Proceedings of the 2016 Winter Simulation Conference
T. M. K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka, and S. E. Chick, eds.

THE EFFECTS OF ESTIMATION OF HETEROSCEDASTICITY ON STOCHASTIC KRIGING

Wenjing Wang
Xi Chen

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

ABSTRACT

In this paper, we study the effects of using smoothed variance estimates in place of the sample variances on the performance of stochastic kriging (SK). Different variance estimation methods are investigated and it is shown through numerical examples that such a replacement leads to improved predictive performance of SK. An SK-based dual metamodeling approach is further proposed to obtain an efficient simulation budget allocation rule and consequently more accurate prediction results.

1 INTRODUCTION

Consider the simulation outputs $\{\mathscr{Y}_j(\mathbf{x}_i)\}_{j=1}^{n_i}$ obtained at design point $\mathbf{x}_i$, generated by running n_i simulation replications at $\mathbf{x}_i \in \mathscr{X} \subseteq \mathbb{R}^d$, for $i = 1,2,\ldots,k$. Suppose that each random output $\mathscr{Y}_j(\mathbf{x}_i)$ can be regarded as being generated by the following heteroscedastic nonparametric model

$$\mathscr{Y}_j(\mathbf{x}_i) = m(\mathbf{x}_i) + \varepsilon_j(\mathbf{x}_i),\ \ j = 1,2,\ldots,n_i, \tag{1}$$

where $\mathscr{Y}_j(\mathbf{x}_i)$ is a random observation of $m(\mathbf{x}_i)$ disturbed by independent and identically distributed (i.i.d.) zero-mean random errors $\varepsilon_j(\mathbf{x}_i)$ with variance $\mathsf{V}(\mathbf{x}_i)$. The variance of the simulation errors, $\mathsf{V}(\mathbf{x})$, typically changes systematically and smoothly with $\mathbf{x}$, rather than stays constant.

Suitable methods of variance modeling are required, for the variance function itself may be of interest on its own right; and more importantly, the regression analysis must be adjusted to account for the variance heterogeneity. A wealth of literature has been devoted to variance estimation. Assuming that the heteroscedasticity is due to some smooth variance function, researchers extend the regression techniques used for mean function estimation to estimate the variance function. Aitkin (1987) and Jobson and Fuller (1980) investigate parametric dual modeling approaches to estimate both mean and variance functions, via a pseudo-likelihood procedure and a weighted joint least squares estimation method, respectively. The nonparametric modeling approaches for variance estimation are not sparse, and the methods typically fall into two categories: difference-based and residual-based. The former type does not consider fitting a curve for the mean function but instead tries to remove the mean-function effect by differencing (Brown and Levine 2007, Gasser et al. 1986); the latter considers estimating the variance by a weighted average of squared residuals, which are obtained after some initial mean function estimation steps (Fan and Yao 1998, Hall and Carroll 1989).

In this paper, we focus on studying the effects of replacing the sample variances with smoothed variance estimates on the predictive performance of stochastic kriging (SK). We further describe an SK-based dual metamodeling approach that can simultaneously estimate the mean and variance functions and achieve highly accurate prediction results. The remainder of the paper is organized as follows. Section 2 provides

a review on the standard SK methodology. Section 3 introduces different variance estimation methods to be investigated. Section 4 presents two examples to demonstrate the effects of using smoothed variance estimates on the predictive performance of SK. Section 5 introduces an SK-based dual metamodeling approach and demonstrates its usefulness through an example.

2 STOCHASTIC KRIGING

Stochastic kriging (SK) methodology proposed by Ankenman et al. (2010) is an effective metamodeling tool for approximating a *mean* response surface implied by a stochastic simulation. SK is an extension of the *kriging* methodology which is highly successful in the design and analysis of deterministic computer experiments (Santner et al. 2003). SK distinguishes itself by its ability to correctly account for sampling uncertainty inherent in a stochastic simulation in addition to the response-surface uncertainty.

Standard SK models the simulation response estimate obtained at a design point $\mathbf{x} \in \mathscr{X} \subset \mathbb{R}^d$ on the jth simulation replication as

$$\mathscr{Y}_j(\mathbf{x}) = \mathsf{Y}(\mathbf{x}) + \varepsilon_j(\mathbf{x}) = \mathbf{f}(\mathbf{x})^\top \beta + \mathsf{M}(\mathbf{x}) + \varepsilon_j(\mathbf{x}) \ , \tag{2}$$

where $\mathsf{Y}(\mathbf{x})$ represents the unknown true response that we intend to estimate at point $\mathbf{x} \in \mathscr{X}$, and the term $\varepsilon_j(\mathbf{x})$ represents the mean-zero simulation error realized on the jth replication. The simulation errors $\varepsilon_1(\mathbf{x}), \varepsilon_2(\mathbf{x}), \ldots$ are assumed to be independent and identically distributed (i.i.d.) across replications at a given design point, and the variance of $\varepsilon_j(\mathbf{x})$, $\mathsf{V}(\mathbf{x})$, may depend on $\mathbf{x}$.

