

## RELATIVE ERROR STOCHASTIC KRIGING

Mustafa H. Tongarlak  
Bruce E. Ankenman  
Barry L. Nelson

Department of Industrial Engineering and Management Sciences  
Northwestern University  
2145 Sheridan Road, Evanston, IL 60208

### ABSTRACT

We use stochastic kriging to build predictors with bounded relative error over the design space. We propose design strategies that guide sequential algorithms with and without adaptation to the data to make allocation and stopping decisions such that a prespecified relative precision is realized with some confidence. We also present an empirical evaluation of the proposed design strategies.

### 1 INTRODUCTION

Because of their ability to capture any level of complexity, computer simulation models are often used to represent complex systems to obtain accurate estimates of performance measures. Advancements in computing technologies and emergence of better design and analysis methodologies has helped spread simulation's use as a decision support tool. Despite recent advancements, however, achieving accurate estimates through simulation can still be very costly in terms of time and effort. If simulation-based estimates of performance for multiple system scenarios are needed rapidly, then there may not be adequate time for a thorough analysis of all scenarios. Our approach is to map a performance measure of interest as a function of decision variables using the information gained by simulation runs made at carefully chosen scenarios.

More specifically, we use the stochastic kriging method proposed by Ankenman et al. (2010) to produce a complete performance response surface (predictor) and propose experiment design strategies to make computational budget allocation and stopping decisions. Our goal is to create a predictor with bounded maximum relative error over the entire design space. This objective is relevant when reducing the magnitude of the prediction error relative to the performance measure is more critical than the absolute error of the estimate, as is often the case when output performance has a wide range as a function of controllable parameters. For instance, a manufacturer may be able to run a factory rather lightly loaded, or at near capacity, depending on current sales. Product cycle times as a function of load might differ by an order of magnitude or more over this range, making relative error control more meaningful than absolute error.

The theory of stochastic kriging advanced in Ankenman et al. (2010) is an extension of kriging. Kriging is an interpolation-based metamodeling technique widely used with deterministic computer experiments to mimic the true response surface when little or no prior knowledge exists about the form of the response surface. Unlike regression models, kriging makes weak assumptions about the true surface and creates flexible prediction surfaces using the spatial structure among design points. In deterministic computer experiments, the true response at a design point is known with certainty and the response at other points in the design space is predicted by the kriging metamodel. In stochastic computer simulations, however, the true response at a design point is uncertain due to the sampling variability and therefore kriging is not directly applicable. Stochastic kriging distinguishes two types of uncertainty about the performance response surface: sampling error, which occurs because the stochastic simulation output is not the same as

the true performance, and metamodel error, which arises when the response at untried settings is estimated. The stochastic kriging method provides accurate metamodels and allows the assessment of uncertainty about them. In addition, experiment designs used with stochastic kriging are capable of controlling both model and sampling uncertainty. This makes it possible to “run until you are done,” i.e., increase simulation effort until the metamodel reaches a desired level of precision.

The remainder of the paper is organized as follows: Section 2 outlines the mathematical framework of stochastic kriging and introduces the notation used in this paper. Section 3 describes our proposed experiment design strategies, such as stopping and design criteria, and two sequential algorithms. In Section 4, we set up experiments to illustrate the ability of the proposed methods to predict the test surfaces with prespecified relative error, while Section 5 reports the outcome of these experiments. We conclude in Section 6.

## 2 BACKGROUND

In this section we outline the formulation of the stochastic kriging method proposed by Ankenman et al. (2010), which sets the foundation for this paper, and much of the notation used in this paper is borrowed from their mathematical framework.

Suppose  $Y(\mathbf{x})$  is the unknown response at a design point  $\mathbf{x} = (x_1, x_2, \dots, x_d)$  in the  $d$ -dimensional experiment design space of interest,  $\mathcal{X}$ . Further suppose that an experiment design that consists of design points  $\mathbf{x}_i, i = 1, 2, \dots, k$ , and a stochastic simulation model is used to predict the response at *any*  $\mathbf{x}_0 \in \mathcal{X}$ . The simulation output from the  $j^{\text{th}}$  replication of the simulation model at design point  $\mathbf{x}$  can be represented as

$$\mathcal{Y}_j(x) = Y(\mathbf{x}) + \varepsilon_j(x), \quad (1)$$

where  $\varepsilon_j(x)$  is a random variable with zero mean representing the sampling error inherent in stochastic simulation output. Although it is independent and identically distributed across replications at a design point  $\mathbf{x}$ , the variance of  $\varepsilon_j(x)$  over the design space is not necessarily constant.

In stochastic kriging (SK), the response  $Y(\mathbf{x})$  at  $\mathbf{x}$  is represented as

$$Y(\mathbf{x}) = f(\mathbf{x})^\top \beta + M(\mathbf{x}), \quad (2)$$

where  $f(\mathbf{x})$  is a vector of known functions of  $\mathbf{x}$ ,  $\beta$  is a vector of unknown parameters of compatible dimension, and  $M$  is a realization of a mean-zero random field with covariance  $\Sigma_M(\mathbf{x}_1, \mathbf{x}_2)$  between  $M(\mathbf{x}_1)$  and  $M(\mathbf{x}_2)$ . Here we assume  $M$  is a stationary Gaussian random field, a typical assumption when kriging is used in the design and analysis of (deterministic) computer experiments, and  $\Sigma_M(\mathbf{x}_1, \mathbf{x}_2) = \tau^2 R_M(\mathbf{x}_1 - \mathbf{x}_2; \theta)$ , where  $\tau^2$  represents the process variance,  $\theta$  is a vector of unknown parameters, and  $R_M(\mathbf{x}_1 - \mathbf{x}_2; \theta)$  is the correlation between  $M(\mathbf{x}_1)$  and  $M(\mathbf{x}_2)$ . The trend term,  $f(\mathbf{x})^\top \beta$ , provides a global model much like in regression and we reduce it in this paper to the constant  $\beta_0$ , as is often done in practice.

