Proceedings of the 2022 Winter Simulation Conference
B. Feng, G. Pedrielli, Y. Peng, S. Shashaani, E. Song, C.G. Corlu, L.H. Lee, E.P. Chew, T. Roeder, and P. Lendermann, eds.

EMPIRICAL UNIFORM BOUNDS FOR HETEROSCEDASTIC METAMODELING

Yutong Zhang
Xi Chen

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

ABSTRACT

This paper proposes pointwise variance estimation-based and metamodel-based empirical uniform bounds for heteroscedastic metamodeling based on the state-of-the-art nominal uniform bound available from the literature by considering the impact of noise variance estimation. Numerical results show that the existing nominal uniform bound requires a relatively large number of design points and a high number of replications to achieve a prescribed target coverage level. On the other hand, the metamodel-based empirical bound outperforms the nominal bound and other competing bounds in terms of empirical simultaneous coverage probability and bound width, especially when the simulation budget is small. However, the pointwise variance estimation-based empirical bound is relatively conservative due to its larger width. When the budget is sufficiently large so that the impact of heteroscedasticity is low, both empirical bounds' performance approaches that of the nominal bound.

1 INTRODUCTION

Simulation metamodeling has been widely used to model and analyze complex stochastic systems (Santner et al. 2003). Heteroscedastic metamodeling has received much attention due to the increasing recognition of the importance of tackling heteroscedastic noise prevalent in simulation models in various applications, e.g., inventory control (Binois et al. 2018) and online management of emerging epidemics (Hu and Ludkovski 2017). Various heteroscedastic metamodeling techniques have emerged over the years, and a considerable number of them are Gaussian process (GP) related. These include, but are not limited to, the Markov chain Monte Carlo-based fully Bayesian approach (Goldberg et al. 1997), maximum a posteriori-based GP (Kersting et al. 2007), stochastic kriging (SK, Ankenman et al. 2010), practical heteroscedastic GP modeling (Binois et al. 2018), and more recently, variational inference-based heteroscedastic GP modeling (Wang et al. 2019).

A GP-based modeling approach can provide predictive mean and variance values at each given input point, based on which one can construct a pointwise confidence bound that covers the true function value at the given input point with a prescribed probability. However, when one is interested in learning the true mean values at more than one input point, a simultaneous confidence region covering the true mean values at an arbitrary set of input points with a prescribed high probability is arguably more desirable. Existing approaches for constructing a simultaneous confidence region typically rely on bootstrapping and the Bonferroni (Kleijnen and Van Beers 2022) and Šidák corrections (De Brabanter et al. 2010); moreover, these methods only apply appropriately when the number of input points is countably many. Recently, methods dedicated to building (nominal) uniform bounds for heteroscedastic metamodeling have emerged, e.g., from the functional analysis (Kirschner and Krause 2018) and the GP modeling perspectives (Xie and Chen 2020). These uniform bounds are intended for containing the true mean values at an arbitrary set of input points with a prescribed high probability. However, such nominal uniform bounds assume

the knowledge of the underlying heteroscedastic noise variances and ignore the impact of using variance estimates, potentially resulting in undercoverage in practice.

In this paper, we propose two approaches for constructing empirical uniform bounds for heteroscedastic metamodeling by incorporating heteroscedastic noise variance estimation. When the simulation budget is relatively small, the empirical bounds outperform the nominal bound in terms of simultaneous coverage probability and bound width. Their performance approaches that of the nominal bound when the budget becomes large. Section 2 reviews the least square estimation in reproducing kernel Hilbert space, establishes its connection to SK, and provides its corresponding nominal uniform bound. Section 3 elaborates on the proposed empirical uniform bounds for heteroscedastic metamodeling. Section 4 provides numerical evaluations and some conclusions.

2 REVIEW ON LEAST SQUARE ESTIMATION IN REPRODUCING KERNEL HILBERT SPACE

Given a separable Hilbert space $\mathscr{H}$, let $\mathbf{V}_0 : \mathscr{H} \to \mathscr{H}$ denote a positive definite operator, and denote the inner product $\langle \cdot, \cdot \rangle_{\mathbf{V}_0} := \langle \cdot, \mathbf{V}_0 \cdot \rangle$ with the corresponding norm $||\cdot||_{\mathbf{V}_0}$. To estimate an unknown function $f \in \mathscr{H}$ of interest, we generate data $\{v_i, \{y_{ij}\}_{j=1}^{n_i}\}_{i=1}^{k}$, where n_i is the replication number allocated to v_i, such that $y_{ij} = \langle v_{ij}, f \rangle_{\mathscr{H}} + \varepsilon_{ij}$, $j = 1, 2, \ldots, n_i$, $i = 1, 2, \ldots, k$, with ε_{ij} following a sub-Gaussian distribution with variance proxy ρ_i^2. Define the operator $\mathscr{M} : \mathscr{H} \to \mathbb{R}^k$ such that for all $v \in \mathscr{H}$ and $i = 1, 2, \ldots, k$, $(\mathscr{M} v)_i = \langle v_i, v \rangle$, and denote its adjoint by $\mathscr{M}^* : \mathbb{R}^k \to \mathscr{H}$. Define the $k \times k$ diagonal matrix $\mathbf{\Sigma}_\varepsilon := \mathrm{diag}(\rho_1^2/n_1, \ldots, \rho_k^2/n_k)$, one can obtain the regularized least squares estimator $\mu := \arg\min_{f \in \mathscr{H}} ||\mathscr{M} f - \bar{\mathbf{y}}||^2_{\mathbf{\Sigma}_\varepsilon^{-1}} + ||f||^2_{\mathbf{V}_0}$. A closed-form expression follows as $\mu = \mathbf{V}^{-1} \mathscr{M}^* \mathbf{\Sigma}_\varepsilon^{-1} \bar{\mathbf{y}}$, where $\bar{\mathbf{y}} = (\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_k)^\top$, $\bar{y}_i = (n_i)^{-1} \sum_{j=1}^{n_i} y_{ij}$, and $\mathbf{V} = \mathscr{M}^* \mathbf{\Sigma}_\varepsilon^{-1} \mathscr{M} + \mathbf{V}_0$.

If $\mathscr{H}$ is a reproducing kernel Hilbert space (RKHS), a special case of the separable Hilbert space over $\mathbb{R}^d$ with kernel function $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, we have the canonical embedding $K_{\mathbf{x}} = K(\mathbf{x}, \cdot)$ for any $\mathbf{x} \in \mathscr{X} \subseteq \mathbb{R}^d$, where $\mathscr{X}$ is the input space. Let $v_i = K_{\mathbf{x}_i} \in \mathscr{H}$ be the embedding of $\mathbf{x}_i \in \mathscr{X}$ such that $\langle v_i, f \rangle_{\mathscr{H}} = \langle K_{\mathbf{x}_i}, f \rangle_{\mathscr{H}} = f(\mathbf{x}_i)$. Hence, $y_{ij} = f(\mathbf{x}_i) + \varepsilon_j(\mathbf{x}_i)$, where $\varepsilon_j(\mathbf{x}_i)$ satisfies Assumption 1 to be stated later. Denote the design-point set as $\mathscr{D} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$ with $\mathscr{D} \subseteq \mathscr{X}$. Notably, for $\mathbf{V}_0 = \kappa \mathscr{I}$, where $\kappa > 0$ and $\mathscr{I}$ is an identity operator, the representer theorem yields the following tractable form of μ:

$$\mu(\mathbf{x}) = \langle \mu, K_{\mathbf{x}} \rangle_{\mathscr{H}} = K(\mathbf{x}, \mathbf{X})^\top (K(\mathbf{X}, \mathbf{X}) + \kappa \mathbf{\Sigma}_\varepsilon)^{-1} \bar{\mathbf{y}}, \tag{1}$$