The terms $\mathbf{f}(\cdot)$ and β are, respectively, a $p \times 1$ vector of known functions of $\mathbf{x}$ and a $p \times 1$ vector of unknown parameters. The term $\mathsf{M}(\cdot)$ represents a second-order stationary mean-zero Gaussian random field (Santner et al. 2003, Kleijnen 2008). That is, the spatial covariance between any two points in the random field is typically modeled as

$$\mathrm{Cov}(\mathsf{M}(\mathbf{x}), \mathsf{M}(\mathbf{y})) = \tau^2 \mathscr{R}(\mathbf{x} - \mathbf{y}; \theta) \ , \tag{3}$$

where τ^2 can be interpreted as the spatial variance of the random process $\mathsf{M}(\cdot)$ at all $\mathbf{x}$. One can think of $\mathsf{M}(\mathbf{x})$ as being sampled from a space of mappings $\mathbb{R}^d \to \mathbb{R}$, in which functions are assumed to exhibit spatial correlation. The spatial correlation function $\mathscr{R}(\cdot; \theta)$ determines the smoothness properties of $\mathsf{M}(\cdot)$ and it depends on $\mathbf{x}$ and $\mathbf{y}$ only through their difference. The parameter vector $\theta = (\theta_1, \theta_2, \ldots, \theta_d)^\top$ controls how quickly the spatial correlation diminishes as the two points become farther apart, and it measures the roughness of the underlying response surface in each direction. Ankenman et al. (2010) refer to the stochastic nature of M as *extrinsic uncertainty*, in contrast to the *intrinsic uncertainty* represented by ε that is inherent in a stochastic simulation output.

An experimental design for SK consists of $\{(\mathbf{x}_i, n_i), i = 1, 2, \ldots, k\}$, a set of design points to run independent simulations and the corresponding numbers of replications to apply. Denote the $k \times 1$ vector of the sample averages of simulation responses by $\bar{\mathscr{Y}} = \left(\bar{\mathscr{Y}}(\mathbf{x}_1), \bar{\mathscr{Y}}(\mathbf{x}_2), \ldots, \bar{\mathscr{Y}}(\mathbf{x}_k)\right)^\top$, in which

$$\bar{\mathscr{Y}}(\mathbf{x}_i) = \frac{1}{n_i} \sum_{j=1}^{n_i} \mathscr{Y}_j(\mathbf{x}_i) = \mathsf{Y}(\mathbf{x}_i) + \bar{\varepsilon}(\mathbf{x}_i), \text{ and } \bar{\varepsilon}(\mathbf{x}_i) = \frac{1}{n_i} \sum_{j=1}^{n_i} \varepsilon_j(\mathbf{x}_i) \quad i = 1, 2, \ldots, k \ . \tag{4}$$

That is, $\bar{\mathscr{Y}}(\mathbf{x}_i)$ is the resulting point estimate of the performance measure of interest obtained at design point $\mathbf{x}_i$ and $\bar{\varepsilon}(\mathbf{x}_i)$ is the simulation error associated with it. We write $\bar{\varepsilon}$ as a shorthand for the vector $(\bar{\varepsilon}(\mathbf{x}_1), \bar{\varepsilon}(\mathbf{x}_2), \ldots, \bar{\varepsilon}(\mathbf{x}_k))^\top$. To do global prediction, standard stochastic kriging prescribes using the best linear predictor of $\mathsf{Y}(\mathbf{x}_0)$ that has the minimum mean squared error among all linear predictors at a given point $\mathbf{x}_0$ (see Ankenman et al. 2010, Appendix EC.1),

$$\widehat{\mathsf{Y}}(\mathbf{x}_0) = \mathbf{f}(\mathbf{x}_0)^\top \beta + \Sigma_{\mathsf{M}}(\mathbf{x}_0, \cdot)^\top \Sigma^{-1} \left(\bar{\mathscr{Y}} - \mathbf{F}\beta\right), \tag{5}$$

whose corresponding mean squared error is given by

$$\mathrm{MSE}\left(\widehat{\mathsf{Y}}(\mathbf{x}_0)\right) = \Sigma_{\mathsf{M}}(\mathbf{x}_0,\mathbf{x}_0) - \Sigma_{\mathsf{M}}(\mathbf{x}_0,\cdot)^\top \Sigma^{-1} \Sigma_{\mathsf{M}}(\mathbf{x}_0,\cdot), \tag{6}$$

