Proceedings of the 2019 Winter Simulation Conference
N. Mustafee, K.-H.G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, eds.

BAYESIAN SEQUENTIAL EXPERIMENTAL DESIGN FOR STOCHASTIC KRIGING WITH JACKKNIFE ERROR ESTIMATES

Guowei Sun

Department of Mathematics
Institute for Systems Research
University of Maryland
College Park, MD 20742, USA

Yunchuan Li

Department of Electrical and Computer Engineering
Institute for Systems Research
University of Maryland
College Park, MD 20742, USA

Michael C. Fu

Robert H. Smith Business School
Institute for Systems Research
University of Maryland
College Park, MD 20742, USA

ABSTRACT

We propose a fully sequential experimental design procedure for stochastic kriging (SK) methodology of fitting unknown response surfaces from simulation experiments. The procedure first estimates the current SK model performance by jackknifing the existing data points. Then, an additional SK model is fitted on the jackknife error estimates to capture the landscape of the current SK model performance. Methodologies for balancing exploration and exploitation trade-off in Bayesian optimization are employed to select the next simulation point. Compared to experimental design procedures, our method is robust to the SK model specifications. We design a dynamic allocation algorithm, which we call kriging-based dynamic stochastic kriging (KDSK), and illustrate its performance through two numerical experiments.

1 INTRODUCTION

Kriging originated in the geostatistics community for analyzing data with spatial correlations (Koehler and Owen 1996; Cressie 1993). It was later extended for constructing metamodels in the design and analysis of deterministic computer experiments (Welch et al. 1990). More recently, the stochastic kriging methodology extended the kriging estimator to modeling outcomes of stochastic simulations by introducing the intrinsic noise, which could be reduced by having more simulation replications at the corresponding design points (Ankenman et al. 2010). Kriging, often referred to as Gaussian process regression in the machine learning community, is also the foundation for Bayesian optimization algorithms, which recently enjoyed great success in machine learning applications (Jones et al. 1998; Snoek et al. 2012).

In the context of simulation metamodeling, stochastic kriging (SK) methods builds a global estimate for the unknown function. A carefully designed experiment is crucial in ensuring the model performance. Common practice is to use static designs, such as the uniform design, Latin Hypercube Design (LHS), and maximum entropy designs (Koehler and Owen 1996). On the other hand, dynamical designs are expected to be much more efficient, as more resources can be allocated to regions where the SK model is believed to have poor performance. The work of Ng and Yin (2012), Chen and Zhou (2014), and Wang and Hu (2018) focuses on utilizing the posterior uncertainty estimates from the fitted SK model for selecting design

points. However, we illustrate in Section 2.3 that the posterior uncertainty estimates in SK models do not properly reflect the roughness of the unknown function and such policies often lead to a near uniform design. Van Beers and Kleijnen (2008) and Kleijnen and Beers (2004) developed bootstrap procedures where new data points are sampled from the fitted kriging model for evaluating potential design choices. Our approach differs from theirs, as we perform re-sampling of existing data points rather than from the established SK models, and therefore is more robust to SK model specifications.

In this paper, we propose a novel approach for sequential experimental design for SK models. We use a jackknife procedure to obtain a rough estimation of model prediction error at existing design points, and construct a second SK model on the error estimates to obtain a smooth landscape for the model performance. Value of information (VoI) measures developed for Bayesian optimization are then used for selecting the next design point. Our approach is robust to the hyper-parameter choice of an SK model, as it relies on jackknifing the existing data points for estimating model prediction error rather than on the estimated posterior distributions. To the best of our knowledge, this is the first such attempt to develop a dynamic experimental design procedure for building SK models.

The rest of this paper is organized as follows. Section 2 briefly reviews the SK model and techniques for its experimental design, and presents an example for motivating our approach. Section 3 introduces the jackknife error estimates and a kriging model that is used to search for new design points. Section 4 presents a sequential design algorithm. Section 5 illustrates its performance with two numerical experiments. Finally, we conclude the paper in Section 6.

2 STOCHASTIC KRIGING AND ITS EXPERIMENTAL DESIGN

2.1 Preliminaries on Stochastic Kriging

SK metamodels construct a response surface of an unknown function $y(\mathbf{x}) \in R$ for $\mathbf{x} \in \mathscr{H}$, where $\mathbf{x}$ is a d dimensional vector and $\mathscr{H}$ is a compact subset of $\mathbb{R}^d$. We assume the analytical form of $y(\mathbf{x})$ is unknown, but independent samples of its noisy observations could be obtained with simulation. The standard SK model assumes the output of the j-th simulation replication at the design point $\mathbf{x}$ can be modeled as

$$y_j(\mathbf{x}) = f(\mathbf{x})^T\beta + M(\mathbf{x}) + \varepsilon_j(\mathbf{x}), \tag{1}$$

where $f(\mathbf{x})$ is a feature vector at the point $\mathbf{x}$ and β is a fixed constant vector. The term $f(\mathbf{x})^T\beta$ describes a fixed trend in the unknown target function y, and empirical evidence shows that a constant term performs well in practice (Welch et al. 1990; Ankenman et al. 2010). The term $M(\cdot)$ is a second-order stationary zero mean Gaussian process, which models the deviation of the true function y from the fixed trend term $f(\mathbf{x})^T\beta$. The term $\varepsilon_j(\mathbf{x})$, often referred to as intrinsic noise, captures randomness from stochastic simulations. We work with the most basic setting where $\varepsilon_j(\mathbf{x})$ can be considered independent and identically distributed (i.i.d.) at each design point $\mathbf{x}$.

