Optimal Parameter Selection for Electronic Packaging Using Sequential Computer Simulations

Abhishek Gupta
Department of Statistics,
The Wharton School,
University of Pennsylvania,
Philadelphia, PA 19104

Yu Ding
Department of Industrial and Systems
Engineering,
Texas A&M University,
3131 TAMU,
College Station, TX 77843-3131

Leon Xu

Tommi Reinikainen

Nokia Inc.,
6000 Connection Drive,
Irving, TX 75039

Optimal parameter selection is a crucial step in improving the quality of electronic packaging processes. Traditional approaches usually start with a set of physical experiments and then employ Design of Experiment (DOE) based response surface methodology (RSM) to find the parameter settings that will optimize a desired system response. Nowadays deterministic computer simulations such as Finite Element Analysis (FEA) are often used to replace physical experiments when evaluating a system response, e.g., the stress level in an electronic packaging. However, FEA simulations are usually computationally expensive due to their inherent complexity. In order to find the optimal parameters, it is not practical to use FEA simulations to calculate system responses over a large number of parameter combinations. Nor will it be effective to blindly use DOE-based response surface methodology to analyze the deterministic FEA outputs. In this paper, we will utilize a spatial statistical method (i.e., the Kriging model) for analyzing deterministic FEA outputs from an electronic packaging process. We suggest a sequential method when using the Kriging model to search for the optimal parameter values that minimize the stress level in the electronic packaging. Compared with the traditional RSM, our sequential parameter selection method entertains several advantages: it can remarkably reduce the total number of FEA simulations required for optimization, it makes the optimal solution insensitive to the choice of the initial simulation setting, and it can also depict the response surface and the associated uncertainty over the entire parameter space. [DOI: 10.1115/1.2193551]

1 Introduction

With the increasingly competitive business environment, there is a need to design and manufacture more technically complex products, with assured product reliability in shorter time than the current practice. In the mobile phones market, this translates into better versatility, portability, and visual and ergonomic appeal among other features. This in turn calls for continuous improvement in electronic packaging.

A potential problem with electronic packaging for mobile phones could be caused by the failure of the solder joints under thermal and mechanical loading. Figure 1 shows the bending process map, where the output response, denoted by y, is the von Mises stresses generated in the joints. It is considered as the key feature to assess solder reliability under bending. The input parameters, denoted by x's, include geometry and material properties associated with the solder joint. Hence, our objective is to find the optimal parameter setting of x that can minimize the stress level in solder joints in an electronic packaging.

The process response y can be generally expressed as $y=f(\mathbf{x})$, where $\mathbf{x}$ is a vector of the input parameters x's. We suppose the domain of y is χ that is a fixed subset of $\mathfrak{R}^d$ and $\mathbf{x} \in \chi$, where d is the number of input parameters under consideration. Given the complexity of the packaging process, it is impossible in most practical cases to express $f(\mathbf{x})$ in closed form. In fact, Finite Element Analysis (FEA) simulations have been widely used to numerically evaluate this $f(\mathbf{x})$ for electronic packaging mechanics with a lot of success [1–3]. These FEA simulations entertain certain advantages over running physical experiments such as being cost effective, easy to try out design alternatives, and thus have a greater impact on product design and manufacturing.

Here we consider a generic model of a second level packaging design (electronic package to board) under bending loading. Figure 2 shows the Chip Scale Package (CSP)-Printed Wiring Board (PWB) model used in this study. Engineers from our collaborating company identified four potentially important parameters related to the material properties of the PWB: PWB in-plane Young's modulus (x_1), component substrate in-plane Young's modulus (x_2), die attach Young's modulus (x_3), and molding compound Young's modulus (x_4).

Table 1 provides the allowable parameter regions. The stress conditions under bending load in the PWB are modeled and calculated using a commercial FEA software, *ANSYS* [4].

Once the FEA computer simulation is established, one may want to include the FEA simulation as a part of an optimization routine (such as any nonlinear optimization routine). That is, to find the optimal parameter settings to minimize the von Mises stresses in solder joints so that the electronic packaging reliability can be improved during a bending process. The difficulty is that due to the complexity of the FEA codes, it is usually computationally expensive to calculate $f_{\text{FEA}}(\mathbf{x})$ (i.e., an evaluation of $f(\mathbf{x})$ using FEA) over a large number of parameter combinations, which may be required to find the optimal parameter settings.

In light of this, people have suggested treating FEA simulations as physical experiments, and then, following the response surface methodology [5], find the optimal system response by using an economical number of experimental runs, or equivalently, FEA simulations. As we will show in the latter comparison, the experiment-based response surface methodology is not effective to handle the complex FEA simulations in several aspects. One key difference is that the model fitting methods need to be different for treating deterministic FEA outputs and random physical experimental data.

Figure 3 shows the difference in data analysis strategies for physical and computer experiments. In physical experiments, random errors are always present and hence the experimenters are
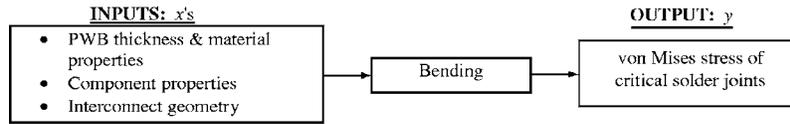
Fig. 1 Bending process map

encouraged to take replicates. The prediction from such replicated data is done through the fitting of trend lines, e.g., from linear regression. In computer experiments, linear regression can still be used to get a trend line. However, it does not hold a clear meaning in the absence of random errors. Recently spatial statistical methods like the Kriging model [6] have been used to analyze the deterministic computer data. The Kriging model is an interpolative model, that will fit a predictor to pass through all the observed points because there is no uncertainty involved in the observed value at a particular parametric level [refer to Fig. 3(b)]. On the other hand, we usually do not employ such an interpolative approach in the case of a physical experiment because each observation is associated with uncertainty and the interpolative approach will cause overfitting and thus poor prediction. The Kriging predictor, as it passes through all the observed points, is more complicated than a first or second order polynomial. For other combinations of parameter levels where an observation (i.e., a computer output) is not available, the predicted value will be obtained from interpolation using the Kriging model developed from the observed points.

As is the case with most modeling approaches, an accurate Kriging model over the entire parameter space may still require a large number of FEA simulations. In this paper, we will devise a sequential strategy to address the issue of how to reduce the number of FEA simulations when searching for the optimal parameter settings. The basic idea is as follows. We are oftentimes more concerned with a subregion or subregions of the parameter space that contain the optimal settings, which suggests that collecting data with the same density throughout the parameter space may be unnecessary. In the sequential strategy, we choose in the initial step to use a small number of FEA simulations, covering the whole parameter space. Subsequently we keep zooming into a smaller subregion of interest based on the predicted values and prediction uncertainty calculated from the previous steps. More FEA simulations will be conducted only for the small subregion so that we can have additional information to fit a more accurate model until a reasonably accurate optimum is found. The benefit of our sequential parameter selection for improving the quality of electronic packaging will be demonstrated by a comparison with a traditional response surface methodology.

Our approach and conclusions are contained in the next four sections. In Sec. 2 we present an overview of the spatial statistic modeling (Kriging) method for handling deterministic computer simulations. In Sec. 3 we present the details of the sequential strategy for optimal parameter selection in an electronic packaging process. In Sec. 4 we provide a comparison of the sequential Kriging strategy with the traditional RSM. Finally, in Sec. 5 we conclude the paper.
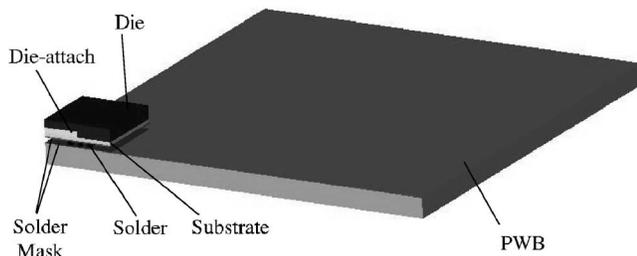
Fig. 2 The CSP-PWB model

