Achieving Uniformity in a Semiconductor Fabrication Process Using Spatial Modeling

Jacqueline M. HUGHES-OLIVER, Jye-Chyi LU, Joseph C. DAVIS, and Ronald S. GYURCSIK

Material is deposited onto the wafer surface during several steps of wafer fabrication. This material must be deposited evenly across the entire wafer surface, close to the targeted thickness, and with little wafer-to-wafer variability. But unequal variances across the wafer and under different process conditions, as well as nonstationary correlation across a wafer, make these goals difficult to achieve, because traditional methods for optimizing deposition processes assume homogeneity and independence. We avoid these assumptions and determine the best settings of process variables using physically motivated statistical models for the mean response, unequal variances, and nonstationary spatial correlation structure. Data from a rapid thermal chemical vapor deposition process is used to illustrate the approach. A simulation exercise demonstrates the advantages of fitting flexible variance models and using appropriate performance measures.

KEY WORDS: Nonstationary correlation; Optimality criteria; Repeatability; Restricted maximum likelihood; Spatial correlation; Variance modeling.

1. INTRODUCTION

Semiconductor fabrication processes are costly, complicated processes whose economic viability demands high yield from *every* location on *each* wafer (Bartelink 1994). A major step in semiconductor fabrication is the deposition of material onto the wafer surface. Deposition processes must provide uniformity, target achievement, and repeatability. A uniform surface is one that has equal deposited thickness across the wafer. Uniformity may sometimes be defined more broadly to mean low within-wafer, or site-to-site, or location-to-location variability. Target achievement is said to occur when the average deposited thickness is within an acceptable range. It is important to note that uniformity and target achievement are not the same. Uniformity may be attained, but at a thickness much different from the target. It is also possible for a surface to be considered close to the target, but not uniform because it is too "bumpy." Finally, a repeatable process yields similar surfaces for wafers processed under identical conditions. Repeatability may also be referred to as low run-to-run or wafer-to-wafer variability, and is the goal of many studies in statistical process control.

Uniformity and target achievement have received much attention in the chemical vapor deposition engineering literature (see, e.g., Guo and Sachs 1993 and Lin and Spanos 1990), but very little attention has been paid to repeatability. This has followed from an implicit assumption that variability comes solely from differences in expected thickness and not from differences in the variance of the response. In other words, emphasis is placed on modeling the mean responses (and measures derived from the means), while assuming equal variances. For uniformity optimization, people typically model statistics related to the sample variance taken over "noise" replicates (such as spatial locations) within a single wafer. The variation between true replicates across wafers is ignored. The issue of repeatability cannot be addressed adequately without investigating the relationship between wafer-to-wafer variability and process settings. A variance model describing this relationship can then be used to determine process settings to minimize the wafer-to-wafer variability, thus giving rise to a more repeatable process.

Additional information is often required for successful variance modeling. For example, consistent estimation of the parameters in a variance model is achievable only if the correlation structure between responses is adequately specified. Yet a common, but very questionable, assumption is that responses on a single wafer are independent. We argue that it is better to allow the data to suggest a correlation model. If no significant correlation is detected, then we can return to the simpler case of independent responses.

Although a simple variance–covariance model such as homogeneity and independence will not usually affect the consistency of estimators of the mean model parameters, it will certainly cause these estimators to be inefficient. An appropriate combination of variance and correlation models can improve the efficiency of the mean parameter estimators and thus reduce the size of the confidence region for the optimum process settings. The net effect is more precise identification of the optimum process settings.

The goal of this article is to review the assumptions and drawbacks of traditional optimization techniques for deposition processes and to present a method that does not make these assumptions, yet is simple to use. In Section 2 we describe the wafer fabrication deposition process that motivates this work. In Section 3 we review what is typically done with these kinds of data; we also suggest a method for handling violations of the usual assumptions of homogeneity and independence, even when no replicates are available. In Section 4 we apply results from Section 3 to the data of

Jacqueline M. Hughes-Oliver is Assistant Professor and Jye-Chyi Lu is Associate Professor, Department of Statistics, North Carolina State University, Raleigh, NC 27695. Joseph C. Davis is Technical Staff member, Semiconductor Process and Device Center, Texas Instruments, Inc., Dallas, TX 75265. Ronald S. Gyurcsik is Associate Professor, Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC 27695. The authors thank the editor, two referees, Cavell Brownie, Marcia Gumpertz, Douglas Nychka, Sastry Pantula, and William Swallow for their comments and suggestions, which led to a much improved manuscript.

Section 2. In Section 5 we provide results from a simulation exercise, and in Section 6 we end with discussion.

2. A SEMICONDUCTOR WAFER FABRICATION PROCESS

This work is motivated by a particular rapid thermal chemical vapor deposition (RTCVD) process used at the Center for Advanced Electronics Material Processing at North Carolina State University. A single silicon wafer is covered with oxide on both sides, placed in a quartz chamber, and heated to a temperature that depends on the processing time. Gaseous polysilicon, which is injected into the chamber, then bonds to the heated surface of the wafer. The wafer is heated by two banks of lamps, one bank of lamps above the wafer surface and the other below the surface. The temperature of the wafer is monitored by a single pyrometer that records the temperature at the wafer's center. This temperature reading is used in a feedback loop that changes the power of the lamps to maintain a constant processing temperature. The goal is to determine process settings that minimize wafer-to-wafer variability and deposit a uniform thickness of 1,000 angstroms of polysilicon across the upper surface of the wafer.

There are several complications, however. First, the single pyrometer "reads" temperature only at the wafer's center, but temperature across the wafer is uneven because heat is lost at the edges (Davis, Gyurcsik, and Lu 1993; Ozturk, Sorrell, and Wortman 1991; Perkins, Riley, and Gyurcsik, 1995). Placing pyrometers at several sites on the wafer would, unfortunately, not remove all temperature variation, because a pyrometer does not actually read temperature, but instead calculates it as a function of the emissivity of the wafer surface. The surface emissivity changes as the polysilicon thickness increases, but thickness is not included in the formula used by a pyrometer. Thus as the polysilicon layer increases, there is less control over the temperature of the wafer. Because there is no easy way to measure the temperature across the wafer, we treat it as an unobservable response that introduces spatial patterns across the wafer. It is expected that sites close to the edge will have smaller average thickness and show more wafer-to-wafer variability in thickness than sites close to the center.

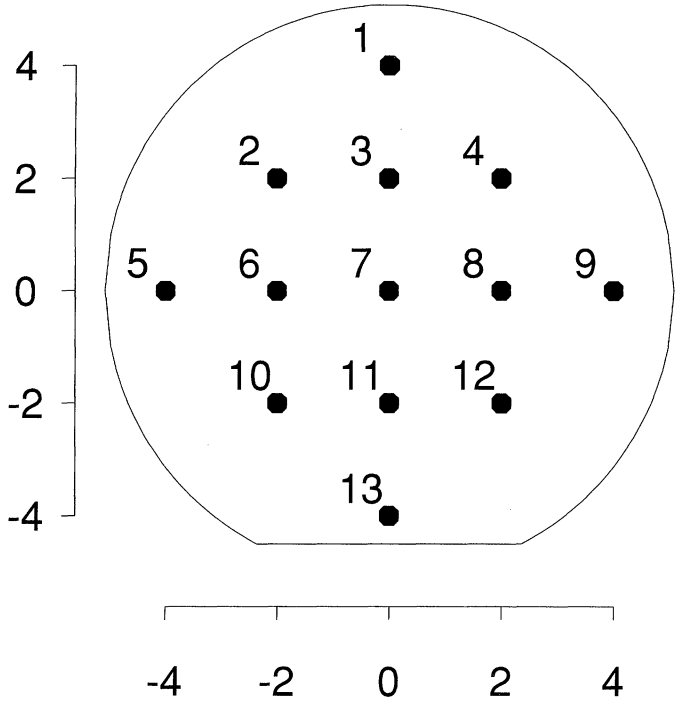

Figure 1. Sites at Which Polysilicon Thickness is Measured (Axes Centimeters). Polysilicon is expected to be thinner and more variable at the edges than at the center.