For any two design points $\mathbf{x}$ and $\mathbf{y}$, the SK model assumes the covariance between $M(\mathbf{x})$ and $M(\mathbf{y})$ is given by

$$\mathrm{Cov}(M(\mathbf{x}), M(\mathbf{y})) = \tau^2 R(d(\mathbf{x},\mathbf{y}), \gamma), \tag{2}$$

where τ is the stationary variance of the Gaussian process $M(\cdot)$, $d(\cdot) : \mathbb{R}^d \to \mathbb{R}^+$ describes the closeness of points $\mathbf{x}$ and $\mathbf{y}$ in the space $\mathscr{H}$ (e.g., a vector norm), the function $R(\cdot)$ is chosen such that $R(0,\gamma) = 1$ and $\lim_{d\to\infty} R(d,\gamma) = 0$, $\forall\gamma$, and the parameter γ controls the smoothness of the random field (i.e., the fitted response surface).

Given experimental design configuration $\{\mathbf{x_i}, n_i\}_{i=1}^k$, where $\mathbf{x_i}$'s are the design points at which to perform simulations and n_i is the number of simulation replications for $\mathbf{x}_i$, the samples $\mathscr{Y}_{ij}, i \le k, j \le n_i$ can be used to compute the sample averages $\bar{\mathscr{Y}} = \{\bar{\mathscr{Y}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\mathscr{Y}_{ij}\}_{i=1}^k$ as the target of SK model fitting. Denote $\mathbf{F} = [f(\mathbf{x}_1), f(\mathbf{x}_2), .., f(\mathbf{x}_k)]$ the matrix of feature vectors at the existing design points. Following the notation in Ankenman et al. (2010), let $\Sigma_M = [\mathrm{Cov}(M(\mathbf{x}_i), M(\mathbf{x}_j))]_{i,j\le k}$ be the $k \times k$ covariance

matrix of random variables $M(\mathbf{x}_i), i \leq k$, $\Sigma_\varepsilon = \text{diag}[\text{Var}(\frac{1}{n_i}\sum_{j=1}^{n_i}\varepsilon_{ij}(\mathbf{x}_i))]_{i\leq k} = \text{diag}[\frac{Var(\varepsilon(\mathbf{x}_i))}{n_i}]_{i\leq k}$ be the $k\times k$ diagonal matrix capturing the intrinsic noise under the iid noise assumption for the given experimental design, and $\Sigma_M(\mathbf{x}_0,\cdot) = [\text{Cov}(M(\mathbf{x}_0),M(\mathbf{x}_1)),\text{Cov}(M(\mathbf{x}_0),M(\mathbf{x}_2)),...,\text{Cov}(M(\mathbf{x}_0),M(\mathbf{x}_k))]^T$ be the $k\times 1$ vector representing the correlation between the random field at a potential new design point $M(\mathbf{x}_0)$ and that at the existing design points $\{M(\mathbf{x}_i), i\leq k\}$. Assuming β, $\Sigma_M, \Sigma_\varepsilon$ and $\Sigma_M(\mathbf{x}_0,\cdot)$ are known fixed quantities, the best predictor of $y(\mathbf{x}_0)$ that minimizes the mean squared error, which we denote as $\hat{y}(\mathbf{x}_0)$, is shown in Ankenman et al. (2010) to be

$$\hat{y}(\mathbf{x}_0) = f(\mathbf{x}_0)^T\beta + \Sigma_M(\mathbf{x}_0,\cdot)^T(\Sigma_M+\Sigma_\varepsilon)^{-1}(\bar{\mathscr{Y}} - F\beta) \tag{3}$$

with optimal Mean Squared Error (MSE)

$$\text{MSE}(\hat{y}(\mathbf{x}_0)) = \Sigma_M(\mathbf{x}_0,\mathbf{x}_0) - \Sigma_M(\mathbf{x}_0,\cdot)^T(\Sigma_M+\Sigma_\varepsilon)^{-1}\Sigma_M(\mathbf{x}_0,\cdot). \tag{4}$$

If the coefficients β are estimated with generalized least square regression, i.e.,

$$\hat{\beta} = (F^T(\Sigma_M+\Sigma_\varepsilon)^{-1}F)^{-1}F^T(\Sigma_M+\Sigma_\varepsilon)^{-1}\bar{\mathscr{Y}},$$

then the optimal predictor becomes

$$\hat{y}(\mathbf{x}_0) = f(\mathbf{x}_0)^T\hat{\beta} + \Sigma_M(\mathbf{x}_0,\cdot)^T(\Sigma_M+\Sigma_\varepsilon)^{-1}(\bar{\mathscr{Y}} - F\hat{\beta})$$

with MSE

$$\begin{aligned}&\text{MSE}(\hat{y}(\mathbf{x}_0)) = \Sigma_M(\mathbf{x}_0,\mathbf{x}_0) - \Sigma_M(\mathbf{x}_0,\cdot)(\Sigma_M+\Sigma_\varepsilon)^{-1}\Sigma_M(\mathbf{x}_0,\cdot)\\ &+(f(\mathbf{x}_0 - F^T(\Sigma_M+\Sigma_\varepsilon)^{-1})\Sigma_M(\mathbf{x}_0,\cdot))^T(F^T(\Sigma_M+\Sigma_\varepsilon)^{-1}F)^{-1}(f(\mathbf{x}_0 - F^T(\Sigma_M+\Sigma_\varepsilon)^{-1})\Sigma_M(\mathbf{x}_0,\cdot)).\end{aligned}$$

In cases where Σ_M and Σ_ε also need to be estimated, the MSE expression becomes intractable. We refer the readers to Welch et al. (1990), Kleijnen (2018), and Ankenman et al. (2010) for reviews of the original kriging methods and its stochastic kriging variation for stochastic simulation experiments. For simplicity, we use $\theta = (\beta,\tau,\gamma)$ to represent the hyperparameters for setting up an SK model in (1).