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. By abuse of notation, we use $K(\mathbf{X}, \mathbf{X})$ to denote the $k \times k$ kernel matrix across the design points with the element in the ith row and the jth column given by $K(\mathbf{X}, \mathbf{X})_{i,j} = K(\mathbf{x}_i, \mathbf{x}_j)$; similarly, $K(\mathbf{x}, \mathbf{X})$ denotes the $k \times 1$ vector $(K(\mathbf{x}, \mathbf{x}_1), K(\mathbf{x}, \mathbf{x}_2), \ldots, K(\mathbf{x}, \mathbf{x}_k))^\top$. In (1), $\bar{\mathbf{y}} = (\bar{y}(\mathbf{x}_1), \bar{y}(\mathbf{x}_2), \ldots, \ \bar{y}(\mathbf{x}_k))^\top$ denotes the $k \times 1$ vector of the sample averages of simulation outputs, with $\bar{y}(\mathbf{x}_i) = f(\mathbf{x}_i) + \bar{\varepsilon}_{n_i}(\mathbf{x}_i)$ and $\bar{\varepsilon}_{n_i}(\mathbf{x}_i) = n_i^{-1} \sum_{j=1}^{n_i} \varepsilon_j(\mathbf{x}_i)$ denoting the average random noise incurred at $\mathbf{x}_i$ for $i = 1, 2, \ldots, k$, which is abbreviated to $\bar{\varepsilon}(\mathbf{x}_i)$ when there is no risk of confusion. The noise term $\varepsilon_j(\mathbf{x}_i)$ follows a sub-Gaussian distribution with variance proxy $\mathsf{V}(\mathbf{x}_i)$ as stated in Assumption 1. In this case, The noise variance-covariance matrix follows as $\mathbf{\Sigma}_\varepsilon = \mathrm{diag}\left(\mathsf{V}(\mathbf{x}_1)/n_1, \ldots, \mathsf{V}(\mathbf{x}_k)/n_k\right)$. By replacing the operators in the separate Hilbert space with their counterparts in the RKHS, we have $||v||^2_{\mathbf{V}^{-1}} = \kappa^{-1}(K(\mathbf{x}, \mathbf{x}) - K(\mathbf{x}, \mathbf{X})^\top (K(\mathbf{X}, \mathbf{X}) + \kappa \mathbf{\Sigma}_\varepsilon)^{-1} K(\mathbf{x}, \mathbf{X}))$.

Assumption 1 For any $\mathbf{x} \in \mathscr{X} \subseteq \mathbb{R}^d$, $\varepsilon_1(\mathbf{x}), \varepsilon_2(\mathbf{x}), \ldots,$ are assumed to be independent and identically distributed (i.i.d.) sub-Gaussian random variables with variance proxy $\mathsf{V}(\mathbf{x})$, written as $\varepsilon_j(\mathbf{x}) \sim \mathrm{subG}(\mathsf{V}(\mathbf{x}))$, where $\sup_{\mathbf{x} \in \mathscr{X}} \mathsf{V}(\mathbf{x}) < \infty$.

We can connect the aforementioned least square estimation in RKHS to heteroscedastic GP modeling, in particular, stochastic kriging. Let the unknown function $f \sim GP(0, \tau^2 K(\cdot, \cdot))$ be a sample from a GP, where $\tau^2 = \kappa^{-1}$ denotes the process variance parameter. Assume that the observation noise terms $\varepsilon_j(\mathbf{x}_i)$, $j = 1, 2, \ldots, n_i$, are i.i.d. normally distributed with mean zero and variance $\mathsf{V}(\mathbf{x}_i)$, for $i = 1, 2, \ldots, k$. Then the posterior distribution of $f(\mathbf{x})$ at any input point $\mathbf{x} \in \mathscr{X}$ given the data set is normal with the predictive

mean as in (1) and the predictive variance given by

$$\sigma^2(\mathbf{x}) = \tau^2 \left(K(\mathbf{x},\mathbf{x}) - K(\mathbf{x},\mathbf{X})^\top \left(K(\mathbf{X},\mathbf{X}) + \tau^{-2}\boldsymbol{\Sigma}_\varepsilon \right)^{-1} K(\mathbf{x},\mathbf{X}) \right). \tag{2}$$

Given this connection, the uniform error bound derived for least square estimation in RKHS can be expressed in the SK notation, which is stated in Lemma 1 below. Notice that the noise distribution assumption stipulated by SK is more restrictive than Assumption 1. Nevertheless, the uniform bound derived in the RKHS setting under Assumption 1 can serve as a uniform bound for SK as well.

Lemma 1 (Lemma 7 in Kirschner and Krause 2018) Under Assumption 1, the following uniform error bound for the least squares estimator $\mu(\cdot)$ given in (1) holds with probability at least $1-\delta$ for all $k \geq 1$ and $\forall \mathbf{x} \in \mathscr{X}$:

$$|\mu(\mathbf{x}) - f(\mathbf{x})| \leq \underbrace{\left(\sqrt{\ln\left(|\mathbf{I}_k + \tau^2 \boldsymbol{\Sigma}_\varepsilon^{-1} K(\mathbf{X},\mathbf{X})|\right) - 2\ln\delta} + \tau^{-1}||f||_K \right)}_{:=\beta_f(\delta)} \sigma(\mathbf{x}),$$

where $\mathbf{I}_k$ is the $k \times k$ identity matrix and $\|f\|_K$ is the RKHS norm associated with the kernel function K.

Remark 1 A uniform bound for the function $f(\cdot)$ follows from Lemma 1 immediately. It holds with probability at least $1-\delta$ for all $k \geq 1$ and $\forall \mathbf{x} \in \mathscr{X}$:

$$\mu(\mathbf{x}) - \beta_f(\delta)\sigma(\mathbf{x}) \leq f(\mathbf{x}) \leq \mu(\mathbf{x}) + \beta_f(\delta)\sigma(\mathbf{x}).$$

This uniform bound can only be regarded as nominal, however, due to its assuming the knowledge of the noise variances at the design points in $\mu(\cdot)$, $\sigma(\cdot)$, and $\beta_f(\delta)$. In practice, one may replace the true noise variances with sample variances (Ankenman et al. 2010) and metamodel-based estimates (Wang and Chen 2016; Wang and Chen 2018). Nevertheless, the resulting bounds with variance estimates directly plugged in may fail to achieve a prescribed coverage probability due to ignoring the impact of noise variance estimation.

3 MAIN RESULTS

In this section, we propose two approaches to build empirical uniform bounds for heteroscedastic metamodeling that take into account the impact of noise variance estimation at all design points. Section 3.1 introduces the pointwise variance estimation-based approach that incorporates time-uniform upper and lower bounds for the noise variances at the design points. Section 3.2 adopts the metamodel-based approach that estimates the noise variance function via GP modeling and leverages the corresponding uniform bound for the noise variance function.

3.1 Pointwise Variance Estimation-Based Empirical Uniform Bound

This section starts with some necessary definitions for introducing time-uniform bounds for pointwise noise variance estimation in Lemmas 2 and 3. Theorem 1 provides the proposed pointwise variance estimation-based empirical uniform bound.

Definition 1 (Page 25 in Wainwright 2019) A random variable $X \in \mathbb{R}$ is said to be *sub-exponential* with two given parameters $v, c > 0$, denoted as $\text{subE}(v,c)$, if the following condition is satisfied:

$$\mathbb{E}e^{\lambda X} \leq \exp\left(\frac{v^2\lambda^2}{2}\right),\ \forall \lambda : |\lambda| < 1/c,\ \text{or}\ \ln \mathbb{E}e^{\lambda X} \leq \frac{(-\ln(1-c\lambda) - c\lambda)v}{c^2},\ \forall \lambda \in [0, 1/c). \tag{3}$$

Definition 2 (Definition 1, sub-ψ condition in Howard et al. 2021). Let $(S_j)_{j=0}^\infty, (V_j)_{j=0}^\infty$ be real-valued scalar processes adapted to an underlying filtration $(\mathscr{F}_j)_{j=0}^\infty$ with $S_0 = V_0 = 0$ and $V_j \geq 0$ for all j. For a function

$\psi : [0, \lambda_{\max}) \to \mathbb{R}$, we say (S_j) is sub-ψ *with variance process* (V_j) if, for each $\lambda \in [0, \lambda_{\max})$, there exists a supermartingale $(L_j(\lambda))_{j=0}^{\infty}$ with respect to $(\mathscr{F}_j)$ such that $\mathbb{E}L_0(\lambda) \leq 1$ and $\exp(\lambda S_j - \psi(\lambda)V_j) \leq L_j(\lambda)$ almost surely for all j.

The ψ functions can be specified for different distributions such as sub-exponential, sub-Gaussian, and sub-Gamma, which are interchangeable under some conditions (Howard et al. 2021). The right-hand side of the second inequality in (3) is an example of ψ_E—the ψ function for the sub-exponential with parameter c and v. Such functions are used in computing pointwise uniform bounds for the noise variance $\mathsf{V}(\mathbf{x})$ at each design point $\mathbf{x}$ in Lemmas 2 and 3.