2 Review of Kriging Method

Sacks et al. [6,7] and Welch et al. [8] first proposed the use of Kriging models for analyzing computer experiments or computer simulation data. The idea of Kriging model for computer simulations was extended from geostatistics [6,9]. Since we are interested in obtaining the response surface of function $f_{FEA}(\mathbf{x})$ based on a set of spatially dispersed simulation outputs, our case is very similar to the problem of establishing correlations among spatially distributed locations in geostatistics. Such spatial correlations are modeled using techniques of spatial statistics, which views the spatial response function as a realization of a random field. Assume that $y(\mathbf{x})$ is a response function of a d-dimensional vector of inputs $\mathbf{x}$ over domain χ. The notation $Y(\cdot)$ is used to distinguish the random function from its realizations $y(\cdot)$. $Y(\mathbf{x})$ is most commonly given by

$$Y(\mathbf{x}) = \mathbf{f}^T(\mathbf{x})\boldsymbol{\beta} + Z(\mathbf{x}) \tag{1}$$

where $\boldsymbol{\beta} = [\beta_1 \ \beta_2 \ \ldots \beta_k]^T$ is the vector of unknown regression coefficients, $\mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}) \ f_2(\mathbf{x}) \ldots f_k(\mathbf{x})]^T$ is a vector of known regression functions, and $Z(\mathbf{x})$ is a zero mean stationary Gaussian random field over χ. The intuition behind the model in Eq. (1) is that while the regression portion of the model approximates the response surface globally, the local deviations are captured by the $Z(\mathbf{x})$ component. In other words, $Z(\mathbf{x})$ tries to capture the systematic departure from the global regression part. For this goal, $Z(\mathbf{x})$ is assumed to have covariance,

$$\sigma^2 R(\mathbf{w}, \mathbf{x}) \tag{2}$$

between $Z(\mathbf{w})$ and $Z(\mathbf{x})$ at two vector-valued inputs $\mathbf{w}$ and $\mathbf{x}$ in χ, where σ^2 is the process variance and $R(\mathbf{w}, \mathbf{x})$ is the correlation function. A common assumption in the Kriging modeling approach is that $Z(\mathbf{x})$ is second-order stationary, i.e., the mean is constant and $\text{Cov}\{Z(\mathbf{x}_1), Z(\mathbf{x}_2)\}$ is independent of $\mathbf{x}$ and depends only on the distance $\mathbf{h}$. This is not a restrictive assumption because the nonstationary part of a system response, i.e., the global trend, has already been modeled by the regression terms in Eq. (1).

This covariance structure of a Kriging model is in contrast to its counterpart in the traditional RSM for physical experiments, where the covariance matrix is assumed to be $\sigma^2\mathbf{I}$. The $\sigma^2\mathbf{I}$ covariance matrix is introduced to represent the randomness associated with the replications at the same input. It, however, fails to capture the spatial correlation between two different inputs. Therefore, the new covariance structure of $Z(\mathbf{x})$ provides the Kriging predictor interpolative capability, as desired for analyzing deterministic computer simulations.

In order to use the Kriging predictor in a practical setting, researchers parametrize the correlation matrix. The most popular family of correlation models in the computer experiments literature is the power exponential correlation family [10]. The product of stationary one-dimensional correlations gives us

$$R(\mathbf{w}, \mathbf{x}) = \prod_{j=1}^{d} \exp(-\theta_j |w_j - x_j|^{p_j}) \tag{3}$$

where $\theta_j \geq 0$, $0 \leq p_j \leq 2$ and $\mathbf{w}, \mathbf{x} \in \chi$, with w_j being the jth component of $\mathbf{w}$ and x_j being the jth component of $\mathbf{x}$. In particular, we will use $p_j = 2$, because $R(\mathbf{w}, \mathbf{x})$ will then be infinitely differentiable at zero.

Table 1 Allowable parameter region for material properties (Unit: GPa)

	PWB in-plane Young's modulus (x_1)	Component substrate in-plane Young's modulus (x_2)	Die attach Young's modulus (x_3)	Molding compound Young's modulus (x_4)
High	35	36	5	30
Low	20	12	1	15

To realize a Kriging model, we need to estimate unknown parameters σ^2, $\boldsymbol{\beta}$, and $\boldsymbol{\theta}=(\theta_1,\ldots\theta_d)$ from the FEA simulation data (in our case we take $p_j=2$ for all $j's$).

We first introduce the following notations. The n parameter combinations used to compute the system response are denoted by $\mathbf{D}=\{\mathbf{x}_1,\cdots,\mathbf{x}_n\}$ and the observed response vector corresponding to $\mathbf{D}$ is $\mathbf{y}_D=[y(\mathbf{x}_1)\cdots y(\mathbf{x}_n)]^T$. Define $\mathbf{R}_D$ as the $n\times n$ matrix of correlations between Z's at the input parameter combinations, where $\mathbf{R}_{D(i,j)}=R(\mathbf{x}_i,\mathbf{x}_j)$ for $1\leq i,j,\leq n$. The $n\times k$ regressor matrix is denoted by $\mathbf{F}=[\mathbf{f}^T(\mathbf{x}_1)\cdots\mathbf{f}^T(\mathbf{x}_n)]^T$.

We will employ the maximum likelihood estimation method to estimate σ^2, $\boldsymbol{\beta}$, and $\boldsymbol{\theta}$. Under a normality assumption, the log-likelihood function of σ^2, $\boldsymbol{\beta}$, and $\boldsymbol{\theta}$ is

$$l(\boldsymbol{\beta},\sigma^2,\boldsymbol{\theta}|\mathbf{y}_D)=-\frac{1}{2}[n\log\sigma^2+\log[\det(\mathbf{R}_D)]+(\mathbf{y}_D-\mathbf{F}\boldsymbol{\beta})^T\mathbf{R}_D^{-1}(\mathbf{y}_D$$
$$-\mathbf{F}\boldsymbol{\beta})/\sigma^2]. \tag{4}$$

The above equation is seldom directly used for obtaining the maximum likelihood estimates (MLE) of σ^2, $\boldsymbol{\beta}$, and $\boldsymbol{\theta}$. It is usually simplified as follows. Assuming that $\boldsymbol{\theta}$ is known, we can get the MLE of $\boldsymbol{\beta}$ and σ^2 as

$$\hat{\boldsymbol{\beta}}=(\mathbf{F}^T\mathbf{R}_D^{-1}\mathbf{F})^{-1}\mathbf{F}^T\mathbf{R}_D^{-1}\mathbf{Y}_D \quad \text{and} \quad \hat{\sigma}^2=\frac{1}{n}(\mathbf{y}_D-\mathbf{F}\hat{\boldsymbol{\beta}})^T\mathbf{R}_D^{-1}(\mathbf{y}_D-\mathbf{F}\hat{\boldsymbol{\beta}}),$$
$$\tag{5}$$

respectively. Substituting them in Eq. (4), we can get a simplified log-likelihood function with $\boldsymbol{\theta}$ as the only unknown parameter vector,

$$l(\boldsymbol{\theta}|\mathbf{y}_D)=-\frac{1}{2}\{n\log\hat{\sigma}^2+\log[\det(\mathbf{R}_D)]+n\}. \tag{6}$$

This equation is usually solved by numerical algorithms such as the Newton-Raphson [11], Nelder-Mead's simplex search [12] or the E-M algorithm [11]. In our implementation, we used the simplex search, which is readily available in MATLAB, to solve Eq. (6) for $\hat{\boldsymbol{\theta}}$, the MLE of $\boldsymbol{\theta}$. Then, substitute $\hat{\boldsymbol{\theta}}$ back to Eq. (5) for $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$. Given $\hat{\sigma}^2$ and $\hat{\boldsymbol{\theta}}$, the covariance matrix for $\{Z(x_i)\}_{i=1}^n$ is $\hat{\sigma}^2\hat{\mathbf{R}}_D$, where $\hat{\mathbf{R}}_D$ is the correlation matrix, with $\hat{\theta}_j$'s in the place of θ_j's.