Combining (1) and (2) yields the SK representation of the simulation output  $\mathcal{Y}_j(x)$ :

$$\mathcal{Y}_j(x) = \beta_0 + M(\mathbf{x}) + \varepsilon_j(x).$$

Define  $\Sigma_\varepsilon$  to be the  $k \times k$  covariance matrix with  $(h, i)$  element given by  $\text{Cov}[\bar{\varepsilon}(\mathbf{x}_h), \bar{\varepsilon}(\mathbf{x}_i)]$ , where  $\bar{\varepsilon}(\mathbf{x})$  is the mean of the sampling error over all replications at a design point  $\mathbf{x}$  and define  $\Sigma_M$  as the  $k \times k$  covariance matrix with  $(h, i)$  element given by  $\tau^2 R_M(\mathbf{x}_h - \mathbf{x}_i; \theta)$ . Also let  $\Sigma_M(\mathbf{x}_0, \cdot)$  be the  $k \times 1$  vector of covariances between  $M(\mathbf{x}_0)$  and  $M(\mathbf{x}_i)$  for  $i = 1, 2, \dots, k$ . Then, the MSE-optimal SK predictor of the response  $Y(\mathbf{x}_0)$  at *any*  $\mathbf{x}_0 \in \mathcal{X}$  can be written as

$$\hat{Y}(\mathbf{x}_0) = \beta_0 + \Sigma_M(\mathbf{x}_0, \cdot)^\top [\Sigma_M + \Sigma_\varepsilon]^{-1} (\bar{\mathcal{Y}} - \beta_0 \mathbf{1}_k), \quad (3)$$

where  $\bar{\mathcal{Y}}$  is the  $k \times 1$  vector of sample mean responses at  $\mathbf{x}$  and  $\mathbf{1}_k$  is the  $k \times 1$  vector of ones. Equation (3) assumes known  $\beta_0$ ,  $\Sigma_M$ , and  $\Sigma_\varepsilon$ . In a real application, they need to be estimated. Let the maximum

likelihood estimators of  $(\beta_0, \tau^2, \theta)$  be  $(\hat{\beta}_0, \hat{\tau}^2, \hat{\theta})$ , conditional on

$$\hat{\Sigma}_\varepsilon = \text{Diag} \left\{ \hat{V}(\mathbf{x}_1)/n_1, \hat{V}(\mathbf{x}_2)/n_2, \dots, \hat{V}(\mathbf{x}_k)/n_k \right\}, \quad (4)$$

where  $\hat{V}(\mathbf{x}_i) = \sum_{j=1}^{n_i} (\mathcal{Y}_j(x_i) - \bar{\mathcal{Y}}(x_i))^2 / (n_i - 1)$  is an estimator of the variance of  $\varepsilon(\mathbf{x}_i)$  and  $n_i$  is the number of simulation replications at the design point  $x_i$ . Then, the SK predictor is

$$\hat{Y}(\mathbf{x}_0) = \hat{\beta}_0 + \hat{\tau}^2 R_M(\mathbf{x}_0, \cdot; \hat{\theta})^\top \left[ \hat{\tau}^2 R_M(\hat{\theta}) + \hat{\Sigma}_\varepsilon \right]^{-1} \left( \bar{\mathcal{Y}} - \hat{\beta}_0 \mathbf{1}_k \right).$$

An MSE estimator is also derived in Ankenman et al. (2010) as

$$\begin{aligned} \widehat{\text{MSE}}(\mathbf{x}_0) &= \hat{\tau}^2 - \hat{\tau}^4 R_M(\mathbf{x}_0, \cdot; \hat{\theta})^\top \left[ \hat{\tau}^2 R_M(\hat{\theta}) + \hat{\Sigma}_\varepsilon \right]^{-1} R_M(\mathbf{x}_0, \cdot; \hat{\theta}) \\ &\quad + \delta^\top \delta \left( \mathbf{1}_k^\top \left[ \hat{\tau}^2 R_M(\hat{\theta}) + \hat{\Sigma}_\varepsilon \right]^{-1} \mathbf{1}_k \right)^{-1}, \end{aligned} \quad (5)$$

where  $\delta = 1 - \mathbf{1}_k^\top \left[ \hat{\tau}^2 R_M(\hat{\theta}) + \hat{\Sigma}_\varepsilon \right]^{-1} R_M(\mathbf{x}_0, \cdot; \hat{\theta}) \hat{\tau}^2$ .

When a candidate design point  $\mathbf{x}_0 \in \mathcal{X}$  that has not yet been simulated is considered for entry to the experiment design, an estimate of  $\hat{V}(\mathbf{x}_0)$  is needed to calculate  $\hat{\Sigma}_\varepsilon$  using (4). To obtain this estimate, we assume the true variance  $V(\mathbf{x}) \equiv \text{Var}[\varepsilon(\mathbf{x})]$  is also represented by a spatial correlation model. Since  $V(\mathbf{x}_i)$  is not observable, even at the design points, we let  $\hat{V}(\mathbf{x}_i)$  stand in for it, and we fit a standard kriging model as if  $\hat{V}$  had no noise. Ankenman et al. (2010) show that the consequences of estimating  $V$  in this way are slight as long as the  $n_i$  are not too small.

### 3 DESIGN STRATEGIES AND ALGORITHMS

The primary focus of this section is to produce stochastic kriging predictors with bounded relative error over the entire experimental region. In this study, only sequential methods are considered because single-stage methods fix computational effort a priori and do not allow for adjusting the simulation effort to ensure that a desired precision level is achieved. Sacks et al. (1989) classify sequential methods into two categories based on whether or not the design adapts to additional knowledge about the response surface. Sequential *adaptive* methods are intended to allocate more computational resources to areas of the experimental region with high variability or intricate features. In contrast, a sequential method *without adaptation* uses static allocation at every stage regardless of what is known. Section 3.1 gives examples for both types of sequential methods that are designed to produce predictors with bounded relative error.

The relative error of the SK predictor at a given  $\mathbf{x}_0 \in \mathcal{X}$  can be written as

$$\frac{|\hat{Y}(\mathbf{x}_0) - Y(\mathbf{x}_0)|}{|Y(\mathbf{x}_0)|}. \quad (6)$$

Let  $\gamma$  be the prespecified precision level; then our objective is to achieve predictors that satisfy

$$\max_{\mathbf{x}_0 \in \mathcal{X}} \frac{|\hat{Y}(\mathbf{x}_0) - Y(\mathbf{x}_0)|}{|Y(\mathbf{x}_0)|} \leq \gamma, \quad (7)$$

conditional on a given response surface  $Y$ .