Lemma 2 (Theorem 1, the polynomial stitched bound in Howard et al. 2021) Under Assumption 1, the following time-uniform upper bound for the noise variance $\mathsf{V}(\mathbf{x})$ at $\mathbf{x} \in \mathscr{D}$ holds:

$$\mathbb{P}\left(\forall n \geq \max\{2, 1 + S_\delta(v)\} : \mathsf{V}(\mathbf{x}) \leq \frac{\sum_{i=1}^{n}(y_i(\mathbf{x}) - \bar{y}_{n-1}(\mathbf{x}))^2}{n - 1 - S_\delta(v)}\right) \geq 1 - \delta, \tag{4}$$

where $\delta \in (0,1)$, $S_\delta(v) = k_1\sqrt{v l_\delta(v)} + c k_2 l_\delta(v)$, $l_\delta(v) = s\ln\ln(\eta v m^{-1}) + \ln(\zeta(s)\delta^{-1})\ln^s \eta$, $k_1 = (\eta^{1/4} + \eta^{-1/4})/\sqrt{2}$, $k_2 = (\sqrt{\eta} + 1)/2$, $\zeta(s)$ is the Riemann zeta function, $\eta > 1, m > 0$, and $s > 1$. Similarly, the following time-uniform lower bound for the noise variance $\mathsf{V}(\mathbf{x})$ at any $\mathbf{x} \in \mathscr{D}$ holds:

$$\mathbb{P}\left(\forall n \geq 2 : \mathsf{V}(\mathbf{x}) \geq \frac{\sum_{i=1}^{n}(y_i(\mathbf{x}) - \bar{y}_{n-1}(\mathbf{x}))^2}{n - 1 + S_\delta(v)}\right) \geq 1 - \delta. \tag{5}$$

Lemma 3 (Propositions 6 and 9, the conjugate mixture bound in Howard et al. 2021) Under Assumption 1, the following time-uniform upper bound for the noise variance $\mathsf{V}(\mathbf{x})$ at $\mathbf{x} \in \mathscr{D}$ holds:

$$\mathbb{P}\left(\forall n \geq \max\{2, 1 + NM_\delta(v)\} : \mathsf{V}(\mathbf{x}) \leq \frac{\sum_{i=1}^{n}(y_i(\mathbf{x}) - \bar{y}_{n-1}(\mathbf{x}))^2}{n - 1 - NM_\delta(v)}\}\right) \geq 1 - \delta, \tag{6}$$

where $\delta \in (0,1)$, $NM_\delta(v) = \sqrt{2(v+\rho)\ln((2\delta)^{-1}\sqrt{(v+\rho)\rho^{-1}} + 1)}$, $\rho > 0$, and $v = 32(j-1)$. On the other hand, the following time-uniform lower bound for the noise variance $\mathsf{V}(\mathbf{x})$ at $\mathbf{x} \in \mathscr{D}$ holds:

$$\mathbb{P}\left(\forall n \geq 2 : \mathsf{V}(\mathbf{x}) \geq \frac{\sum_{i=1}^{n}(y_i(\mathbf{x}) - \bar{y}_{n-1}(\mathbf{x}))^2}{n - 1 + GE_\delta(v)}\right) \geq 1 - \delta,$$

where $GE_\delta(v) = \sup\{s \geq 0 : m(s,v) < \delta^{-1}\}$,

$$m(s,v) = \frac{(\frac{\rho}{c^2})^{\frac{\rho}{c^2}}}{\Gamma(\frac{\rho}{c^2})\gamma(\frac{\rho}{c^2}, \frac{\rho}{c^2})} \frac{\Gamma(\frac{v+\rho}{c^2})\gamma(\frac{v+\rho}{c^2}, \frac{cs+v+\rho}{c^2})}{(\frac{cs+v+\rho}{c^2})^{\frac{v+\rho}{c^2}}} \exp\left(\frac{cs+v}{c^2}\right),$$

$c = 4$, $v = 32(n-1)$, $\rho > 0$, and $s \geq 0$.

The proofs of Lemmas 2 and 3 rely on Lemma 5 (in Appendix A) and follow from Example 1 and the proofs of Theorem 1, Proposition 6 (upper bound), and Proposition 9 (lower bound) given in Howard et al. (2021). For the sake of brevity, we omit them here.

Remark 2 As stated in Lemmas 2 and 3, special restrictions on the number of replications at each design point are needed when constructing the pointwise noise variance upper bounds according to (4) and (6) because $n - 1 - S_\delta(v)$ and $n - 1 - NM_\delta(v)$ must be positive to ensure correct inequality directions. Since $S_\delta(v)$ and $NM_\delta(v)$ depend on c and v which are determined by the noise variance assumption stipulated, there is a minimum number of replications at each design point implicitly required by the noise variance assumption stipulated for constructing these pointwise time-uniform noise variance upper bounds.

Remark 3 A common assumption stipulated on the noise variance in the literature is that $\varepsilon_j(\mathbf{x})$'s are i.i.d. normal (Howard et al. 2021), which is more restrictive than Assumption 1. The normality of $\varepsilon_j(\mathbf{x})$'s leads to lower values of c and v compared to the sub-Gaussianity stipulated in Assumption 1; consequently, fewer replications at each design point are required for constructing appropriate pointwise time-uniform noise variance upper bounds under normality.

In light of Lemmas 2 and 3 and Lemma 6 in Appendix A, we are now in a position to construct empirical uniform bounds for the function of interest.

Theorem 1 Denote $\widehat{\mu}(\cdot)$ as the empirical predicted mean function obtained by using $\widehat{\mathbf{\Sigma}}_\varepsilon = \text{diag}(\widehat{\mathsf{V}}(\mathbf{x}_1)/n_1, \widehat{\mathsf{V}}(\mathbf{x}_2)/n_2, \ldots, \widehat{\mathsf{V}}(\mathbf{x}_k)/n_k)$ in lieu of $\mathbf{\Sigma}_\varepsilon$ in (1). Under Assumption 1, the following bound holds with probability at least $1-\delta_1-\delta_2-\delta_3$ for all $k \geq 1$ and $\forall \mathbf{x} \in \mathscr{X}$:

$$|\widehat{\mu}(\mathbf{x}) - f(\mathbf{x})| \leq \tau^{-1}||f||_K \widehat{\sigma} + \widehat{\beta}_k \overline{\sigma}(\mathbf{x}),$$

where δ_1 is the error probability level for the nominal uniform bound of $f(\cdot)$, δ_2 and δ_3 are respectively the error probability levels allocated for building the lower and upper time-uniform bounds of noise variances at all design points,

$$\widehat{\sigma}(\mathbf{x}) = \tau\sqrt{K(\mathbf{x},\mathbf{x}) - K(\mathbf{x},\mathbf{X})^\top (K(\mathbf{X},\mathbf{X}) + \tau^{-2}\widehat{\mathbf{\Sigma}}_\varepsilon)^{-1} K(\mathbf{x},\mathbf{X})}, \quad \widehat{\mathbf{\Sigma}}_\varepsilon = \text{diag}(\widehat{\mathsf{V}}(\mathbf{x}_1)/n_1, \ldots, \widehat{\mathsf{V}}(\mathbf{x}_k)/n_k),$$

$$\widehat{\beta}_k = \sqrt{2\ln(\delta_1^{-1}\det(\mathbf{I}_k + \tau^2 \underline{\mathbf{\Sigma}}_\varepsilon K(\mathbf{X},\mathbf{X})))}, \quad \underline{\mathbf{\Sigma}}_\varepsilon = \text{diag}(\underline{\mathsf{V}}(\mathbf{x}_1)/n_1, \ldots, \underline{\mathsf{V}}(\mathbf{x}_k)/n_k), \quad (7)$$

$$\overline{\sigma}(\mathbf{x}) = \tau\sqrt{K(\mathbf{x},\mathbf{x}) - K(\mathbf{x},\mathbf{X})^\top (K(\mathbf{X},\mathbf{X}) + \tau^{-2}\overline{\mathbf{\Sigma}}_\varepsilon)^{-1} K(\mathbf{x},\mathbf{X})}, \quad \overline{\mathbf{\Sigma}}_\varepsilon = \text{diag}(\overline{\mathsf{V}}(\mathbf{x}_1)/n_1, \ldots, \overline{\mathsf{V}}(\mathbf{x}_k)/n_k), \quad (8)$$

$\widehat{\mathsf{V}}(\mathbf{x}_i)$ is the sample variance at $\mathbf{x}_i$, $i = 1, 2, \ldots, k$, and $\underline{\mathsf{V}}(\mathbf{x}_i)$ and $\overline{\mathsf{V}}(\mathbf{x}_i)$ are respectively the time-uniform lower and upper bounds for $\mathsf{V}(\mathbf{x}_i)$ given in (5) and (6).