Since it is not an option to exhaust the responses for all possible parameter combinations using expensive FEA simulations, we would like to predict y_0 using the Kriging model developed from the observed data $\mathbf{y}_D$. Because we are using Gaussian random fields, both observed and unobserved values can be pooled together in a multivariate Gaussian vector. Thus the joint distribution of Y_0 and $\mathbf{Y}_D$ is

$$\binom{Y_0}{\mathbf{Y}_D} \sim N_{1+n}\left[\binom{\mathbf{f}_0^T}{\mathbf{F}}\boldsymbol{\beta},\sigma^2\binom{1\quad\mathbf{r}_0^T}{\mathbf{r}_0\quad\mathbf{R}_D}\right] \tag{7}$$

where $\mathbf{f}_0=\mathbf{f}(\mathbf{x}_0)$ and $\mathbf{r}_0^T=[R(\mathbf{x}_0-\mathbf{x}_1),\ldots,R(\mathbf{x}_0-\mathbf{x}_n)]$ is the $n\times 1$ vector of correlations of Z's at the simulated input $\mathbf{x}_i$ ($i=1,2,\ldots,n$) and the untried input $\mathbf{x}_0$. Substituting σ^2, $\boldsymbol{\beta}$, and $\boldsymbol{\theta}$ with their MLEs and assuming $\hat{\mathbf{R}}_D$ as positive definite, Santner et al. [10] derives the best empirical mean square error (EMSE) predictor of y_0 as

$$\hat{Y}_0=\mathbf{f}_0^T\hat{\boldsymbol{\beta}}+\hat{\mathbf{r}}_0^T\hat{\mathbf{R}}_D^{-1}(\mathbf{Y}_D-\mathbf{F}\hat{\boldsymbol{\beta}}). \tag{8}$$

The prediction in (8) can be thought to be a sum of the generalized least squares predictor $\mathbf{f}_0^T\hat{\boldsymbol{\beta}}$ and the second term $\hat{\mathbf{r}}_0^T\hat{\mathbf{R}}_D^{-1}(\mathbf{y}_D-\mathbf{F}\hat{\boldsymbol{\beta}})$, which is the smooth of the residuals $(\mathbf{y}_D-\mathbf{F}\hat{\boldsymbol{\beta}})$. It has also been demonstrated that when predicting the value for an existing parameter combination, i.e., let $\mathbf{x}_0=\mathbf{x}_i$, $i=1,\ldots,n$, one will have $\hat{Y}_0=y(\mathbf{x}_i)$ [10]. This result verifies that the Kriging model indeed interpolates the existing simulation data, as shown in Fig. 3(b). In other words, the choice of Kriging model is better suited for deterministic FEA simulations.

Finally, we present the mean square error (MSE) for the Kriging predictor as

$$\text{MSE}[\hat{Y}_0]=\sigma^2\left[1-[\mathbf{f}^T(\mathbf{x}_0)\;\hat{\mathbf{r}}_0^T(\mathbf{x}_0)]\binom{\mathbf{0}\quad\mathbf{F}^T}{\mathbf{F}\quad\hat{\mathbf{R}}_D}^{-1}\binom{\mathbf{f}(\mathbf{x}_0)}{\hat{\mathbf{r}}_0(\mathbf{x}_0)}\right].$$
$$\tag{9}$$

For details of the derivations leading to (9), please refer to [6] or [10]. The subsequent section will develop a sequential strategy, using a Kriging predictor and evaluating its uncertainty recursively.
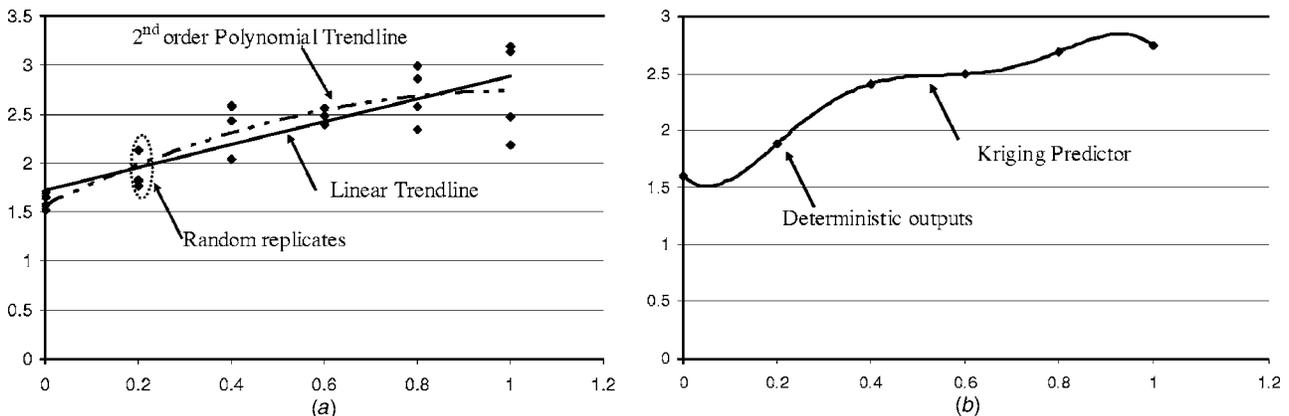
Fig. 3 Difference between (a) physical and (b) computer experiments
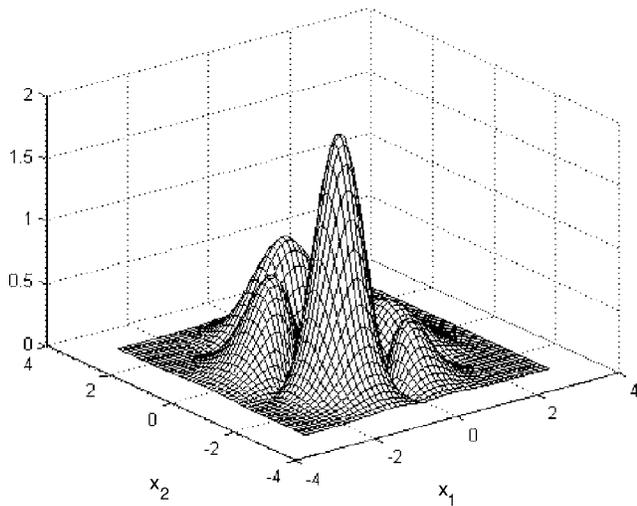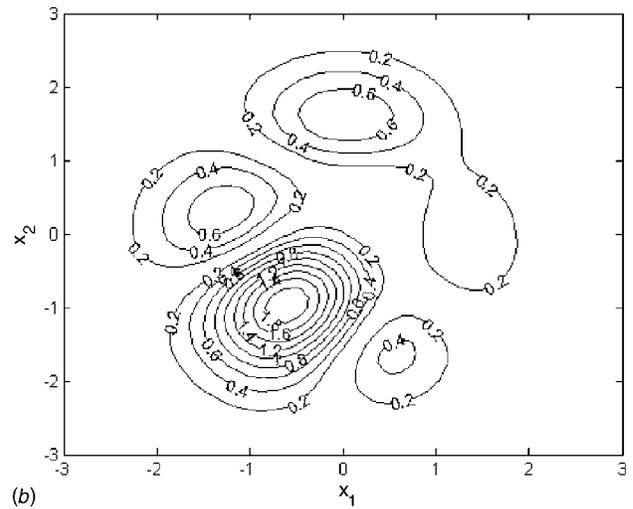
Fig. 4 An illustration of a complicated response surface

3 A Sequential Strategy for Electronic Packaging Design

3.1 Overview of the Sequential Strategy. Suppose we are searching for the maximal value over a response surface, as shown in Fig. 4, where multiple peaks (or valleys for a minimization problem) are present. If we employ a traditional RSM, its success in finding the global maximum depends on the starting point. That is to say, the traditional RSM will be easily entrapped in a local maximum (or minimum) by ascending (or descending) the wrong hill (or valley). This brings us to the important issue of the lack of prior knowledge of the response surface. Before the first simulation, we usually have no information about the response surface and the current operating condition is often chosen as the starting point. Such a starting point has no guarantee to take us to the optimal setting.

The Kriging model does not depend on any starting point. It tries to have an approximating envelope pass through a sample of response outputs. Obviously, in order not to miss the optimal areas, we should choose the input parameter combinations to be spread as evenly as possible in the parameter space, i.e., we need to employ a so-called space filling design. Suppose to model the

above surface we choose a 16 simulation Latin hypercube sampling design (this design will be described in detail later), then using a Kriging model we can fit a response surface. Figure 5(a) shows the contour plot of the response surface generated by the Kriging model. This fitted response offers a reasonably good representation of the overall surface and captures the peak area.

Intuitively it appears that if we run more simulations in a subregion and then revise our Kriging model based on the modified design, it should be able to represent the true surface better than the previous one. The major decision is to select an appropriate zooming-in subregion. There are two primary considerations. One is to zoom in to the areas of interest, i.e., a sub-region where we find large values of the response function (in a maximization problem). Another is to consider the uncertainty associated with the Kriging prediction and add new simulations to the subregions with high uncertainty values. The second consideration is to reduce the likelihood that an initial rough estimation may miss potential optimum areas—for example, other peaks in Fig. 5(a) are not very clearly identified. One can certainly choose to add new simulations at subregions to satisfy both considerations.

