Proceedings of the 2020 Winter Simulation Conference
K.-H. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, T. Roeder, and R. Thiesing, eds.

UNIFORM ERROR BOUNDS FOR STOCHASTIC KRIGING

Guangrui Xie
Xi Chen

Grado Department of Industrial and Systems Engineering
Virginia Tech
1145 Perry Street
Blacksburg, VA 24061, USA

ABSTRACT

In this paper, we propose an approach to construct uniform error bounds (or confidence intervals) for stochastic kriging with a prescribed confidence level. The theoretical development sheds some light on the impact of simulation experimental designs and budget allocation schemes as well as their relative importance on the large-sample properties of stochastic kriging. Through numerical evaluations, we demonstrate the superiority of the uniform error bounds to the simultaneous confidence intervals obtained by applying Bonferroni correction under various experimental settings.

1 INTRODUCTION

In the recent decades, research on metamodeling techniques for stochastic simulation experiments has received considerable attention from the stochastic simulation research community (Ankenman et al. 2010; Dellino et al. 2012; Ng and Yin 2012). Several metamodeling methodologies, such as stochastic kriging (SK, Ankenman et al. (2010)), have been proposed for approximating the *mean response surface* implied by a stochastic simulation under the impact of heteroscedasticity (i.e., the simulation output variance varies across the input space).

Built on Gaussian process (GP), SK is a popular approach suitable for heteroscedastic simulation metamodeling (Chen and Zhou 2017; Wang and Chen 2018). Similar to kriging, a popular methodology for design and analysis of deterministic computer experiments (Santner et al. 2003), a key element of SK prediction is the use of conditional inference based on GP. At each prediction point in the input space, the conditional distribution of a GP is normal with mean and variance in closed form. A pointwise confidence interval of the SK predictor with a prescribed coverage probability can be constructed based on this conditional distribution. In many applications, however, it is often desirable to have a joint confidence region of the SK predictor over an arbitrary set of prediction points with a prescribed simultaneous coverage probability. Existing approaches to construct a joint confidence region typically rely on bootstrapping and correction methods such as Bonferroni or Šidák (De Brabanter et al. 2011) and only apply to a finite number of prediction points. It would be more desirable to develop a method capable of providing reasonable bounds on the maximum approximation error achieved by the SK predictor across the input space. Such a bound can be useful in constructing confidence regions with a guaranteed simultaneous coverage probability, though somewhat conservatively.

In this paper, we propose a method to construct uniform error bounds of the SK predictor. A key implication of this work is that the large-sample properties of SK depend on the number of design points and their locations determined by the experimental design adopted as well as the budget allocation scheme implemented; moreover, the impact of the experimental design dominates that of the budget allocation scheme. The rest of the paper is organized as follows. In Section 2, we provide a brief review on SK.

978-1-7281-9499-8/20/$31.00 ©2020 IEEE

In Section 3, we provide the main results on constructing uniform error bounds for SK. In Section 4, we present some numerical experiments to demonstrate the performance of the proposed uniform error bounds.

2 REVIEW OF STOCHASTIC KRIGING

In SK, the system output obtained at a point $\mathbf{x} \in \mathscr{X} \subseteq \mathbb{R}^d$ on the jth simulation replication $\mathscr{Y}_j(\mathbf{x})$ is modeled as

$$\mathscr{Y}_j(\mathbf{x}) = f(\mathbf{x}) + \varepsilon_j(\mathbf{x}), \tag{1}$$

where $f(\mathbf{x})$ denotes the unknown true mean response that we intend to estimate at $\mathbf{x} \in \mathscr{X}$ and it is assumed that $f : \mathscr{X} \to \mathbb{R}$ represents a second-order stationary mean-zero Gaussian process (GP, Santner et al. (2003)). The spatial covariance between any two points in the GP is given by $\Sigma_{\mathsf{M}}(\mathbf{x}, \mathbf{x}')$, where $\Sigma_{\mathsf{M}} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$ denotes the kernel or covariance function. The simulation errors incurred at $\mathbf{x}$ on different replications, $\varepsilon_j(\mathbf{x})$'s, are assumed to be independent and identically distributed (i.i.d.) random variables with mean zero and input-dependent variance $\mathsf{V}(\mathbf{x}) \equiv \mathrm{Var}(\varepsilon_j(\mathbf{x}))$. We assume the normality of $\varepsilon_j(\mathbf{x})$ which could be anticipated since in a discrete-event simulation the simulation output $\mathscr{Y}_j(\mathbf{x})$ typically represents the average of a large number of more basic random variables obtained on the jth simulation replication.

Given a fixed simulation budget B to expend for approximating the mean response surface via SK, an experimental design for performing the simulation runs can be given as $\{(\mathbf{x}_i, n_i)_{i=1}^k : \sum_{i=1}^k n_i = B\}$, where k denotes the number of distinct design points selected from the input space $\mathscr{X}$, $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ denote the k design-point locations, and n_i represents the number of replications to apply at $\mathbf{x}_i$, $i = 1, 2, \ldots, k$. With the simulation outputs generated, one can obtain the SK predictor of $f(\mathbf{x}_0)$ at any $\mathbf{x}_0 \in \mathscr{X}$ as follows:

$$\mu_{k,B}(\mathbf{x}_0) = \Sigma_{\mathsf{M}}(\mathbf{x}_0, \mathbf{X})^\top (\Sigma_{\mathsf{M}}(\mathbf{X}, \mathbf{X}) + \Sigma_\varepsilon)^{-1} \bar{\mathscr{Y}},$$

and its corresponding predictive variance is given by

$$\sigma^2_{k,B}(\mathbf{x}_0) = \Sigma_{\mathsf{M}}(\mathbf{x}_0, \mathbf{x}_0) - \Sigma_{\mathsf{M}}(\mathbf{x}_0, \mathbf{X})^\top (\Sigma_{\mathsf{M}}(\mathbf{X}, \mathbf{X}) + \Sigma_\varepsilon)^{-1} \Sigma_{\mathsf{M}}(\mathbf{x}_0, \mathbf{X}),$$

where $\mathbf{X} = \left(\mathbf{x}_1^\top, \mathbf{x}_2^\top, \ldots, \mathbf{x}_k^\top\right)^\top$ denotes the $k \times d$ design matrix; $\bar{\mathscr{Y}} = \left(\bar{\mathscr{Y}}(\mathbf{x}_1), \bar{\mathscr{Y}}(\mathbf{x}_2), \ldots, \bar{\mathscr{Y}}(\mathbf{x}_k)\right)^\top$ denotes the $k \times 1$ vector of the sample averages of simulation outputs, with $\bar{\mathscr{Y}}(\mathbf{x}_i) = f(\mathbf{x}_i) + \bar{\varepsilon}(\mathbf{x}_i)$ and $\bar{\varepsilon}(\mathbf{x}_i)$ denoting the average random error incurred at $\mathbf{x}_i$, $i = 1, 2, \ldots, k$. With a slight abuse of notation, we use $\Sigma_{\mathsf{M}}(\mathbf{X}, \mathbf{X})$ to denote the $k \times k$ matrix that records the spatial covariances across the k design points and $\Sigma_{\mathsf{M}}(\mathbf{x}_0, \mathbf{X})$ to represent the $k \times 1$ vector that contains the spatial covariances between the k design points and the prediction point $\mathbf{x}_0$. The $k \times k$ diagonal matrix Σ_ε denotes the variance-covariance matrix of the $k \times 1$ vector of average random errors $\bar{\varepsilon} = (\bar{\varepsilon}(\mathbf{x}_1), \bar{\varepsilon}(\mathbf{x}_2), \ldots, \bar{\varepsilon}(\mathbf{x}_k))^\top$, and $\Sigma_\varepsilon = \mathrm{diag}(\mathsf{V}(\mathbf{x}_1)/n_1, \mathsf{V}(\mathbf{x}_2)/n_2, \ldots, \mathsf{V}(\mathbf{x}_k)/n_k)$.