where $\Sigma = \Sigma_{\mathsf{M}} + \Sigma_\varepsilon$, and $\mathbf{F} = \left(\mathbf{f}(\mathbf{x}_1)^\top, \mathbf{f}(\mathbf{x}_2)^\top, \ldots, \mathbf{f}(\mathbf{x}_k)^\top\right)^\top$. The $k \times k$ matrix Σ_{M} records spatial covariances across the design points, i.e., its (i,h)th entry $\Sigma_{\mathsf{M}}(\mathbf{x}_i,\mathbf{x}_h)$ gives $\mathrm{Cov}(\mathsf{M}(\mathbf{x}_i),\mathsf{M}(\mathbf{x}_h))$ as specified in (3). The $k \times 1$ vector $\Sigma_{\mathsf{M}}(\mathbf{x}_0,\cdot)$ contains the spatial covariances between the k design points and a given prediction point $\mathbf{x}_0$. The $k \times k$ matrix Σ_ε is the variance-covariance matrix of the vector of simulation errors associated with the vector of point estimates $\bar{\mathscr{Y}}$, $\bar{\varepsilon}$. As Chen et al. (2012) show that the use of common random numbers (CRN) does not necessarily help improve the performance of the SK predictor, in this paper we assume that CRN is not applied in simulation experiments. In this case Σ_ε reduces to a $k \times k$ diagonal matrix $\mathrm{diag}\{\mathsf{V}(\mathbf{x}_1)/n_1, \mathsf{V}(\mathbf{x}_2)/n_2, \ldots, \mathsf{V}(\mathbf{x}_k)/n_k\}$ with $\mathsf{V}(\mathbf{x}_i) := \mathrm{Var}(\varepsilon_j(\mathbf{x}_i))$, $j = 1,2,\ldots,n_i$.

To implement SK for the prediction, the standard practice is to first substitute $\widehat{\Sigma}_\varepsilon$ into $\Sigma = \Sigma_{\mathsf{M}} + \Sigma_\varepsilon$, with the ith diagonal entry of $\widehat{\Sigma}_\varepsilon$ specified by the simulation output sample variances for $i = 1,2,\ldots,k$. Prediction then follows (5) and (6) upon obtaining parameter estimates for τ^2, θ and β through maximizing the log-likelihood function formed under the assumption that $(\mathsf{Y}(\mathbf{x}_0), \bar{\mathscr{Y}}^\top)^\top$ follows a multivariate normal distribution (see, e.g., Ankenman et al. 2010, Chen and Kim 2014).

3 HETEROSCEDASTIC VARIANCE ESTIMATION

In this section, we consider several variance estimation methods that provide a smoothed estimator $\widehat{\mathsf{V}}(\mathbf{x})$ for $\mathbf{x} \in \mathscr{X}$. First, we present a result similar to Theorem 1 of Ankenman et al. (2010) based on Assumption 1 given below. That is, estimating Σ_ε with smoothed variance estimates $\widehat{\mathsf{V}}(\mathbf{x}_i)$'s based on the sample variances s_i^2 introduces no prediction bias. The proof is similar to that of Theorem 1 of Ankenman et al. (2010), hence it is omitted.

Assumption 1 (Ankenman et al. (2010)) The random field M is a stationary Gaussian random field, and $\varepsilon_1(\mathbf{x}_i), \varepsilon_j(\mathbf{x}_i), \ldots$ are i.i.d. $\mathscr{N}(0, \mathsf{V}(\mathbf{x}_i))$, independent of $\varepsilon_j(\mathbf{x}_h)$ for all j and $h \neq i$ (i.e., no CRN), and independent of M.

Theorem 1 Let $\widehat{\Sigma}_\varepsilon = \mathrm{diag}\{\widehat{\mathsf{V}}(\mathbf{x}_1), \widehat{\mathsf{V}}(\mathbf{x}_2), \ldots, \widehat{\mathsf{V}}(\mathbf{x}_k)\}$ and define

$$\widehat{\widehat{\mathsf{Y}}}(\mathbf{x}_0) = \mathbf{f}(\mathbf{x}_0)^\top \beta + \Sigma_{\mathsf{M}}(\mathbf{x}_0,\cdot)^\top \left(\Sigma_{\mathsf{M}} + \widehat{\Sigma}_\varepsilon\right)^{-1} \left(\bar{\mathscr{Y}} - \mathbf{F}\beta\right), \tag{7}$$

where the $\widehat{\mathsf{V}}(\mathbf{x}_i)$ is an estimator of $\mathsf{V}(\mathbf{x}_i)$ given by a metamodel built on the sample variances s_i^2 obtained at design point $\mathbf{x}_i$, $i = 1,2,\ldots,k$. If Assumption 1 holds, then $\mathrm{E}[\widehat{\widehat{\mathsf{Y}}}(\mathbf{x}_0) - \mathsf{Y}(\mathbf{x}_0)] = 0$ for any $\mathbf{x}_0 \in \mathscr{X}$.