After the wafers are processed, polysilicon thickness is measured at 13 sites on each wafer (see Fig. 1). Site 7 is called the center point; it would be the center of the wafer if the wafer were a perfect circle. In the absence of random variation, one would expect the polysilicon to form an asymmetric mound-shaped surface on the wafer, centered at site 7. The mound occurs at site 7 because the wafer temperature is greatest at the center. The asymmetry is because the wafer is not circular, causing site 13 to be closer to the edge than, say, site 9. Similarly, the wafer-to-wafer variance surface is expected to be asymmetric and bowl-shaped. Finally, the (unobservable) temperatures at two sites are related not only by their proximity, but also by their distances from the center. This is because heat dissipates in radial bands from the warmest location on the wafer (the center) to the coolest locations on the wafer (the edges). Consequently, is that typical correlation models, which assume that correlation in polysilicon thickness is simply a function of distance between sites, may not suitably reflect this temperature effect.

Wafers for this experiment were selected from two lots, A and B. The lots come from the same boule (the silicon crystal before it is sliced) and are not expected to show any differences. Our analysis assumes no lot-to-lot differences.

There are two process or controllable variables: oxide thickness, the amount of oxide deposited on the wafer prior to deposition of the polysilicon, and deposition time, the length of time the wafer is heated for polysilicon deposition. Oxide thickness ranges from 458 to 1,498 angstroms. Deposition time ranges from 18 to 52 seconds. The experimental design points are selected according to a "quasi-replicated" central composite design (see Fig. 2). Wafer A21, for example, was covered with 986 angstroms of oxide before being placed in the quartz chamber for 35 seconds. Wafer A17 was also to receive 986 angstroms of oxide, but in fact was covered with 1,049 angstroms of oxide. The process used for depositing oxide onto the silicon wafer is not precise enough to guarantee exact replication of the oxide thickness from one run to the next. Consequently, we use the term "quasi-replicated" to indicate that the design points are close, although not identical.

Figure 3 shows the relationship between polysilicon thickness and distance from the wafer center for two of the 22 wafers. Measurement site number is used as the plotting symbol for sites of interest. Several features of the data are evident in Figure 3. First, polysilicon thickness decreases as the distance from the center increases. Second, site 13 has the smallest thickness, and site 7 (the center point) has

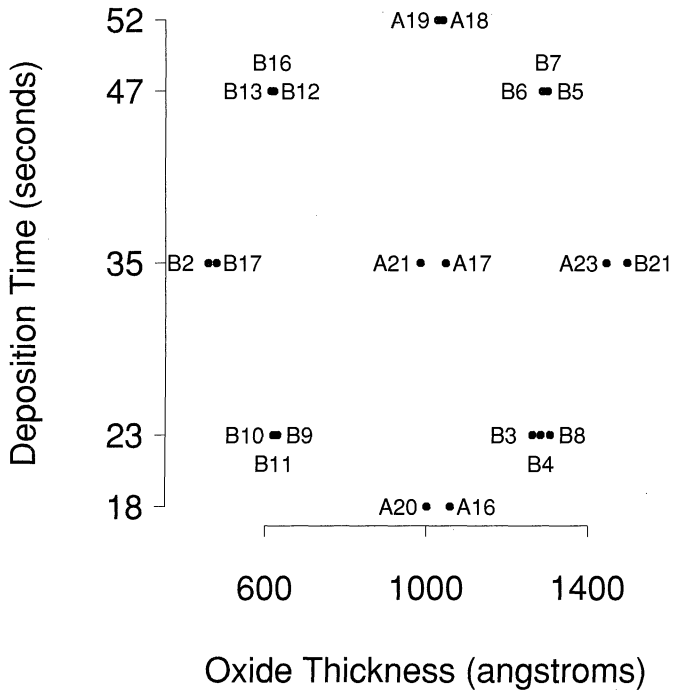

Figure 2. Experiment Design, With Wafer Identifiers. Wafers preceded by the letter A are from lot A, and wafers preceded by the letter B are from lot B. No lot-to-lot differences are expected. Oxide thickness and deposition time are expected to affect both the response mean and variance.

the largest thickness; this was expected, as discussed earlier. Third, the differences in polysilicon thicknesses on the edge (sites 1, 9, 5, and 13) far exceed the differences for sets of points closer to the center (plotting symbol *), suggesting that site-to-site variability increases with distance from the center. Fourth, polysilicon thickness is affected by the choice of processing conditions; deposition is much greater for wafer B16 than for wafer A16. Fifth, wafer A16 is much flatter than wafer B16, suggesting that uniformity is affected by processing conditions.

The complete dataset is available in Statlib.

3. METHODS FOR OPTIMIZING A DEPOSITION PROCESS

Let $\mathbf{y}(\mathbf{x})$ be a vector of responses obtained from a single wafer having covariate vector $\mathbf{x}$. Suppose that the responses are further indexed within experimental units as

$$\mathbf{y}(\mathbf{x}, \mathbf{Z}_\mathbf{x}) = [y(\mathbf{x}, \mathbf{z}_1), \ldots, y(\mathbf{x}, \mathbf{z}_{S_\mathbf{x}})]',$$

where $\mathbf{z}_k$ is a d-dimensional vector of within-unit covariates for the kth response, $S_\mathbf{x}$ is the number of responses for the unit with covariates $\mathbf{x}$, and $\mathbf{Z}_\mathbf{x} = [\mathbf{z}_1', \ldots, \mathbf{z}_{S_\mathbf{x}}']'$ is the $S_\mathbf{x} \times d$ matrix of within-unit covariates. In the RTCVD process of Section 2, $\mathbf{x}$ represents settings of process variables oxide thickness and deposition time, $\mathbf{z}_k$ is the coordinate of the kth wafer measurement site, and there are $S_\mathbf{x} \equiv S = 13$ measurements for each wafer. Because all wafers share common measurement sites, the 13×2 matrix $\mathbf{Z}_\mathbf{x} \equiv \mathbf{Z}$ is the same for all $\mathbf{x}$'s. For convenience, throughout the article we use the simpler S and $\mathbf{Z}$ instead of $S_\mathbf{x}$ and $\mathbf{Z}_\mathbf{x}$.

Suppose that the functional relationship between the stochastic response $\mathbf{y}$ and deterministic predictors $\mathbf{x}$ and $\mathbf{Z}$ is

$$\mathbf{y}(\mathbf{x}, \mathbf{Z}) = \mathbf{f}(\mathbf{x}, \mathbf{Z}) + \varepsilon(\mathbf{x}, \mathbf{Z}), \tag{1}$$

where $\mathbf{f}(\mathbf{x}, \mathbf{Z})$ is deterministic and $\varepsilon(\mathbf{x}, \mathbf{Z})$ is stochastic. This relationship may be used to predict responses at far more process settings than can reasonably be done using experimentation. Each of the predicted responses will have an associated loss or cost value, depending on desirability and/or manufacturing cost issues. The goal of (robust) process optimization is to determine the value of $\mathbf{x}$ that minimizes the expected loss, irrespective of $\mathbf{Z}$.