A pointwise confidence interval (CI) for the SK predictor at a given prediction point $\mathbf{x}_0 \in \mathscr{X}$ with a prescribed confidence level $(1-\alpha)$ can be constructed as $\mu_{k,B}(\mathbf{x}_0) \pm z_{1-\alpha/2} \sigma_{k,B}(\mathbf{x}_0)$, where $\alpha \in (0,1)$ and $z_{1-\alpha/2}$ denotes the $(1-\alpha/2)$-quantile of the standard normal distribution (Kleijnen 2015). One way to construct a joint confidence region of the SK predictor at N prediction points (i.e., $\mathbf{x}_{0,i} \in \mathscr{X}$, $i = 1, 2, \ldots, N$) with confidence level $(1-\alpha)$ is to apply Bonferroni correction and obtain a joint confidence region comprising N pointwise confidence intervals: $\mu_{k,B}(\mathbf{x}_{0,i}) \pm z_{1-\alpha/(2N)} \sigma_{k,B}(\mathbf{x}_{0,i})$, $i = 1, 2, \ldots, N$. Nevertheless, the literature on constructing a joint confidence region of the SK predictor over an arbitrary set of prediction points is sparse.

3 MAIN RESULTS

In this section, we present three main results regarding the construction of probabilistic uniform error bounds for SK.

Theorem 1 Consider a zero mean Gaussian process defined through the continuous covariance function $\Sigma_{\mathsf{M}}(\cdot,\cdot)$ with Lipschitz constant L_Σ on the compact set $\mathscr{X} \subset \mathbb{R}^d$. Also consider a continuous unknown function $f(\mathbf{x}) : \mathscr{X} \to \mathbb{R}$ with Lipschitz constant L_f, and its noisy observations $\mathscr{Y}(\cdot)$'s satisfy the assumptions stipulated under model (1). Then, the predictive mean function $\mu_{k,B}(\cdot)$ and standard deviation $\sigma_{k,B}(\cdot)$ obtained based on a simulation dataset $\mathbb{D}_{k,B} = \{\mathbf{x}_i, \{\mathscr{Y}_j(\mathbf{x}_i)\}_{j=1}^{n_i}, i = 1,2,\ldots,k : \sum_{i=1}^{k} n_i = B\}$ are continuous with Lipschitz constant $L_{\mu_{k,B}}$ and modulus of continuity $\omega_{\sigma_{k,B}}(\cdot)$ on $\mathscr{X}$, which are respectively given by

$$\begin{aligned} L_{\mu_{k,B}} &\leq L_\Sigma \sqrt{k} \left\| \left(\Sigma_{\mathsf{M}}(\mathbf{X},\mathbf{X}) + \Sigma_\varepsilon\right)^{-1} \bar{\mathscr{Y}} \right\|, \\ \omega_{\sigma_{k,B}}(\tau) &\leq \sqrt{2\tau L_\Sigma \left(1 + k \left\| \left(\Sigma_{\mathsf{M}}(\mathbf{X},\mathbf{X}) + \Sigma_\varepsilon\right)^{-1} \right\| \max_{\mathbf{x},\mathbf{x}' \in \mathscr{X}} \Sigma_{\mathsf{M}}(\mathbf{x},\mathbf{x}')\right)}, \end{aligned}$$

where $\|\cdot\|$ denotes the Euclidean norm throughout the paper; and L_Σ, the Lipschitz constant of the covariance function $\Sigma_{\mathsf{M}}(\cdot,\cdot)$, is defined as

$$L_\Sigma \equiv \max_{\mathbf{x},\mathbf{x}' \in \mathscr{X}} \left\| \left(\frac{\partial \Sigma_{\mathsf{M}}(\mathbf{x},\mathbf{x}')}{\partial x_1} \; \frac{\partial \Sigma_{\mathsf{M}}(\mathbf{x},\mathbf{x}')}{\partial x_2} \; \cdots \; \frac{\partial \Sigma_{\mathsf{M}}(\mathbf{x},\mathbf{x}')}{\partial x_d} \right)^\top \right\|.$$

Given any $\alpha \in (0,1)$, choose $\tau \in \mathbb{R}_+$ and set

$$\beta(\tau) = 2\log\left(M(\tau,\mathscr{X})/\alpha\right) \text{ and } \gamma_B(\tau) = (L_{\mu_{k,B}} + L_f)\tau + \sqrt{\beta(\tau)}\,\omega_{\sigma_{k,B}}(\tau), \tag{2}$$

then it holds true that

$$\mathrm{P}\left(|\mu_{k,B}(\mathbf{x}) - f(\mathbf{x})| \leq \sqrt{\beta(\tau)}\,\sigma_{k,B}(\mathbf{x}) + \gamma_B(\tau), \forall \mathbf{x} \in \mathscr{X} \right) \geq 1 - \alpha, \tag{3}$$

where L_f denotes a Lipschitz constant of $f(\cdot)$ that holds with probability of at least $1-\alpha_L$ with $\alpha_L \in (0,1)$ and can be given as

$$L_f = \left\| \begin{bmatrix} \sqrt{2\log\left(\frac{2d}{\alpha_L}\right)} \max_{\mathbf{x}\in\mathscr{X}} \sqrt{\Sigma_{\mathsf{M}}^{\partial 1}(\mathbf{x},\mathbf{x})} + 12\sqrt{6d} \max\left\{ \max_{\mathbf{x}\in\mathscr{X}} \sqrt{\Sigma_{\mathsf{M}}^{\partial 1}(\mathbf{x},\mathbf{x})}, \sqrt{rL_\Sigma^{\partial 1}} \right\} \\ \vdots \\ \sqrt{2\log\left(\frac{2d}{\alpha_L}\right)} \max_{\mathbf{x}\in\mathscr{X}} \sqrt{\Sigma_{\mathsf{M}}^{\partial d}(\mathbf{x},\mathbf{x})} + 12\sqrt{6d} \max\left\{ \max_{\mathbf{x}\in\mathscr{X}} \sqrt{\Sigma_{\mathsf{M}}^{\partial d}(\mathbf{x},\mathbf{x})}, \sqrt{rL_\Sigma^{\partial d}} \right\} \end{bmatrix} \right\|;$$

$L_\Sigma^{\partial i}$ denotes the Lipschitz constant of the partial derivative kernel $\Sigma_{\mathsf{M}}^{\partial i}(\mathbf{x},\mathbf{x})$ on the set $\mathscr{X}$ for $i = 1,2,\ldots,d$, and $r \equiv \max_{\mathbf{x},\mathbf{x}' \in \mathscr{X}} ||\mathbf{x} - \mathbf{x}'||$.

Theorem 1 is inspired by Theorem 3.1 of Lederer and Umlauft (2019), and its proof is given in Xie (2020). Below we make some remarks on Theorem 1. First, the parameter τ denotes the grid constant used in the derivation of Theorem 1 and $M(\tau,\mathscr{X})$ is the minimum number of points in a grid over $\mathscr{X}$ with the grid constant τ. For example, consider a hypercubic set $\mathscr{X} \subseteq \mathbb{R}^d$, an upper bound of $M(\tau,\mathscr{X})$ can be given as $(1 + r/\tau)^d$, where r denotes the edge length of the hypercube. Notice that $\beta(\tau)$ and $\gamma_B(\tau)$ in (2) can be obtained analytically given a simulation dataset $\mathbb{D}_{k,B}$ and the covariance function $\Sigma_{\mathsf{M}}(\cdot,\cdot)$. Therefore, a probabilistic uniform error bound that holds true simultaneously for all $\mathbf{x} \in \mathscr{X}$ can be constructed with a prescribed error level α and a choice of the grid constant τ. Second, we note that the upper bound of the approximation error achieved at $\mathbf{x}$, $\sqrt{\beta(\tau)}\sigma_{k,B}(\mathbf{x}) + \gamma_B(\tau)$, consists of two parts; and τ can be chosen arbitrarily small so that the first part dominates the second part. We will use this fact to study the asymptotic performance of the uniform bound as the number of design points k and the total simulation budget B approach infinity.