However, since  $Y$  is unknown, relative error given by (6) can not be calculated. Recognizing this, Law (2007) describes a method to statistically bound relative error at a single design point. In Section 3.2, we build on this method to suggest a stopping criterion, which is based on the mean squared error (MSE) estimator of SK prediction given in Equation (5), to bound the relative error with some confidence over the entire design space. Finally, in Section 3.3 we formulate an optimization problem for allocation of computational resources at each stage of a sequential adaptive method.

### 3.1 Two Algorithms

In this section, we outline two algorithms, a sequential adaptive method and a sequential method without adaptation, each with two phases. The initial phase consists of a fixed design  $(\mathbf{x}_i, n_i)$ ,  $i = 1, 2, \dots, k$ , simulation of it, and a resulting SK predictor. If the initial SK fit produces a satisfactory predictor then there is no need for the sequential phase. The sequential phase allocates an incremental computational budget of  $N_t$  on iteration  $t$ , then runs simulations and updates the predictor. Allocation of the computational budget is restricted to candidate design points consisting of two sets of space-filling points,  $\mathcal{S}_1$  and  $\mathcal{S}_2$ . In this study, both  $\mathcal{S}_1$  and  $\mathcal{S}_2$  are Latin hypercube designs (LHD) with good space-filling properties (e.g., maximin LHD). Space-filling designs are desirable to enable exploration and discovery of surprising behavior when the form of the response surface is unknown. In addition, constraining candidate design points to  $\mathcal{S}_1 \cup \mathcal{S}_2$  avoids clustering often seen in sequential methods. Particulars of the algorithms are given next.

#### Sequential adaptive method

- 1 Allocate  $n_0$  replications to each point in the initial design  $\mathcal{S}_1$ .
- 2 Make initial simulation runs and fit SK to produce the initial fit  $\hat{Y}_0$ . If the stopping criterion is satisfied, accept the current fit and terminate the algorithm. Otherwise, set  $t = 1$  and proceed to the next step.
- 3 Allocate  $N_t$  additional observations to one of the candidate design points in  $\mathcal{S}_1 \cup \mathcal{S}_2$ , based on the design criterion advanced in Section 3.2.
- 4 Make additional simulation runs and fit SK to produce  $\hat{Y}_t$ . If the predictor satisfies the stopping criterion, then accept it and terminate the algorithm. Otherwise, increment  $t$  and go to Step 3.

#### Sequential method without adaptation

- 1 Allocate  $n'_0$  replications to each point in  $\mathcal{S}_1 \cup \mathcal{S}_2$ .
- 2 Make initial simulation runs and fit stochastic kriging to produce the initial fit  $\hat{Y}_0$ . If the stopping criterion is satisfied, accept the current fit and terminate the algorithm. Otherwise, set  $t = 1$  and proceed to the next step.
- 3 Allocate  $n'_t$  additional replications to each point in  $\mathcal{S}_1 \cup \mathcal{S}_2$ .
- 4 Make additional simulation runs and fit SK to produce  $\hat{Y}_t$ . If the predictor satisfies the stopping criterion, then accept it and terminate the algorithm. Otherwise, increment  $t$  and go to Step 3.

### 3.2 Stopping Criterion

We define a measure based on the MSE estimator of the metamodel called maximum relative root mean squared error, which at the end of iteration  $t$  is given by

$$\max_{\mathbf{x}_0 \in \mathcal{X}} \frac{\sqrt{\widehat{\text{MSE}}_t(\mathbf{x}_0)}}{|\hat{Y}_t(\mathbf{x}_0)|}.$$

Using a fine grid of points  $\mathcal{P} \in \mathcal{X}$  this can be approximated as

$$\text{MrRMSE}_t = \max_{\mathbf{x}_0 \in \mathcal{P}} \frac{\sqrt{\widehat{\text{MSE}}_t(\mathbf{x}_0)}}{|\hat{Y}_t(\mathbf{x}_0)|}. \quad (8)$$

Based on (8) stop when

$$z_{\alpha/2} \text{MrRMSE}_t \leq \gamma', \quad (9)$$

where  $\gamma'$  is a function of  $\gamma$  such that a prespecified relative precision is realized pointwise with some confidence.

Following Law (2007), we write

$$1 - \alpha \approx \Pr \left( |\hat{Y}_t(\mathbf{x}_0) - Y(\mathbf{x}_0)| \leq z_{\alpha/2} \sqrt{\widehat{\text{MSE}}_t(\mathbf{x}_0)} \right),$$

where  $z_{\alpha/2} \sqrt{\widehat{\text{MSE}}_t(\mathbf{x}_0)}$  is the half-width of an approximate  $(1 - \alpha)100\%$  percent ( $0 < \alpha < 1/2$ ) confidence interval for  $Y(\mathbf{x}_0)$  on iteration  $t$  and  $z_{\alpha/2}$  is the standard normal quantile corresponding to the probability of  $1 - \alpha/2$ . Thus,

$$\begin{aligned} 1 - \alpha &\approx \Pr \left( |\hat{Y}_t(\mathbf{x}_0) - Y(\mathbf{x}_0)| / |\hat{Y}_t(\mathbf{x}_0)| \leq z_{\alpha/2} \sqrt{\widehat{\text{MSE}}_t(\mathbf{x}_0)} / |\hat{Y}_t(\mathbf{x}_0)| \right) \\ &\leq \Pr \left( |\hat{Y}_t(\mathbf{x}_0) - Y(\mathbf{x}_0)| \leq z_{\alpha/2} \text{MrRMSE}_t |\hat{Y}_t(\mathbf{x}_0)| \right) \quad [\text{from (8)}] \\ &\leq \Pr \left( |\hat{Y}_t(\mathbf{x}_0) - Y(\mathbf{x}_0)| \leq \gamma' |\hat{Y}_t(\mathbf{x}_0)| \right) \quad [\text{from (9)}] \\ &= \Pr \left( |\hat{Y}_t(\mathbf{x}_0) - Y(\mathbf{x}_0)| \leq \gamma' |\hat{Y}_t(\mathbf{x}_0) - Y(\mathbf{x}_0) + Y(\mathbf{x}_0)| \right) \quad [\text{add, subtract } Y(\mathbf{x}_0)] \\ &\leq \Pr \left( |\hat{Y}_t(\mathbf{x}_0) - Y(\mathbf{x}_0)| \leq \gamma' (|\hat{Y}_t(\mathbf{x}_0) - Y(\mathbf{x}_0)| + |Y(\mathbf{x}_0)|) \right) \quad [\text{triangle inequality}] \\ &= \Pr \left( |\hat{Y}_t(\mathbf{x}_0) - Y(\mathbf{x}_0)| / |Y(\mathbf{x}_0)| \leq \gamma' / (1 - \gamma') \right). \end{aligned}$$

