to slice large-diameter silicon crystal ingots into wafers. Wire sawing can lead to undesirable waviness in final product (Yasunaga et al., 1997), which may be due to the vibrations present in the wire (Zhao et al., 2011). Flattening procedures, which include lapping, grinding, and chemical–mechanical polishing (CMP), are used to reduce the waviness in the final product. For more details on the processes involved in the production of wafers, see Evans et al. (2003) and Pei et al. (2003).

Significant work has been carried out in the study of these flattening procedures, and majority of work has been focused in two areas:

Deterministic differential equation models An example of the differential equation methods can be seen in Fu et al. (2001), which develops a model that uses plasticity properties of the surface material to study CMP. Similar developments have been made in both lapping (Chang et al., 2000) and grinding (Pietsch & Kerstan, 2005).

Feature extraction This method consists of first extracting a feature from each surface. After feature extraction, traditional analysis of experiments is performed where the extracted feature is modeled as the response (e.g., Zhang et al. (2007)). An example of an extracted feature is total thickness variation (TTV), which is computed by subtracting the smallest thickness measurement from the largest thickness measurement on a single wafer.

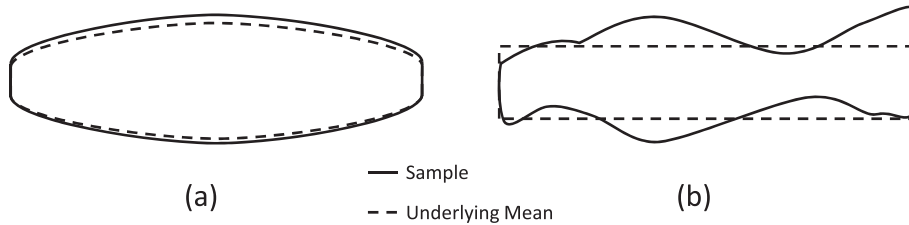


Figure 1. Exaggerated cross sections of semiconductor wafers with similar total thickness variation (TTV) caused by (a) mean structure and (b) stochasticity.

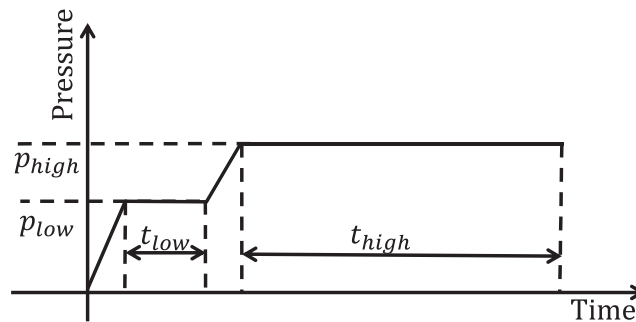


Figure 2. Pressure profile from the lapping process described in Section 4.

However, current research has not yet resolved the issues discussed in this paper. Deterministic differential equation models lack the ability to describe stochastic outputs. Feature extraction techniques do not take full advantage of the structure of the data, and those techniques lack the ability to compensate for known elements such as sensor noise. As an example of the limits of feature extraction, compare a surface that consists of high thickness in the middle and low thickness on the edges (Figure 1(a)) to a surface that has sharp peaks and troughs in thickness (Figure 1(b)). With feature extraction, differentiation between these surfaces is impossible.

Furthermore, analogies between modern metrology for surfaces and our model indicate the importance of modeling stochastic effects. As an example, the Birmingham set parameters, outlined in Dong et al. (1994), are widely used as descriptive parameters for surface topology (Messner et al., 2003). These parameters are specifically designed to extract features related to the *spatial* and *amplitude* properties of the surface. These quantities are directly comparable with the variance and correlation parameters in the proposed model.

Motivated by the gap between current research in flattening procedures and the described goals, a designed experiment was performed on a lapping process (outlined in Table I). Thickness measurements were taken at over 150 points after a lapping process. The lapping pressure varies throughout the process, and the lapping pressure as a function of time can be seen in Figure 2.