2.2 Experimental Design For SK

Experimental design refers to the placement of $\{\mathbf{x}_i, i\leq T\}$ in the design space $\mathscr{H}$ and the corresponding number of replication n_i, given a total simulation budget of T with $\sum n_i = T$. In this work, we focus on the search of $\mathbf{x}_i$ and assume $n_i = 1$. Let $D_k = \{(\mathbf{x}_i, y_i),\ 1\leq i\leq k\}$ denote a set of k observed data points. The performance of the SK estimator $\hat{y}$ fitted on D_k can be evaluated using the integrated mean squared error (IMSE) defined as

$$IMSE = \int_{\mathscr{H}}(\hat{y}(\mathbf{x}) - y(\mathbf{x}))^2 d\mathbf{x}. \tag{5}$$

Wang and Hu (2018) proved that IMSE will monotonically decrease if more data are inserted to D_k for SK models with known fixed θ. The experimental design problem can be formulated as an optimization problem for minimizing the IMSE with respect to (w.r.t.) the design choice. In both Wang and Hu (2018) and Chen and Zhou (2014), IMSE is estimated using MSE from fitted SK models. In Section 2.3, we illustrate the limit of such approaches through a motivating example: MSE from fitted Gaussian processes often fail to capture the observed shape of existing data points, therefore providing little information on the landscape of model performance.

2.3 A Motivating Example: Uninformative MSE in SK

Consider the problem of fitting the unknown function $y = \sin(3x)e^{-250(x-0.25)^2}$ on the interval $(0,1)$. For simplicity, we assume the observations are noiseless, i.e., $\varepsilon(x) = 0$ w.p. 1. We fit an SK model with a simple uniform design with design points $\{0, 0.1, \ldots, 0.9, 1\}$. The implementation details of the SK model are listed in Section 5.1. The same setup is used for generating Figures 1 to 3. Despite $y(0.1), y(0.2), y(0.3)$ having larger jumps in function values compared to $y(0.8), y(0.9), y(1)$, as illustrated in Figure 1 (b), the MSE from the fitted SK model is roughly uniform across the design space $(0,1)$. The observed shape at $\{y_i\}_{i=1}^{11}$ is not reflected in the posterior belief of prediction errors in the fitted SK model. The observation is consistent with the formula for MSE in Equation (4), since $\bar{\mathscr{Y}}$ does not appear in the expression. For addressing the issue, we use jackknife error estimates to capture the landscape of model prediction performance.

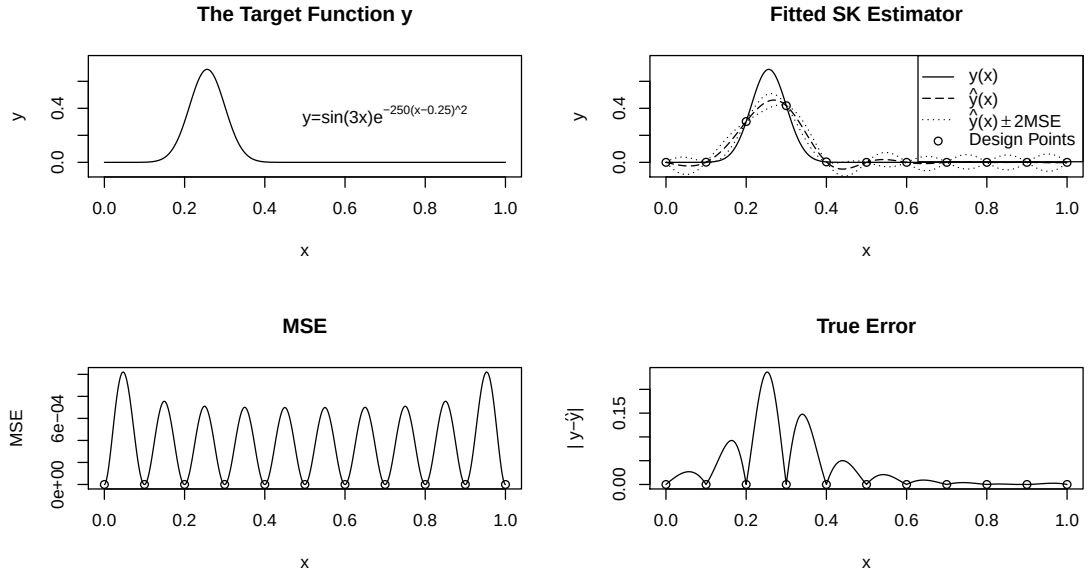

Figure 1: The SK model uses an exponential correlation kernel function ($\mathrm{Cov}(M(\mathbf{x}), M(\mathbf{y})) = \tau \exp(M(\mathbf{x}) - M(\mathbf{y}))^2/\sigma^2$) with hyperparameters estimated through maximum likelihood estimation. In (c), the MSE is uniform across $(0,1)$. In (d), the true prediction error is larger in regions with larger variation in the observed function values, which is not captured by MSE.

3 BAYESIAN EXPERIMENTAL DESIGN WITH JACKKNIFE ERROR ESTIMATES

Given current observations $D_k = \{(\mathbf{x}_i, y_i)\}_{i=1}^k$ and total budget T, we formulate our approach for dynamically allocating the remaining budget $T-k$. The intuition is to place the budget on regions where the current model is believed to have larger prediction error either due to insufficient sampling or higher roughness in the underlying target function.

3.1 Jackknife Prediction Error Estimates

Jackknife is a resampling technique where typically one observation is left out from an existing dataset for computing an estimate of an unknown target (Efron and Stein 1981). The true prediction error of an SK predictor $\hat{y}$ fitted with the data points D_k at the point $\mathbf{x}$, which we denote as $\delta(\mathbf{x})$, is

$$\delta(\mathbf{x}) = |\hat{y}(\mathbf{x}) - y(\mathbf{x})|.$$

Computation of $\delta(x)$ requires evaluating $y(\mathbf{x})$, which is assumed to be expensive. Therefore, we use a jackknife procedure by leaving $(\mathbf{x}_i, y_i)$ out from D_k and use the readily available y_i for computing an estimate of $\delta(\mathbf{x}_i)$.