For the example in Fig. 4, we choose to use the first criterion. Figure 5(a) shows the sub-region selected and Fig. 5(b) shows the
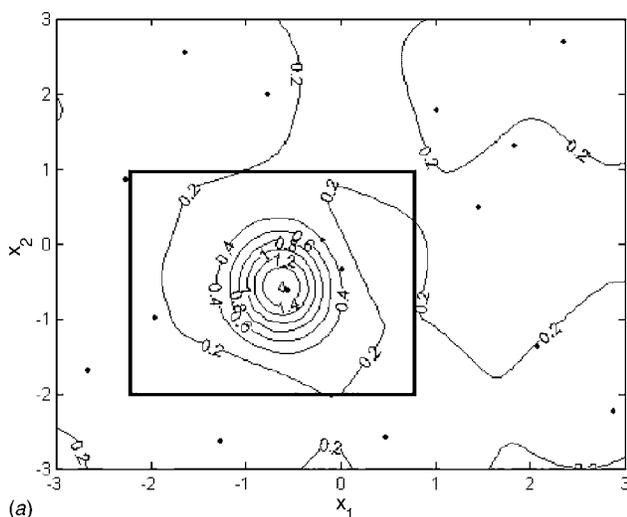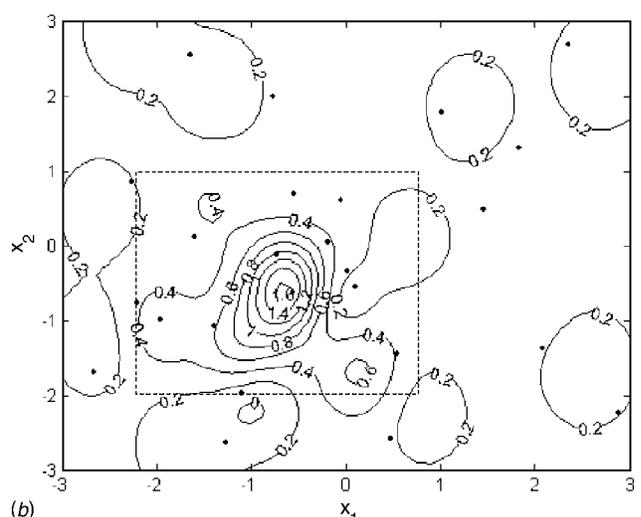
Fig. 5 An illustration of the sequential strategy

contour plot after we have added nine simulations in the zoom-in subregion. Together with the initial 16 simulations, the 25 simulations make the prediction in that selected subregion more accurate. Actually, the region outside the subregion is also benefited, as can be seen through the clearer boundaries that have started to form for the other peak area. The procedure can be iteratively carried out until we see no significant improvement.

The general idea of sequential experiments has also been applied to other applications, such as in integrated circuit design by Bernardo et al. [13] and artificial joint design by Chang et al. [14]. The details of our sequential strategy will be illustrated using the electronic packaging process. Also note that all variable values are reported coded (0 for low and 1 for high), unless indicated otherwise.

3.2 Initial Step. The first step in the sequential strategy is very important, as all our subsequent decisions are based on the model we get after this step. Thus, we want to ensure that this model is as good a representation of the response surface as possible. In this step we need to do the following.

(1) Choose a Kriging model, i.e., the regression polynomial and the covariance structure, that will model the functional response.
(2) Choose an experimental design plan to obtain a sample of simulation outputs from the parameter space.
(3) Construct the Kriging model using the FEA simulation outputs.

The regression polynomial and the covariance structure for the Kriging model are often chosen based on the experimenter's experience or prior knowledge about the process. Different from ordinary regression analysis, the Kriging model is largely a nonparametric method and thus the prediction result does not rely heavily on the choice of regression polynomials. For this reason, we choose a constant for the global regression part in Eq. (1) for our electronic packaging problem. The constant is selected in lieu of any polynomial equation just as a starting expression. Our aim in this procedure is to keep the number of unknowns as low as possible. Based on the first step analysis, we will have a better sense if we should add polynomial terms (the first order or the second order). If we find that satisfactory results are not obtained from just a constant regression component, then we will have to add the appropriate polynomial terms. In fact, we will observe from the latter analysis, the Kriging model with a constant polynomial term is rather capable of capturing a complicated surface topology with deterministic trends. Thus, for this electronic packaging problem, we have $f_1(\mathbf{x})=1$ and β_1.

As for the covariance function, we choose the one given in Eq. (3), and for reasons discussed in Sec. 2 we assume $p_j=2$. Thus, for the electronic packaging model we have six unknown parameters, namely σ^2, $\boldsymbol{\theta} \equiv (\theta_1, \theta_2, \theta_3, \theta_4)$, and β_1.

Once the Kriging model structure is chosen, we need an experimental plan to indicate how many FEA simulations will be performed and at what parameter combinations. In comparison to the design of physical experiments, where the data points are often discrete factorial combinations and replicates are used to account for the random variation, the design for deterministic computer simulations should be spread throughout the design space without any replications—since the output is deterministic, it does not make a difference with or without replication. Such spread-out designs are known as space-filling design, [10].

Introduced by McKay et al. [15], Latin Hypercube Sampling (LHS) design is one of the most commonly used space-filling designs. An LHS design yielding n design points involves stratifying the parameter space into n equal probability intervals for each dimension, randomly selecting a value in each stratum and combining them to get a design point. In a high parameter space

(more than one dimension), LHS design itself may not be necessarily space filling. Researchers introduced other criteria such as minimizing the maximum distance between design points (the *minimax* criterion) or maximizing the minimum distance between design points (the *maximin* criterion) to improve the space-filling property [10]. These two criteria are considered equivalent but the *maximin* criterion is more popular because it puts a direct restriction on the *minimum* distance so that any two design points will not be too close to each other. Hence, in our simulation plan, we adopt the *maximin* LHS design, where a design point is a specific parameter combination.

One critical question for running time-expensive FEA simulations is how to decide the number of simulations (i.e., the number of design points). We would always like to run a lesser number of FEA simulations, while keeping in mind not to affect the fidelity of the model appreciably. The least number of simulations we need to perform is the number of unknown parameters in the Kriging model; we will otherwise have a singular problem. Bernardo et al. [13] suggested as a rule of the thumb choosing the number of design points to be three times the number of unknown parameters for superior prediction performance. This means for a case like ours, where we have six unknown parameters, we need to run at least 6, but could have as many as 18 simulations, depending on the prediction requirement.

We will have subsequent simulations to revise the first step model, yet the first step should provide a good global view. We recommend choosing the number of simulations in the initial step to be 1.5 times the number of unknown parameters in our sequential simulation procedure. This implies that we will run 9 FEA simulations in the initial step and we choose a 9×4 ($d=4$) *maximin* LHS design; the design matrix is shown in the Appendix. Simulations corresponding to this design were carried out in *ANSYS* and the corresponding simulated von Mises values (y) are included in the Appendix as well. A Kriging model was constructed after calculating $\hat{\sigma}^2$, $\hat{\boldsymbol{\theta}}$, and $\hat{\beta}$ as $\hat{\sigma}^2=37.81$, $\hat{\boldsymbol{\theta}} = (2.5341, 0.6286, 0.4502, 1.0967)$, and $\hat{\beta}=170.47$.

Finally, the Kriging model is used to generate predicted values over the parameter space. We can also use Eq. (9) to calculate the uncertainty associated with the model. In order to have a rough idea of what the response surface looks like, we display the contour plot of the response and its MSE value in Fig. 6 using two input variables x_1 and x_2 (other variable pairs can be used in the same way), where the black dots indicate the locations of the nine parameter combinations.

We use this Kriging predictive model to explore the four-dimensional design surface. Optimization routines such as the simplex search are often employed to find the optimal region. In order to avoid being entrapped into local optima, random search based algorithms (genetic algorithm or simulated annealing) are also used. In our case, since we have a relatively low dimension, we simply choose to generate predicted values over a lattice of the parameter space for obtaining the minimum. The predicted and actual minima together with the corresponding parameter values are given in Table 2.