Our experiment studies the effect of altering different aspects from this pressure profile, which includes the low pressure setting, the time at both the high and low pressures, and the rotation speed (measured in rotations per minute) of the lapping pads. The process consists of a batch setup, with over 40 wafers being produced during each run.

4.2. Exploratory data analysis

The measurement locations are shown in Figure 3 and are indexed in a counterclockwise spiral direction from inside to out starting at angle $\pi/2$. The lapping process grinds fine particles on a rotating surface. Because of this flattening mechanism, the use of Cartesian coordinates to analyze the mean and correlation structures would produce cryptic results. Therefore, we use polar coordinates to examine the mean and correlation functions.

To study the properties of the mean and correlation functions, we first directly calculate the sample mean vector and the sample correlation matrices from each of the experimental indices i . Using this estimate directly for analysis is not considered a stable procedure because the number of points on a surface (150) exceeds the number of surface observations (≈ 40) in this case study. This is a common problem in applications relating to surface engineering and other functional responses. However, these computations are still useful for exploratory analysis.

Figure 4 shows the sample mean profile for all nine experimental indices. From this figure, a clear negative slope relationship between the thickness measurement and the radial distance is present, along with local deviations from the general trend. The negative relationship between radial distance and thickness has been observed in mechanical polishing of wafers (Pietsch & Kerstan, 2005) and a similar trend is confirmed here. The global mean trend is modeled with the function $\rho(\mathbf{x})$, where $\rho(\mathbf{x})$ represents the radial distance from the center of the wafer. To model the local mean function, a set of cubic-spline basis functions, defined on polar coordinates, is used with knot points shown in Figure 3. These two functions comprise a basis set of functions for each surface, that is, $\mathbf{f}_x(\mathbf{x})$ discussed in Section 2.

The sample correlation is useful for investigating properties of the correlation functions. Here, we consider empirical approximations of the correlation functions computed by averaging the estimated correlation coefficients of any two points that differ by the same distance, yielding empirical measurements of a stationary correlation function, termed $\tilde{R}(\Delta\psi, \Delta\rho)$. We examine at the correlation function by studying the empirical angular correlation, $\tilde{R}_\psi(\Delta\psi) = \tilde{R}(\Delta\psi, 0)$, and the empirical radial correlation, $\tilde{R}_\rho(\Delta\rho) = \tilde{R}(0, \Delta\rho)$.

Figure 5 shows the empirical angular correlation function over several experimental indices. As shown in Gaspari & Cohn (1999), for a surface in two dimensions, a valid stationary correlation function of $\Delta\psi$, where $\Delta\psi$ is the difference in the angle between two points, must take the form of $R(\Delta\psi) = \sum_{k=0}^{\infty} a_k \cos(k\Delta\psi)$ with $\sum_{k=0}^{\infty} a_k = 1$.

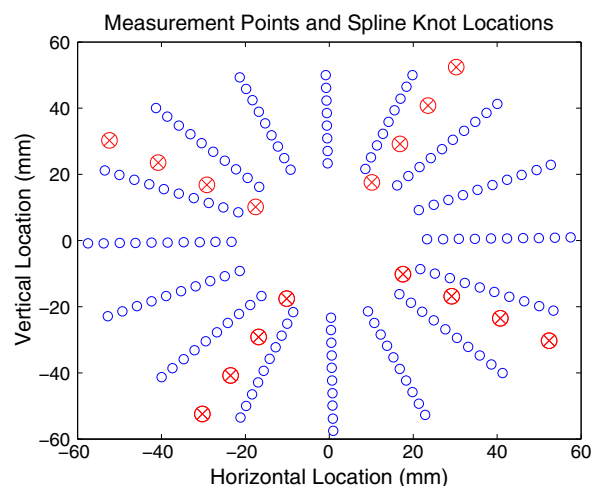


Figure 3. Measurement locations for each wafer marked with \circ . Spline knot locations marked with \otimes .