Let $D_k[-i] = \{(\mathbf{x}_1, y_1), ..., (\mathbf{x}_{i-1}, y_{i-1}), (\mathbf{x}_{i+1}, y_{i+1}), .., (\mathbf{x}_k, y_k)\}$ denote the data set with $(\mathbf{x}_i, y_i)$ left out from the D_k. By fitting an SK model with the same θ as that in fitting $\hat{y}$ on data set $D_k[-i]$, a prediction of $y(\mathbf{x}_i)$ can be obtained, which we denote as $\tilde{y}_i(\mathbf{x}_i)$ Then, the jackknife error estimate at points x_i can be computed as

$$\Delta_i = |\tilde{y}_i(\mathbf{x}_i) - y_i|. \tag{6}$$

By using the same θ as that in fitting $\hat{y}$, Δ_i captures the performance of such SK model under the data set $D_k[-i]$. There are two issues with such an error estimator: (1) $\tilde{y}(\mathbf{x}_i)$ are estimated based on a smaller sample size, therefore Δ_i generally overestimates the prediction error of $\hat{y}$, and (2) SK model could be extremely sensitive to the data point $(\mathbf{x_i}, y_i)$ and return drastically different $\tilde{y}_i(\mathbf{x})$ and $\hat{y}(\mathbf{x})$. Despite the above issues, $\{\Delta_i\}_{i=1}^k$ still provide an indication of SK model performance on $\mathscr{H}$ and could help in searching for the next design point $\mathbf{x}_{k+1}$. As illustrated in Figure 2 (a), $\{\Delta_i\}_{i=1}^k$ align well with the unknown true prediction error of $\hat{y}$ for our motivating problem.

3.2 Modeling the Prediction Error Landscape

With $\{(\mathbf{x}_i, \Delta_i)\}_{i=1}^k$ obtained from the jackknife step, we construct a kriging model to represent our belief on the unknown prediction errors

$$\Delta(\mathbf{x}) = \mu^\Delta + M^\Delta(\mathbf{x}) + \varepsilon^\Delta(\mathbf{x}), \tag{7}$$

where $M^\Delta(\cdot)$ and $\tau^\Delta R^\Delta(\cdot, \theta^\Delta)$ follow the standard SK model setup outlined in Section 2.1. Using similar notation as that in the stochastic kriging model in Equations (3) and (4), given the existing design points and jackknife error estimates $\{(\mathbf{x}_i, \Delta_i)\}_{i=1}^k$, the MSE-optimal estimates of $\Delta(\mathbf{x}_0), \mathbf{x}_0 \in \mathscr{H},$ for $\mathbf{x}_0 \notin \{\mathbf{x}_i\}_{i=1}^k$ from the model in (7), denoted by $\tilde{\Delta}(\mathbf{x}_0)$, would be

$$\tilde{\Delta}(\mathbf{x}_0) = \mu^\Delta + \Sigma_{M^\Delta}(\mathbf{x}_0, \cdot)^T (\Sigma_{M^\Delta} + \Sigma_{\varepsilon^\Delta})^{-1} (\Delta - \mu^\Delta), \tag{8}$$

where

$$\begin{aligned} \Delta &= [\Delta_1, \Delta_2, \cdots, \Delta_k]^T, \\ \mu^\Delta &= [\mu^\Delta, \mu^\Delta, \cdots, \mu^\Delta]^T, \end{aligned}$$

with the MSE

$$\mathrm{MSE}(\tilde{\Delta}(\mathbf{x}_0)) = \Sigma_{M^\Delta}(\mathbf{x}_0, \mathbf{x}_0) - \Sigma_{M^\Delta}(\mathbf{x}_0, \cdot)^T (\Sigma_{M^\Delta} + \Sigma_{\varepsilon^\Delta})^{-1} \Sigma_{M^\Delta}(\mathbf{x}_0, \cdot). \tag{9}$$

A superscript of Δ is placed on all covariance matrices and vectors for clarification. $\tilde{\Delta}$ can be viewed as a posterior belief of the upper bounds for model prediction error of $\hat{y}$. When selecting the next design point $\mathbf{x}_{k+1}$, we use a myopic policy of the form

$$\mathbf{x}_{k+1} = \underset{\mathbf{z} \in \mathscr{H}, \mathbf{z} \notin \{\mathbf{x}_i\}_{i=1}^k}{\arg\max} \; g\left(\tilde{\Delta}(\mathbf{z}), \mathrm{MSE}(\tilde{\Delta}(\mathbf{z}))\right), \tag{10}$$

where g is a function for measuring the benefit of drawing an additional sample at $\mathbf{z}$. Following the terminology in the machine learning community, we call g the acquisition function. Instead of the greedy approach of setting the next design point to the maximizer of $\tilde{\Delta}$, we borrow ideas from Bayesian Optimization for balancing exploitation (sampling in regions where $\tilde{\Delta}(x)$ is large) and exploration (sampling in regions where $\tilde{\Delta}(x)$ has higher uncertainty). We introduce two common choices of acquisition functions: the probability of improvement (PI) and expected improvement (EI).

3.3 Probability of Improvement and Expected Improvement

Let $\Delta^* = \max\{\Delta_i\}_{i=1}^k$ denote the current maximum among the jackknife error estimates. Under the model in (7), for any $\mathbf{x} \in \mathscr{H}$, $\Delta(\mathbf{x})$ conditioned on $\{\mathbf{x}_i, \Delta_i\}_{i=1}^k$ is a Gaussian random variable with mean $\tilde{\Delta}(\mathbf{x})$ and variance $\text{MSE}(\tilde{\Delta}(\mathbf{x}))$ in Equations (8) and (9), respectively. Let $\Phi(\cdot)$ and $\phi(\cdot)$ be the cumulative distribution function and density function for the standard normal distribution, respectively, and g^{PI} and g^{EI} denote the two acquisition functions, respectively.