Proof. Denote $\mathbf{A} = K(\mathbf{X},\mathbf{X}) + \tau^{-2}\mathbf{\Sigma}_\varepsilon$, $\hat{\mathbf{A}} = K(\mathbf{X},\mathbf{X}) + \tau^{-2}\widehat{\mathbf{\Sigma}}_\varepsilon$, and we can write $\widehat{\mu}(\mathbf{x}) = K(\mathbf{x},\mathbf{X})^\top \hat{\mathbf{A}}^{-1}\bar{\mathbf{y}} = K(\mathbf{x},\mathbf{X})^\top \hat{\mathbf{A}}^{-1}(\mathbf{f} + \bar{\boldsymbol{\varepsilon}})$, where $\mathbf{f} = (f(\mathbf{x}_1), f(\mathbf{x}_2), \ldots, f(\mathbf{x}_k))^\top$ and $\bar{\boldsymbol{\varepsilon}} = (\bar{\varepsilon}(\mathbf{x}_1), \bar{\varepsilon}(\mathbf{x}_2), \ldots, \bar{\varepsilon}(\mathbf{x}_k))^\top$ denote the $k \times 1$ vector of true function values and the $k \times 1$ vector of average noise terms at the k design points. Hence, we have

$$|\widehat{\mu}(\mathbf{x}) - f(\mathbf{x})| = |K(\mathbf{x},\mathbf{X})\hat{\mathbf{A}}^{-1}(\mathbf{f} + \bar{\boldsymbol{\varepsilon}}) - f(\mathbf{x})| \leq \underbrace{|K(\mathbf{x},\mathbf{X})^\top \hat{\mathbf{A}}^{-1}\mathbf{f} - f(\mathbf{x})|}_{:=U_a} + \underbrace{|K(\mathbf{x},\mathbf{X})^\top \hat{\mathbf{A}}^{-1}\bar{\boldsymbol{\varepsilon}}|}_{:=U_b}, \quad (9)$$

Following Appendix C of Chowdhury and Gopalan (2017), the term U_a in (9) can be upper bounded as follows:

$$\begin{aligned}
U_a &= |\boldsymbol{\phi}(\mathbf{x})^\top f - \boldsymbol{\phi}(\mathbf{x})^\top \mathbf{\Phi}_k^\top (\mathbf{\Phi}_k \mathbf{\Phi}_k^\top + \tau^{-2}\widehat{\mathbf{\Sigma}}_\varepsilon)^{-1} \mathbf{\Phi}_k f| \\
&= |\boldsymbol{\phi}(\mathbf{x})^\top f - \boldsymbol{\phi}(\mathbf{x})^\top (\mathbf{\Phi}_k^\top \mathbf{\Phi}_k + \tau^{-2}\widehat{\mathbf{\Sigma}}_\varepsilon)^{-1} \mathbf{\Phi}_k^\top \mathbf{\Phi}_k f| \\
&= |\boldsymbol{\phi}(\mathbf{x})^\top (\mathbf{\Phi}_k^\top \mathbf{\Phi}_k + \tau^{-2}\widehat{\mathbf{\Sigma}}_\varepsilon)^{-1} \tau^{-2}\widehat{\mathbf{\Sigma}}_\varepsilon f| \\
&\leq ||\boldsymbol{\phi}(\mathbf{x})^\top (\mathbf{\Phi}_k^\top \mathbf{\Phi}_k + \tau^{-2}\widehat{\mathbf{\Sigma}}_\varepsilon)^{-1} \widehat{\mathbf{\Sigma}}_\varepsilon||_K \tau^{-1}||f||_K \\
&= \|f\|_K \sqrt{\tau^{-2}\boldsymbol{\phi}(\mathbf{x})^\top (\mathbf{\Phi}_k^\top \mathbf{\Phi}_k + \tau^{-2}\widehat{\mathbf{\Sigma}}_\varepsilon)^{-1} \widehat{\mathbf{\Sigma}}_\varepsilon \tau^{-2}\widehat{\mathbf{\Sigma}}_\varepsilon (\mathbf{\Phi}_k^\top \mathbf{\Phi}_k + \tau^{-2}\widehat{\mathbf{\Sigma}}_\varepsilon)^{-1} \boldsymbol{\phi}(\mathbf{x})} \\
&\leq \|f\|_K \sqrt{\tau^{-2}\boldsymbol{\phi}(\mathbf{x})^\top (\mathbf{\Phi}_k^\top \mathbf{\Phi}_k + \tau^{-2}\widehat{\mathbf{\Sigma}}_\varepsilon)^{-1} \widehat{\mathbf{\Sigma}}_\varepsilon (\mathbf{\Phi}_k^\top \mathbf{\Phi}_k + \tau^{-2}\widehat{\mathbf{\Sigma}}_\varepsilon)(\mathbf{\Phi}_k^\top \mathbf{\Phi}_k + \tau^{-2}\widehat{\mathbf{\Sigma}}_\varepsilon)^{-1} \boldsymbol{\phi}(\mathbf{x})} \\
&= \|f\|_K \sqrt{\tau^{-2}\boldsymbol{\phi}(\mathbf{x})^\top (\mathbf{\Phi}_k^\top \mathbf{\Phi}_k + \tau^{-2}\widehat{\mathbf{\Sigma}}_\varepsilon)^{-1} \widehat{\mathbf{\Sigma}}_\varepsilon \boldsymbol{\phi}(\mathbf{x})} \\
&= \tau^{-1}||f||_K \widehat{\sigma}(\mathbf{x}),
\end{aligned}$$

where $\mathbf{\Phi}_k$ and $\phi(\mathbf{x})$ are defined in Lemma 6 in Appendix A. The term U_b in (9) can be upper bounded as:

$$U_b = \left| K(\mathbf{x},\mathbf{X})^\top \left((\hat{\mathbf{A}}^{-1}\mathbf{A} - \mathbf{I}_k)\mathbf{A}^{-1} + \mathbf{A}^{-1} \right) \bar{\boldsymbol{\varepsilon}} \right| \leq \underbrace{\left| K(\mathbf{x},\mathbf{X})^\top (\hat{\mathbf{A}}^{-1}\mathbf{A} - \mathbf{I}_k)\mathbf{A}^{-1}\bar{\boldsymbol{\varepsilon}} \right|}_{:=U_c} + \left| K(\mathbf{x},\mathbf{X})^\top \mathbf{A}^{-1}\bar{\boldsymbol{\varepsilon}} \right|. \quad (10)$$

Regarding U_c, $\hat{\mathbf{A}} \to \mathbf{A}$ almost surely as $n_i \to \infty, \forall i = 1,2,\ldots,k$. By Weyl's eigenvalue perturbation theorem, the eigenvalues of $\hat{\mathbf{A}}^{-1}\mathbf{A}$ are consistent estimators of $\mathbf{I}_k$. Hence, for $\forall \varepsilon' > 0, \exists N_0 > 0$, such that for all $n_i > N_0, i = 1,2,\ldots,k$, $\mathbb{P}(\mathscr{A}) \geq 1-\varepsilon'$, where $\mathscr{A} := |\lambda_{\max}(\hat{\mathbf{A}}^{-1}\mathbf{A} - \mathbf{I}_k)| \leq \varepsilon'$. Hence, we have

$$\mathbb{P}(U_c \leq |\varepsilon' K(\mathbf{x},\mathbf{X})^\top \mathbf{A}^{-1}\bar{\varepsilon}|) \geq 1-\varepsilon', \quad (11)$$

when n_i is sufficiently large for $i = 1,2,\ldots,k$. Plugging (11) back into (10) yields $U_b \leq (1+\varepsilon')|K(\mathbf{x},\mathbf{X})^\top \mathbf{A}^{-1}\bar{\varepsilon}|$. By the proof of Lemma 1, the following bound holds with probability at least $1-\delta_1$ for all $k \geq 1$ and $\forall \mathbf{x} \in \mathscr{X}$:

$$K(\mathbf{x},\mathbf{X})^\top \mathbf{A}^{-1}\bar{\varepsilon} \leq \underbrace{\left(\sqrt{\ln\left(|\mathbf{I}_k + \tau^2 \boldsymbol{\Sigma}_\varepsilon^{-1} K(\mathbf{X},\mathbf{X})|\right) - 2\ln\delta_1} \right)}_{:=\beta_k(\delta_1)} \sigma(\mathbf{x}). \quad (12)$$