3.3 Subsequent Steps. Naturally, we will add more simulations in the following steps to improve our model prediction and find the minimal response value. The key question is on which parameter combination should we run additional simulations. We will use both zoom-in criteria: (i) run more simulations over a region of interest and (ii) run extra simulations in a region of high uncertainty. The selection of zoom-in region is subjective in nature, depending on design objectives and the designers' understanding of underlying physics. We recommend using graphic tools such as main effect plots or response contour plots to facilitate designers in selecting a region of interest.

3.3.1 Reduce The Overall Uncertainty Level—The 2nd Stage. The next step is to reduce the prediction uncertainty because the
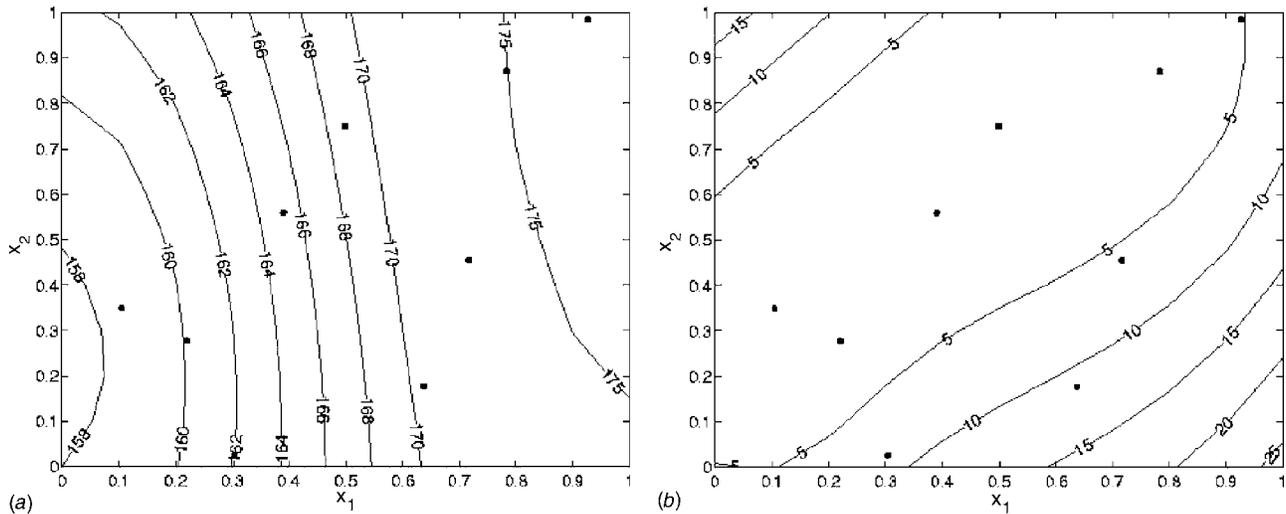
Fig. 6 Contour plots of the response (a) and its MSE value (b)

existence of a high uncertainty will mislead us to a region where the true response value is not small at all, and thus will taint our efforts in finding the minimum region. Figures 7(a)–7(f) show MSE contour plots for pairwise design variables; given four design variables, we have six such plots.

We can certainly observe some regions with high uncertainty. For instance, given the true response around 170, an MSE of 17 accounts roughly for 10% of the response value. We thus identify zoom-in subregions with a 10% error, which translates to a boundary with a MSE value between 15 and 20. Following this thought and also utilizing our graphic plot, we marked (with a thicker dash line) on each contour plot our zoom-in sub-region, e.g., it is $0.8 \leq x_1 \leq 1.0$ and $0 \leq x_2 \leq 0.3$ in Fig. 7(a). A similar procedure applies to other plots in Fig. 7, and we eventually take the union of all the subregions as the region to be zoomed in.

$$0.8 \leq x_1 \leq 1.0; \quad 0 \leq x_2 \leq 1.0; \quad 0 \leq x_3 \leq 1.0; \quad 0 \leq x_4 \leq 1.0 \tag{10}$$

We then choose a six-simulation (six is the number of unknown parameters) LHS design over the zoom-in region and concatenate the data points to our previous design to get a modified 15-simulation design. Please refer to the Appendix for the design matrix.

With the 6 additional simulations, a new Kriging model was developed using the FEA outputs from a total of 15 simulations. The updated model parameters are $\hat{\sigma}^2 = 27.2614$, $\hat{\theta} = (4.1054, 0.5849, 0.3153, 0.8876)$, and $\hat{\beta} = 171.82$. The prediction of the minimum from the second stage model is $\hat{y} = 157.57$, corresponding to the same design parameters as in Table 2 and the true response value of $y = 154.44$. Obviously, the minimum point has not changed, but we have successfully reduced the prediction uncertainty to be much less than 5% in the majority of the regions, as evidenced by the MSE plots in Fig. 8. With this small level of uncertainty, the minimal value region indicated by the prediction model can be deemed trustworthy.

3.3.2 Refine the Location of the Minimum Point—The 3rd Stage. The second stage Kriging model predicts that the minimum response value is about $y = 154.44$, which sets a target for our

zoom-in effort. Also noticing that among the first 15 simulations, the minimum y is 158.62, we decide to set the subregion boundary to be the contour line of 158. We select the subregion by using interaction contour plots between a pair of variables, as shown in Fig. 9. In those graphic plots, we are slightly conservative in drawing the region so that the 158 line is included.

As done earlier, we get the zoom-in region from Fig. 9 by taking the union of subregions from each of the contour plots. The following region is obtained as the region to be zoomed in.

$$0 \leq x_1 \leq 0.1; \quad 0 \leq x_2 \leq 0.5; \quad 0 \leq x_3 \leq 0.8; \quad 0 \leq x_4 \leq 0.5 \tag{11}$$

We again choose a 6-simulation LHS design over this region (the design matrix is also in the Appendix) and concatenate the data points to our previous design to get a modified 21-simulation design. A new third stage Kriging model was fitted using these 21 simulation outputs and the updated model parameters are $\hat{\sigma}^2 = 56.89$, $\hat{\theta} = (2.2029, 0.1449, 0.8688, 4.7316)$, and $\hat{\beta} = 167.83$. This Kriging model finds a smaller value of the response function and the corresponding design variables are different. The results are given in Table 3.

3.4 Final Step and Stopping Rule. The "zooming-in" procedure as described above will usually be continued repeatedly until the model so obtained is sufficiently accurate and the area of interest (e.g., a minimum region) is located. We will use the MSE plot to characterize the model accuracy and use the change in the minimum response values in two consecutive steps to benchmark if an area of interest is located. If the increment/decrement in the response values is not significant and the MSE plot shows no large uncertainty value, then we can stop the sequential strategy.

For the problem at hand, the decrease in the minimum response values found by the third stage model is only 1% from the second stage model. The largest MSE value is only about 7% (the MSE plot for x_1, x_2 is shown in Fig. 10(b)). Both criteria are satisfied. Therefore, we choose to stop at this stage. Up to now, we have established an adequate Kriging model to represent the FEA simulation model. The optimal parameter settings in Table 3 are our final solution.

Table 2 Results after the initial step (in GPa units, coded values for *x*'s in brackets)