The next result reveals that under appropriate conditions, the approximation error as quantified by the uniform bound vanishes as $k, B \to \infty$. Below we make the choice of grid constant τ depend on the number of design points k, so $\beta(\tau)$ and $\gamma_B(\tau)$ depend on k as well. The notation $\tau(k)$, $\beta_k(\tau)$, and $\gamma_{k,B}(\tau)$ will be used below to emphasize their dependence on k.

Theorem 2 Suppose that the assumptions of Theorem 1 are satisfied. Furthermore, consider an infinite sequence of observations are obtained at a growing number of design points and assume that the absolute value of the unknown mean function $f(\mathbf{x}) : \mathscr{X} \to \mathbb{R}$ is bounded above by $\bar{f} \in \mathbb{R}_+$. If there exists a $\zeta > 0$ such that $\sigma_{k,B}(\mathbf{x}) = \mathscr{O}\left(\log(k)^{-\frac{1}{2}-\zeta}\right)$ at $\forall \mathbf{x} \in \mathscr{X}$, then it holds for every $\alpha \in (0,1)$ that

$$\mathrm{P}\left(\sup_{\mathbf{x}\in\mathscr{X}} |\mu_{k,B}(\mathbf{x}) - f(\mathbf{x})| = \mathscr{O}\left(\log(k)^{-\zeta}\right)\right) \geq 1-\alpha.$$

Proof. Due to Theorem 1 with $\beta_k(\tau) = 2\log\left(M(\tau(k),\mathscr{X})/(\pi_k\alpha)\right)$ such that $\sum_{k=1}^{\infty}\pi_k = 1/2$ and the union bound over all $k>0$, it follows that the following event is true with probability of at least $1-\alpha/2$:

$$\sup_{\mathbf{x}\in\mathscr{X}} |\mu_{k,B}(\mathbf{x}) - f(\mathbf{x})| \leq \sqrt{\beta_k(\tau)}\sup_{\mathbf{x}\in\mathscr{X}}\sigma_{k,B}(\mathbf{x}) + \gamma_{k,B}(\tau),\ \forall k>0. \tag{4}$$

A trivial bound for the covering number can be obtained by considering a uniform grid over the cube containing $\mathscr{X}$. This approach leads to $M(\tau(k),\mathscr{X}) \leq (1+r/\tau(k))^d$, where $r = \max_{\mathbf{x},\mathbf{x}'\in\mathscr{X}}\|\mathbf{x}-\mathbf{x}'\|$. Therefore, we have

$$\beta_k(\tau) \leq 2d\log\left(1+r/\tau(k)\right) - 2\log(\pi_k) - 2\log(\alpha). \tag{5}$$

Furthermore, due to Theorem 1, we have

$$L_{\mu_{k,B}} \leq L_\Sigma\sqrt{k}\left\|\left(\Sigma_{\mathsf{M}}(\mathbf{X},\mathbf{X})+\Sigma_\varepsilon\right)^{-1}\bar{\mathscr{Y}}\right\|.$$

Since the matrix $\Sigma_{\mathsf{M}}(\mathbf{X},\mathbf{X})+\Sigma_\varepsilon$ is positive definite and $f(\cdot)$ is bounded above by $\bar{f}$, we have

$$\left\|\left(\Sigma_{\mathsf{M}}(\mathbf{X},\mathbf{X})+\Sigma_\varepsilon\right)^{-1}\bar{\mathscr{Y}}\right\| \leq \left\|\bar{\mathscr{Y}}\right\|/\lambda_{\min}\left(\Sigma_{\mathsf{M}}(\mathbf{X},\mathbf{X})+\Sigma_\varepsilon\right) \leq (\sqrt{k}\bar{f}+\|\bar{\varepsilon}\|)/\mathsf{V}_{k,\min},$$

where $\lambda_{\min}(\mathbf{A})$ is the minimum eigenvalue of a square symmetric matrix $\mathbf{A}$, $\mathsf{V}_{k,\min} \equiv \min_{1\leq i\leq k}\mathsf{V}(\mathbf{x}_i)/n_i$, and recall $\bar{\varepsilon} = (\bar{\varepsilon}(\mathbf{x}_1),\bar{\varepsilon}(\mathbf{x}_2),\ldots,\bar{\varepsilon}(\mathbf{x}_k))^\top$. Given that $\bar{\varepsilon}$ is multivariate normally distributed with mean zero and variance-covariance matrix Σ_ε, $\|\bar{\varepsilon}\|^2$ is equal to $\sum_{i=1}^{k} a_i Z_i^2$ in distribution, where Z_i's are i.i.d. standard normal random variables and $a_i = \mathsf{V}(\mathbf{x}_i)/n_i$, for $i=1,2,\ldots,k$, i.e., $\|\bar{\varepsilon}\|^2 \stackrel{\mathscr{D}}{=} \sum_{i=1}^{k} a_i Z_i^2$. Then by Lemma 1 of Laurent and Massart (2000), we have

$$\mathrm{P}\left(\|\bar{\varepsilon}\|^2 \geq 2\left(\sum_{i=1}^{k}\frac{\mathsf{V}^2(\mathbf{x}_i)}{n_i^2}\right)^{\frac{1}{2}}\sqrt{\eta_{k,B}} + 2\mathsf{V}_{k,\max}\eta_{k,B} + \sum_{i=1}^{k}\frac{\mathsf{V}(\mathbf{x}_i)}{n_i}\right) \leq \exp(-\eta_{k,B})$$

for any $\eta_{k,B}>0$, where $\mathsf{V}_{k,\max} \equiv \max_{1\leq i\leq k}\mathsf{V}(\mathbf{x}_i)/n_i$. Therefore, with probability of at least $1-\exp(-\eta_{k,B})$, we have

$$\|\bar{\varepsilon}\|^2 \leq 2\left(\sum_{i=1}^{k}\frac{\mathsf{V}^2(\mathbf{x}_i)}{n_i^2}\right)^{\frac{1}{2}}\sqrt{\eta_{k,B}} + 2\mathsf{V}_{k,\max}\eta_{k,B} + \sum_{i=1}^{k}\frac{\mathsf{V}(\mathbf{x}_i)}{n_i}.$$

Hence, if we set $\eta_{k,B} = \log(1/(\pi_k \alpha))$ so that $\sum_{k=1}^{\infty} \pi_k = 1/2$, then applying the union bounds over all $k > 0$ yields

$$\left\| (\Sigma_{\mathsf{M}}(\mathbf{X},\mathbf{X}) + \Sigma_{\varepsilon})^{-1} \bar{\mathscr{Y}} \right\| \leq \left(\sqrt{k}\bar{f} + \left(2 \left(\sum_{i=1}^{k} \frac{\mathsf{V}^2(\mathbf{x}_i)}{n_i^2} \right)^{\frac{1}{2}} \sqrt{\eta_{k,B}} + 2\mathsf{V}_{k,\max}\eta_{k,B} + \sum_{i=1}^{k} \frac{\mathsf{V}(\mathbf{x}_i)}{n_i} \right)^{\frac{1}{2}} \right) \mathsf{V}_{k,\min}^{-1}$$

for all $k > 0$ with probability of at least $1 - \alpha/2$. Hence, by Theorem 1, the Lipschitz constant of the predictive mean function $\mu_{k,B}(\cdot)$ satisfies

$$L_{\mu_{k,B}} \leq L_{\Sigma} \underbrace{\sqrt{k} \left(\sqrt{k}\bar{f} + \left(2 \left(\sum_{i=1}^{k} \frac{\mathsf{V}^2(\mathbf{x}_i)}{n_i^2} \right)^{\frac{1}{2}} \sqrt{\eta_{k,B}} + 2\mathsf{V}_{k,\max}\eta_{k,B} + \sum_{i=1}^{k} \frac{\mathsf{V}(\mathbf{x}_i)}{n_i} \right)^{\frac{1}{2}} \right) \mathsf{V}_{k,\min}^{-1}}_{\equiv U_{k,B}}, \ \forall k > 0,$$