Thus, for (7) to hold for  $\mathbf{x}_0 \in \mathcal{P}$ , we set  $\gamma = \gamma' / (1 - \gamma')$  implying  $\gamma' = \gamma / (1 + \gamma)$  and the stopping criterion in (9) can be restated as

$$z_{\alpha/2} \text{MrRMSE}_t \leq \frac{\gamma}{1 + \gamma}. \quad (10)$$

### 3.3 Design Criterion

MrRMSE is used to determine when to stop the sequential methods. In particular, iterations continue until MrRMSE is low enough to satisfy the stopping criterion given by (10). Thus, it is natural to focus on reducing MrRMSE while making allocation decisions at every stage of the sequential adaptive method (sequential methods without adaptation use static allocation rather than allocation based on a design criterion).

A simple design strategy targeted to reducing MrRMSE is to allocate  $N_{t+1}$ , the incremental budget on iteration  $t + 1$ , to the candidate design point that had the highest relative RMSE in the last iteration; i.e., to choose

$$\operatorname{argmax}_{\mathbf{x}_i \in \mathcal{S}_1 \cup \mathcal{S}_2} \frac{\sqrt{\widehat{\text{MSE}}_t(\mathbf{x}_i)}}{|\hat{Y}_t(\mathbf{x}_i)|}.$$

Notice that the candidate point with the highest relative RMSE may not be the point with the maximum actual relative error. Also note that, even choosing the point with the maximum relative error would not necessarily lead to the greatest reduction on the left-hand side of (7), the real objective. The following design strategy picks the point that, we believe, will reduce MrRMSE the most.

Suppose that there are  $K$  candidate points in  $\mathcal{S}_1 \cup \mathcal{S}_2$  and let  $n_t = (n_{1t}, n_{2t}, \dots, n_{Kt})$  be a vector of the number of replications allocated to these points up to iteration  $t$ . The design problem on iteration  $t + 1$  is to choose the candidate point that is expected to improve MrRMSE the most when the incremental budget  $N_{t+1}$  is allocated to it. We consider (but do not yet simulate) allocating the budget to each candidate point separately and calculate

$$\max_{\mathbf{x}_0 \in \mathcal{P}} \frac{\sqrt{\widehat{\text{MSE}}_t(\mathbf{x}_0; n_{t+1})}}{|\hat{Y}_t(\mathbf{x}_0)|}$$

where  $\widehat{Y}_t(\mathbf{x}_0)$  is the predictor from iteration  $t$  and  $\widehat{MSE}_t(\mathbf{x}_0; n_{t+1})$  is the prediction of what  $\widehat{MSE}_t(\mathbf{x}_0)$  would be on iteration  $t + 1$  at a given allocation  $n_{t+1}$ . Notice that even though no simulations are run yet at the allocation step of iteration  $t + 1$ ;  $\widehat{MSE}_t(\mathbf{x}_0; n_{t+1})$  can be calculated by updating  $R_M(\mathbf{x}_0, \cdot; \widehat{\theta})$ ,  $R_M(\widehat{\theta})$  and  $\widehat{\Sigma}_\epsilon$  in (5) with  $n_{t+1}$ .

Then, the design problem on iteration  $t + 1$  can be formulated as a mathematical program:

$$\begin{aligned} \underset{n_{t+1}}{\text{minimize}} \quad & \max_{\mathbf{x}_0 \in \mathcal{P}} \frac{\sqrt{\widehat{MSE}_t(\mathbf{x}_0; n_{t+1})}}{|\widehat{Y}_t(\mathbf{x}_0)|} & (11) \\ \text{subject to} \quad & n_{i,t+1} \leq n_{it} + N_{t+1} I_{i,t+1} \quad (i = 1, 2, \dots, K) \\ & \sum_{i=1}^K I_{i,t+1} = 1 \\ & I_{i,t+1} \in \{0, 1\} \quad (i = 1, 2, \dots, K). \end{aligned}$$

The denominator of the objective function is independent of the decision variables  $n_{t+1}$ , because we used the predictor from step  $t$  of the surface.

An alternative approach would be to use  $\widetilde{Y}_t(\mathbf{x}_0; n_{t+1})$ , the prediction of what the surface would be on iteration  $t + 1$  at a given allocation  $n_{t+1}$ , in the denominator of (11). If no additional points enter the design on iteration  $t + 1$ , this prediction can be written as follows:

$$\widetilde{Y}_t(\mathbf{x}_0; n_{t+1}) = \beta_0 + \Sigma_M(\mathbf{x}_0, \cdot)^\top [\Sigma_M + \Sigma_\epsilon(n_{t+1})]^{-1} (\mathcal{Y} - \beta_0 \mathbf{1}_k). \quad (12)$$

Notice that only  $\Sigma_\epsilon(n_{t+1})$  changes with different allocations of the incremental budget. That is, our prediction of the surface is revised simply by reweighting the current responses,  $\mathcal{Y}$ , with no new information. Similarly, if an additional point is going to enter the design, the prediction at that point will also get some weight in forming the prediction  $\widetilde{Y}_t(\mathbf{x}_0; n_{t+1})$ , even though it is not based on any new information. This allows for the possibility of achieving the goal in (11) by reweighting the data so as to inflate  $\widetilde{Y}_t(\mathbf{x}_0; n_{t+1})$  and (apparently) reduce relative error. This is clearly a fictitious improvement. The following numerical example illustrates for one particular case that this is indeed likely to occur.