1. **PI** at point $\mathbf{x}$ is defined as $P\{\Delta(\mathbf{x}) \geq \Delta^*\}$, and has the analytical expression

$$g^{PI}(\mathbf{x}) = \Phi\left(\frac{\tilde{\Delta}(\mathbf{x}) - \Delta^*}{\sqrt{\text{MSE}(\tilde{\Delta}(\mathbf{x}))}}\right).$$

2. **EI** computes the expected value of improvement $(\tilde{\Delta}(\mathbf{x}) - \Delta^*)^+$ and is given by (Jones et al. 1998)

$$g^{EI}(\mathbf{x}) = (\Delta^* - \tilde{\Delta}(\mathbf{x}))\Phi\left(\frac{\Delta^* - \tilde{\Delta}(\mathbf{x})}{\sqrt{\text{MSE}(\tilde{\Delta}(\mathbf{x}))}}\right) + \sqrt{\text{MSE}(\tilde{\Delta}(\mathbf{x}))}\,\phi\left(\frac{\Delta^* - \tilde{\Delta}(\mathbf{x})}{\sqrt{\text{MSE}(\tilde{\Delta}(\mathbf{x}))}}\right),$$

where $(\cdot)^+$ is $\max(\cdot, 0)$.

Both PI and EI are popular choices of acquisition functions for balancing exploration and exploitation trade-off, and are shown to be successful in many stochastic optimization problems such as global optimization (or Bayesian optimization in the machine learning community) (Jones et al. 1998; Snoek et al. 2012), multi-armed bandits (Srinivas et al. 2012), and ranking and selection (Ryzhov 2016). In Figure 2 (b), both g^{PI} and g^{EI} select the next design points in regions where the true error is higher for our motivating example, whereas the point with maximum MSE of $\hat{y}(\mathbf{x})$ (i.e., $x^{MSE} = \arg\max_x y(\hat{\mathbf{x}})$), x^{MSE}, lies in the region where the true error is small.

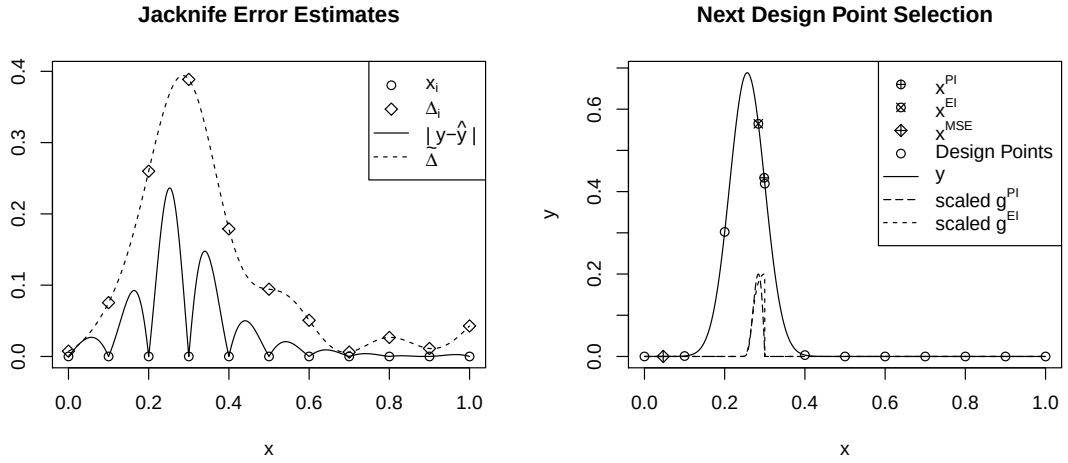

Figure 2: **(a)** Δ_i and $\tilde{\Delta}$ overestimate the true error $|y - \hat{y}|$ but capture the overall trend of $|y - \hat{y}|$. The regions where y is observed to be flat have smaller values of $\tilde{\Delta}$. **(b)** The g^{EI} and g^{PI} criteria choose the next point in the more interesting region, compared to the MSE based selection.

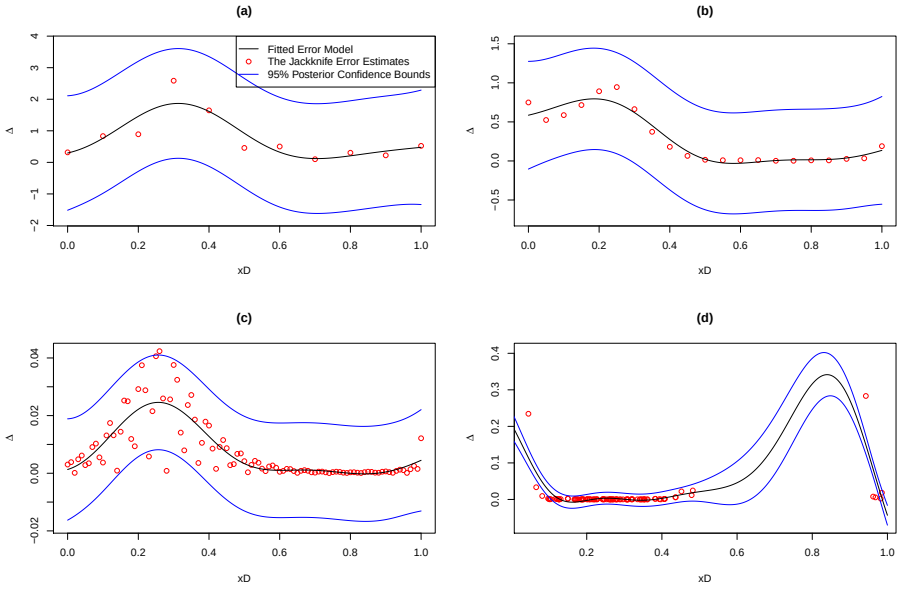

Figure 3: **(a):** 10 points uniformly spaced on the interval $[0,1]$. **(b):** 20 uniform design points. **(c):** 100 uniform design points. **(d):** 100 design points randomly selected to focus around 0.25 and 1.