x_1	x_2	x_3	x_4	y	$\hat{y}$
20.0 (0.0)	16.8 (0.2)	3.0 (0.5)	16.5 (0.1)	154.44	157.53
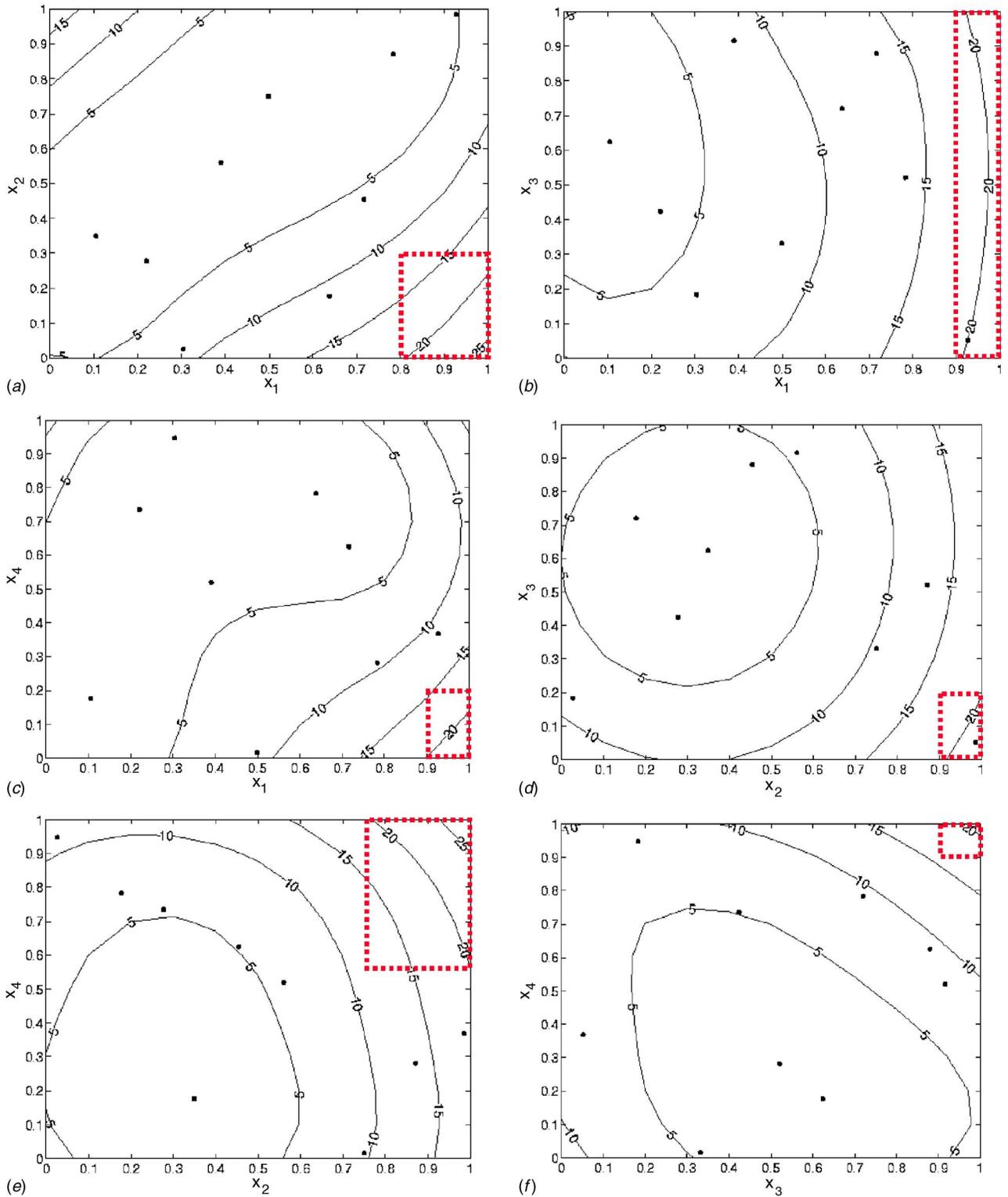
Fig. 7 MSE plots for pairwise design variables

Figure 10 shows the contour plot of the response function and the MSE value for the final Kriging model (only the one for x_1, x_2 are shown). The Kriging contour provides a global view of the response surface over the design space; the whole response surface has a decreasing trend along its reverse diagonal direction and the minimum region can be easily located to be around the left-bottom corner. Recall that the Kriging model we used in this example only has a constant polynomial term β_1. However, it

appears that we do not have to include a first order term because the resulting Kriging model (mainly through its correlation term Z) is able to capture the surface change over the design space adequately.

4 Comparison With Classical Designs

Traditional response surface methodology also follows a sequential strategy. It usually starts with a first order fractional fac-
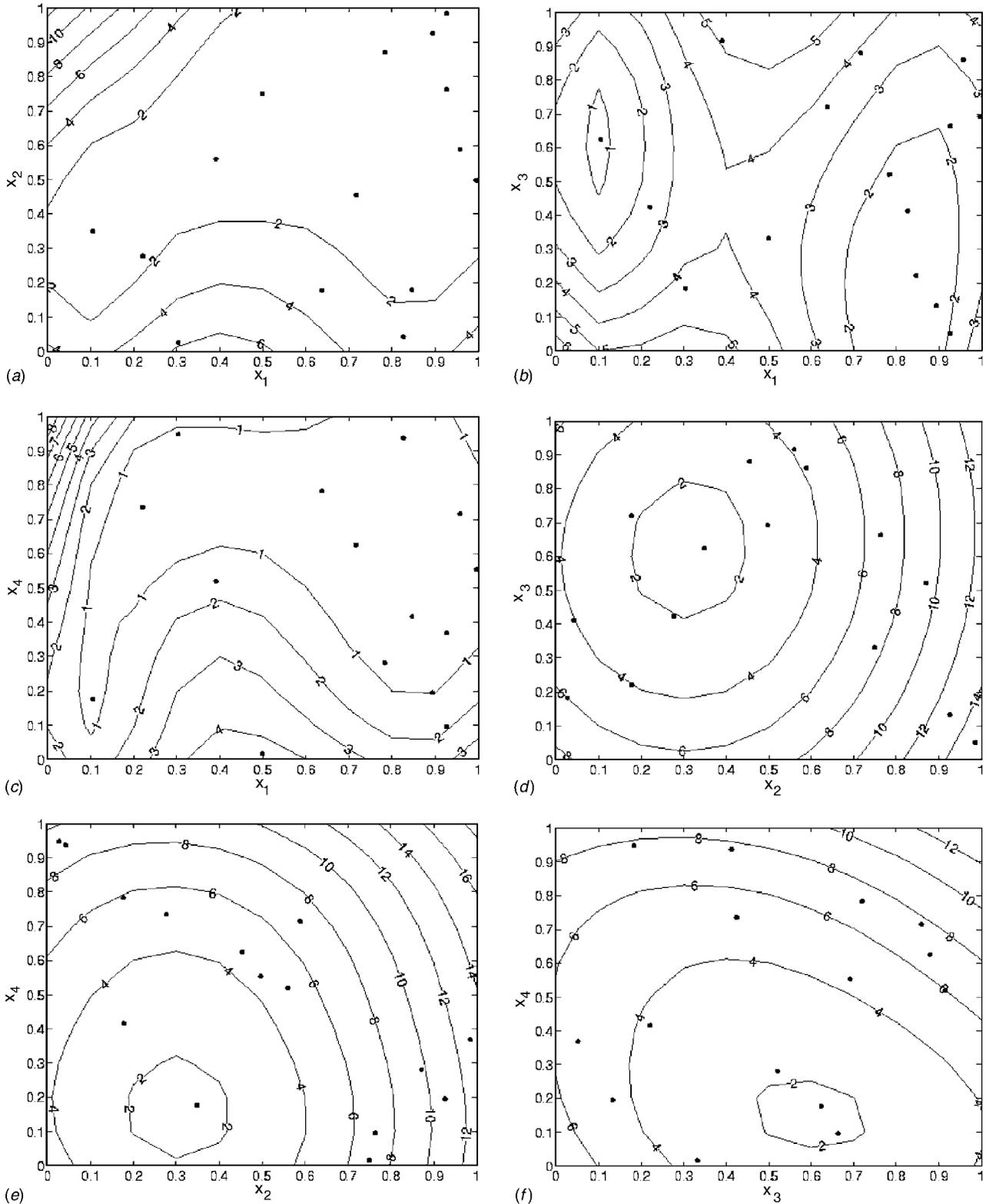
Fig. 8   MSE plots for pairwise design variables after the second stage

torial design, then finds the probing direction based on the steepest decent method, and finally switches to a second order design (such as a Central Composite Design) to locate the optimal point when it approaches an optimum (likely a local one) [5,16].