where $U_{k,B}$ denotes all the terms to the right of L_{Σ} in the inequality above. Since $\eta_{k,B}$ grows slowly (typically logarithmically with k for some commonly used $\{\pi_k\}$ sequences), it is not hard to see that $L_{\mu_{k,B}} \in \mathscr{O}(k)$ with probability of at least $1 - \alpha/2$. The modulus of continuity $\omega_{\sigma_{k,B}}(\cdot)$ of the predictive standard deviation satisfies

$$\omega_{\sigma_{k,B}}(\tau) \leq \left[2L_{\Sigma}\tau(k) \left(k \max_{\tilde{\mathbf{x}},\tilde{\mathbf{x}}' \in \mathscr{X}} \Sigma_{\mathsf{M}}(\tilde{\mathbf{x}},\tilde{\mathbf{x}}') \Big/ \mathsf{V}_{k,\min} + 1 \right) \right]^{\frac{1}{2}}$$

as $\left\| (\Sigma_{\mathsf{M}}(\mathbf{X},\mathbf{X}) + \Sigma_{\varepsilon})^{-1} \right\| \leq 1 \big/ \mathsf{V}_{k,\min}$. Due to the union bound, (4) holds with probability of at least $1 - \alpha$ with

$$\gamma_{k,B}(\tau) \leq \left[2L_{\Sigma}\tau(k)\beta_k(\tau) \left(k \max_{\tilde{\mathbf{x}},\tilde{\mathbf{x}}' \in \mathscr{X}} \Sigma_{\mathsf{M}}(\tilde{\mathbf{x}},\tilde{\mathbf{x}}')/\mathsf{V}_{k,\min} + 1 \right) \right]^{\frac{1}{2}} + L_f\tau(k) + L_{\Sigma}U_{k,B}\tau(k). \tag{6}$$

$\gamma_{k,B}(\tau)$ must converge to zero as $k, B \to \infty$ to guarantee a vanishing approximation error at $\forall \mathbf{x} \in \mathscr{X}$. This can be achieved if the grid constant as a function of k, $\tau(k)$, decreases faster than $\mathscr{O}\left((k\log(k))^{-1}\right)$. We can set $\tau(k) = \mathscr{O}(k^{-2})$ to guarantee $\lim_{k,B\to\infty} \gamma_{k,B}(\tau) = 0$ in light of (6). This choice implies that $\beta_k(\tau) = \mathscr{O}(\log(k))$ due to (5). Given that there exists a $\zeta > 0$ such that $\sigma_{k,B}(\mathbf{x}) = \mathscr{O}(\log(k)^{-\frac{1}{2}-\zeta})$ at $\forall \mathbf{x} \in \mathscr{X}$. The proof is complete by noticing that $\sqrt{\beta_k(\tau)}\sigma_{k,B}(\mathbf{x}) = \mathscr{O}(\log(k)^{-\zeta})$ at $\forall \mathbf{x} \in \mathscr{X}$. □

We remark on Theorem 2 and its proof with respect to a desirable simulation experiment design for SK. First and foremost, given a fixed total budget B to allocate at k distinct design points, the budget allocation scheme adopted impacts the width of the uniform error bound through $\sigma_{k,B}(\mathbf{x})$ and $\gamma_{k,B}(\tau)$ as manifested by (4) and (6). In particular, efficient unequal budget allocation schemes are expected to reduce the magnitudes of $\sigma_{k,B}(\mathbf{x})$ and $\gamma_{k,B}(\tau)$ as compared to an equal allocation scheme hence lead to a smaller uniform error bound. Second, we see from (5) that the smaller the grid constant $\tau(k)$, the greater the coefficient $\beta_k(\tau)$; furthermore, to guarantee a vanishing approximation error at $\forall \mathbf{x} \in \mathscr{X}$, the choice of $\tau(k)$ must satisfy $\tau(k)\beta_k(\tau)k \to 0$ as $k \to \infty$ and $\sup_{\mathbf{x}\in\mathscr{X}} \sqrt{\beta_k(\tau)}\sigma_{k,B}(\mathbf{x}) \to 0$ as $k, B \to 0$. Third, the number of replications allocated to all design points, n_i's, must not be too small to ensure that the multivariate normality of $\bar{\varepsilon}_k$ holds approximately. Lastly, the assumption on the convergence rate $\sigma_{k,B}(\mathbf{x}) = \mathscr{O}(\log(k)^{-\frac{1}{2}-\zeta})$ for $\forall \mathbf{x} \in \mathscr{X}$ is mild, which can be shown to hold when the commonly used Lipschitz continuous covariance functions such as Gaussian and Matérn (with the parameter controlling the smoothness of sample paths $\nu > 2$) are adopted for metamodeling; see details from Xie (2020).

The last result sheds some light on conditions with respect to simulation experimental designs that ensure a vanishing pointwise prediction variance (or equivalently, standard deviation) as the number of design points k and the total simulation budget B approach infinity.

Theorem 3 Suppose that the assumptions of Theorem 1 are satisfied. Denote a simulation dataset obtained as a result of allocating a total of B simulation replications at k design points by $\mathbb{D}_{k,B} = \{\{\mathbf{x}_i, \{\mathscr{Y}_j(\mathbf{x}_i)\}_{j=1}^{n_i}\}, i = 1,2,\ldots,k : \sum_{i=1}^{k} n_i = B\}$. Let $\mathbb{D}_k^{\mathbf{x}}$ denote the set of design points and $\mathbb{B}_\rho(\mathbf{x}) = \{\mathbf{x}' \in \mathbb{D}_k^{\mathbf{x}} : \|\mathbf{x}' - \mathbf{x}\| \leq \rho\}$ denote the set of design points restricted to a ball around $\mathbf{x}$ with radius $\rho > 0$. Then, for each $\mathbf{x} \in \mathscr{X}$ and $\rho \leq \Sigma_{\mathsf{M}}(\mathbf{x},\mathbf{x})/L_\Sigma$, an upper bound for the predictive variance at $\mathbf{x}$ can be given as

$$\sigma_{k,B}^2(\mathbf{x}) \leq \left(\Sigma_{\mathsf{M}}(\mathbf{x},\mathbf{x}) \max_{\mathbf{x}_i \in \mathbb{B}_\rho(\mathbf{x})} \frac{\mathsf{V}(\mathbf{x}_i)}{n_i} + |\mathbb{B}_\rho(\mathbf{x})| \left(4L_\Sigma \rho \Sigma_{\mathsf{M}}(\mathbf{x},\mathbf{x}) - L_\Sigma^2 \rho^2\right)\right) \left(|\mathbb{B}_\rho(\mathbf{x})| \left(\Sigma_{\mathsf{M}}(\mathbf{x},\mathbf{x}) + 2L_\Sigma \rho\right) + \max_{\mathbf{x}_i \in \mathbb{B}_\rho(\mathbf{x})} \frac{\mathsf{V}(\mathbf{x}_i)}{n_i}\right)^{-1},$$

where $|\mathbb{B}_\rho(\mathbf{x})|$ denotes the cardinality of $\mathbb{B}_\rho(\mathbf{x})$.