3.4 Practical Model Fitting for Jackknife Error Estimates

In our empirical tests we found $\{\Delta_i\}$ to be extremely rough, even for a smooth $y(x)$. For standard SK models, the underlying parameters governing the assumed Gaussian process are often estimated through maximum likelihood estimation (Ankenman et al. 2010). Such approach tends to lead to a $\tilde{\Delta}$ that overfits to the jackknife error estimates $\{\Delta_i\}$. As the jackknife procedure only provides a noisy indication of model prediction error, we recommend building a smooth model on $\{(\mathbf{x}_i, \Delta_i)\}$ by setting stronger correlation matrix Σ_{M^Δ} and using noise covariance matrix $\Sigma_{\varepsilon^\Delta}$ with larger diagonal components. In Figure 3, we illustrate the jackknife error estimates for various design choices for the motivating problem outlined in Section 2.3. The model implementation details are listed in Section 5.1. By manually setting a smooth model for the jackknife error estimates, the model $\tilde{\Delta}$ captures the regions that would benefit the most from additional data sample in the 4 designs illustrated in Figure 3. For uniform designs in (a),(b),(c) of Figure 3, $\tilde{\Delta}$ captures the need for more samples at regions around 0.3, as the underlying target function exhibits sharp changes. For the design in (d) where budget is allocated around 0.3 and 1, $\tilde{\Delta}$ captures the need for more samples around 0.8, as we have not learned much about y in that region. A smooth $\tilde{\Delta}$ is often sufficient for capturing the overall landscape of model performance.

4 THE KDSK ALGORITHM

We summarize our approach and propose the KDSK algorithm for sequential experimental design for SK models. In Algorithm 1, superscript (t) represents the allocation steps and $D_m^{(t)}$ represents a data set with m data points. At the t-th iteration, t SK models will be constructed to obtain the jackknife error estimates, each with $(n_0+t)^3$ computational complexity, where n_0 is the number of initial samples. The optimization of the acquisition function could be non-trivial, especially for SK models on a higher dimensional space

(Jones et al. 1998). However, under the standard assumption that obtaining an output from the underlying target function y is expensive, the computation overhead of KDSK is justified for obtaining a better $\hat{y}$.

Algorithm 1: KDSK

Input: The initialization budget n_0 and the initial data $D_{n_0}^{(0)} = \{x_i, \mathscr{Y}_i\}_{i=1}^{n_0}$. The total remaining design budget k. SK model parameter choices θ for $\hat{y}$ and θ^{Δ} for $\tilde{\Delta}$. The acquisition function g.

Output: Final experimental design and observed values $\{x_i, \mathscr{Y}_i\}_{i=1}^{k}$ and a fitted SK model $\hat{y}$

Set $t \leftarrow 1$,
while $t \leq k$ **do**
  Jackknife $D_{n_0+t}^{(t)}$ to obtain data sets $\left\{D_{n_0+t-1}^{(t)}[-i]\right\}_{i=1}^{n_0+t}$
  Compute the $\{\Delta_i\}_{i=1}^{n_0+t}$ with Equation (6) with $\tilde{y}$ constructed according to θ
  Fit an SK model on $\{(x_i, \Delta_i)\}_{i=1}^{n_0+t}$ with model specification θ^{Δ}
  Select $\mathbf{x}_{t+n_0}$ using Equation (10)
  Evaluate the unknown function at x_{t+n_0} and obtain $\mathscr{Y}_{t+n_0}$
  Set $D_{n_0+t+1}^{(t+1)} \leftarrow \{(\mathbf{x}_{t+n_0}, \mathscr{Y}_{t+n_0})\} \cup D_{t+n_0}^{(t)}$
  Set $t \leftarrow t+1$
Construct an SK model $\hat{y}$ with data set $D_{n_0+k}^{k}$ and model parameter θ. Return $\hat{y}$.

5 NUMERICAL EXPERIMENTS

In this section, we illustrate the effectiveness of the proposed KDSK algorithm through two numerical experiments. Let the domain of interest $\mathscr{H}$ be $(0,1)$ and $\hat{y}$ denote the fitted SK model based on a design choice D. IMSE defined in Equation (5) is used to evaluate the quality of $\hat{y}$. For comparison purpose, we also implement two naive allocation policies, the UNIFORM policy and min-MSE policy. For UNIFORM allocation, the current total budget is uniformly allocated on $\mathscr{H}$; therefore it is not dynamic and has an advantage, as it has no initialization overhead. We use it as a benchmark and observe the benefit of having dynamic allocation algorithms. The *min-MSE* methods selects the next point to be the maximizer of the posterior variance of the current SK model. In both experiments, the dynamic algorithms are initialized with a uniform 11 design points, and IMSE is computed at each allocation step for illustrating the effectiveness of allocation algorithms.

5.1 The Motivating Deterministic Function

We test with the target function $y(x) = sin(3x)e^{-250*(x-0.25)^2}$ for $x \in (0,1)$. $\hat{y}$ is constructed in two ways: (1) with fixed and known θ, and (2) with θ estimated through maximum likelihood. The following setups are applied:

1. **model choices for $\hat{y}$:** For the fixed parameter experiment, the correlation is chosen to have the Gaussian kernel $R(d(x,y)) = \tau e^{-(x-y)^2/\sigma}$ with σ set to be 10 and τ equal to 1. The constant trend term μ is set to be 0. The noise covariance matrix is set to be $\Sigma_{\varepsilon} = diag(0.1)$. When constructing $\hat{y}$ with estimated parameter values, we obtain the fitted SK models with the **GPfit** software package which estimates θ by maximizing the maximum likelihood with a multi-start gradient based search (L-BFGS-B) algorithm (MacDonald et al. 2015). Note that in the second setting, the error covariance matrix is set to be 0; therefore the fitted SK model interpolates the existing data points; and
2. **model Choices for $\tilde{\Delta}$:** The correlation kernel is set to be Gaussian with $\tau^{\Delta} = 1, \sigma^{\Delta} = 1$. The trend term is set to be 0. We also include a noise term with $\Sigma_{\varepsilon^{\Delta}} = diag(0.005)$.