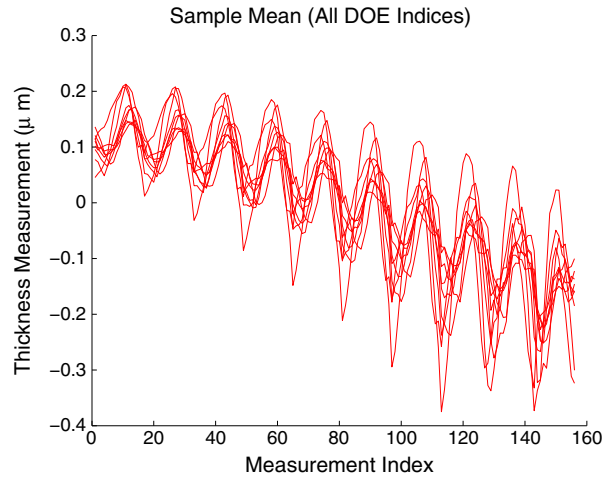


Figure 4. Sample mean profile for each experimental index.

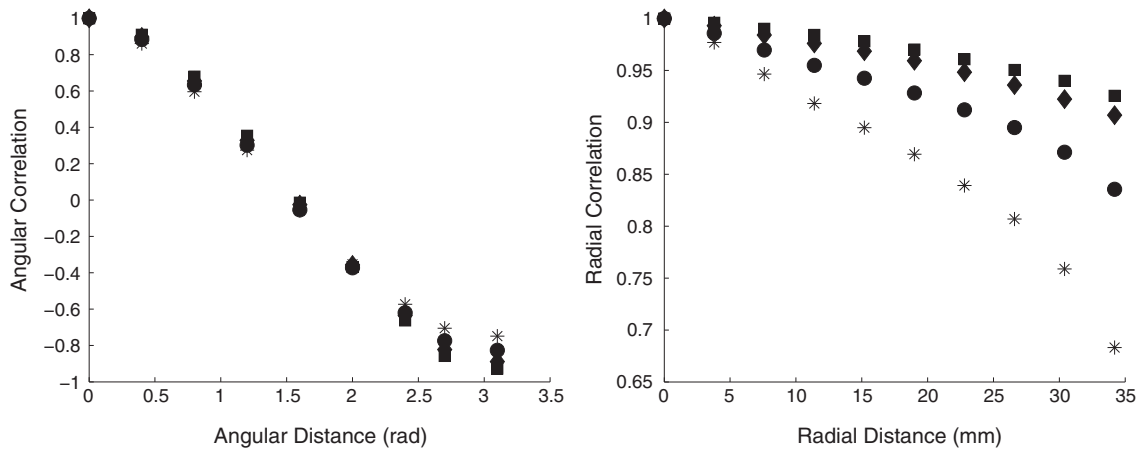


Figure 5. Empirical angular (left) and radial (right) correlation functions estimated from the sample correlation matrices for experimental indices 1(circles), 3(squares), 5(diamonds) and 8(stars).

With this in mind, the correlation function in the angular direction will modeled with a second order approximation,

$$R_{\psi}(\Delta\psi) = \theta_{\psi} \cos(|\Delta\psi|) + (1 - \theta_{\psi}) \cos(2|\Delta\psi|),$$

with a shape parameter θ_{ψ} , which appears to be largely unaffected by the control variables \mathbf{u} in Figure 5.