Optimization for wafer deposition processes is usually done in two steps. In the first step, covariates affecting uniformity are determined, and their values are set to provide the most uniform surface, irrespective of how thick that uniform surface is. From the remaining covariates (assuming that some are left), those affecting thickness are determined, and their values are set to adjust the thickness to target. As such, the lion's share of the work is devoted to optimization for uniformity. A commonly used uniformity loss metric (Guo and Sachs 1993; Lin and Spanos 1990) is the coefficient of variation across a single wafer. Differences in optimization approaches arise from how this metric is calculated.

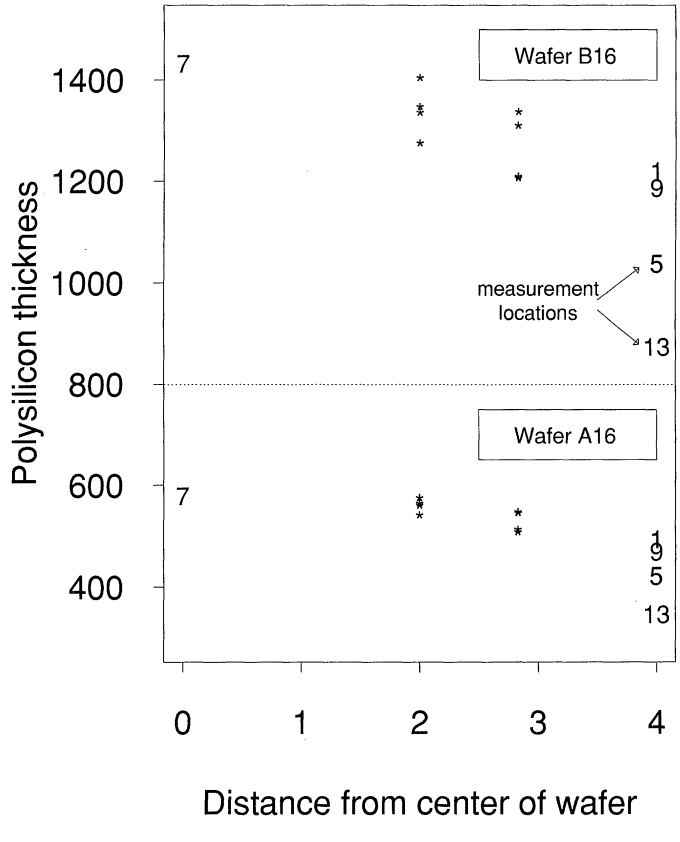

Figure 3. Polysilicon Thickness Versus Distance of Measurement Site From the Center of the Wafer, With Locations Coded as in Figure 1, for Wafers A16 and B16.

A simple approach (Lin and Spanos 1990) is to use the observed responses to calculate

$$U_{\mathrm{SUM}}(\mathbf{x}) = \frac{\sqrt{\frac{1}{S-1}\sum_{k=1}^{S}[y(\mathbf{x},\mathbf{z}_k) - \overline{y}(\mathbf{x},\cdot)]^2}}{\overline{y}(\mathbf{x},\cdot)} \times 100,$$

where $\overline{y}(\mathbf{x},\cdot) = (1/S)\sum_{k=1}^{S} y(\mathbf{x},\mathbf{z}_k)$ is the average observed thickness on a single wafer. This provides a single summary statistic for each wafer. These statistics are then regressed against the controllable $\mathbf{x}$ variables, and the optimum $\mathbf{x}$ is chosen to minimize the predicted $U_{\mathrm{SUM}}(\mathbf{x})$. We refer to this as summary statistic (SUM) modeling.

Realizing that modeling the uniformity metric may cause loss of information, create difficulty in modeling and interpretation, and cause violation of modeling assumptions (see Nair 1992), Guo and Sachs (1993) suggested obtaining models for the raw responses, one model for each measurement site on the wafer. That is, they obtained models

$$y(\mathbf{x},\mathbf{z}_k) = f_k(\mathbf{x}) + \varepsilon_k(\mathbf{x}), \qquad k = 1,2,\ldots,S, \quad (2)$$

where $f_k(\mathbf{x})$ is a known function of some unknown parameters and the errors $\varepsilon_k(\mathbf{x})$ are assumed to be independent zero-mean random variables with possibly unequal variances over k but constant variance over all $\mathbf{x}$ for a given k. The ordinary least squares predicted values from these models are then used to calculate the uniformity metric as

$$U_{\mathrm{MRS}}(\mathbf{x}) = \frac{\sqrt{\frac{1}{S-1}\sum_{k=1}^{S}[\widehat{f}_k(\mathbf{x}) - \bar{\widehat{f}}(\mathbf{x})]^2}}{\bar{\widehat{f}}(\mathbf{x})} \times 100,$$

where $\bar{\widehat{f}}(\mathbf{x}) = (1/S)\sum_{k=1}^{S}\widehat{f}_k(\mathbf{x})$ is the average predicted thickness on a single wafer. The optimum $\mathbf{x}$ is chosen to minimize $U_{\mathrm{MRS}}(\mathbf{x})$. We follow Guo and Sachs (1993) and refer to this as multiple response surface (MRS) modeling.

Although MRS modeling has advantages over SUM modeling, it also has several disadvantages. First, if the S models in (??) are similar, then a much more efficient overall model is obtained by combining all the data. Second, while different wafer sites are allowed to have different variances (since each model is site specific), the variances are assumed constant for different levels of $\mathbf{x}$. This is not always a reasonable assumption, as seen in the RTCVD process of Section 2. Third, by following Guo and Sachs' approach that obtains a separate model for each wafer site, the correlation between responses on a single wafer are ignored. As discussed in Section 1, this may lead to inefficient estimation of parameters and optimality criteria, and overstatement of importance of model terms. And fourth, the site-specific models of (??) depend greatly on sampling schemes. Poor sampling plans can produce information not representative of the wafer as a whole.

The two-step process of first optimizing for uniformity and then adjusting for target is often impractical, because the same covariates may affect both uniformity and thickness. For this reason, some combined measure of uniformity and target achievement is often needed to simultaneously optimize for these goals. The mean squared error (MSE) is commonly recommended for this purpose (Nair 1992). The MSE can be used with either the SUM or the MRS modeling approach. We simply replace U_{SUM} and U_{MRS} with

$$\mathrm{MSE}_{\mathrm{SUM}}(\mathbf{x}) = \frac{1}{S-1}\sum_{k=1}^{S}[y(\mathbf{x},\mathbf{z}_k) - \overline{y}(\mathbf{x},\cdot)]^2 + [\overline{y}(\mathbf{x},\cdot) - T]^2 \quad (3)$$

and

$$\mathrm{MSE}_{\mathrm{MRS}}(\mathbf{x}) = \frac{1}{S-1}\sum_{k=1}^{S}[\widehat{f}_k(\mathbf{x}) - \bar{\widehat{f}}(\mathbf{x})]^2 + [\bar{\widehat{f}}(\mathbf{x}) - T]^2, \quad (4)$$

where T is the target thickness. The disadvantages associated with SUM and MRS modeling, as described in the previous paragraph, remain.