Suppose there are two design points  $x_1$  and  $x_2$  at the two extremes of an interval,  $[0, 1]$ . Further suppose that there is constant variance  $V(x) = 0.25$  over the interval, the process variance  $\tau^2$  is 1, there exists Gaussian correlation  $r(h) = \exp(-h^2)$  between points that are a distance of  $h$  apart, and prior to iteration  $t + 1$ , both points have had 10 replications with sample mean responses of  $\mathcal{Y}(x_1) = 5$  and  $\mathcal{Y}(x_2) = 1$ . On iteration  $t + 1$ , we consider 3 allocation possibilities:  $n_{t+1} = (30, 10)$ ,  $n_{t+1} = (10, 30)$  or  $n_{t+1} = (20, 20)$ , i.e., allocating  $N_{t+1} = 20$  only to  $x_1$ , only to  $x_2$ , or splitting it equally between the two. Figure 1 shows predictions  $\widetilde{Y}_t(x_0; n_{t+1})$  for these 3 alternatives and illustrates that allocating a higher percentage of the replication budget to  $x_1$  elevates  $\widetilde{Y}_t(x_0; n_{t+1})$ .

## 4 EXPERIMENT DESIGN FOR EVALUATION

In this section, we illustrate our design strategy by producing SK predictors for one and two-dimensional surfaces using the algorithms outlined in Section 3.

### 4.1 One-dimensional Surfaces

We defined four one-dimensional surfaces,  $Y(x)$ , in the interval  $\mathcal{X} = [1, 2]$  for experimentation: (1) a monotone increasing function, (2) a concave parabola, (3) a convex parabola and (4) a function with both a local maximum and a local minimum (See Figure 2). An experiment also requires specification of the variance over the design space. For each surface, we define variance surfaces  $V(x)$  that are (directly/inversely) proportional to  $Y(x)$  as well as constant variance  $V$  over the interval.

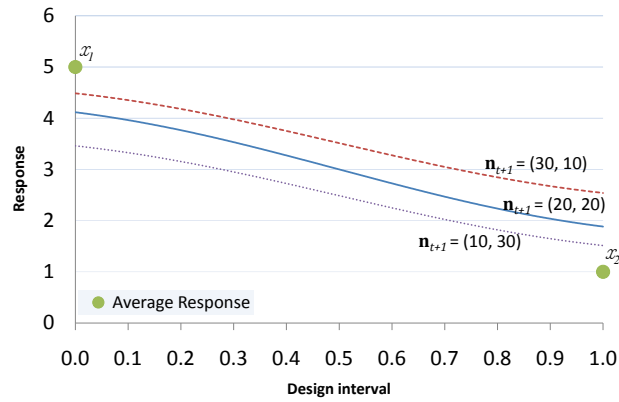


Figure 1: SK predictors for 3 allocation possibilities to 2 candidate design points on iteration  $t + 1$ .

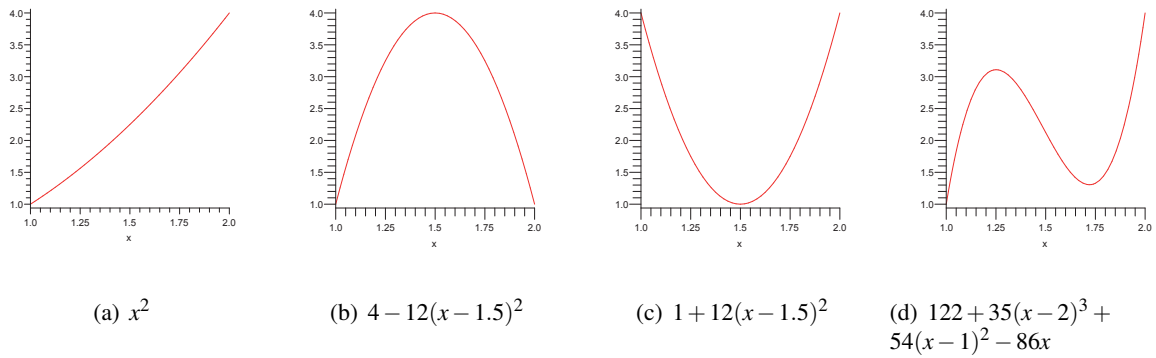


Figure 2: Four one-dimensional surfaces.

The design points are selected from candidate sets  $\mathcal{S}_1 = \{1, 1.25, 1.5, 1.75, 2\}$  and  $\mathcal{S}_2 = \{1.125, 1.375, 1.625, 1.875\}$ . Note that a LHD in one dimension is a set of evenly distributed points. In the initial phase of the sequential adaptive method, each point in  $\mathcal{S}_1$  receives 6 replications and on each iteration of the sequential phase, a budget of  $N_t = 6$  replications is allocated to one of the points in  $\mathcal{S}_1 \cup \mathcal{S}_2$ . On the other hand, each point in  $\mathcal{S}_1 \cup \mathcal{S}_2$  receives 6 replications in the initial phase of the sequential method without adaptation, and 1 more replication on each iteration of the sequential phase.

In this illustration, the output of a replication at a design point  $x_i$  is a value drawn from a normal distribution with mean  $Y(x_i)$  and variance  $V(x_i)$ . Depending on the form of the variance, the standard deviation at a design point  $x_i$  is one of the following:

$$\begin{aligned} \sigma(x_i) &= c_1, && \text{if variance is constant,} \\ \sigma(x_i) &= c_2 \cdot Y(x_i), && \text{if } V(x_i) \text{ is directly proportional to } Y(x_i), \\ \sigma(x_i) &= c_3 / Y(x_i), && \text{if } V(x_i) \text{ is inversely proportional to } Y(x_i), \end{aligned} \tag{13}$$

where  $c_1, c_2$  and  $c_3$  are constants. To keep the magnitude of the variance surface relatively stable while changing the form of the variance, we fix  $c_2$  and choose  $c_1$  and  $c_3$  such that the integral of the coefficient of variation over the design space is equal to  $c_2$ , i.e.,

$$\int_1^2 \frac{\sigma(x_0)}{|Y(x_0)|} dx_0 = c_2.$$

For each surface in Figure 2, two sets of constants corresponding to  $c_2 = 0.10$  and  $c_2 = 0.25$ , respectively, are used to generate a total of 6 variance surfaces. These constants are reported in Table 1. Experiments are made for each response surface–variance surface combination, totaling 24 experiments, and results are presented in Section 5.