We employ the traditional RSM for the electronic packaging problem. Because of the lack of prior knowledge of where to start

our search, we simply choose the starting point as the current operating conditions, i.e., the center point of our parameter space. A linear model was fitted around this starting point. Because the assumption of linearity will hold only over a small region, we will have to limit the first step design to a smaller hypercube whose edge size is one-fourth the size of the design range. This choice of
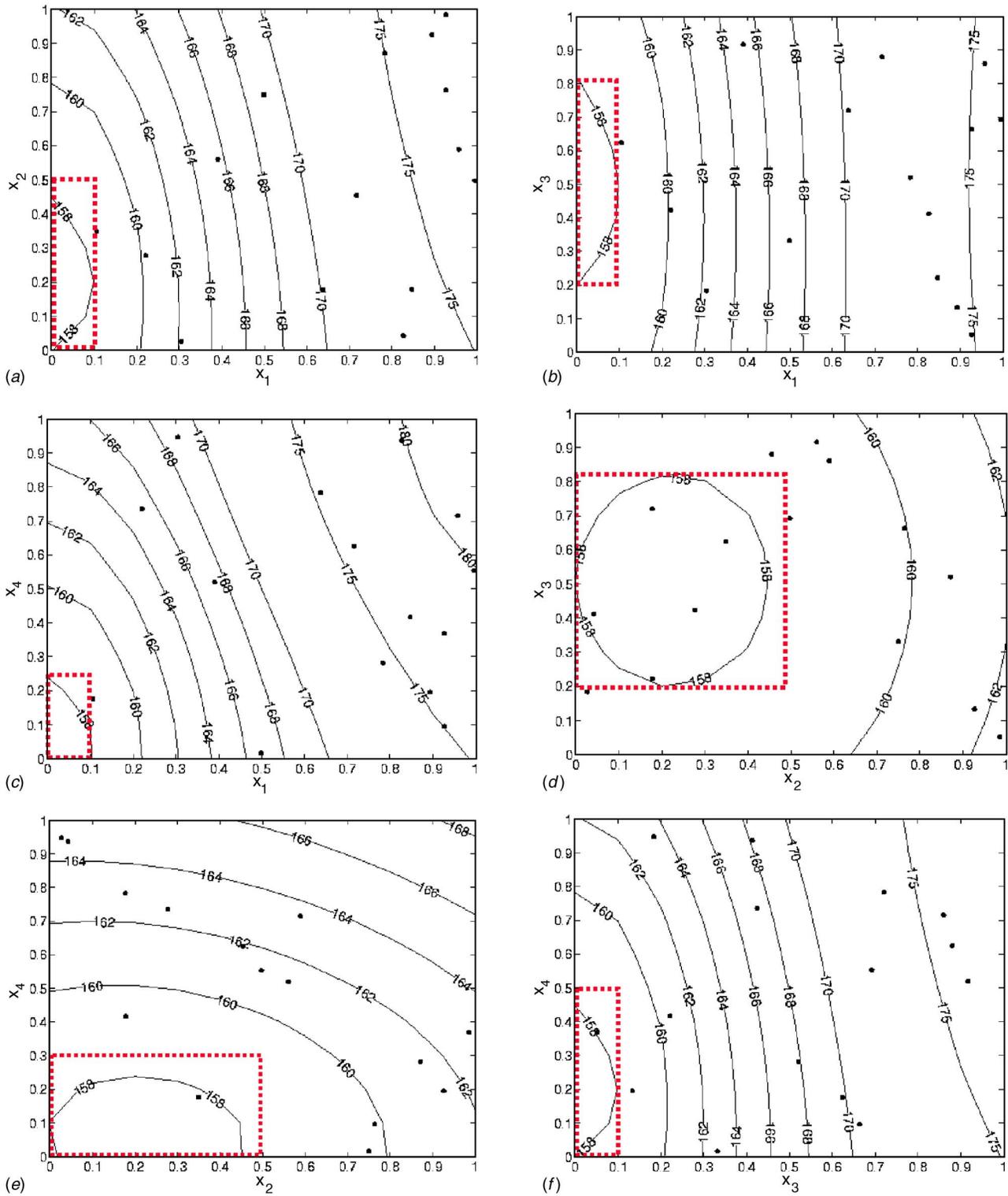
Fig. 9 Pairwise interaction plots from the second stage model

Table 3 Results after the third step (in GPa units, coded values for x's in brackets)

x_1	x_2	x_3	x_4	y	$\hat{y}$
20.0 (0.0)	12.0 (0.0)	2.2 (0.3)	16.5 (0.1)	153.47	154.31

one-fourth is empirical since there is no established rule in the literature. Given four design parameters, we use a 2^{4-1} fractional factorial design, which will guarantee that the main effects can be reasonably estimated, since the aliased three-factor interactions are usually not significant. From this experiment, a first order model is obtained as

$$\hat{y} = 118.3 + 1.4788x_1 + 0.122\,085x_2 - 0.0955x_3 + 0.442x_4.$$

(12)
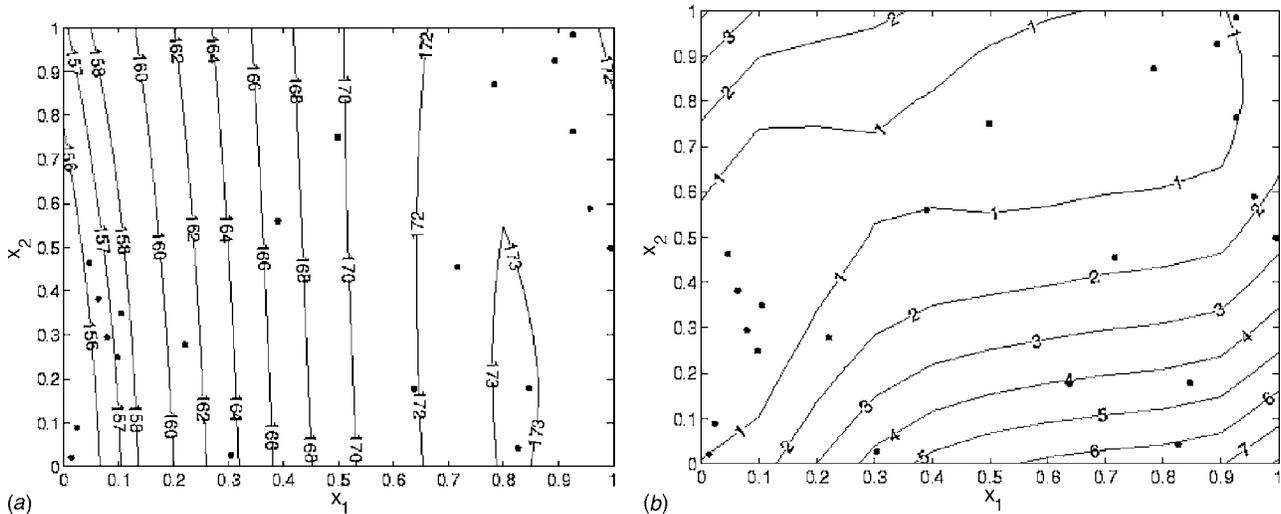
Fig. 10   Contour plots: (a) response surface, (b) MSE value for the third stage model

The coefficients of the first order model indicate the steepest direction. Following the direction, we will change the settings of design variables by taking a few sequential steps. Since the coefficient of x_1 is the greatest, we can see that the greatest decrease will occur if we change the values with respect to x_1. If we take the step sizes relative to x_1 (when coded a unit step corresponds to 1.875 of x_1) then, the steepest descent vector is (−1.875, −0.3963, +0.0086, −0.5604), which will be added to the current design setting sequentially. For each step, we get the corresponding simulation value, as shown in Fig. 11(a), which gives us an idea of the validity. We go along this direction unless we either hit the boundary of the parameter space or the simulation values start increasing instead of decreasing (i.e., we encounter an inflection point). In our case after five steps, we reach the parameter space boundary.

Then, we fit a second order model around this point using a Small Composite Design [5], one of the second order design methods similar to Face Centered Design. Its design settings for three variables are illustrated in Fig. 11(b). This Small Composite Design is based on a 2^{4-1} fractional factorial but includes eight more design points at the centers of each hyper-surface and one design point at the center of the hypercube (Fig. 11(b) is an illustration for three design variables). The edge size of the hypercube for the Small Composite Design is also chosen as one-fourth the size of the design range. After performing FEA simulations according to this second order design, we fit a second order model as

$$\hat{y} = 66.85 + 4.83x_1 + 1.0x_2 + 1.267x_3 + 0.0466x_4 - 0.0067x_1x_2$$
$$- 0.0012x_1x_3 + 0.0242x_1x_4 - 0.03091x_2x_3 - 0.017x_2x_4$$
$$- 0.0236x_3x_4 - 0.0738x_1^2 - 0.004x_2^2 + 0.0059x_3^2 + 0.0054x_4^2$$

$$(13)$$

Based on the above model, we can find the minimum value of the response function and the corresponding design parameters; they are shown in Table 4.

Compared with the minimum response value found by the Kriging model in Table 3, the difference is negligible, meaning both methods can successfully find the optimal region for our electronic packaging problem. However, the sequential Kriging method, as presented in this paper, takes 22 simulations (1st stage: 9; 2nd stage: 6; 3rd stage: 6; and the final solution: 1). The traditional RSM, on the other hand, requires 31 simulations (8 for a 2^{4-1} design, 5 for the steepest descent, 17 for the Face Centered Design, and 1 simulation at the final solution). The number of FEA simulations required by the sequential Kriging model is 41% less than the traditional RSM, which is a remarkable reduction.