We now suggest a flexible modeling strategy that is less dependent on the sampling scheme and is able to detect and estimate heterogeneity and correlation patterns. We apply (??) directly to the raw responses to get a mean model, and at the same time we model the variance and correlation structures through the error component. Suppose that $\varepsilon(\mathbf{x},\mathbf{Z})$ has mean vector $\mathbf{0}$, and variance-covariance matrix $\mathbf{V}_{\mathbf{X}} = \mathbf{\Lambda}_{\mathbf{X}}\mathbf{R}\mathbf{\Lambda}_{\mathbf{X}}$, and that $\varepsilon(\mathbf{x}_i,\mathbf{Z})$ and $\varepsilon(\mathbf{x}_j,\mathbf{Z})$ are independent for all $i \neq j$. The (k,l) entry of $\mathbf{R}$ is the correlation between $\varepsilon(\mathbf{x},\mathbf{z}_k)$ and $\varepsilon(\mathbf{x},\mathbf{z}_l)$. $\mathbf{\Lambda}_{\mathbf{X}}$ is the diagonal matrix containing the standard deviations of $\varepsilon(\mathbf{x},\mathbf{z}_k)$, for $k = 1,\ldots,S$. This modeling approach uses a single combined model to predict at all sites on the wafer, allows changing variances across $\mathbf{z}$ (sites in the example of Sec. 2) and across $\mathbf{x}$ (levels of controllable variables), and accounts for dependence among measurements on a single wafer, while assuming independence from one wafer to the next.

If in addition we assume a linear form for the mean function $\mathbf{f}(\mathbf{x},\mathbf{Z})$ and assume normality for the stochastic $\varepsilon(\mathbf{x},\mathbf{Z})$, then (??) becomes the well-known multivariate normal linear model, and for this reason we refer to this as the MVN method. Simultaneous modeling of mean, variance, and correlation is quite popular. It is used, for example, in the analysis of repeated measurements (Diggle, Liang, and Zeger 1994), and in spatial data analysis (Cressie 1991). In robust parameter design, where the goal is to determine levels of controllable variables to minimize the effect of noise variables, it is important to model the mean and the variability of the response over different levels of the noise variables.

Estimation without using distributional information such as normality is often done using estimated generalized least squares, generalized estimating equations, or some other iterative procedure. These methods all require a fair amount of replication to estimate the variances. For example, if repeated measurements for a single individual are correlated but all have the same variance, then they may be combined to form an estimate of the common variance, while still allowing estimation of the correlation. The same sort of approach is taken in robust parameter design. A sum-

mary statistic $s_{\mathbf{x}}^2 = [1/(S-1)]\sum_{k=1}^{S}[y(\mathbf{x},\mathbf{z}_k) - \overline{y}(\mathbf{x},\cdot)]^2$ is formed, and a model, is obtained for $\log(s_{\mathbf{x}}^2)$ as a function of $\mathbf{x}$.

These techniques are, unfortunately, not applicable to the example of Section 2. One reason is that the responses across a single wafer are heterogeneous, and so a variability estimate based on combining these responses would be meaningless. Moreover, we have no true replicates, and at most three quasi-replicates, for each design point; this is not enough to provide a reasonable estimate of variance. Without replication for estimating variance, we rely on physically motivated models and distributional assumptions to complete the estimation process. Maximum likelihood or any of its variants such as restricted maximum likelihood (REML) (Diggle et al. 1994) is one way of incorporating distributional assumptions to allow the modeling of variances in the absence of replicates. In Section 4 we apply REML estimation of model (**??**) to the RTCVD data of Section 2.

4. APPLICATION TO RTCVD PROCESS

The RTCVD process described in Section 2 has two controllable variables (oxide thickness and deposition time), 22 wafers, and 13 measurement sites on each wafer. The fitting criterion used here is REML.

More specifically, consider the multivariate normal linear model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{5}$$

where $\mathbf{y} = (\mathbf{y}'(\mathbf{x}_1,\mathbf{Z}),\mathbf{y}'(\mathbf{x}_2,\mathbf{Z}),\ldots,\mathbf{y}'(\mathbf{x}_{22},\mathbf{Z}))'$ is the vector of all 286 (13 measurements for each of 22 wafers) responses, $\mathbf{X}$ is made up of $\mathbf{x}_1,\ldots,\mathbf{x}_{22},\mathbf{Z}$ and is the $286 \times p$ matrix of across- and within-wafer covariates, $\boldsymbol{\beta}$ is the unknown p-vector of mean parameters, and $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}'(\mathbf{x}_1,\mathbf{Z}),\boldsymbol{\varepsilon}'(\mathbf{x}_2,\mathbf{Z}),\ldots,\boldsymbol{\varepsilon}'(\mathbf{x}_{22},\mathbf{Z}))'$ is multivariate normal with expectation $\mathbf{0}$ and block-diagonal variance–covariance matrix $\mathbf{V} = \operatorname{diag}(\mathbf{V}_{\mathbf{x}_1},\mathbf{V}_{\mathbf{x}_2},\ldots,\mathbf{V}_{\mathbf{x}_{22}})$. In addition, for the variance–covariance matrix of the ith wafer we allow $\mathbf{V}_{\mathbf{x}_i} = \boldsymbol{\Lambda}_{\mathbf{x}_i}\mathbf{R}\boldsymbol{\Lambda}_{\mathbf{x}_i}$, where $\boldsymbol{\Lambda}_{\mathbf{x}_i}$ is a 13×13 diagonal matrix of standard deviations of $\mathbf{y}(\mathbf{x}_i,\mathbf{Z})$ and is a function of variance parameters $\boldsymbol{\theta}$, whereas $\mathbf{R}$ is a 13×13 correlation matrix that is a function of correlation parameters $\boldsymbol{\delta}$. Hence the overall variance matrix is a function of $\boldsymbol{\theta}$ and $\boldsymbol{\delta}$, so we could write $\mathbf{V} = \mathbf{V}(\boldsymbol{\theta},\boldsymbol{\delta})$, although for simplicity we use the notation $\mathbf{V}$ throughout the article. The block-diagonal structure of $\mathbf{V}$ is based on the belief that responses obtained from different wafers are independent of each other. We allow correlation only within a wafer, and this correlation pattern, as represented by $\mathbf{R}$, is assumed to be the same for all wafers.

Given $\boldsymbol{\theta}$ and $\boldsymbol{\delta}$, the maximum likelihood estimator (MLE) of $\boldsymbol{\beta}$ is known to be

$$\widehat{\boldsymbol{\beta}}(\mathbf{V}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}.$$

Substituting this into the likelihood function for (**??**), we obtain the profile log-likelihood (as a function of $\boldsymbol{\theta}$ and $\boldsymbol{\delta}$ only, and omitting the unnecessary constant)

$$\ln \mathcal{L}_p(\boldsymbol{\theta},\boldsymbol{\delta}) = -\sum_{i=1}^{22}\ln|\boldsymbol{\Lambda}_{\mathbf{x}_i}| - 11\ln|\mathbf{R}| - \frac{1}{2}(\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\beta}})'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\beta}})'.$$

The corresponding REML objective function is given by

$$f_{\mathrm{REML}}(\boldsymbol{\theta},\boldsymbol{\delta}) = -2\ln\mathcal{L}_p(\boldsymbol{\theta},\boldsymbol{\delta}) + \ln|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|,$$

and the REML estimates are obtained by minimizing $f_{\mathrm{REML}}(\boldsymbol{\theta},\boldsymbol{\delta})$ with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\delta}$.