Figure 5 shows the approximate radial correlation function over several experimental indices. Unlike the angular correlation function, this indicates the control variables (\mathbf{u}) greatly affect the correlation structure. The unknown correlation function $R_{\rho}(\Delta\rho)$ is therefore modeled using a Matérn class of correlation functions,

$$R_{\rho}(\Delta\rho) = \frac{1}{2^{\phi-1}\Gamma(\phi)} \left(|\Delta\rho| \frac{2\sqrt{\phi}}{\theta_{\rho}(\mathbf{u})} \right)^{\alpha} \mathcal{K}_{\phi} \left(|\Delta\rho| \frac{2\sqrt{\phi}}{\theta_{\rho}(\mathbf{u})} \right),$$

where \mathcal{K}_ϕ is the modified Bessel function of order ϕ . Here, θ_ρ is strongly affected by \mathbf{u} ; therefore, this will be modeled by the linear function described Section 2. In total, the correlation function will be modeled by the outer product of R_ψ and R_ρ ,

$$R(\Delta\psi, \Delta\rho) = R_\psi(\Delta\psi)R_\rho(\Delta\rho).$$

4.3. Results from parameter estimation and verification of model significance

The focal point of this paper is the introduction of control variables, \mathbf{u} , to the stochastic parameters of the Gaussian process error term. This section will justify the need for this modeling through a comparison with a model with constant stochastic parameters. For simplicity, we compare two models of increasing complexity:

[M0] Model (1) with assumptions $\sigma^2(\mathbf{u}) = \sigma^2$ and $\theta_\rho(\mathbf{u}) = \theta_\rho$, that is, both are constant functions.

[M1] Model (1) where the functions $\sigma^2(\mathbf{u})$ and $\theta_\rho(\mathbf{u})$ will be assumed to be a linear functions of \mathbf{u} , as outlined in previous sections.

Both models will have parameters chosen via the maximum likelihood criterion.

To check for a misrepresented surface, we compare examples of measured surfaces and simulated surfaces from M0 and M1 with estimated parameters. Figure 6 displays thickness maps with the scale removed at the request of the data provider. While the surfaces do not look identical due to stochasticity in the process, the measured surfaces appear to be similar to simulated surfaces. Furthermore, the model M1 seems to better capture the stochastic behavior of the surface, although the difference is subtle.

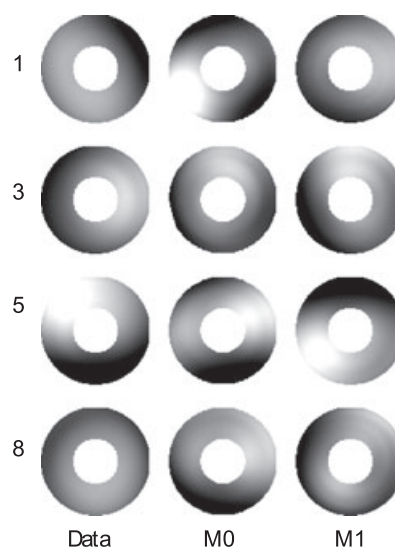


Figure 6. Examples of measured and simulated surfaces; darker regions indicate lower thickness. Scale is removed at the request of the data provider.

Table II. Model selection criteria.		
Model	M0	M1
$\ln L$	0	4482.6
BIC	413.8388	-1742.2

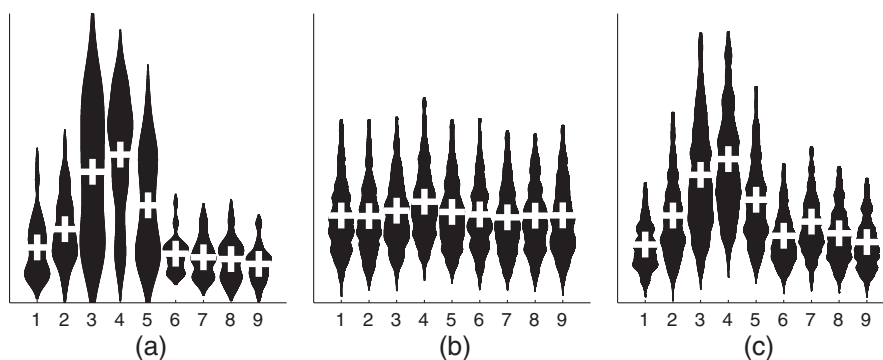


Figure 7. Violin plot of total thickness variation (TTV) at the nine experimental indices listed in Table I from different models: (a) TTV from surfaces collected in the experiment and (b) and (c) TTV from surfaces generated from model *M0* and *M1* with estimated parameters, respectively. The black area represents a kernel smoothed density estimate and the white “+” represents the median.