Proof. The proof is along the lines of Theorem 3.1 of Lederer et al. (2019). Since $\Sigma_{\mathsf{M}}(\mathbf{X},\mathbf{X}) + \Sigma_\varepsilon$ is positive definite, it follows that

$$\begin{aligned} \sigma_{k,B}^2(\mathbf{x}) &= \Sigma_{\mathsf{M}}(\mathbf{x},\mathbf{x}) - \Sigma_{\mathsf{M}}(\mathbf{x},\mathbf{X})^\top \left(\Sigma_{\mathsf{M}}(\mathbf{X},\mathbf{X}) + \Sigma_\varepsilon\right)^{-1} \Sigma_{\mathsf{M}}(\mathbf{x},\mathbf{X}) \\ &\leq \Sigma_{\mathsf{M}}(\mathbf{x},\mathbf{x}) - \|\Sigma_{\mathsf{M}}(\mathbf{x},\mathbf{X})\|^2 / \lambda_{\max}(\Sigma_{\mathsf{M}}(\mathbf{X},\mathbf{X}) + \Sigma_\varepsilon) \\ &\leq \Sigma_{\mathsf{M}}(\mathbf{x},\mathbf{x}) - \|\Sigma_{\mathsf{M}}(\mathbf{x},\mathbf{X})\|^2 / \left(\lambda_{\max}(\Sigma_{\mathsf{M}}(\mathbf{X},\mathbf{X})) + \lambda_{\max}(\Sigma_\varepsilon)\right), \end{aligned} \tag{7}$$

as $a^\top \mathbf{H}^{-1} a \leq \lambda_{\min}(\mathbf{H}^{-1})\|a\|^2$ and $\lambda_{\min}(\mathbf{H}^{-1}) = 1/\lambda_{\max}(\mathbf{H})$, with $\lambda_{\min}(\mathbf{H})$ and $\lambda_{\max}(\mathbf{H})$ respectively denoting the minimum and maximum eigenvalues of a square symmetric matrix $\mathbf{H}$.

By Geršhgorin's theorem (Geršhgorin 1931), we have

$$\lambda_{\max}(\Sigma_{\mathsf{M}}(\mathbf{X},\mathbf{X})) \leq k \max_{\mathbf{x}',\mathbf{x}'' \in \mathbb{D}_k^{\mathbf{x}}} \Sigma_{\mathsf{M}}(\mathbf{x}',\mathbf{x}'')$$

Furthermore, we have $\lambda_{\max}(\Sigma_\varepsilon) = \max_{1 \leq i \leq k} \mathsf{V}(\mathbf{x}_i)/n_i$. Due to the definition of $\Sigma_{\mathsf{M}}(\mathbf{x},\mathbf{X})$, we have

$$\|\Sigma_{\mathsf{M}}(\mathbf{x},\mathbf{X})\|^2 \geq k \min_{\mathbf{x}' \in \mathbb{D}_k^{\mathbf{x}}} \Sigma_{\mathsf{M}}^2(\mathbf{x}',\mathbf{x}). \tag{8}$$

Hence, it follows from (7)–(8) that

$$\sigma_{k,B}^2(\mathbf{x}) \leq \Sigma_{\mathsf{M}}(\mathbf{x},\mathbf{x}) - k \min_{\mathbf{x}' \in \mathbb{D}_k^{\mathbf{x}}} \Sigma_{\mathsf{M}}^2(\mathbf{x}',\mathbf{x}) \left(k \max_{\mathbf{x}',\mathbf{x}'' \in \mathbb{D}_k^{\mathbf{x}}} \Sigma_{\mathsf{M}}(\mathbf{x}',\mathbf{x}'') + \max_{1 \leq i \leq k} \mathsf{V}(\mathbf{x}_i)/n_i\right)^{-1}. \tag{9}$$

The bound in (9) can be further simplified by using the monotonicity of the predictive variance as a function of the number of design points (Wang and Hu 2018), i.e., $\sigma_k^2(\mathbf{x}) \leq \sigma_{k'}^2(\mathbf{x})$ at $\forall \mathbf{x} \in \mathscr{X}$ if $k > k'$ and hence considering only data inside the ball $\mathbb{B}_\rho(\mathbf{x})$ with radius $\rho > 0$.

Using this reduced set of design points inside the ball $\mathbb{B}_\rho(\mathbf{x})$ instead of $\mathbb{D}_k^{\mathbf{x}}$ and writing the right-hand side of (9) into a single fraction yields

$$\sigma_{k,B}^2(\mathbf{x}) \leq \left(\Sigma_{\mathsf{M}}(\mathbf{x},\mathbf{x}) \max_{\mathbf{x}_i \in \mathbb{B}_\rho(\mathbf{x})} \frac{\mathsf{V}(\mathbf{x}_i)}{n_i} + |\mathbb{B}_\rho(\mathbf{x})| \xi(\mathbf{x},\rho)\right) \left(|\mathbb{B}_\rho(\mathbf{x})| \max_{\mathbf{x}',\mathbf{x}'' \in \mathbb{B}_\rho(\mathbf{x})} \Sigma_{\mathsf{M}}(\mathbf{x}',\mathbf{x}'') + \max_{\mathbf{x}_i \in \mathbb{B}_\rho(\mathbf{x})} \frac{\mathsf{V}(\mathbf{x}_i)}{n_i}\right)^{-1}, \tag{10}$$

where

$$\xi(\mathbf{x},\rho) = \Sigma_{\mathsf{M}}(\mathbf{x},\mathbf{x}) \max_{\mathbf{x}',\mathbf{x}'' \in \mathbb{B}_\rho(\mathbf{x})} \Sigma_{\mathsf{M}}(\mathbf{x}',\mathbf{x}'') - \min_{\mathbf{x}' \in \mathbb{B}_\rho(\mathbf{x})} \Sigma_{\mathsf{M}}^2(\mathbf{x}',\mathbf{x}). \tag{11}$$

Under the assumption that $\rho \leq \Sigma_{\mathsf{M}}(\mathbf{x},\mathbf{x})/L_{\Sigma}$, it follows from the Lipschitz continuity of $\Sigma_{\mathsf{M}}(\cdot,\cdot)$ that

$$\min_{\mathbf{x}' \in \mathbb{B}_{\rho}(\mathbf{x})} \Sigma_{\mathsf{M}}^2(\mathbf{x}',\mathbf{x}) \geq (\Sigma_{\mathsf{M}}(\mathbf{x},\mathbf{x}) - L_{\Sigma}\rho)^2.$$

Furthermore,

$$\max_{\mathbf{x}',\mathbf{x}'' \in \mathbb{B}_{\rho}(\mathbf{x})} \Sigma_{\mathsf{M}}(\mathbf{x}',\mathbf{x}'') \leq \Sigma_{\mathsf{M}}(\mathbf{x},\mathbf{x}) + 2L_{\Sigma}\rho, \tag{12}$$

which follows from the fact that

$$|\Sigma_{\mathsf{M}}(\mathbf{x}',\mathbf{x}'') - \Sigma_{\mathsf{M}}(\mathbf{x},\mathbf{x})| \leq |\Sigma_{\mathsf{M}}(\mathbf{x}',\mathbf{x}'') - \Sigma_{\mathsf{M}}(\mathbf{x}',\mathbf{x})| + |\Sigma_{\mathsf{M}}(\mathbf{x}',\mathbf{x}) - \Sigma_{\mathsf{M}}(\mathbf{x},\mathbf{x})| \leq 2L_{\Sigma}\rho,\ \forall \mathbf{x}',\mathbf{x}'' \in \mathbb{B}_{\rho}(\mathbf{x}).$$

Hence, by (11)–(12), we have

$$\begin{aligned} \xi(\mathbf{x},\rho) &\leq \Sigma_{\mathsf{M}}(\mathbf{x},\mathbf{x})(\Sigma_{\mathsf{M}}(\mathbf{x},\mathbf{x}) + 2L_{\Sigma}\rho) - (\Sigma_{\mathsf{M}}(\mathbf{x},\mathbf{x}) - L_{\Sigma}\rho)^2 \\ &= 4L_{\Sigma}\rho\Sigma_{\mathsf{M}}(\mathbf{x},\mathbf{x}) - L_{\Sigma}^2\rho^2. \end{aligned} \tag{13}$$