4.1 Model Fitting

Our approach to the model-fitting process is to start with the simplest model, perform a variety of residual diagnostics to determine whether a more complicated model might be helpful, fit this more complicated model, and then test for significant improvement in the fit.

For our simplest case, case 1, we assume independence ($\mathbf{R} = \mathbf{I}$) and homogeneity of all responses ($\boldsymbol{\Lambda}_{\mathbf{x}_i} = \sigma^2\mathbf{I}$ for all $i = 1,\ldots,22$). Under these assumptions, the fits are easily obtained from any statistical package, and we find 18 variables to be important predictors of the mean response. These variables are shown in Table 1. The resulting value of $f_{\mathrm{REML}}(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\delta}})$ is 2,150, based on 19 estimated parameters.

Regression diagnostics from the case 1 fit (not shown here) clearly show signs of heterogeneity. The residuals from measurement site 13, for example, are far more variable than residuals from other sites. This is indeed what

Table 1. RTCVD Data Predictor Variables for the Mean and Standard Deviation Models, and Corresponding Case 5 Parameter Estimates With Their Estimated Standard Deviations

	Predictor Variables																	
	1	*ox*[a]	*ti*[b]	*ox* ∗ *ti*	*ox*^{2}	*ti*^{2}	*ox*^{3}	*d*[c]	*ca*[d]	*sa*[e]	*d* ∗ *ca*	*d* ∗ *sa*	*d*^{2}	*d*^{3}	*ox* ∗ *d*	*ti* ∗ *d*	*ox* ∗ *ca*	*ti* ∗ *ca*
$\widehat{\boldsymbol{\beta}}$	930.0	−147.2	292.0	−74.0	70.1	−73.3	−98.6	−79.2	−25.9	4.0	−29.5	−.2	−30.7	−6.4	22.7	−11.0	10.6	−6.9
$\widehat{\sigma(\widehat{\boldsymbol{\beta}})}$	6.6	7.7	3.1	6.1	9.3	7.0	11.6	5.3	1.5	1.1	2.4	1.9	2.2	4.5	3.8	2.8	2.1	1.5
$\widehat{\boldsymbol{\theta}}$	2.85	−.56				−.17	−.39				.68		.05	.77	.03	.21		
$\widehat{\sigma(\widehat{\boldsymbol{\theta}})}$	.11	.20				.15	.28				.09		.10	.10	.11	.10		

[a] Coded oxide thickness, [−1 = 458, 1 = 1498 angstroms]
[b] Coded deposition time, [−1 = 18, 1 = 52 seconds]
[c] Coded distance from the wafer center, [−1 = 0, 1 = 4]
[d] Cosin of angle from vertical, cos(π, 5π/4, π, 3π/4, 3π/2, 3π/2, π, π/2, π/2, 7π/4, 2π, π/4, 2π)
[e] Sine of angle from vertical, sin(π, 5π/4, π, 3π/4, 3π/2, 3π/2, π, π/2, π/2, 7π/4, 2π, π/4, 2π)

we expect to find, so we add a standard deviation model and call this enhanced version case 2. In case 2 we maintain independence ($\mathbf{R} = \mathbf{I}$) but model $\mathbf{\Lambda}_{\mathbf{x}_i}$ as a function of across- and within-wafer covariates to allow heterogeneity across and/or within wafers.

The regression diagnostics from case 1 suggest predictors for the standard deviation model, but to avoid omitting or overlooking important predictors, we use a log-linear model based on *all* 18 predictors from Table 1. That is, we model

$$\log(\text{diag}(\mathbf{\Lambda}_{\mathbf{x}_1}, \mathbf{\Lambda}_{\mathbf{x}_2}, \ldots, \mathbf{\Lambda}_{\mathbf{x}_{22}})) = \mathbf{X}\boldsymbol{\theta},$$

where $\mathbf{X}$ is also the design matrix for the mean model. The optimization for this (and all future cases) is coded in S-PLUS using starting values obtained from the previous case. The resulting $f_{\text{REML}}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\delta}})$ is 1,897, based on 36 estimated parameters. This fit used the same 18 predictor variables for the mean model as did case 1, although some of these variables were no longer statistically significant. We keep the larger mean model to assure nesting of the cases for lack-of-fit testing.

A likelihood ratio test (LRT) of case 1 (independent and homogeneous) versus case 2 (independent and heterogeneous) compares the observed reduction in $f_{\text{REML}}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\delta}})$ of 153 to the chi-squared distribution with 17 df. The p value below 10^{-16} strongly supports heterogeneity.

A careful look at the residuals (not shown here) from case 1 suggests that perhaps only nine predictors are needed for the standard deviation model. Our case 3 fits this simpler standard deviation model (see Table 1 for the predictors used; they are the same as for cases 4 and 5 discussed later) to get $f_{\text{REML}}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\delta}})$ of 1,905, based on 27 estimated parameters. The LRT of case 3 versus case 2 has a p value of .5341, suggesting that the smaller standard deviation model adequately describes the heterogeneity of the responses. We keep this standard deviation model for all future cases, even if some terms lose their statistical significance.

Our attention now turns to correlation among the responses. Based on plots of the raw data and residuals from case 3, as well as findings in the chemical vapor deposition literature, we expect strong correlation within wafers but independence across wafers. To capture the within-wafer correlation, we apply the commonly used exponential decay correlation model, $\mathbf{R}_{kl} = \exp(-\delta_1 h_{kl})$, where h_{kl} is the Euclidean distance between measurement sites k and l. This correlation model allows decreasing correlation as measurement sites become further apart and does not pay attention to angular or radial orientation from the wafer center. Case 4 fits this correlation model, the standard deviation based on nine predictors, and the mean model based on 18 predictors. The resulting value of $f_{\text{REML}}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\delta}})$ is 1,816, based on 28 estimated parameters. The p value below 10^{-16} for the LRT of case 3 (heterogeneous but independent) versus case 4 (heterogeneous and exponential correlation) strongly supports within-wafer dependence.

The determination of an appropriate form for the within-wafer correlation leads us to case 5. Based on thermal nonuniformity patterns reported in the chemical vapor deposition literature, a stationary correlation model such as the exponential model given earlier seems inappropriate. The description given in Section 2 for the RTCVD process points to three requirements: (a) the correlation between responses at two measurement sites decreases as the distance between the sites increases, (b) the correlation between two sites depends on whether the sites fall in the same radial (distance) band from the center of the wafer, and (c) sites within radial bands close to the center of the wafer are more correlated than sites within radial bands far from the center.

The first requirement could be met by any of a number of widely used isotropic correlation functions (Cressie 1991; Handcock and Wallis 1994; Mardia and Marshall 1984), including the exponential correlation function used in case 4. These isotropic functions model the correlation between responses at two sites (k and l) as a function of only the length of the distance vector between the sites. The second and third requirements, on the other hand, disallow consideration of all isotropic or even second-order stationary correlation functions. The individual distances from the two sites to the center of the wafer must be included in the correlation function. We propose the following alternative correlation model:

$$\mathbf{R}_{kl} = \exp\{-\delta_1 h_{kl} \exp[\delta_2 |h_{k7} - h_{l7} + \delta_3 \min(h_{k7}, h_{l7})]\}, \tag{6}$$

where site 7 is the center of the wafer. The third requirement is met only if $\delta_3 \geq 0$, so that pairs of sites on the same radial band but at fixed distance apart become less correlated as the bands get further from the center. Necessary, but not sufficient, conditions for positive semidefiniteness include $\delta_1 \geq 0, \delta_2 \geq 0$ (see the Appendix). Setting $\delta_2 = \delta_3 = 0$, we get the usual exponential decay correlation model used in case 4.