Table II demonstrates that the likelihood is higher for the proposed model *M1* versus the null comparison *M0*, indicating that the proposed model better explains the data.[†] However, *M1* contains more parameters than *M0*, so a comparison with Bayesian information criteria (BIC) is more appropriate. BIC was chosen for its strong penalty for excessive parameters and is given by $-2 \ln L + k \ln n$, where L is the likelihood, k is the number of parameters, and n is the number of observations. A smaller value of BIC represents a superior representation relative to the number of parameters. In Table II, the values of BIC indicate that *M1* is a superior explanatory model.

To add to the preceding results, an engineering verification is to extract a feature from data generated from the models. The chosen metric for this situation is TTV, which was discussed earlier in this section. This value was chosen because of its pervasive use in wafer manufacturing (Tso & Teng, 2001). Figure 7 shows violin plots of TTV from simulated data compared the TTV from the collected data. The full model, *M1*, best mimics the behavior of the original data. Fundamentally, the need for quantifying changes in the stochastic portion of the model is demonstrated in this analysis.

5 Conclusions

This paper has demonstrated a model for systems that produce surface responses by incorporating not only surface variables but also control variables in a Gaussian process error structure. Incorporating control variables into the stochastic portion of the model allows for interpretation of the behavior of a system under different operating environments. The use of the model is justified in the described case study on wafer thickness surfaces. In future applications,

[†] The log-likelihood of *M0* is stated as null, or 0, and the log-likelihood of *M1* is the value after subtracting the log-likelihood from *M0*. This normalized log-likelihood is also used to compute BIC for both models. This normalization does not affect the comparison of likelihood or BIC among the two models.

using different aspects of this model, but not the full model, might be beneficial. For example, Chang et al. (2012) studied nanoparticle dispersion, where the correlation structure is considered a function of the physical properties of the atom and does not change with control variables.

Acknowledgement

This research was supported by the US National Science Foundation grant CMMI-1030125.

References

- Bhushan, B (2003), 'Adhesion and stiction: mechanisms, measurement techniques, and methods for reduction', *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures*, **21**, 2262–2296.
- Chang, CJ, Xu, L, Huang, Q & Shi, J (2012), 'Quantitative characterization and modeling strategy of nanoparticle dispersion in polymer composites', *IIE Transactions*, **44**, 523–533.
- Chang, Y, Hashimura, M & Dornfeld, D (2000), 'An investigation of material removal mechanisms in lapping with grain size transition', *Journal of Manufacturing Science and Engineering*, **122**, 413–419.
- Dempster, A, Laird, N & Rubin, D (1977), 'Maximum likelihood from incomplete data via the EM algorithm', *Journal of the Royal Statistical Society. Series B*, **39**, 1–38.
- Dong, WP, Sullivan, PJ & Stout, KJ (1994), 'Comprehensive study of parameters for characterising three-dimensional surface topography IV: parameters for characterising spatial and hybrid properties', *Wear*, **178**, 45–60.
- Evans, CJ, Paul, E, Dornfeld, D, Lucca, DA, Byrne, G, Tricard, M, Klocke, F, Dambon, O & Mullany, BA (2003), 'Material removal mechanisms in lapping and polishing', *CIRP Annals-Manufacturing Technology*, **52**, 611–633.
- Faraway, JJ (1997), 'Regression analysis for a functional response', *Technometrics*, **39**, 254–261.
- Fu, G, Chandra, A, Guha, S & Subhash, G (2001), 'A plasticity-based model of material removal in chemical-mechanical polishing (CMP)', *IEEE Transactions on Semiconductor Manufacturing*, **14**, 406–417.
- Gaspari, G & Cohn, S (1999), 'Construction of correlation functions in two and three dimensions', *Quarterly Journal of the Royal Meteorological Society*, **125**, 723–757.
- Gelfand, A, Diggle, P, Fuentes, M & Guttorp, P (2010), *Handbook of Spatial Statistics*, CRC Press.
- Gneiting, T & Schlather, M (2004), 'Stochastic models that separate fractal dimension and the Hurst effect', *SIAM Review*, **46**, 269–282.
- Matérn, B (1986), *Spatial Variation*, 2nd edn., Springer-Verlag, Berlin.
- Matheron, G (1963), 'Principles of geostatistics', *Economic Geology*, **58**, 1246–1266.
- Messner, C, Silberschmidt, V & Werner, E (2003), 'Thermally induced surface roughness in austenitic-ferritic duplex stainless steels', *Acta Materialia*, **51**, 1525–1537.
- Neal, RM & Hinton, GE (1998), 'A view of the EM algorithm that justifies incremental, sparse, and other variants', in *Learning in Graphical Models*, Springer, pp. 355–368.