The proof is complete upon plugging (13) into (10). □

We make some remarks on Theorem 3. First, the radius parameter ρ defines how far away from a prediction point $\mathbf{x}$ the design points are considered to be informative. Second, it is easy to show that the upper bound for $\sigma_{k,B}^2(\mathbf{x})$ in Theorem 3 is an increasing function of $\max\limits_{\mathbf{x}_i \in \mathbb{B}_{\rho}(\mathbf{x})} \mathsf{V}(\mathbf{x}_i)/n_i$. Hence, reducing the local maximum noise variance around the prediction point $\mathbf{x}$ helps reduce $\sigma_{k,B}^2(\mathbf{x})$. Moreover, it is easy to see upon some algebraic manipulation of the upper bound that given any fixed $\rho > 0$, as long as $|\mathbb{B}_{\rho}(\mathbf{x})| \to \infty$, the impact of simulation noise can be mitigated. To guarantee $\sigma_{k,B}^2(\mathbf{x}) \to 0$ as $k,B \to 0$, it is sufficient to require that the radius $\rho : k \to \mathbb{R}_+$, as a function of the number of design points k, satisfy $\rho(k) \leq \Sigma_{\mathsf{M}}(\mathbf{x},\mathbf{x})/L_{\Sigma}$ for $\forall k \geq 0$, $\lim\limits_{k\to\infty} \rho(k) = 0$ and $\lim\limits_{k\to\infty} |\mathbb{B}_{\rho(k)}(\mathbf{x})| = \infty$. The above implies that compared to efficient unequal budget allocation schemes, an experimental design consisted of ever more densely located design points plays a more important role in ensuring a vanishing prediction variance, given a sufficiently large simulation budget.

4 NUMERICAL EXPERIMENTS

In this section we numerically evaluate the performance of the proposed uniform error bound/confidence interval (referred to as $\mathrm{CI_u}$) in comparison with the simultaneous confidence interval obtained with Bonferroni correction (referred to as $\mathrm{CI_b}$).

Two numerical examples are considered. The first one is related to an M/M/1 queue and the second one is a 2-dimensional synthetic example. We start with describing the common experimental setup used in both examples. A simulation experiment is performed with a total budget of B observations to collect at k distinct design points, with n_i observations to obtain at design point $\mathbf{x}_i$, for $i = 1,2,\ldots,k$. Three budget allocation schemes, i.e., the equal allocation, unequal allocation 1, and unequal allocation 2, are considered. Specifically, the equal budget allocation scheme sets $n_i = \lceil B/k \rceil$, where $\lceil a \rceil$ denotes the least integer not less than a. Assuming that the noise variance function is known, the unequal allocation 1 sets $n_i = \left\lceil \frac{\mathsf{V}(\mathbf{x}_i)}{\sum_{i=1}^k \mathsf{V}(\mathbf{x}_i)} B \right\rceil$, while the unequal allocation 2 sets $n_i = \left\lceil \frac{\sqrt{\mathsf{V}(\mathbf{x}_i)}}{\sum_{i=1}^k \sqrt{\mathsf{V}(\mathbf{x}_i)}} B \right\rceil$. We consider distinct experimental settings comprising various combinations of the total budget B, the number of design points k, and the budget allocation scheme adopted. The nominal confidence level is set to 0.95.

Specifically, we repeat the simulation experiment under each experimental setting for $M = 100$ independent macro-replications and calculate the simultaneous coverage probability (SCP) of a given confidence interval defined as follows:

$$\mathrm{SCP} = \frac{1}{M} \sum_{m=1}^{M} \mathbf{1}\{f(\mathbf{x}_{0,i}) \in \mathrm{CI}(\mathbf{x}_{0,i}) \text{ for each } i = 1,2,\ldots,N \text{ on the } m\text{th macro-replication}\},$$

where $\text{CI}(\mathbf{x}_{0,i})$ refers to either CI_u or CI_b obtained at prediction point $\mathbf{x}_{0,i}$, $i = 1,2,\ldots,N$; and $\mathbf{1}\{A\}$ is 1 if event A is true and 0 otherwise. Moreover, we compare the half-widths of CI_u and CI_b constructed for the N prediction points based on their respective maximum half-widths $H_{\max}$ achieved on each macro-replication.

4.1 An M/M/1 Queueing Example

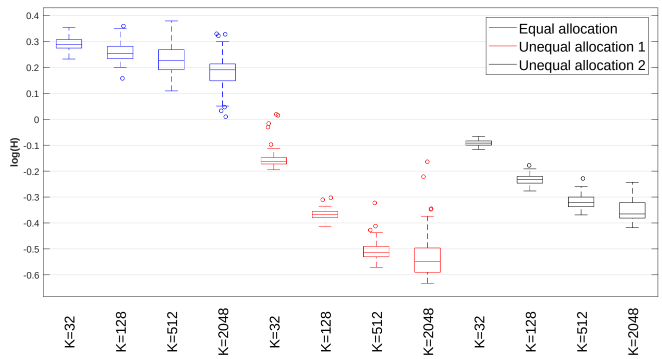
Consider simulating an M/M/1 queue with service rate 1 and arrival rate x with $x \in \mathscr{X} = [0.3, 0.9]$. It is well known from queueing theory that the steady-state mean number of customers in the queue is $f(x) = 1/(1-x)$. This is the function we intend to estimate. The general experiment setup is as given at the beginning of Section 4, and the specific setup for this example is given as follows. The grid constant τ is set to $10^{-10}/k^2$ to ensure that $\sqrt{\beta_k(\tau)}\sigma_{k,B}(x)$ dominates $\gamma(\tau)$ in (3). We consider using a total budget $B \in \{2560, 25600\}$ and varying the number of distinct design points $k \in \{16, 32, 128, 512\}$ when $B = 2560$ and $k \in \{32, 128, 512, 2048\}$ when $B = 25600$. The three budget allocation schemes mentioned at the beginning of Section 4 are applied given each combination of B and k. Notice that the true noise variance function is $\mathsf{V}(x) \approx 2x(1+x)/(T(1-x)^4)$ for large T and we set $T = 1000$ in this example. The k design points are evenly spaced in $\mathscr{X}$, and a grid of $N = 1000$ equispaced prediction points is selected from $\mathscr{X}$.

Summary of results. Table 1 shows the SCPs of the uniform confidence intervals (CI_u) and the confidence intervals obtained with Bonferroni correction (CI_b) under various experimental settings. We have the following observations. First, given a fixed budget B, CI_u can typically achieve SCPs that are higher than the nominal level 0.95, whereas the SCPs of CI_b are close to 0.95 only when the number of design points k is relatively small. Second, the simultaneous coverage performance of CI_u stays satisfactory as the number of design points k increases until it reaches a point that only a few replications are allocated to the design points according to a given budget allocation scheme; see, e.g., the SCPs of CI_u obtained under the equal allocation scheme given $B = 2560$. The simultaneous coverage performance of CI_b, however, deteriorates rapidly as k increases given a fixed budget B; this is observed under all three budget allocation schemes. Third, the SCPs of CI_u and CI_b increase with the budget B when the budget is applied to a fixed number of design points k according to a given budget allocation scheme. Lastly, applying an unequal budget allocation scheme helps improve the simultaneous coverage performance of CI_b when the number of design points k is not too large relative to the budget B given. The impact on SCPs of CI_u is not as obvious though.

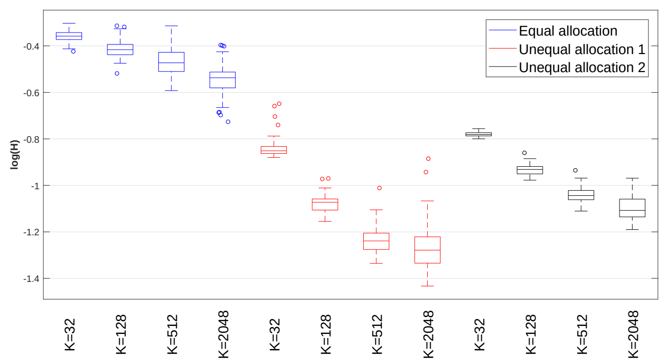
Table 1: SCPs of CI_u and CI_b obtained for the M/M/1 example under different experimental settings.