We use EI as the acquisition function. Additional 12 allocation steps are performed, with the numerical results illustrated in Figure 4. In both implementations, KDSK outperforms UNIFORM and min-MSE.

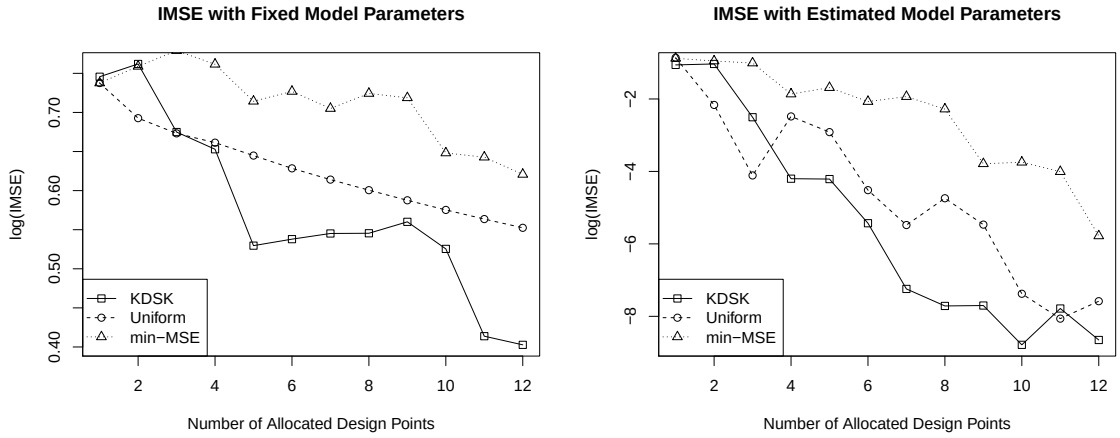

Figure 4: (a) $\hat{y}$ constructed with fixed parameters with noise. (b) $\hat{y}$ constructed with estimated parameters without error term.

5.2 The Steady-State M/M/1 Queue with Noise

Taking the example from Ankenman et al. (2010) for estimating the number of customers in an M/M/1 queueing system with service rate 1.02 and arrival rate $x \in (0,1)$. For simplicity, we use the known steady state results $y(x) = \frac{x}{1.02-x} + \varepsilon$ where ε is a normal noise with standard deviation 0.1. In this experiment, we test the performance of the KDSK algorithm with both the EI and PI activation function with the following setups:

1. **implementation of $\hat{y}$:** The correlation function is chosen to be the Gaussian Kernel $R(d(x,y)) = \tau e^{-(x-y)^2/\sigma}$ with $\tau = 1$ and $\sigma = 1$ and the constant term is set to be 0. The error covariance matrix is $\Sigma_\varepsilon = diag(0.01)$; and
2. **Implementation of $\tilde{\Delta}$:** We use the same setup as for $\hat{y}$. And we test with two choices of $\Sigma_{\varepsilon^\Delta}$: $diag(0.05)$ and $diag(0.01)$.

20 dynamic allocation steps are performed and the IMSE of the fitted $\hat{y}$ at each step is shown in Figure 5. The min-MSE approach has the worst performance. KDSK-PI and KDSK-EI have the best performance with KDSK-PI slightly outperforming KDSK-EI in both tests. We list the design choices of KDSK-PI and KDSK-EI in Table 1 with x_t denoting the choice of x at allocation step t. The allocation budget is placed heavily in the region where $y(x)$ has sharp changes.

Table 1: The dynamical design choices of KDSK algorithms.

	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	x_{20}	x_{21}
KDSK-EI	0.999	0.858	0.866	0.866	0.998	0.758	0.637	0.873	0.863	0.999	0.001
KDSK-PI	0.994	0.889	0.878	0.999	0.725	0.890	0.998	0.811	0.0142	0.892	0.868

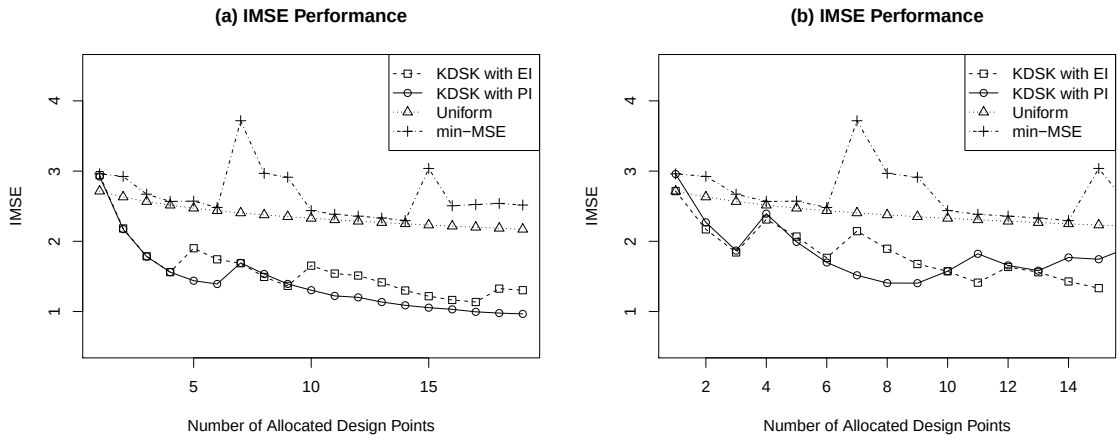

Figure 5: (a): $\Sigma_{\varepsilon^\Delta}$: $diag(0.05)$ in the error model G^Δ. (b)$\Sigma_{\varepsilon^\Delta}$: $diag(0.01)$.