Table 1: Constants that specify the variance surface.

$Y(x)$	Low Variance			High Variance		
	$c_1$	$c_2$	$c_3$	$c_1$	$c_2$	$c_3$
$x^2$	0.20	0.10	0.34	0.50	0.25	0.86
$4 - 12(x - 1.5)^2$	0.26	0.10	0.58	0.66	0.25	1.45
$1 + 12(x - 1.5)^2$	0.17	0.10	0.23	0.41	0.25	0.59
$122 + 35(x - 2)^3 + 54(x - 1)^2 - 86x$	0.20	0.10	0.35	0.50	0.25	0.88

### 4.2 Two-dimensional Surface

A second set of experiments involve a six-hump camel-back surface Branin (1972), i.e.,

$$Y(\mathbf{x}) = Y(x_1, x_2) = \left(4 - 2.1x_1^2 + \frac{x_1^4}{3}\right)x_1^2 + x_1x_2 + (-4 + 4x_2^2)x_2^2,$$

where  $\mathcal{X} = [-1, 1] \times [-1, 1]$ . The true surface is plotted in Figure 3 on  $200 \times 200 = 40,000$  equally spaced points. Candidate sets  $\mathcal{S}_1$  and  $\mathcal{S}_2$  are maximin LHDs with 25 and 81 candidate design points respectively, and constructed using MATLAB’s `lhsdesign` function. Two sample sets are displayed in Figure 4.

In the initial phase of the sequential adaptive method, each point in  $\mathcal{S}_1$  receives 20 replications and on each iteration of the sequential phase, a budget of  $N_t = 20$  replications is allocated to one of the points in  $\mathcal{S}_1 \cup \mathcal{S}_2$ . On the other hand, each point in  $\mathcal{S}_1 \cup \mathcal{S}_2$  receives 20 replications in the initial phase of the sequential method without adaptation, and 1 more replication on each iteration of the sequential phase.

As in Section 4.1, we define variance surfaces  $V(\mathbf{x})$  that are (directly/inversely) proportional to  $Y(\mathbf{x})$  as well as constant variance  $V$  over the interval. In this two dimensional surface we set  $c_1 = c_2 = c_3 = c$



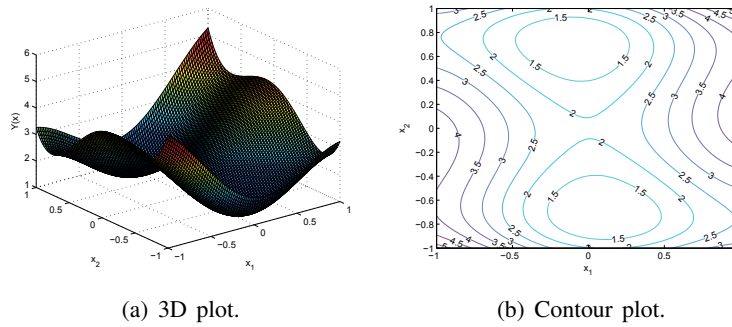


Figure 3: Six-hump camel-back surface.

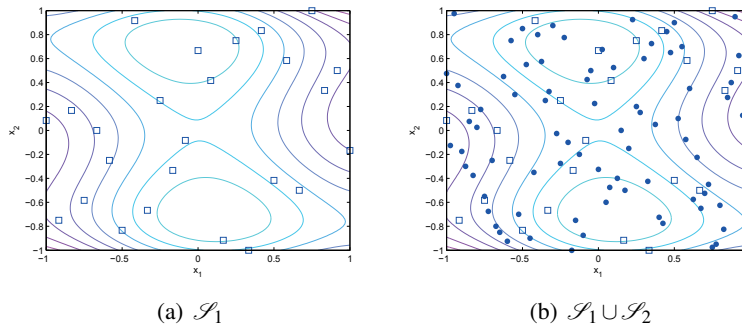


Figure 4: Candidate design points.

and experiment with  $c = 0.2$ , and  $c = 0.4$ , i.e., for each variance form, we experiment with 2 variance magnitudes, totaling 6 variance surfaces. Since the performance of the algorithms also depend on the choice of candidate design points, 2 different  $\mathcal{S}_1$  and  $\mathcal{S}_2$  sets are generated. Results from 12 experiments are also presented in Section 5.

## 5 RESULTS

For each of the scenarios specified in Section 4, we ran 10 macroreplications of both algorithms. Substituting a precision level of  $\gamma = 0.10$  and a confidence level of  $(1 - \alpha)100\% = 99\%$  in (10), the stopping criterion we used is given by

$$2.5758 \text{MrRMSE}_t \leq 0.0909. \tag{14}$$

Each macroreplication is stopped as soon as the condition in (14) is satisfied and MrRMSE, the maximum relative error, and the total computational resources used are recorded. For each scenario, the average and the half-width of the 99% CI of these metrics are reported.

Tables 2 and 4 present the output from our numerical study with 4 one-dimensional surfaces using, respectively, the sequential adaptive method and the sequential method without adaptation. Each row of these tables corresponds to one of the 24 response surface–variance surface pairs. The first column shows which response surface is selected, while the second and the third columns identify the variance surface created by (13) based on the constants reported in Table 1. The other three columns are, respectively, the MrRMSE, the maximum relative error, and the total number of replications needed to stop. Notice that both algorithms consistently achieve maximum true relative error that is on average less than the prespecified precision level of 10%. It is also evident from the last column that the sequential adaptive method stops sooner than the sequential method without adaptation, using fewer resources. Finally, the required precision is achieved faster when the variance is low.