B	k	Equal allocation		Unequal allocation 1		Unequal allocation 2	
		CI_u	CI_b	CI_u	CI_b	CI_u	CI_b
2560	16	1	0.80	1	0.97	1	0.98
	32	1	0.68	1	0.93	1	0.90
	128	0.98	0.46	1	0.52	1	0.70
	512	0.92	0.33	0.88	0.14	1	0.57
25600	32	1	0.98	1	1	1	1
	128	1	0.69	1	0.96	1	0.93
	512	1	0.54	1	0.81	1	0.78
	2048	1	0.50	0.99	0.33	1	0.72

We next examine the maximum half-widths of CI_u and CI_b. Figure 1 (a) and (b) respectively summarize the magnitudes of $H_{\max}$'s (in a logarithmic scale) of CI_u and CI_b with $B = 25600$ obtained on 100 independent macro-replications. As similar conclusions can be reached regarding the $H_{\max}$'s obtained with $B = 2560$, we omit the details to economize on space. The following observations can be made from Figure 1 (a) and (b). First, the $H_{\max}$'s of CI_u significantly dominate those of CI_b under each experimental setting studied. Second, for both CI_u and CI_b, the magnitude of $H_{\max}$'s tends to decrease with the number of design points k when a fixed budget B is applied according to a given budget allocation scheme. Third, for both CI_u and CI_b, when a given budget B is allocated to a fixed number of design points, the magnitude of $H_{\max}$'s



(a) $\mathrm{CI_u}$

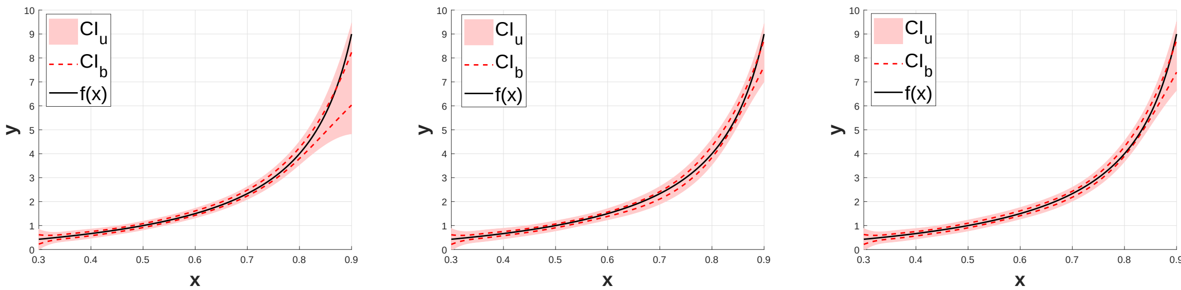


(b) $\mathrm{CI_b}$

Figure 1: Boxplots of $\log(H_{\max})$ of $\mathrm{CI_u}$ and $\mathrm{CI_b}$ for the M/M/1 example obtained on 100 independent macro-replications with $B = 25600$.

produced under the equal allocation scheme is the highest; the magnitudes of $H_{\max}$'s produced under the two unequal allocation schemes are similar, with unequal allocation scheme 2 leading to higher $H_{\max}$'s than unequal allocation scheme 1. The above helps explain the slightly higher SCPs of $\mathrm{CI_u}$ observed under unequal allocation scheme 2 as compared to unequal allocation scheme 1 as shown in Table 1. It is worth noting that, regarding $\mathrm{CI_u}$, the two unequal budget allocation schemes lead to much shorter half-widths than the equal budget allocation scheme without compromising the simultaneous coverage performance as observed in Table 1.

To illustrate the differences between $\mathrm{CI_u}$ and $\mathrm{CI_b}$, in Figure 2 we show the true mean function $f(x)$ together with $\mathrm{CI_u}$ and $\mathrm{CI_b}$ obtained on an arbitrarily chosen macro-replication by allocating a budget of $B = 2560$ to $k = 128$ design points according to the three budget allocation schemes. We observe that under all three budget allocation schemes, the $\mathrm{CI_u}$ obtained can cover the true mean function $f(x)$ at all prediction points. In strong contrast, the $\mathrm{CI_b}$ obtained fails to cover $f(x)$ where the M/M/1 queue is in "heavy traffic" (i.e., where x is close to 0.9). Furthermore, it is interesting to see that for both $\mathrm{CI_u}$ and $\mathrm{CI_b}$, the two unequal allocation schemes can significantly reduce the predictive uncertainty in the "heavy traffic" region and lead to confidence bounds with more even widths throughout the input space $\mathscr{X}$. In contrast, the equal allocation scheme results in much wider confidence bounds in the "heavy traffic" region.

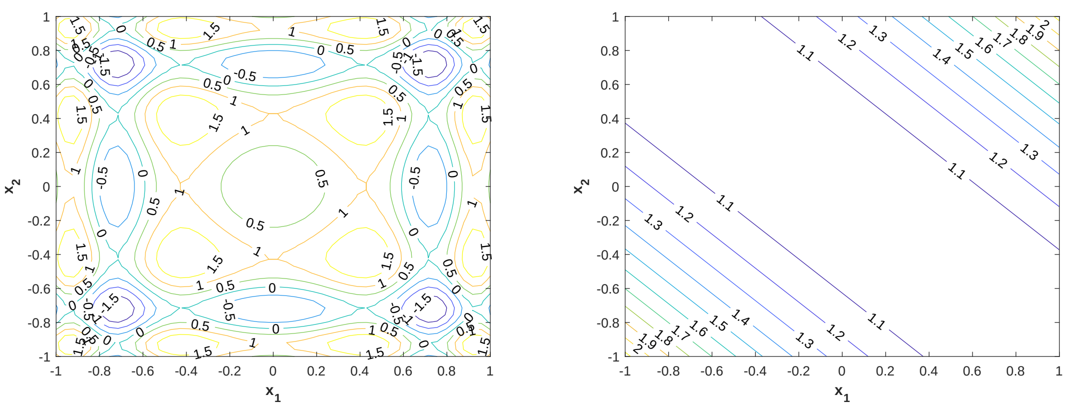


(a) Equal allocation (b) Unequal allocation 1 (c) Unequal allocation 2

Figure 2: An illustration of the $\mathrm{CI_u}$ and $\mathrm{CI_b}$ obtained for the M/M/1 example on an arbitrarily chosen macro-replication with $B = 2560$ and $k = 128$; $f(x)$ denotes the true mean function to estimate.

4.2 A Two-Dimensional Example

Consider the following 2-D example where we try to estimate the mean function $f(\mathbf{x}) = \sin(9x_1^2) + \sin(9x_2^2)$ with $\mathbf{x} = (x_1, x_2)^\top \in \mathscr{X} = [-1,1] \times [-1,1]$. Specifically, the simulation output at design point $\mathbf{x}$ on the jth replication is generated according to model (1), with the noise variance function given by $\mathsf{V}(\mathbf{x}) = (2 + \cos(\pi + (x_1 + x_2)/2))^2$, $\mathbf{x} \in \mathscr{X}$. The contour plots of the true mean and noise variance functions are shown in Figure 3 (a) and (b), respectively. We give the specific setup for this example next and refer the reader for the general experiment setup to the beginning of Section 4. The grid constant τ is set to $10^{-4}/k^2$ in this example to ensure that $\sqrt{\beta_k(\tau)}\sigma_{k,B}(\mathbf{x})$ dominates $\gamma(\tau)$ in (3). We consider using a total budget $B \in \{2560, 10240\}$ and varying the number of distinct design points $k \in \{32, 64, 128, 256, 512\}$ when $B = 2560$ and $k \in \{32, 64, 128, 256, 512, 1024, 2048\}$ when $B = 10240$. The three budget allocation schemes mentioned at the beginning of Section 4 are applied given each combination of B and k. A set of k design points is selected from $\mathscr{X}$ via Latin hypercube sampling and a grid of $N = 2500$ equispaced points in $\mathscr{X}$ is used as the prediction points.