By Lemma 6, $\beta_k(\delta_1)$ can be upper bounded by $\widehat{\beta}_k(\delta_1) := \sqrt{\ln\left(|\mathbf{I}_k + \tau^2 \underline{\boldsymbol{\Sigma}}_\varepsilon^{-1} K(\mathbf{X},\mathbf{X})|\right) - 2\ln\delta_1}$, which is abbreviated to $\widehat{\beta}_k$ for ease of notation; this upper bound is obtained by replacing $\boldsymbol{\Sigma}_\varepsilon$ with $\underline{\boldsymbol{\Sigma}}_\varepsilon := \mathrm{diag}\left(\underline{\mathsf{V}}(\mathbf{x}_1)/n_1, \ldots, \underline{\mathsf{V}}(\mathbf{x}_k)/n_k\right)$, where the $\underline{\mathsf{V}}(\mathbf{x}_i)$'s are the pointwise time-uniform lower bounds of the noise variances at the k design points obtained with error probability δ_2/k each; hence, the confidence level is $(1-\delta_2/k)^k \geq 1-\delta_2$. Similarly, $\sigma(\mathbf{x})$ can be upper bounded by $\overline{\sigma}(\mathbf{x}) := \tau\sqrt{K(\mathbf{x},\mathbf{x}) - K(\mathbf{x},\mathbf{X})^\top (K(\mathbf{X},\mathbf{X}) + \tau^{-2}\overline{\boldsymbol{\Sigma}}_\varepsilon)^{-1} K(\mathbf{x},\mathbf{X})}$ which is obtained by replacing $\boldsymbol{\Sigma}_\varepsilon$ with $\overline{\boldsymbol{\Sigma}}_\varepsilon$. The error probability allocated for constructing the pointwise time-uniform upper bound of the noise variance at each design point is δ_3/k, hence the confidence level is $(1-\delta_3/k)^k \geq 1-\delta_3$. Finally, by combining the bounds for U_a and U_b with (9) and taking $\varepsilon' \to 0$, we have $|\widehat{\mu}(\mathbf{x}) - f(\mathbf{x})| \leq \tau^{-1}||f||_K \widehat{\sigma}(\mathbf{x}) + \widehat{\beta}_k \overline{\sigma}(\mathbf{x})$ with probability at least $(1-\delta_1)(1-\delta_2)(1-\delta_3)$ which is further bounded below by $1-\delta_1-\delta_2-\delta_3$. □

Remark 4 Recall that Lemmas 2 and 3 provide two types of methods for constructing the pointwise time-uniform lower and upper bounds of the noise variance at a given design point. Theorem 1 selects the stitched lower bound because this bound has a closed form while the conjugate bound does not. Similarly, Theorem 1 adopts the conjugate upper bound because this bound requires fewer replications at each design point for the time-uniform upper bound construction.

3.2 Metamodel-Based Empirical Uniform Bound

This section proposes a metamodel-based empirical uniform bound. Specifically, we adopt a GP for modeling the logarithm of the noise variance function and construct its corresponding uniform bound for the noise variance function. Let $r(\mathbf{x}) := \ln(\widehat{\mathsf{V}}(\mathbf{x}))$, which is modeled as $r(\mathbf{x}) = g(\mathbf{x}) + \varepsilon_g(\mathbf{x})$, where $g(\mathbf{x}) := \ln(\mathsf{V}(\mathbf{x})) \sim \mathrm{GP}(0, \tau_g^2 K_g(\cdot,\cdot))$ and $\varepsilon_g(\mathbf{x})$ is sub-Gaussian with R^2 as the variance proxy.

Lemma 4 (Theorem 2 in Chowdhury and Gopalan 2017) Let $g : \mathscr{X} \to \mathbb{R}$ be a member of the RKHS of real-valued functions on $\mathscr{X}$ with kernel K_g. The following uniform error bound for $g(\cdot)$ holds with probability at least $1-\delta$ for all $k \geq 1$ and $\forall \mathbf{x} \in \mathscr{X}$:

$$|\hat{g}(\mathbf{x}) - g(\mathbf{x})| \leq \underbrace{\left(\frac{R}{\tau_g} \sqrt{\ln\left(|\mathbf{I}_k + \frac{\tau_g^2}{R^2} K_g(\mathbf{X},\mathbf{X})| \right) + 2 + 2\ln(1/\delta)} + \tau_g^{-1}\|g\|_{K_g} \right)}_{:=\beta_g(\delta)} \sigma_g(\mathbf{x}),$$

where $\hat{g}(\mathbf{x})$ and $\sigma_g^2(\mathbf{x})$ are the predictive mean and the predictive variance of $g(\mathbf{x})$ at $\mathbf{x}$, respectively given by

$$\hat{g}(\mathbf{x}) = K_g(\mathbf{x},\mathbf{X})^\top (K_g(\mathbf{X},\mathbf{X}) + \lambda \mathbf{I}_k)^{-1}\mathbf{r},$$
$$\sigma_g^2(\mathbf{x}) = \tau_g^2 \left(K_g(\mathbf{x},\mathbf{x}) - K_g(\mathbf{x},\mathbf{X})^\top \left(K_g(\mathbf{X},\mathbf{X}) + \tau_g^{-2} R^2 \mathbf{I}_k\right)^{-1} K_g(\mathbf{x},\mathbf{X}) \right),$$

where $\mathbf{r} = (r(\mathbf{x}_1), r(\mathbf{x}_2), \ldots, r(\mathbf{x}_k))^\top = \left(\ln \widehat{\mathsf{V}}(\mathbf{x}_1), \ln \widehat{\mathsf{V}}(\mathbf{x}_2), \ldots, \ln \widehat{\mathsf{V}}(\mathbf{x}_k)\right)^\top$.

Remark 5 In light of Lemma 4, the following uniform bound for the noise variance function $\mathsf{V}(\cdot)$ holds with probability at least $1-\delta$ for all $k \geq 1$ and $\forall \mathbf{x} \in \mathscr{X}$:

$$\underbrace{\tilde{\mathsf{V}}(\mathbf{x}) / \exp\left(\beta_g(\delta)\sigma_g(\mathbf{x})\right)}_{:=\mathsf{V}_l} \leq \mathsf{V}(\mathbf{x}) \leq \underbrace{\tilde{\mathsf{V}}(\mathbf{x}) \exp\left(\beta_g(\delta)\sigma_g(\mathbf{x})\right)}_{:=\mathsf{V}_u}, \tag{13}$$

where $\tilde{\mathsf{V}}(\mathbf{x}) =: \exp(\hat{g}(\mathbf{x}))$ is the predictive noise variance at any $\mathbf{x} \in \mathscr{X}$.

Theorem 2 Denote $\widehat{\mu}(\cdot)$ as the empirical predicted mean function obtained by using $\widehat{\mathbf{\Sigma}}_\varepsilon = \mathrm{diag}(\tilde{\mathsf{V}}(\mathbf{x}_1)/n_1, \tilde{\mathsf{V}}(\mathbf{x}_2)/n_2, \ldots, \tilde{\mathsf{V}}(\mathbf{x}_k)/n_k)$ in lieu of $\mathbf{\Sigma}_\varepsilon$ in (1). Under Assumption 1, the following bound holds with probability at least $1-\delta_1-\delta_2$ for all $k \geq 1$ and $\forall \mathbf{x} \in \mathscr{X}$:

$$|\widehat{\mu}(\mathbf{x}) - f(\mathbf{x})| \leq \tau^{-1} ||f||_K \widehat{\sigma} + \widehat{\beta}_k \overline{\sigma}(\mathbf{x}), \tag{14}$$

where δ_1 denotes the error probability level for the nominal uniform bound of $f(\cdot)$ and δ_2 is the error probability level for the uniform bound of $\mathsf{V}(\cdot)$. The terms $\widehat{\sigma}$, $\widehat{\beta}_k$, and $\overline{\sigma}(\mathbf{x})$ are defined similarly as in Theorem 1 but with $\underline{\mathsf{V}}(\mathbf{x})$ and $\overline{\mathsf{V}}(\mathbf{x})$ at each design point $\mathbf{x} \in \mathscr{X}$ respectively replaced by V_l and V_u defined in (13).