Fitting the models of case 5 results in $f_{\text{REML}}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\delta}}) =$ 1,793, based on 30 estimated parameters. The p value of .00001 for LRT of case 4 (heterogeneous and exponential correlation) versus case 5 [heterogeneous and correlation model (**??**)] supports the nonstationary correlation form given in (**??**). We feel that case 5 represents well all the salient features of the data in statistically significant ways above and beyond cases 1–4. We now give some details on the case 5 fit.

The sum of squared standardized residuals is 268.1, indicating a very good fit over the 286 observed responses. The average of estimated standard deviations of the responses is 28.02, with a minimum standard deviation of 5.95 and a maximum standard deviation of 194.02. The maximum standard deviation is always associated with measurement site 13. The eigenvalues of $\widehat{\mathbf{R}}$ are 5.22, 1.34, 1.22, 1.22, .92, .88, .66, .66, .39, .23, .13, .06, and .06, which are all positive, indicating that $\widehat{\mathbf{R}}$ is positive definite. The standardized residuals show no pattern of dependence on either the process or spatial variables, and indeed appear to be a sample from a Gaussian distribution.

The estimate of $\boldsymbol{\beta}$, and associated standard deviations obtained from $(\mathbf{X}'\widehat{\mathbf{V}}^{-1}\mathbf{X})^{-1}$, are displayed in Table 1. The estimate of $\boldsymbol{\theta}$, along with parametric bootstrap estimates of

the associated standard deviations, are also given in Table 1. (For the parametric bootstrap, the pseudovariates are generated according to (??), using the observed matrix $\mathbf{X}$ and the case 5 estimates of $\boldsymbol{\beta}$, $\boldsymbol{\theta}$, and $\boldsymbol{\delta}$. We generated 100 samples, each containing 286 observations, and obtained case 5 fits for each sample. The bootstrap variances are based on these 100 replicates of the estimates.) The fitted correlation model is

$$\widehat{\mathbf{R}}_{kl} = \exp\{ \underset{(.008)}{-.024}\ h_{kl} \exp[\underset{(.115)}{1.054}\ |h_{k7} - h_{l7}| + \underset{(.156)}{.720}\ \min(h_{k7}, h_{l7})]\}, \quad (7)$$

where the bootstrap estimates of standard deviation are shown in parentheses below (??).

These fits capture several features of the data. As the oxide layer thickens, the amount of deposited polysilicon decreases, and the process repeatability improves. On the other hand, as deposition time increases, the amount of deposited polysilicon also increases, and the process becomes less repeatable. There is also an oxide–time interaction that affects the amount of deposited polysilicon. However, this interaction is not strong enough to alter the qualitative statements made earlier, provided that the values of (ox, ti) remain in the $[-1, 1]^2$ square design space.

Next, both the amount of deposited polysilicon and the repeatability of the process are affected by location of measurement site on the wafer. Polysilicon deposition decreases with movement either away from the wafer center or toward the straight edge of the wafer, with a greater rate of decrease if both conditions are met. Process repeatability also worsens with movement out from the wafer center and/or toward the straight edge. Much of this is unaffected by choice of oxide layer or deposition time and can be remedied only by technological advances. The existence of significant interactions between (ox, ti) and (d, ca) in both the mean and standard deviation models does, however, allow uniformity and equality of response variance within a wafer to be affected to some degree by choice of oxide thickness and deposition time.

Finally, the fitted correlation model supports many of the conjectured thermal uniformity patterns found in the chemical vapor deposition literature. In addition, by recognizing and fitting this correlation we have obtained more efficient estimators of the mean-model parameters and better estimates of variances.

4.2 Process Optimization

Using the fitted models obtained for the RTCVD data, we can separately determine the optimum settings of ox and ti for achieving the desired uniformity, target thickness, and repeatability. These settings are $(ox, ti) = (1, -1)$ for uniformity (minimize $(1/12)\sum_{k=1}^{13}[\widehat{y}(\mathbf{x}, \mathbf{z}_k) - \overline{\widehat{y}}(\mathbf{x}, \cdot)]^2$), (ox, ti) on the curve $ti = 1.97 - .50ox + 1.36\sqrt{1.25 - 2.09ox + .65ox^2 - .72ox^3}$ with $-1 \le ox \le .5$ for target thickness (minimize $[\overline{\widehat{y}}(\mathbf{x}, \cdot) - T]^2$), and $(ox, ti) = (1, -1)$ for repeatability [minimize $(1/13)\sum_{k=1}^{13}\widehat{\sigma^2}(\mathbf{x}, \mathbf{z}_k)$]. Because the settings for uniformity and repeatability are exactly the same, it is reasonable to believe that almost any modeling approach would lead to about the same suggested optimum processing condition, and indeed this is the case.

The SUM modeling approach, where $\mathrm{MSE}_{\mathrm{SUM}}$ is regressed against controllable inputs, gives unacceptable negative predicted values for $\mathrm{MSE}_{\mathrm{SUM}}$. As a result, we instead regress $\log(\mathrm{MSE}_{\mathrm{SUM}})$ against the controllable inputs. The controllable inputs we use are the first seven predictor variables in Table 1. The R^2 for this fit is .98. The predicted $\log(\mathrm{MSE}_{\mathrm{SUM}})$ is then minimized at $(ox, ti) = (.622, 1.000)$.

In the MRS approach, all predicted thicknesses are positive and $\mathrm{MSE}_{\mathrm{MRS}}$ is minimized at $(ox, ti) = (.569, .998)$. The controllable inputs used in the regressions are the same as for the SUM modeling, and the R^2 for the 13 regressions range from .95 to .998.

To allow better comparison to the SUM and MRS approaches, in the MVN approach we select optimum conditions to minimize a MSE. The predictions from the mean model in (??) are used in the same way as the predictions from the MRS approach to calculate

$$\mathrm{MSE}_{\mathrm{UT}}(\mathbf{x}) = \frac{1}{12}\sum_{k=1}^{13}[\widehat{y}(\mathbf{x}, \mathbf{z}_k) - \overline{\widehat{y}}(\mathbf{x}, \cdot)]^2 + [\overline{\widehat{y}}(\mathbf{x}, \cdot) - T]^2, \quad (8)$$

where $\widehat{y}(\mathbf{x}, \mathbf{z}_k)$ is the predicted thickness at measurement site k for a wafer processed using conditions $\mathbf{x}$, $\overline{\widehat{y}}(\mathbf{x}, \cdot) = (1/13)\sum_{k=1}^{13}\widehat{y}(\mathbf{x}, \mathbf{z}_k)$ is the average prediction for the wafer, and $T = 1{,}000$ is the target thickness. This $\mathrm{MSE}_{\mathrm{UT}}$ is minimized at $(ox, ti) = (.495, .947)$.

The optimum conditions determined by the three methods SUM, MRS, and MVN are very similar. So then, what advantage does MVN provide? We address this question in the next section.

5. A SIMULATION

We used the RTCVD data to illustrate the SUM, MRS, and MVN methods. The findings are that two of the three processing regions that separately optimize for uniformity, target achievement, and repeatability are so close to each other that any method capable of isolating just one of these regions will give very good overall results.