Tables 3 and 5 are similar in format to Tables 2 and 4 except that there is only one response surface (a six-hump camel-back surface) with 2 distinct choices of candidate design points labeled  $DOE_1$  and  $DOE_2$ ; the first column identifies this choice. In this study, neither algorithm consistently achieves maximum true relative error that is always less than prespecified precision level of 10%. Nevertheless, when they miss the target, the highest relative error observed is not much larger than the prespecified precision level. Observe that the choice of candidate design points is the most significant factor in achieving bounded maximum relative error with minimal computational effort. The form and the magnitude of the variance surface influences how many replications are needed to stop, but is secondary to the choice of candidate sets  $\mathcal{S}_1$  and  $\mathcal{S}_2$  in effecting whether the maximum relative error is less than the prespecified precision level when the algorithm stops. Recall that the candidate sets are composed of space-filling points and not constructed by a design criterion. Rather, the design criterion is used to allocate replications. Therefore, if no candidate design point corresponds to a critical portion of the response surface or to a high variance region, then regardless of the computational effort the maximum relative error is not reduced to that required by the precision level (see, for example, the scenarios in which  $DOE_1$  is chosen and the variance surface is proportional to the response surface). It is difficult to consistently choose a good experimental design because the true relative error is not known to the experimenter. However, one way to ensure that bounded maximum relative error is achieved is to choose the number of points in  $\mathcal{S}_1 \cup \mathcal{S}_2$  large to allow for budget allocation to more of the critical regions.

Table 2: Sequential adaptive method used to fit predictors for 4 one-dimensional surfaces. Average MrRMSE, max relative error and replication budget spent are presented with their half-widths.

Surface	c	Variance Form	MrRMSE (%)	Max. Rel. Error(%)	Number of Reps
$x^2$	Low	constant	$3.32 \pm 0.04$	$4.12 \pm 0.31$	$59 \pm 4$
		proportional	$2.49 \pm 0.42$	$4.58 \pm 0.89$	$32 \pm 6$
		inverse	$3.41 \pm 0.02$	$4.65 \pm 0.70$	$118 \pm 10$
	High	constant	$3.48 \pm 0.01$	$5.82 \pm 0.42$	$292 \pm 11$
		proportional	$3.47 \pm 0.02$	$5.97 \pm 0.85$	$156 \pm 10$
		inverse	$3.51 \pm 0.00$	$5.93 \pm 0.65$	$678 \pm 15$
$4 - 12(x - 1.5)^2$	Low	constant	$3.38 \pm 0.04$	$7.10 \pm 0.85$	$187 \pm 13$
		proportional	$3.30 \pm 0.06$	$6.51 \pm 0.55$	$65 \pm 5$
		inverse	$3.49 \pm 0.01$	$5.97 \pm 0.85$	$593 \pm 21$
	High	constant	$3.48 \pm 0.01$	$7.00 \pm 0.58$	$970 \pm 63$
		proportional	$3.49 \pm 0.01$	$6.86 \pm 0.90$	$310 \pm 23$
		inverse	$3.50 \pm 0.01$	$6.43 \pm 0.66$	$3682 \pm 150$
$1 + 12(x - 1.5)^2$	Low	constant	$3.39 \pm 0.04$	$4.91 \pm 0.62$	$64 \pm 5$
		proportional	$3.04 \pm 0.34$	$6.16 \pm 0.81$	$51 \pm 7$
		inverse	$3.37 \pm 0.05$	$6.49 \pm 0.70$	$74 \pm 6$
	High	constant	$3.50 \pm 0.01$	$6.06 \pm 0.34$	$278 \pm 13$
		proportional	$3.51 \pm 0.00$	$6.43 \pm 0.74$	$294 \pm 12$
		inverse	$3.47 \pm 0.02$	$7.92 \pm 0.69$	$422 \pm 22$
$122 + 35(x - 2)^3$ $+54(x - 1)^2 - 86x$	Low	constant	$3.36 \pm 0.07$	$7.77 \pm 0.38$	$104 \pm 6$
		proportional	$3.26 \pm 0.06$	$8.15 \pm 0.53$	$85 \pm 6$
		inverse	$3.44 \pm 0.02$	$7.58 \pm 0.39$	$180 \pm 6$
	High	constant	$3.46 \pm 0.02$	$8.76 \pm 0.56$	$730 \pm 20$
		proportional	$3.50 \pm 0.01$	$8.85 \pm 0.68$	$558 \pm 17$
		inverse	$3.37 \pm 0.05$	$9.24 \pm 0.57$	$1349 \pm 72$

Table 3: Sequential adaptive method used to fit predictors for six-hump camel-back surface. Average MrRMSE, max relative error and replication budget spent are presented with their half-widths.

Design	c	Variance Form	MrRMSE (%)	Max. Rel. Error(%)	Number of Reps
DOE <sub>1</sub>	Low	constant	3.50 ± 0.01	11.30 ± 0.60	3750 ± 110
		proportional	3.52 ± 0.00	11.93 ± 0.62	16358 ± 596
		inverse	3.27 ± 0.05	11.94 ± 0.82	982 ± 57
	High	constant	3.52 ± 0.00	10.71 ± 0.56	14826 ± 550
		proportional	3.53 ± 0.00	11.28 ± 0.29	64392 ± 1539
		inverse	3.50 ± 0.03	12.19 ± 0.37	3536 ± 212
DOE <sub>2</sub>	Low	constant	3.47 ± 0.02	8.70 ± 0.36	2048 ± 116
		proportional	3.51 ± 0.01	8.28 ± 0.31	8854 ± 369
		inverse	3.31 ± 0.03	8.70 ± 0.24	818 ± 8
	High	constant	3.52 ± 0.00	8.67 ± 0.33	8500 ± 250
		proportional	3.52 ± 0.00	8.65 ± 0.52	36628 ± 1347
		inverse	3.48 ± 0.02	8.68 ± 0.17	3286 ± 98

## 6 CONCLUSION

This paper has provided a methodology that sequentially allocates simulation runs to preselected candidate design points and has a potential to produce a complete performance response surface with bounded relative error using the stochastic kriging method proposed by Ankenman et al. (2010). We tested this methodology with various one-dimensional response surface–variance surface pairs as well as with a two-dimensional surface and different candidate design points. We learned that this methodology is able to produce SK predictors either with bounded maximum relative error or with maximum relative error that is close to the prespecified precision level, and the choice of candidate design points is critically important in the performance.