- Nelder, JA & Wedderburn, RW (1972), 'Generalized linear models', *Journal of the Royal Statistical Society. Series A*, **135**, 370–384.
- Pei, Z (2002), 'A study on surface grinding of 300 mm silicon wafers', *International Journal of Machine Tools and Manufacture*, **42**, 385–393.
- Pei, ZJ, Xin, XJ & Liu, W (2003), 'Finite element analysis for grinding of wire-sawn silicon wafers: a designed experiment', *International Journal of Machine Tools and Manufacture*, **43**, 7–16.
- Pietsch, GJ & Kerstan, M (2005), 'Understanding simultaneous double-disk grinding: operation principle and material removal kinematics in silicon wafer planarization', *Precision Engineering*, **29**, 189–196.
- Qiu, P, Zou, C & Wang, Z (2010), 'Nonparametric profile monitoring by mixed effects modeling', *Technometrics*, **52**, 265–277.
- Sansó, B & Guenni, L (2000), 'A nonstationary multisite model for rainfall', *Journal of the American Statistical Association*, **95**, 1089–1100.
- Santner, TJ, Williams, BJ & Notz, W (2003), *The Design and Analysis of Computer Experiments*, Springer Verlag.
- Tso, PL & Teng, CC (2001), 'A study of the total thickness variation in the grinding of ultra-precision substrates', *Journal of Materials Processing Technology*, **116**, 182–188.
- Wahba, G (1978), 'Improper priors, spline smoothing and the problem of guarding against model errors in regression', *Journal of the Royal Statistical Society. Series B (Methodological)*, **40**, 364–372.
- Wu, CFJ & Hamada, M (2009), *Experiments: Planning, Analysis, and Optimization*, 2nd edn., Wiley New York.
- Yasunaga, N, Takashina, M & Itoh, T (1997), 'Development of sequential grinding-polishing process applicable to large-size Si wafer finishing', in *Proceedings of the International Symposium on Advances in Abrasive Technology, Sydney, Australia*, pp. 96–100.
- Zeger, S & Diggle, P (1994), 'Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters', *Biometrics*, **50**, 689–699.
- Zhang, JZ, Chen, JC & Kirby, ED (2007), 'Surface roughness optimization in an end-milling operation using the Taguchi design method', *Journal of Materials Processing Technology*, **184**, 233–239.
- Zhao, H, Jin, R, Wu, S & Shi, J (2011), 'PDE-constrained Gaussian process model on material removal rate of wire saw slicing process', *Journal of Manufacturing Science and Engineering*, **133**.