However, our primary interest lies in comparing the methods when we know that the processing regions that separately optimize for uniformity, target achievement, and repeatability are far apart. To investigate this, we perform a simulation exercise for which the true optimum processing regions are $(ox, ti) = (-.75, .75)$ for uniformity, $(.75, .75)$ for achieving a target of 1,000, and $(0, -.75)$ for repeatability. The models that we use are

$$E[y(ox, ti, \mathbf{z}_k)] = 1{,}000 + 100(ox - 1.3)^2 + 100(ti - .75)^2 + 100d_k(ox + .75)^2 + 300ca_k(ti - .75)^2, \quad (9)$$

and

$$\log\{\mathrm{SD}[y(ox, ti, \mathbf{z}_k)]\} = 5 + .4ox^2 + .4(ti + .75)^2, \quad (10)$$

where d_k and ca_k are the coded site-to-center distance and cosine of angle from vertical of measurement site $\mathbf{z}_k$, for $k = 1, 2, \ldots, 13$, as in the RTCVD process. In addition to assuming that the responses are normally distributed, for simplicity we also assume that all of the responses are independent.

We used the values observed for ox and ti in the RTCVD process in (9) and (10) to simulate 13 responses for each of 22 wafers, to yield a sample of size 286. We obtained 1,000 of these Monte Carlo samples and applied the three methods SUM, MRS, and MVN (where MSE_{UT} is minimized) to determine optimum conditions. We also considered a fourth method: model according to MVN, but minimize

$$\text{MSE}_{\text{UTR}(\mathbf{x})} = \frac{1}{12}\sum_{k=1}^{13}[\hat{y}(\mathbf{x}, \mathbf{z}_k) - \bar{\hat{y}}(\mathbf{x}, \cdot)]^2 + [\bar{\hat{y}}(\mathbf{x}, \cdot) - T]^2 + \frac{1}{13}\sum_{k=1}^{13}\widehat{\sigma^2}(\mathbf{x}, \mathbf{z}_k), \quad (11)$$

where $\widehat{\sigma^2}(\mathbf{x}, \mathbf{z}_k)$ is the predicted variance of the thickness at measurement site k for a wafer processed using conditions $\mathbf{x}$. MSE_{UTR} finds the optimum condition to simultaneously satisfy all of uniformity, target achievement, and repeatability and is meaningful only with the MVN modeling approach. The SUM approach offers a common estimate of $\text{var}(y(\mathbf{x}, \mathbf{z}_k))$, irrespective of the values of $\mathbf{x}$ and k, which, of course, has no effect on minimizing $\text{MSE}_{\text{UTR}}(\mathbf{x})$. The MRS approach offers different estimators for different k's, but not for different $\mathbf{x}$'s; again, this has no effect on minimizing $\text{MSE}_{\text{UTR}}(\mathbf{x})$. We expect that if the region of repeatability is far from the regions of uniformity and target achievement, then MSE_{UTR} and MSE_{UT} will give different optima.

Based on simulation models (9) and (10) *without* random variation, and using the criterion function MSE_{UT} to simultaneously optimize for both uniformity and target achievement, we determine an optimum processing region of $(ox, ti) = (.43, .75)$, which we will refer to as UT. Similarly, using the criterion function MSE_{UTR} to simultaneously optimize for uniformity, target achievement, and repeatability, we determine an optimum processing region of $(ox, ti) = (.29, .19)$, which we will refer to as UTR. These optima do not fall in the middle of the individual optima because of the shapes of the different criteria functions. The MVN approach where MSE_{UT} is minimized is designed to estimate UT. The MVN approach where MSE_{UTR} is minimized is designed to estimate UTR.

The simulation results are displayed in Figure 4 as scatterplots of the optima obtained from the four methods. Reference lines indicate UT and UTR. The MVN with MSE_{UT} method clearly pays no attention to repeatability. The MVN with MSE4_{UTR} does a very good job of estimating UTR. Somewhat surprising is the fact that SUM comes very close to estimating UTR, whereas MRS estimates neither UT nor UTR. SUM and MRS appear to apply unequal weight to the three components in (11). It is also important to note that there is far less spread in the regions determined by MVN; that is, using the MVN modeling gives increased efficiency for determining optimum conditions.

6. DISCUSSION

This article recommends that multivariate modeling techniques be used when modeling the multivariate responses obtained from semiconductor deposition processes. In striving for simplicity, the traditional approaches make unreasonable a priori assumptions of homogeneity and independence of responses. A full multivariate treatment, on the other hand, is very computationally achievable, does not assume homogeneity or independence, and can provide some useful information. First, it provides models (mean, variance, and correlation) useful for understanding the process being studied. These models may be useful in contexts other than process optimization. Second, because the modeling is done independently of the optimization, there is flexibility in choosing an appropriate optimality criterion. For example, if repeatability is deemed more important than target achievement, then weights may be applied to the different components of (11) to reflect this imbalance. Flexibility in choosing a criterion is very important because, as demonstrated in Section 5, the selected processing conditions can vary considerably depending on the criterion used. Third, if a partial multivariate approach such as MRS is taken, then it is unclear what the process is being optimized to achieve, as illustrated in the simulation exercise. Fourth, the selected conditions will not vary widely as a result of adding random variation to the observed sample. This increased efficiency gives comfort that the true optimum processing condition is close to the one selected.

As a direct consequence of the RTCVD application, we have also suggested a nonstationary correlation structure that includes the exponential decay model as a special case. Although designed for the RTCVD process, this correlation model may be applicable to situations for which it is believed that the dependence structure on an experimental unit has a bull's eye pattern. Properties of this nonstationary correlation structure are under investigation.

Finally, while performing the model fitting steps for the RTCVD process, we made several observations that the reader may find useful in personal implementations of these or similar models. These are ideas that may have appeared elsewhere but are worth repeating. Parsimony in the variance model is essential. If the variance model is overparameterized, one may end up with large variances associated with very poor fitted values. The standardized residual will not show this to be a problem and will in fact yield a very small sum of squared standardized residuals. This might mislead one into believing that a good fit has been obtained. Likelihood fits (ML or REML) are supposed to correct these problems because $\ln|\mathbf{V}|$ is a part of the minimizing criterion function. If independence is assumed, then minimizing $\ln|\mathbf{V}|$ controls how large the variances can be, and so problems such as the one just mentioned will not occur. But if the observations are allowed to be dependent, then $\ln|\mathbf{V}|$ can be very small even when the variances are very large. This is because if the correlation matrix $\mathbf{R}$ is close to singular (but still positive definite) — for example,