## REFERENCES

- Ankenman, B. E., B. L. Nelson, and J. Staum. 2010. “Stochastic kriging for simulation metamodeling”. *Operations Research* 58:371–382.
- Bratin, F. H. 1972. “Widely convergent method for finding multiple solutions of simultaneous nonlinear equations”. *IBM Journal of Research and Development* 16:504–522.
- Law, A. M. 2007. *Simulation Modeling and Analysis*. Fourth ed. McGraw-Hill, NY.
- Sacks, J., W. J. Welch, T. J. Mitchell, and H. Wynn. 1989. “Design and analysis of computer experiments”. *Statistical Science* 4:409–423.

## AUTHOR BIOGRAPHIES

**MUSTAFA H. TONGARLAK** is a Postdoctoral Fellow in the Department of Managerial Economics & Decision Sciences at Kellogg School of Management. He holds a PhD in Industrial Engineering & Management Sciences from Northwestern University. His research interests include design and analysis of stochastic simulation experiments, and sustainable operations.

**BRUCE ANKENMAN** is an Associate Professor in the Department of Industrial Engineering & Management Sciences at Northwestern University. His research interests include the statistical design and analysis of experiments. Although much of his work has been concerned with physical experiments, recent research has focused on computer simulation experiments.

*Tongarlak, Ankenman, and Nelson*

**BARRY L. NELSON** is the Walter P. Murphy Professor and Chair of the Department of Industrial Engineering and Management Sciences at Northwestern University. His research centers on the design and analysis of computer simulation experiments on models of stochastic systems.

Table 4: Sequential method without adaptation used to fit predictors for 4 one-dimensional surfaces. Average MrRMSE, max relative error and replication budget spent are presented with their half-widths.

Surface	c	Variance Form	MrRMSE (%)	Max. Rel. Error(%)	Number of Reps
$x^2$	Low	constant	$3.45 \pm 0.04$	$3.30 \pm 0.73$	$199 \pm 26$
		proportional	$2.71 \pm 0.45$	$3.67 \pm 0.86$	$54 \pm 10$
		inverse	$3.51 \pm 0.00$	$3.94 \pm 0.95$	$536 \pm 55$
	High	constant	$3.52 \pm 0.00$	$3.56 \pm 0.61$	$1355 \pm 62$
		proportional	$3.48 \pm 0.01$	$5.26 \pm 0.72$	$408 \pm 24$
		inverse	$3.53 \pm 0.00$	$2.09 \pm 0.50$	$3612 \pm 110$
$4 - 12(x - 1.5)^2$	Low	constant	$3.50 \pm 0.01$	$4.77 \pm 0.36$	$563 \pm 29$
		proportional	$3.42 \pm 0.02$	$5.65 \pm 0.65$	$104 \pm 11$
		inverse	$3.52 \pm 0.00$	$4.18 \pm 0.75$	$2297 \pm 92$
	High	constant	$3.52 \pm 0.00$	$4.34 \pm 0.69$	$3130 \pm 89$
		proportional	$3.50 \pm 0.00$	$5.67 \pm 0.52$	$484 \pm 14$
		inverse	$3.53 \pm 0.00$	$4.58 \pm 0.64$	$13830 \pm 300$
$1 + 12(x - 1.5)^2$	Low	constant	$3.45 \pm 0.03$	$5.39 \pm 0.68$	$99 \pm 9$
		proportional	$3.08 \pm 0.34$	$5.58 \pm 0.84$	$67 \pm 8$
		inverse	$3.46 \pm 0.02$	$4.38 \pm 0.52$	$143 \pm 14$
	High	constant	$3.52 \pm 0.00$	$3.95 \pm 0.39$	$617 \pm 29$
		proportional	$3.51 \pm 0.00$	$5.52 \pm 0.47$	$477 \pm 19$
		inverse	$3.52 \pm 0.00$	$4.92 \pm 0.85$	$1047 \pm 41$
$122 + 35(x - 2)^3 + 54(x - 1)^2 - 86x + 122$	Low	constant	$3.44 \pm 0.02$	$6.15 \pm 0.55$	$226 \pm 25$
		proportional	$3.46 \pm 0.02$	$7.83 \pm 0.45$	$118 \pm 7$
		inverse	$3.51 \pm 0.00$	$3.90 \pm 0.36$	$851 \pm 35$
	High	constant	$3.52 \pm 0.00$	$6.10 \pm 0.45$	$1748 \pm 63$
		proportional	$3.52 \pm 0.00$	$8.78 \pm 0.76$	$651 \pm 14$
		inverse	$3.53 \pm 0.00$	$3.45 \pm 0.14$	$5362 \pm 133$

Table 5: Sequential method without adaptation used to fit predictors for six-hump camel-back surface. Average MrRMSE, max relative error and replication budget spent are presented with their half-widths.

Design	c	Variance Form	MrRMSE (%)	Max. Rel. Error(%)	Number of Reps
DOE <sub>1</sub>	Low	constant	$3.51 \pm 0.00$	$12.60 \pm 0.54$	$2583 \pm 85$
		proportional	$3.53 \pm 0.00$	$12.85 \pm 0.45$	$19162 \pm 606$
		inverse	$2.51 \pm 0.01$	$9.40 \pm 0.46$	$2120 \pm 0$
	High	constant	$3.52 \pm 0.00$	$12.12 \pm 0.47$	$10181 \pm 322$
		proportional	$3.53 \pm 0.00$	$12.46 \pm 0.50$	$72585 \pm 2004$
		inverse	$3.51 \pm 0.00$	$11.87 \pm 0.33$	$3169 \pm 125$
DOE <sub>2</sub>	Low	constant	$3.37 \pm 0.05$	$8.99 \pm 0.30$	$2268 \pm 69$
		proportional	$3.53 \pm 0.00$	$9.19 \pm 0.45$	$11837 \pm 383$
		inverse	$2.16 \pm 0.02$	$6.42 \pm 0.22$	$2120 \pm 0$
	High	constant	$3.52 \pm 0.00$	$10.06 \pm 0.56$	$7724 \pm 528$
		proportional	$3.53 \pm 0.00$	$10.22 \pm 0.53$	$45177 \pm 1626$
		inverse	$3.48 \pm 0.02$	$8.06 \pm 0.20$	$2321 \pm 83$