Proof. The proof is identical to that of Theorem 1 until upper bounding the term U_c. According to Lemma 4, we have for $\forall \delta_2 > 0$, $\mathbb{P}(\mathscr{B}) \geq 1-\delta_2$, where $\mathscr{B} := \{|\tilde{\mathsf{V}}(\mathbf{x}_i) - \mathsf{V}(\mathbf{x}_i)| \leq \tilde{\mathsf{V}}(\mathbf{x}_i)\beta_{\mathsf{V}_i},\ i = 1,2,\ldots,k\}$ with $\beta_{\mathsf{V}_i} = \max\{\exp\left(\beta_g(\delta)\sigma_g(\mathbf{x}_i)\right) - 1, \exp\left(-\beta_g(\delta)\sigma_g(\mathbf{x}_i)\right) - 1\}$. For $\forall n_i \geq 2$, $i = 1,2,\ldots,k$, we have that $\mathscr{B}' := \{|\tilde{\mathsf{V}}(\mathbf{x}_i) - \mathsf{V}(\mathbf{x}_i)|/n_i \leq \tilde{\mathsf{V}}(\mathbf{x}_i)\beta_{\mathsf{V}_i}/n_i,\ i = 1,2,\ldots,k\}$ is true almost surely on $\mathscr{B}$. Then, for $\forall \varepsilon' > 0$, $\exists N_0 > 0$, such that for $\forall n_i \geq N_0$, $i = 1,2,\ldots,k$, the event $\mathscr{A} := |\lambda_{\max}(\hat{\mathbf{A}}^{-1}\mathbf{A} - \mathbf{I}_k)| \leq \varepsilon'$ is true almost surely on $\mathscr{B}'$. Hence, we have $\mathbb{P}(\mathscr{A}) = \mathbb{P}(\mathscr{A}|\mathscr{B}')\mathbb{P}(\mathscr{B}'|\mathscr{B})\mathbb{P}(\mathscr{B}) \geq 1 \times 1 \times (1-\delta_2) = 1-\delta_2$. The remaining proof of Theorem 2 proceeds as that of Theorem 1. The only difference is that we adopt metamodel-based uniform upper and lower bounds for the noise variance function estimation. By Lemma 6 in Appendix A, $\beta_k(\delta_1)$ and $\sigma(\mathbf{x})$ can be simultaneously upper bounded by $\widehat{\beta}_k$ and $\overline{\sigma}(\mathbf{x})$ as given in (7) and (8) with $\underline{\mathsf{V}}(\mathbf{x})$ and $\overline{\mathsf{V}}(\mathbf{x})$ replaced by V_l and V_u defined in (13) with probability at least $1-\delta_2$. Combining the error probabilities allocated to (12) and (13), we have (14) hold with probability $(1-\delta_1)(1-\delta_2)$, which is further bounded below by $1-\delta_1-\delta_2$. □

4 NUMERICAL EVALUATIONS

In this section, we adopt the following M/M/1 queueing example to demonstrate the performance of the empirical uniform bounds proposed in Sections 3.1 and 3.2. The input variable is the arrival rate, $x \in \mathscr{X} \subseteq (0,1)$, and the mean response surface of interest is the steady-state mean number of customers in the queue, $f(x) = x/(1-x)$. The noise variance function is $\mathsf{V}(x) \approx 2x(1+x)/(T(1-x)^4)$ for T large and we set $T = 1000$ in this example.

Experimental setup. The input space $\mathscr{X} = [0.3, 0.9]$. The design-point set comprises k equispaced points in $\mathscr{X}$ with the value of k varying in $\mathscr{K} = \{8, 16, 32, 64, 128, 256\}$. An equal number of replications n_i is used at all k design points and we consider two replication allocation sets: the low-replication set

where $n_i \in \{5, 10, 15\}$ and the high-replication set where $n_i \in \{45, 55, 65\}$. Notice that, to be constructed appropriately, the pointwise uniform bounds for noise variances require that n_i be sufficiently large.

Methods in comparison. Two benchmarking methods are considered to be compared with the proposed pointwise variance estimation-based empirical uniform bound (denoted by $\mathrm{CI_p}$) and the metamodel-based empirical uniform bound (denoted by $\mathrm{CI_m}$). The first one is to apply the Bonferroni correction to construct the uniform bound (denoted by $\mathrm{CI_b}$) that holds simultaneously at all prediction points of interest (Kleijnen and Van Beers 2022). The other corresponds to the nominal uniform bound for heteroscedastic metamodeling (denoted by $\mathrm{CI_u}$) as in Lemma 1 but with the unknown noise variances at the design points directly replaced by the corresponding sample variances.

Implementation configurations. The prediction-point set comprises a grid of $N = 1000$ equispaced points in $\mathscr{X}$. For constructing the pointwise variance estimation-based empirical bound, we adopt $\eta = 2$, $s = 1.4$, and $m = 1$ in building the lower stitched bounds given in (5), as suggested by Howard et al. (2021). In building the upper conjugate bounds given in (6), we use $\rho = 5$, which requires the smallest number of replications at each design point. The overall error probability level for a given uniform bound of $f(\cdot)$ is set to 0.05. Regarding the two replication allocation sets, the low-replication set is used for constructing $\mathrm{CI_b}$, $\mathrm{CI_u}$, and $\mathrm{CI_p}$, and the high-replication set is applied to constructing all four bounds.

Evaluation metrics. To assess the four bounds' performance under each experimental setting, we adopt two metrics: the empirical simultaneous coverage probability (SCP, Xie and Chen 2020) and the average interval width (AIW, Lam and Zhang 2021) of a uniform bound, obtained over a total of $M = 100$ macro-replications. The SCP is defined as follows:

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

where $\mathrm{CI}(x_{0,i})$ denotes a given bound at prediction point $x_{0,i}$. The AIW obtained on each macro-replication is given by

$$\mathrm{AIW}_m = \frac{1}{2N}\sum_{i=1}^{N} \left(U_m(x_{0,i}) - L_m(x_{0,i})\right), \quad m = 1, 2, \ldots, M,$$

where $U_m(x_{0,i})$ and $L_m(x_{0,i})$ are the upper and lower limits of a given bound at $x_{0,i}$ on the mth macro-replication.

Summary of results. The numerical results of the M/M/1 queueing example are shown in Tables 1 to 4. Tables 1 and 2 give the SCP results obtained using the low- and the high-replication sets, and Tables 3 and 4 provide a summary of the AIWs.

Table 1 shows the SCPs achieved by $\mathrm{CI_b}$, $\mathrm{CI_u}$, and $\mathrm{CI_m}$ using the low-replication set. The following observations are made. First, the Bonferroni bounds have the lowest SCPs, followed by the nominal uniform bounds, and the proposed metamodel-based empirical bounds have the best SCPs. Second, given a fixed number of design points k, the SCPs of all bounds tend to increase with n_i; similarly, given a fixed number of replications n_i, the SCPs tend to increase with k. Lastly, the Bonferroni bounds never achieve the target level of 0.95, while the nominal uniform bounds can achieve the target level when k and n_i are relatively large. In contrast, the metamodel-based empirical bounds can almost always achieve the target level under the settings considered in Table 1.

Table 2 shows the SCPs achieved by all four bounds using the high-replication set. We have the following observations. First, as seen in Table 1, the Bonferroni bounds yield the lowest SCPs, followed by the nominal uniform bounds and the metamodel-based uniform bounds, and the pointwise variance estimation-based empirical bounds deliver the highest SCPs. Second, due to the relatively large number of replications allocated to each design point, except for the Bonferroni bounds, the SCPs for the other three bounds are always higher than the target level. Third, the SCPs for the pointwise variance estimation-based empirical bounds are always 1; we conjecture that such conservativeness is caused by their large widths, which will be examined next.

Table 1: The SCPs achieved by the Bonferroni bounds (CI_b), the nominal uniform bounds (CI_u), and the metamodel-based empirical uniform bounds (CI_m), obtained using the low-replication set.

k \ n_i	5			10			15		
	CI_b	CI_u	CI_m	CI_b	CI_u	CI_m	CI_b	CI_u	CI_m
8	0.15	0.46	0.95	0.23	0.67	0.91	0.31	0.74	0.95
16	0.26	0.72	1	0.28	0.8	0.99	0.25	0.85	0.95
32	0.19	0.78	0.98	0.4	0.94	0.99	0.36	0.89	1
64	0.24	0.87	0.99	0.26	0.96	1	0.44	0.98	1
128	0.17	0.89	1	0.31	0.98	1	0.33	0.98	1
256	0.15	0.88	1	0.32	0.96	1	0.36	0.99	1

Table 2: The SCPs achieved by CI_b, CI_u, CI_m, and the pointwise variance estimation-based empirical bound (CI_p), obtained using the high-replication set.

k \ n_i	45				55				65			
	CI_b	CI_u	CI_m	CI_p	CI_b	CI_u	CI_m	CI_p	CI_b	CI_u	CI_m	CI_p
8	0.48	0.96	0.96	1	0.65	1	0.98	1	0.57	0.98	0.99	1
16	0.38	0.95	1	1	0.45	1	1	1	0.41	0.98	0.99	1
32	0.53	0.99	1	1	0.44	1	0.99	1	0.5	1	1	1
64	0.46	0.98	1	1	0.42	0.99	1	1	0.35	1	0.99	1
128	0.36	1	1	1	0.41	1	1	1	0.37	1	1	1
256	0.43	1	1	1	0.46	1	1	1	0.45	1	1	1

Table 3 summarizes the AIWs obtained under the same setting as considered in Table 1. We have the following observations. First, the Bonferroni bounds have the smallest AIWs, followed by the nominal uniform bounds and the metamodel-based empirical bounds when using the low-replication set. Second, as k or n_i increases, the AIWs of all bounds show a decreasing trend. Third, the differences in the AIWs of different bounds diminish as k or n_i increases.