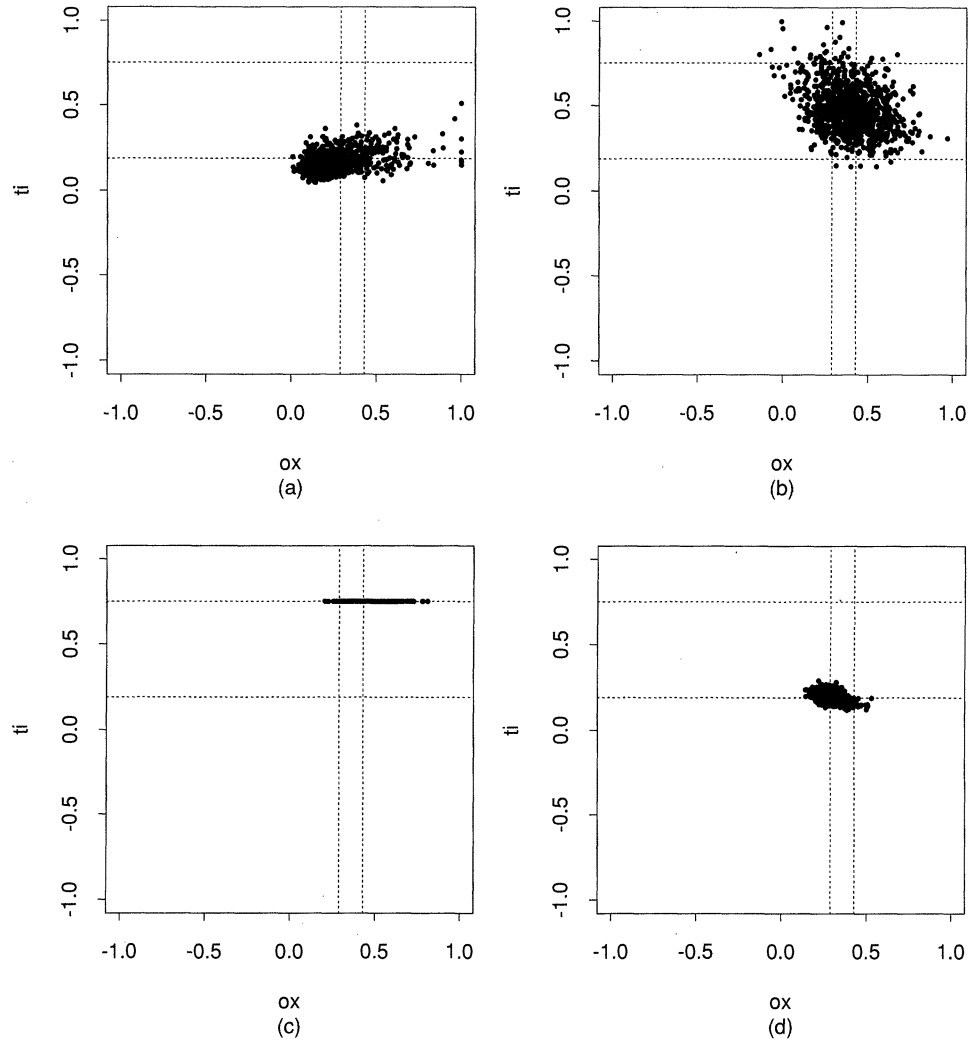

Figure 4. Optimum Processing Conditions Obtained From the (a) SUM, (b) MRS, (c) MVN With MSE_{UT}, and (d) MVN with MSE_{UTR} Methods. The data are simulated according to (9) and (10), such that (ox, ti) selected from MVN with MSE_{UT} estimates UT = (.43, .75) and (ox, ti) selected from MVN with MSE_{UTR} estimates UTR = (.29, .19). Reference lines indicate UT and UTR.

all correlations close to 1 — then $\mathbf{V}$ is also very close to singular (but still positive definite), and thus $\ln|\mathbf{V}|$ is very small. Because ML and REML estimation can reward singularity of $\mathbf{V}$, it is important to check that $\mathbf{V}$ is positive definite. The following matrices must also be positive definite: $\mathbf{\Lambda}$, $\mathbf{R}$, and $\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}$.

APPENDIX: NECESSARY CONDITIONS FOR POSITIVE SEMIDEFINITENESS

If $R(\mathbf{z}_k, \mathbf{z}_l) \equiv R_{kl}$, given by equation (6), is a positive semidefinite kernel, then any matrix obtained by evaluating this kernel at a given set of sites in R^2 must be positive semidefinite (Stewart 1976). Consider the three sites $(0,0)$ (the center), $(\lambda_1, 0)$, and $(\lambda_1 + \lambda_2, 0)$, where $0 < \lambda_1, \lambda_2$. Then $R((0,0),(\lambda_1,0)) = \rho_{01} = \exp\{-\delta_1\lambda_1 \exp[\delta_2\lambda_1]\}$, $R((0,0),(\lambda_1+\lambda_2,0)) = \rho_{02} = \exp\{-\delta_1(\lambda_1+\lambda_2)\exp[\delta_2(\lambda_1+\lambda_2)]\}$, and $R((\lambda_1,0),(\lambda_1+\lambda_2,0)) = \rho_{12} = \exp\{-\delta_1\lambda_2\exp[\delta_2\lambda_2 + \delta_3\lambda_1]\}$. The resulting matrix,

$$C = \begin{bmatrix} 1 & \rho_{01} & \rho_{02} \\ \rho_{01} & 1 & \rho_{12} \\ \rho_{02} & \rho_{12} & 1 \end{bmatrix},$$

is positive semidefinite if and only if $\rho_{01} \le 1$, $\rho_{12} \le 1$, and $\det \mathbf{C} = 1 - \rho_{01}^2 - \rho_{02}^2 - \rho_{12}^2 + 2\rho_{01}\rho_{02}\rho_{12} \ge 0$ for all $\lambda_1, \lambda_2 > 0$. The first two conditions require $\delta_1 \ge 0$, irrespective of the values of δ_2 and δ_3. But, if $\delta_2 < 0$, λ_1 is held fixed, and $\lambda_2 \to \infty$, then $\rho_{02} \to 1$ and $\rho_{12} \to 1$, so that $\det \mathbf{C} \to -(\rho_{01} - 1)^2$, which does not meet the necessary condition of being nonnegative. Hence the third condition requires $\delta_2 \ge 0$.

[Received August 1995. Revised September 1997.]

REFERENCES

Bartelink, D. J. (1994), "Statistical Metrology: At the Root of Manufacturing Control," *Journal of Vacuum Science Technology* Ser. B, 12, 2785–2794.

Cressie, N. A. C. (1991), *Statistics for Spatial Data*, New York: Wiley.

Davis, J. C., Gyurcsik, R. S., and Lu, J. C. (1993), "Application of Semi-Empirical Model Building to the RTCVD of Polysilicon," *in Statistics in the Semiconductor Industry: Case Studies of Process/Equipment Characterization*, Vol. 2, eds. V. Czitrom and P. Spagon, Austin, TX: Sematech, pp. 202–214.

Diggle, P. J., Liang, K., and Zeger, S. L. (1994), *Analysis of Longitudinal Data*, London: Oxford University Press.

Guo, R., and Sachs, E. (1993), "Modeling, Optimization and Control of Spatial Uniformity in Manufacturing Processes," *IEEE Transactions on Semiconductor Manufacturing*, 6, 41–57.

Handcock, M. S., and Wallis, J. R. (1994), "An Approach to Statistical Spatial-temporal Modeling of Meteorological Fields," *Journal of the American Statistical Association*, 89, 368–378.

Lin, K. K., and Spanos, C. J. (1990), "Statistical Equipment Modeling for VLSI Manufacturing: An Application to LPCVD," *IEEE Transactions. on Semiconductor Manufacturing*, 3, 216–229.

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

Nair, V. J. (1992), "Taguchi's Parameter Design: A Panel Discussion," *Technometrics*, 34, 127–161.

Ozturk, M. C., Sorrell, F. Y., and Wortman, J. J. (1991), "Manufacturability Issues in Rapid Thermal Chemical Vapor Deposition," *Transactions on Semiconductor Manufacturing*, 4, 155–165.

Perkins, R. H., Riley, T. J., and Gyurcsik, R. S. (1995), "Thermal Uniformity and Stress Minimization During Rapid Thermal Processes," *Transactions on Semiconductor Manufacturing*, 8, 272–279.

Stewart, J. (1976), "Positive Definite Functions and Generalizations: An Historical Survey," *Rocky Mountain Journal of Mathematics*, 6, 409–434.