In this electronic packaging example, the response surface of the von Mises stress is actually not complicated—it has a global descending trend toward its left-bottom corner and it has only one minimum point. That is the reason why the traditional RSM gives us the same optimum values as the Kriging model. Should we have a complicated surface as in Fig. 4, the chance of RSM to
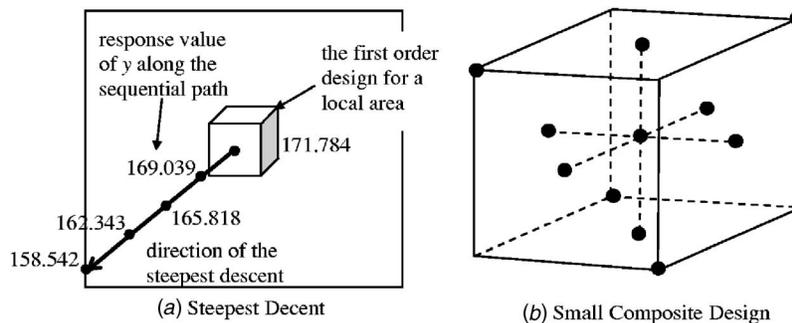
Fig. 11   Traditional response surface methodology for electronic packaging problem

Table 4 Results from the steepest descent method (in GPa units)

x_1	x_2	x_3	x_4	y	$\hat{y}$
20.0	12.0	1.0344	17.2416	153.48	151.50

locate the global optimum successfully is in fact not high. However, our sequential strategy will not be bounded by the complexity of the surface due to its space filling nature at each stage.

The advantages of the proposed sequential strategy also manifest in other aspects. One is that the final Kriging model will provide a global view of the response surface over the parameter space. By contrast, the traditional RSM will fall short of doing so since it fits polynomial models to a few small areas. It will be difficult to piece together the global response surface of the von Mises stress.

Another advantage is the uncertainty evaluation. The Kriging model provides an uncertainty evaluation of the von Mises stress prediction over the entire parameter space. For the traditional RSM, uncertainty can be evaluated based on linear regression theory. However, with the absence of random errors, the uncertainty evaluation from linear regression is not guaranteed to be the interpolation uncertainty in a FEA simulation. Once again, those uncertainty evaluations are only available for a few small areas in the design space. A high uncertainty could exist elsewhere without being detected simply because the steepest descent method does not lead our FEA simulation there.

5 Concluding Remarks

A sequential Kriging approach is presented in this paper to find the optimal parameter setting for electronic packaging. The sequential strategy will conduct additional FEA simulations in the region where more information may be needed to revise the subsequent Kriging models for prediction. The sequential strategy demonstrates a superior performance by finding the minimum von Mises stress in an electronic packaging process with 40% less FEA simulations compared to a classical RSM approach. It also provides the global view of the response surface and the uncertainty level over the entire region, of which the traditional RSM falls short.

Despite the current advancements in optimization methodologies using deterministic computer simulations, we would also like to point out the challenges ahead of us. One is the way we are deciding the zoom-in subregions. Currently we are dependent on a visual technique that assesses the interaction contour plots to gain information. This visual technique is intuitive and easy to use for practitioners. However, it can become less effective when the parameter space is high dimensional. It would be desirable to develop a searching procedure to locate the area of interest in a high dimension parameter space.

That naturally leads us to the issues of screening out the important factors. In this study, our collaborating industrial partner has already identified the four key parameters for investigation. In other circumstances, however, screening itself could be an issue, especially considering that a computer simulation often has many more input factors, each of which could possibly take multiple levels. The idea of sequential simulations would be a natural choice for such a screening task, while the detailed procedure is yet to be developed.

Acknowledgment

Appendix

Design table for the sequential simulations in Sec. 3 (parameters and responses are expressed in their actual physical units)

Runs		x_1 (GPa)	x_2 (Gpa)	x_3 (GPa)	x_4 (GPa)	y (GPa)
First stage	1	21.58	20.37	3.50	17.64	158.62
	2	25.85	25.44	4.66	22.79	169.37
	3	30.74	22.91	4.52	24.39	176.75
	4	23.31	18.65	2.70	26.03	165.25
	5	33.90	35.65	1.21	20.52	180.24
	6	31.75	32.90	3.08	19.22	176.60
	7	29.56	16.25	3.88	26.74	175.45
	8	24.56	12.63	1.73	29.21	167.88
	9	27.48	29.99	2.33	15.24	168.63
Second stage	10	34.93	23.94	3.77	23.29	181.38
	11	32.39	13.013	2.65	29.05	179.08
	12	33.40	34.22	1.53	17.92	178.09
	13	32.69	16.27	1.89	21.24	176.49
	14	34.36	26.14	4.44	25.73	182.02
	15	33.90	30.33	3.65	16.43	177.14
Third stage	16	21.47	17.97	3.22	19.15	158.78
	17	20.20	12.51	2.32	22.03	156.82
	18	20.70	23.11	1.43	16.46	156.61
	19	20.35	14.14	1.71	18.12	155.59
	20	21.18	19.07	3.75	20.36	158.89
	21	20.95	21.16	3.12	15.71	156.35

References

[1] Darveaux, R., 2000, "Effect of Simulation Methodology on Solder Joint Crack Growth Correlation," *Proceedings of the 50th IEEE Electronic Components and Technology Conference (ECTC'00)*, Las Vegas, NV.

[2] Shetty, S., and Reinikainen, T., 2001, "Three- and Four-Point Bend Testing for Electronic Packages," *Proceedings of the 2001 EuroSimE*, Paris, France, 9–11 April.

[3] Rodgers, B., Punch, J., Jarvis, J., Myllykoski, P., and Reinikainen, T., 2002, "Finite Element Modeling of a BGA Package Subjected to Thermal and Power Cycling," Inter Society Conference on Thermal Phenomena in Electronic Systems, San Diego, CA, May 30–June 1.

[4] ANSYS, Inc., 2002, *ANSYS 6.1 Documentation*, ANSYS, Inc., Canonsburg, PA.

[5] Myers, R. H., and Montgomery, D. C., 1995, *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, John Wiley and Sons, New York.

[6] Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P., 1989, "Design and Analysis of Computer Experiments (with discussion)," Stat. Sci., **4**, pp. 409–435.

[7] Sacks, J., Schiller, S. B., and Welch, W. J., 1989, "Design for Computer Experiments," Technometrics, **31**, pp. 41–47.

[8] Welch, W. J., Buck, R. J., Sacks, J., Wynn, H. P., Mitchell, T. J., and Morris, M. D., 1992, "Screening, Predicting and Computer Experiments," Technometrics, **34**, pp. 15–25.

[9] Ripley, B. D., 1981, *Spatial Statistics*, John Wiley and Sons, New York.

[10] Santner, T. J., Williams, B. J., and Notz, W. I., 2003, *Design and Analysis of Computer Experiments, Springer Series in Statistics*, Springer Verlag, Berlin.

[11] Lindstrom, M. J., and Bates, D. M., 1988, "Newton-Raphson and EM Algorithms for Linear Mixed-Effects Models for Repeated-Measures Data," J. Am. Stat. Assoc., **83**, pp. 1014–1022.

[12] Nelder, J. A., and Mead, R., 1965, "A Simplex Method for Function Minimization," Comput. J., **7**, pp. 308–313.

[13] Bernardo, M. C., Buck, R., Liu, L., Nazaret, W. A., Sacks, J., and Welch, W. J., 1992, "Integrated Circuit Design Optimization Using a Sequential Strategy," IEEE Trans. Comput.-Aided Des., **11**, pp. 361–372.

[14] Chang, P. B., Williams, B. J., Bhalla, K. S. B., Belknap, T. W., Santner, T. J., Notz, W. I., and Bartel, D. L., 2001, "Design and Analysis of Robust Total Joint Replacements: Finite Element Model Experiments with Environmental Variables," J. Biomech. Eng., **123**, pp. 239–246.

[15] Mckay, M. D., Beckman, R. J., and Conover, W. J., 1979, "A Comparison of Three Methods for Selecting Values of Input Variables In the Analysis of Output From a Computer Code," Technometrics, **21**, pp. 239–245.

[16] Montgomery, D. C., 1997, *Design and Analysis of Experiments*, 4th ed., John Wiley and Sons, New York.