(a) $f(\mathbf{x})$ (b) $\mathsf{V}(\mathbf{x})$

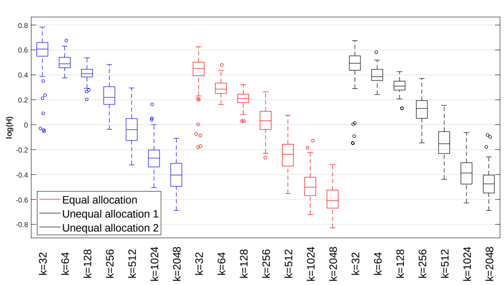
Figure 3: Contour plots of the true mean and noise variance functions for the 2-D example.

Summary of results. Table 2 presents the SCPs of the uniform confidence intervals ($\mathrm{CI_u}$) and the confidence intervals obtained with Bonferroni correction ($\mathrm{CI_b}$) under various experimental settings. We have the following observations. First, given a fixed budget B, $\mathrm{CI_u}$ can achieve SCPs that are higher than the nominal level 0.95 in most of the cases, whereas all SCPs of $\mathrm{CI_b}$ are lower than the nominal level. Second, the simultaneous coverage performance of $\mathrm{CI_u}$ obtained under all three budget allocation schemes is satisfactory unless the number of design points k is relatively small—an insufficient number of design points is inadequate to capture the rapidly changing behavior of $f(\cdot)$ across $\mathscr{X}$ hence results in relatively poor predictive performance. The simultaneous coverage performance of $\mathrm{CI_b}$ improves slightly with k given a fixed budget B but drops again when k becomes too large; the above is observed under all three budget allocation rules. Lastly, applying an unequal budget allocation scheme helps improve the simultaneous coverage performance of $\mathrm{CI_b}$ in most of the cases when the number of design points k is not too large relative to the budget B given. The impact on SCPs of $\mathrm{CI_u}$ is not as obvious.

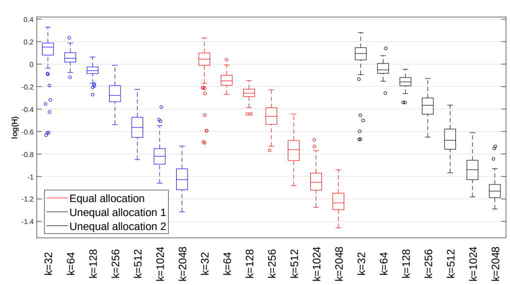
Table 2: SCPs of $\mathrm{CI_u}$ and $\mathrm{CI_b}$ obtained for the 2-D example under different experimental settings.

B	k	Equal allocation		Unequal allocation 1		Unequal allocation 2	
		$\mathrm{CI_u}$	$\mathrm{CI_b}$	$\mathrm{CI_u}$	$\mathrm{CI_b}$	$\mathrm{CI_u}$	$\mathrm{CI_b}$
2560	32	0.58	0.22	0.61	0.24	0.62	0.23
	64	1	0.63	1	0.70	1	0.66
	128	1	0.69	1	0.66	1	0.63
	256	1	0.64	1	0.69	1	0.60
	512	1	0.09	1	0.08	1	0.11
10240	32	0.58	0.31	0.62	0.30	0.63	0.31
	64	0.97	0.59	0.98	0.58	0.99	0.56
	128	1	0.56	1	0.59	1	0.51
	256	1	0.55	1	0.69	1	0.59
	512	1	0.75	1	0.77	1	0.75
	1024	1	0.66	1	0.61	1	0.65
	2048	1	0.04	0.99	0.03	1	0.03

Figure 4 (a) and (b) respectively summarize the magnitudes of $H_{\max}$'s (in a logarithmic scale) of $\mathrm{CI_u}$ and $\mathrm{CI_b}$ with $B = 10240$ obtained on 100 independent macro-replications. The following observations can be made from 4 (a) and (b). First, the $H_{\max}$'s of $\mathrm{CI_u}$ significantly dominate those of $\mathrm{CI_b}$ under each experimental setting studied. Second, for both $\mathrm{CI_u}$ and $\mathrm{CI_b}$, the magnitude of $H_{\max}$'s tends to decrease with the number of design points k when a fixed budget B being applied according to a given budget allocation scheme. Third, for both $\mathrm{CI_u}$ and $\mathrm{CI_b}$, when a given budget B is allocated to a fixed number of design points, the magnitude of $H_{\max}$'s produced under the equal allocation scheme is the highest; the magnitudes of $H_{\max}$'s produced under the two unequal allocation schemes are similar, with unequal allocation scheme 2 leading to higher $H_{\max}$'s than those obtained under unequal allocation scheme 1. It is worth noting that, regarding $\mathrm{CI_u}$, the two unequal budget allocation schemes lead to shorter half-widths as compared to the equal budget allocation scheme without compromising the simultaneous coverage performance as evidenced by Table 2.



(a) $\mathrm{CI_u}$



(b) $\mathrm{CI_b}$

Figure 4: Boxplots of $\log(H_{\max})$ of $\mathrm{CI_u}$ and $\mathrm{CI_b}$ for the 2-D example obtained on 100 independent macro-replications with $B = 10240$.

ACKNOWLEDGMENTS

This paper is based upon work supported by the National Science Foundation under Grants No. CMMI-1846663 and IIS-1849300.

References

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

Chen, X., and Q. Zhou. 2017. "Sequential Design Strategies for Mean Response Surface Metamodeling via Stochastic Kriging with Adaptive Exploration and Exploitation". *European Journal of Operational Research* 262:575—-585.

De Brabanter, K., J. De Brabanter, and J. A. K. Suykens. 2011. "Approximate Confidence and Prediction Intervals for Least Squares Support Vector Regression". *IEEE Transactions on Neural Networks* 22:110–120.

Dellino, G., J. P. C. Kleijnen, and C. Meloni. 2012. "Robust Optimization in Simulation: Taguchi and Krige combined". *INFORMS Journal on Computing* 24:471–484.

Geršhgorin, S. 1931. " Uber die Abgrenzung der Eigenwerte einer Matrix". *Bulletin de l'Académie des Sciences de l'URSS. Classe des sciences mathématiques et na* 6:749–754.

Kleijnen, J. P. C. 2015. *Design and Analysis of Simulation Experiments*. 2nd ed. New York: Springer.

Laurent, B., and P. Massart. 2000. "Adaptive Estimation of a Quadratic Functional by Model Selection". *The Annals of Statistics* 28:1302–1338.

Lederer, A., and J. Umlauft. 2019. "Uniform Error Bounds for Gaussian Process Regression with Application to Safe Control". In *Proceedings of the 33rd Conference on Neural Information Processing Systems*, 1–17. Vancouver, Canada.

Lederer, A., J. Umlauft, and S. Hirche. 2019. "Posterior Variance Analysis of Gaussian Processes with Application to Average Learning Curves". *arXiv:1906.01404v1 [cs.LG]*.

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

Santner, T. J., B. J. Williams, and W. I. Notz. 2003. *The Design and Analysis of Computer Experiments*. New York: Springer.

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

Wang, W., and X. Chen. 2018. "An Adaptive Two-Stage Dual Metamodeling Approach for Stochastic Simulation Experiments". *IISE Transactions* 50:820–836.

Xie, G. 2020. *Robust and Data-Efficient Metamodel-Based Approaches for Online Analysis of Time-Dependent Systems*. Ph. D. thesis, Grado Department of Industrial and Systems Engineering, Virginia Tech.

AUTHOR BIOGRAPHIES

GUANGRUI XIE is a PhD candidate in the Grado Department of Industrial and Systems Engineering at Virginia Tech. He received his B.E. degree in Industrial Engineering at Tianjin University in China. His research interests lie in stochastic optimization and simulation. His email address is guanx92@vt.edu.

XI CHEN is an Associate Professor in the Grado Department of Industrial and Systems Engineering at Virginia Tech. Her research interests include stochastic modeling and simulation, applied probability and statistics, computer experiment design and analysis, and simulation optimization. Her email address is xchen6@vt.edu and her web page is https://sites.google.com/vt.edu/xi-chen-ise/home.