Table 4 summarizes the AIWs obtained under the same settings as considered in Table 2. We have the following observations. First, the Bonferroni bounds have the smallest AIWs, followed by the nominal uniform bounds and the metamodel-based empirical bounds. The pointwise variance estimation-based empirical bounds have the largest AIWs, which verifies our conjecture made from Table 2. Second, as k or n_i increases, the AIWs of all four bounds tend to decrease.

In brief, we conclude with the following comments. The Bonferroni bounds yield the worst performance—their SCPs are typically too low to be satisfactory. The nominal uniform bounds require large k and n_i to achieve the target coverage level. The metamodel-based empirical bounds can almost always achieve the target level, and as k and n_i become large, the difference in their widths from those of the nominal bounds diminishes. The pointwise variance estimation-based empirical bounds are conservative in giving high SCPs due to their large widths, especially when k and n_i are small.

ACKNOWLEDGMENTS

This paper is based upon work supported by the National Science Foundation [IIS-1849300] and NSF CAREER [CMMI-1846663].

Table 3: Summary of the mean and standard error (in parentheses) of the AIWs of CI_b, CI_u, and CI_m obtained using the low-replication set; the symbol "*" indicates that the value in the cell is less than 0.01.

k \ n_i	5			10			15		
	CI_b	CI_u	CI_m	CI_b	CI_u	CI_m	CI_b	CI_u	CI_m
8	0.57 (0.01)	1.08 (0.03)	4.8 (0.40)	0.63 (0.14)	1.35 (0.39)	2.72 (0.48)	0.93 (0.24)	2.2 (0.63)	2.06 (0.41)
16	0.45 (0.01)	0.94 (0.03)	3.85 (0.51)	0.37 *	0.8 (0.01)	1.56 (0.04)	0.34 *	0.73 (0.01)	1.26 (0.03)
32	0.36 (0.01)	0.81 (0.02)	3.45 (0.59)	0.3 *	0.71 (0.01)	1.36 (0.04)	0.26 *	0.6 (0.01)	0.98 (0.02)
64	0.27 *	0.65 (0.02)	2.67 (0.11)	0.22 *	0.53 (0.01)	1.10 (0.03)	0.2 *	0.51 (0.01)	0.85 (0.02)
128	0.21 *	0.55 (0.02)	2.18 (0.08)	0.17 *	0.44 (0.01)	0.89 (0.02)	0.15 *	0.39 *	0.64 (0.01)
256	0.16 *	0.43 (0.01)	1.69 (0.08)	0.13 *	0.34 *	0.73 (0.02)	0.11 *	0.31 *	0.50 (0.01)

Table 4: Summary of the mean and standard error (in parentheses) of the AIWs of all four bounds obtained using the high-replication set; the symbol "*" indicates that the value in the cell is less than 0.01.

k \ n_i	45				55				65			
	CI_b	CI_u	CI_m	CI_p	CI_b	CI_u	CI_m	CI_p	CI_b	CI_u	CI_m	CI_p
8	1 (0.3)	2.67 (0.85)	1.67 (0.49)	3.37 (0.89)	1.68 (0.4)	4.65 (1.13)	2.91 (0.80)	5.29 (1.17)	1.13 (0.31)	3.11 (0.9)	3.37 (0.86)	3.57 (0.93)
16	0.23 *	0.56 (0.01)	0.68 (0.01)	1.14 (0.02)	0.22 *	0.53 (0.01)	0.62 (0.01)	0.94 (0.01)	0.2 *	0.51 (0.01)	0.58 (0.01)	0.82 (0.01)
32	0.18 *	0.46 (0.01)	0.55 (0.01)	1.06 (0.01)	0.16 *	0.42 *	0.48 (0.01)	0.79 (0.01)	0.15 *	0.4 *	0.44 *	0.68 (0.01)
64	0.13 *	0.36 *	0.42 *	0.94 (0.01)	0.12 *	0.33 *	0.37 *	0.68 (0.01)	0.11 *	0.31 *	0.35 *	0.55 *
128	0.1 *	0.27 *	0.33 *	0.91 (0.01)	0.09 *	0.26 *	0.29 *	0.58 (0.01)	0.09 *	0.25 *	0.27 *	0.46 *
256	0.07 *	0.22 *	0.26 *	1.12 (0.01)	0.07 *	0.2 *	0.23 *	0.5 *	0.06 *	0.19 *	0.22 *	0.39 *

A APPENDIX

Lemma 5 Under Assumption 1, given $\mathbf{x} \in \mathscr{D}$, for $\forall n \geq 2$, define $S_n = \mathsf{V}(\mathbf{x})^{-1} \sum_{i=1}^{n} (y_i(\mathbf{x}) - \bar{y}_n(\mathbf{x}))^2 - (n-1)$. Then S_n is sub-exponential with its cumulant generating function (CGF) upper bounded as follows:

$$\ln \mathbb{E} e^{\lambda S_n} \leq \psi_E := \frac{(-\ln(1-c\lambda) - c\lambda)v}{c^2}, \quad \forall \lambda \in [0, 1/c),$$

where $c = 4$ and $v = 32(n-1)$.

Proof. Rewrite $\varepsilon_i(\mathbf{x}) = \mathsf{V}(\mathbf{x}) Z_i$, where the Z_i's are i.i.d. subG(1). For $n \geq 2$, $S_n = \mathsf{V}(\mathbf{x})^{-1} \sum_{i=1}^{n} (y_i(\mathbf{x}) - \bar{y}_n(\mathbf{x}))^2 - (n-1) = \mathsf{V}(\mathbf{x})^{-1} \sum_{i=1}^{n} (\varepsilon_i(\mathbf{x}) - \bar{\varepsilon}_n(\mathbf{x}))^2 - (n-1) = \sum_{i=1}^{n} (Z_i - \bar{Z}_n)^2 - (n-1)$, where $\bar{y}_n(\mathbf{x}) = n^{-1} \sum_{i=1}^{n} y_i(\mathbf{x})$, $\bar{\varepsilon}_n(\mathbf{x}) = n^{-1} \sum_{i=1}^{n} \varepsilon_i(\mathbf{x})$, and $\bar{Z}_n = n^{-1} \sum_{i=1}^{n} Z_i$. Direct calculations yield $\Delta S_n = S_n - S_{n-1} = \frac{n-1}{n} (Z_n - \bar{Z}_{n-1})^2 - 1 =: Y_n^2 - 1$, where $Y_n := \sqrt{(n-1)/n}(Z_n - \bar{Z}_{n-1})$ for $n \geq 2$. Notice that $Z_n \sim \text{subG}(1)$ is independent of $\bar{Z}_{n-1} \sim \text{subG}((n-1)^{-1})$. Hence, for all $2 \leq j \leq n$, $Y_j \sim \text{subG}(1)$, which leads to $\Delta S_j \sim \text{subE}(32, 4)$. It follows that

$$\begin{aligned}
\mathbb{E}(\exp(\lambda S_n)) &= \mathbb{E}(\exp(\lambda(\Delta S_2 + \Delta S_3 + \cdots + \Delta S_n))) \\
&\leq \mathbb{E}\left(\exp\left(\lambda \Delta S_2 \frac{(n-1)32}{32}\right)\right)^{\frac{32}{(n-1)32}} \cdots \mathbb{E}\left(\exp\left(\lambda \Delta S_n \frac{(n-1)32}{32}\right)\right)^{\frac{32}{(n-1)32}} \\
&\leq \left(\exp\left(\frac{1}{2}\lambda^2 (32(n-1))^2\right)\right)^{\frac{32}{(n-1)32}} \cdots \left(\exp\left(\frac{1}{2}\lambda^2 (32(n-1))^2\right)\right)^{\frac{32}{(n-1)32}} \\
&= \exp\left(\frac{1}{2}\lambda^2 (32(n-1))^2\right), \quad \forall \lambda \in [0, 1/4),
\end{aligned}$$

where the first inequality follows from the generalized Hölder's inequality. Hence, S_n is subE$(32(n-1), 4)$. According to (3), we see that the CGF of S_n is upper bounded by ψ_E with $c = 4$ and $v = 32(n-1)$. □

Lemma 6 Denote $h(\mathsf{V}(\mathbf{x}_1), \ldots, \mathsf{V}(\mathbf{x}_k)) = \ln(\det(\mathbf{I}_k + \tau^2 \boldsymbol{\Sigma}_\varepsilon^{-1} K(\mathbf{X}, \mathbf{X})))$, where recall that $\boldsymbol{\Sigma}_\varepsilon = \text{diag}(\mathsf{V}(\mathbf{x}_1)/n_1, \ldots, \mathsf{V}(\mathbf{x}_k)/n_k)$. Then h is nonincreasing in $\mathsf{V}(\mathbf{x}_i)$, $i = 1, 2, \ldots, k$.