6 CONCLUSION

In this paper, we propose a novel approach for the experimental design for the kriging methodology of fitting a global response surface for expensive black-box functions. Instead of relying on the posterior error estimates, which are subject to parameter tuning and model choice, we propose the idea of using a jackknife sampling procedure for establishing a landscape of model performance and perform sequential design point selection with Bayesian information criterion. The performance of our approach is illustrated through two numerical experiments. We discuss challenges for implementing the proposed KDSK algorithm, including the smoothness of jackknife error estimates and scaling issue due to the computational complexity. Our approach successfully captures the observed shape of the target function and adjusts the design choices accordingly, and simulation experiments indicate it is more efficient compared with uniform and MSE-based design methods.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation under Grant CMMI-1434419 and by the Defense Advanced Research Projects Agency (DARPA) under Grant N660011824024. The views, opinions and/or findings expressed are those of the authors and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

REFERENCES

Ankenman, B., B. L. Nelson, and J. Staum. 2010. "Stochastic Kriging for Simulation Metamodeling". *Operations Research* 58(2):371–382.

Chen, X., and Q. Zhou. 2014. "Sequential Experimental Designs for Stochastic Kriging". In *Proceedings of the 2014 Winter Simulation Conference*, edited by A. Tolk, L. Yilmaz, S. Y. Diallo, and I. O. Ryzhov, 3821–3832. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers Inc.

Cressie, N. 1993. *Statistics for Spatial Data*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. New York: J. Wiley.

Efron, B., and C. Stein. 1981, 05. "The Jackknife Estimate of Variance". *The Annals of Statistics* 9(3):586–596.

Jones, D. R., M. Schonlau, and W. J. Welch. 1998. "Efficient Global Optimization of Expensive Black-box Functions". *Journal of Global Optimization* 13(4):455–492.

Kleijnen, J. P., and W. v. Beers. 2004. "Application-driven Sequential Designs for Simulation Experiments: Kriging Metamodelling". *Journal of the Operational Research Society* 55(8):876–883.

Kleijnen, J. P. C. 2018. "Design and Analysis of Simulation Experiments". In *Statistics and Simulation*, edited by J. Pilz, D. Rasch, V. B. Melas, and K. Moder, 3–22. Cham: Springer.

Koehler, J., and A. Owen. 1996. "Computer Experiments". In *Handbook of Statisticss, 13: Design and Analysis of Experiments*, edited by S. Ghosh and C. R. Rao, Chapter 9, 261–308. Amsterdam: Elsevier.

MacDonald, B., P. Ranjan, and H. Chipman. 2015. "GPfit: An R Package for Fitting a Gaussian Process Model to Deterministic Simulator Outputs". *Journal of Statistical Software, Articles* 64(12):1–23.

Ng, S. H., and J. Yin. 2012, August. "Bayesian Kriging Analysis and Design for Stochastic Simulations". *ACM Transactions on Modeling and Computer Simulation* 22(3):17:1–17:26.

Ryzhov, I. O. 2016. "On the Convergence Rates of Expected Improvement Methods". *Operations Research* 64(6):1515–1528.

Snoek, J., H. Larochelle, and R. P. Adams. 2012. "Practical Bayesian Optimization of Machine Learning Algorithms". In *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'12, 2951–2959. Lake Tahoe, Nevada: Curran Associates Inc.

Srinivas, N., A. Krause, S. M. Kakade, and M. W. Seeger. 2012. "Information-theoretic Regret Bounds for Gaussian Process Optimization in the Bandit Setting". *IEEE Transactions on Information Theory* 58(5):3250–3265.

Van Beers, W. C., and J. P. Kleijnen. 2008. "Customized Sequential Designs for Random Simulation Experiments: Kriging Metamodeling and Bootstrapping". *European Journal of Operational Research* 186(3):1099–1113.

Wang, B., and J. Hu. 2018. "Some Monotonicity Results for Stochastic Kriging Metamodels in Sequential Settings". *INFORMS Journal on Computing* 30(2):278–294.

Welch, W. J., T.-K. Yu, S. M. Kang, and J. Sacks. 1990. "Computer Experiments for Quality Control by Parameter design". *Journal of Quality Technology* 22(1):15–22.

AUTHOR BIOGRAPHIES

GUOWEI SUN is a Ph.D. candidate in Applied Mathematics, Statistics and Scientific Computation in the Department of Mathematics at the University of Maryland, College Park. His interests lie in the broad area of stochastic optimization and applied probability. His email is gwsun@math.umd.edu.

YUNCHUAN LI is a Ph.D. candidate in the Department of Electrical and Computer Engineering at the University of Maryland, College Park. His research focuses on optimization and control with applications to operations research problems. His email is yli93@umd.edu.

MICHAEL C. FU holds the Smith Chair of Management Science in the Robert H. Smith School of Business, with a joint appointment in the Institute for Systems Research, A. James Clark School of Engineering, University of Maryland, College Park. He has a Ph.D. in applied math from Harvard and degrees in math and EECS from MIT. He attended his first WSC in 1988 and served as WSC2011 Program Chair, NSF Operations Research Program Director, *Management Science* Stochastic Models and Simulation Department Editor, and *Operations Research* Simulation Area Editor. He is a Fellow of **INFORMS** and **IEEE**. His e-mail address is mfu@umd.edu, and his Web page is https://www.rhsmith.umd.edu/directory/michael-fu.