Proof. The proof is in the same vein as that in Appendix B of Chowdhury and Gopalan (2017). Denote $\mathbf{B}_k = \mathbf{I}_k + \tau^2 \boldsymbol{\Phi}_k^\top \boldsymbol{\Sigma}_\varepsilon^{-1} \boldsymbol{\Phi}_k$, where $\boldsymbol{\Phi}_k = (\phi(\mathbf{x}_1), \ldots, \phi(\mathbf{x}_k))^\top$ is a $k \times \infty$ matrix which satisfies $K(\mathbf{X}, \mathbf{X}) = \boldsymbol{\Phi}_k \boldsymbol{\Phi}_k^\top$, and $\phi(\mathbf{x})$ equals $K(\mathbf{x}, \cdot)$, which is an operator mapping any input point $\mathbf{x}$ in $\mathscr{X}$ to $\mathscr{H}$ associated with kernel function K. It follows that

$$\begin{aligned}
\mathbf{B}_k \boldsymbol{\Phi}_k^\top &= (\mathbf{I}_k + \tau^2 \boldsymbol{\Sigma}_\varepsilon^{-1} K(\mathbf{X}, \mathbf{X})) \boldsymbol{\Phi}_k^\top = \boldsymbol{\Phi}_k^\top + \tau^2 \boldsymbol{\Phi}_k^\top \boldsymbol{\Sigma}_\varepsilon^{-1} \boldsymbol{\Phi}_k \boldsymbol{\Phi}_k^\top \\
&= \boldsymbol{\Phi}_k^\top (\mathbf{I}_k + \tau^2 \boldsymbol{\Sigma}_\varepsilon^{-1} \boldsymbol{\Phi}_k \boldsymbol{\Phi}_k^\top) = \boldsymbol{\Phi}_k^\top (\mathbf{I}_k + \tau^2 \boldsymbol{\Sigma}_\varepsilon^{-1} K(\mathbf{X}, \mathbf{X})).
\end{aligned}$$

By $\det(AB) = \det(A)\det(B)$ for two matrices A and B with compatible dimensions, we have $\det(\mathbf{B}_k)\det(\boldsymbol{\Phi}_k^\top) = \det(\boldsymbol{\Phi}_k^\top) \cdot \det(\mathbf{I}_k + \tau^2 \boldsymbol{\Sigma}_\varepsilon^{-1} K(\mathbf{X}, \mathbf{X}))$. Hence, it follows that

$$\det(\mathbf{B}_k) = \det(\mathbf{I}_k + \tau^2 \boldsymbol{\Sigma}_\varepsilon^{-1} K(\mathbf{X}, \mathbf{X})). \tag{15}$$

Plugging (15) into $h(\mathsf{V}(\mathbf{x}_1), \ldots, \mathsf{V}(\mathbf{x}_k))$ yields

$$\begin{aligned}
h(\mathsf{V}(\mathbf{x}_1), \ldots, \mathsf{V}(\mathbf{x}_k)) &= \ln(\det(\mathbf{B}_k)) = \ln(\det(\mathbf{B}_{k-1} + \tau^2 \frac{n_k}{\mathsf{V}(\mathbf{x}_k)} \phi(\mathbf{x}_k) \phi(\mathbf{x}_k)^\top)) \\
&= \ln(\det(\mathbf{B}_{k-1}) \det(1 + \tau^2 \frac{n_k}{\mathsf{V}(\mathbf{x}_k)} ||\phi(\mathbf{x}_k)||_{\mathbf{B}_{k-1}^{-1}})) \\
&= \ln(\det(\mathbf{B}_0) \Pi_{i=1}^{k} (1 + \tau^2 \frac{n_i}{\mathsf{V}(\mathbf{x}_i)} ||\phi(\mathbf{x}_k)||_{\mathbf{B}_{i-1}^{-1}})) = \sum_{i=1}^{k} \ln(1 + \tau^2 \frac{n_i}{\mathsf{V}(\mathbf{x}_i)} K_{i-1}(\mathbf{x}_i, \mathbf{x}_i)),
\end{aligned}$$

where $\mathbf{B}_0 = \mathbf{I}_k$ and $K_{i-1}(\mathbf{x}_i, \mathbf{x}_i)$ is the predictive variance at $\mathbf{x}_i$ obtained using all observations up to the $(i-1)$th design point. Hence, $h(\mathsf{V}(\mathbf{x}_1), \ldots, \mathsf{V}(\mathbf{x}_k))$ is nonincreasing in the $\mathsf{V}(\mathbf{x}_i)$'s. □

REFERENCES

Ankenman, B., B. L. Nelson, and J. Staum. 2010. "Stochastic kriging for simulation metamodeling". *Operations Research* 58:371–382.

Binois, M., R. B. Gramacy, and M. Ludkovski. 2018. "Practical heteroscedastic Gaussian process modeling for large simulation experiments". *Journal of Computational and Graphical Statistics* 27:808–821.

Chowdhury, S. R., and A. Gopalan. 2017. "On kernelized multi-armed bandits". In *International Conference on Machine Learning*, 844–853. PMLR.

De Brabanter, K., J. De Brabanter, J. A. Suykens, and B. De Moor. 2010. "Approximate confidence and prediction intervals for least squares support vector regression". *IEEE Transactions on Neural Networks* 22:110–120.

Goldberg, P., C. Williams, and C. Bishop. 1997. "Regression with input-dependent noise: A Gaussian process treatment". *Advances in neural information processing systems* 10:1–7.

Howard, S. R., A. Ramdas, J. McAuliffe, and J. Sekhon. 2021. "Time-uniform, nonparametric, nonasymptotic confidence sequences". *The Annals of Statistics* 49:1055–1080.

Hu, R., and M. Ludkovski. 2017. "Sequential design for ranking response surfaces". *SIAM/ASA Journal on Uncertainty Quantification* 5:212–239.

Kersting, K., C. Plagemann, P. Pfaff, and W. Burgard. 2007. "Most likely heteroscedastic Gaussian process regression". In *Proceedings of the 24th international conference on Machine learning*, 393–400. Corvallis, OR: International MachineLearning Society.

Kirschner, J., and A. Krause. 2018. "Information directed sampling and bandits with heteroscedastic noise". In *Proceedings of Machine Learning Research*, 358–384. PMLR.

Kleijnen, J. P., and W. C. Van Beers. 2022. "Statistical tests for cross-validation of kriging models". *INFORMS Journal on Computing* 34:607–621.

Lam, H., and H. Zhang. 2021. "Neural predictive intervals for simulation metamodeling". In *Proceedings of the 2021 Winter Simulation Conference*, edited by S. Kim, B. Feng, K. Smith, S. Masoud, Z. Zheng, C. Szabo, and M. Loper. Phoenix,AZ: Institute of Electrical and Electronics Engineers, Inc.

Santner, T. J., B. J. Williams, and W. I. Notz. 2003. *The design and analysis of computer experiments*. New York: Springer.

Wainwright, M. J. 2019. *High-dimensional statistics: A non-asymptotic viewpoint*, Volume 48. Cambridge University Press.

Wang, W., N. Chen, X. Chen, and L. Yang. 2019. "A variational inference-based heteroscedastic Gaussian process approach for simulation metamodeling". *ACM Transactions on Modeling and Computer Simulation* 29:1–22.

Wang, W., and X. Chen. 2016. "The effects of estimation of heteroscedasticity on stochastic kriging". In *Proceedings of the 2016 Winter Simulation Conference*, edited by P. Frazier, R. Roeder, T.M. andSzechtman, and E. Zhou, 326–337. Institute of Electrical and Electronics Engineers, Inc.

Wang, W., and X. Chen. 2018. "An adaptive two-stage dual metamodeling approach for stochastic simulation experiments". *IISE Transactions* 50:820–836.

Xie, G., and X. Chen. 2020. "Uniform error bounds for stochastic kriging". In *Proceedings of the 2020 Winter Simulation Conference*, edited by K.-H. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, T. Roeder, and R. Thiesing, 361–372. Institute of Electrical and Electronics Engineers, Inc.

AUTHOR BIOGRAPHIES

YUTONG ZHANG is a Ph.D. candidate in the Grado Department of Industrial and Systems Engineering at Virginia Tech. Her research interest lies in machine learning, stochastic modeling, and simulation methodology. Her email address is yutongz@vt.edu and her webpage is https://yutong2018.github.io/.

XI CHEN is an associate professor in the Grado Department of Industrial and Systems Engineering at Virginia Tech. Her research interests include simulation modeling and analysis, 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.