3.1 Variance Estimation via SK with Sample Variances

We consider building an SK metamodel for variance function estimation. According to Section 2, at design point $\mathbf{x}_i$ we need to obtain not only a point estimate of the simulation variance (e.g., the sample variance $s_i^2 = (n_i - 1)^{-1} \sum_{j=1}^{n_i} \left(\mathscr{Y}_j(\mathbf{x}_i) - \bar{\mathscr{Y}}(\mathbf{x}_i)\right)^2$ can be used) but also a variance estimate of such a point estimator.

Under Assumption 1, $(n_i - 1)s_i^2$ follows a chi-squared distribution with degrees of freedom $n_i - 1$. In this case, the variance of the sample variance s_i^2 can be estimated by $\widehat{\mathsf{V}}_{s^2}(\mathbf{x}_i) = 2s_i^4/(n_i - 1)$. With the sample variance s_i^2 and its corresponding variance estimate $\widehat{\mathsf{V}}_{s^2}(\mathbf{x}_i)$ available at design point $\mathbf{x}_i$ for $i = 1,2,\ldots,k$, a variance SK metamodel can be built, and a pointwise predictor of the simulation variance $\mathsf{V}(\mathbf{x}_0)$ at any $\mathbf{x}_0 \in \mathscr{X}$, $\widehat{\mathsf{V}}(\mathbf{x}_0)$, follows conveniently from (5). Lastly, we note that this approach has also been studied by Kamiński (2015).

3.2 Variance Estimation via SK with Log-Transformed Sample Variances

We consider building an SK metamodel for the log-transformed variance function, as such a transformation ensures the non-negativity of the resulting variance estimates. When the sample size at a design point is not large, however, the log transformation of the sample variance may introduce bias which is not negligible. Standard theory (e.g., Cox and Solomon 2003, Section 6.6.4) can be applied to estimate the bias and variance of such a logarithm transformed variance estimator. Specifically, under Assumption 1, the sampling variance s_i^2 obtained with n_i i.i.d. simulation outputs can be written as $s_i^2 := \mathsf{V}(\mathbf{x}_i)X_\nu/\nu$, where X_ν is a chi-squared random variable with $\nu = n_i - 1$ degrees of freedom. The mean and variance of the log transformed sample variance $\ln(s_i^2)$ follow as

$$\mathrm{E}[\ln(s_i^2)] = \ln \mathsf{V}(\mathbf{x}_i) + \psi(\nu/2) - \ln(\nu/2), \quad \mathrm{Var}[\ln(s_i^2)] = \psi_1(\nu/2).$$

where $\nu = n_i - 1$, the digamma function ψ is defined by $\psi(u) = \frac{d \log \Gamma(u)}{du}$ and ψ_1 is the trigamma function (Abramowitz and Stegun (1964), p. 943). It follows immediately that $\mathrm{Bias}[\ln(s_i^2)] = \psi(\nu/2) - \ln(\nu/2)$.

In light of the aforementioned discussion, a bias-corrected estimator of the log-transformed variance function value at $\mathbf{x}_i$ can be given as $\ln(s_i^2) - \psi\left(\frac{(n_i-1)}{2}\right) + \ln\left(\frac{(n_i-1)}{2}\right)$, and its corresponding variance estimator follows as $\widehat{\mathsf{V}}_{\ln s^2}(\mathbf{x}_i) = \psi_1\left(\frac{(n_i-1)}{2}\right)$.

With the point estimate of log-transformed variance and the corresponding variance estimate $\widehat{\mathsf{V}}_{\ln s^2}(\mathbf{x}_i)$ obtained at design point $\mathbf{x}_i$ for $i = 1, 2, \ldots, k$, an SK metamodel for the log-variance function can be built. A pointwise predictor of the log-variance $\ln(\mathsf{V}(\mathbf{x}_0))$, $\widehat{\ln \mathsf{V}}(\mathbf{x}_0)$, then follows from (5). A pointwise predictor of the simulation variance at $\mathbf{x}_0$ can be obtained subsequently as $\widehat{\mathsf{V}}(\mathbf{x}_0) = \exp\{\widehat{\ln \mathsf{V}}(\mathbf{x}_0)\}$.

3.3 Variance Estimation via SK with Bootstrap Sample Variances

We consider employing the bootstrap sampling method to provide smoothed variance estimates based on the sample variances s_i^2, $i = 1, 2, \ldots, k$ (see, e.g., Cheng 2006 and references therein). Specifically, at design point $\mathbf{x}_i$, denote the set of simulation outputs generated by $\mathbf{S}(\mathbf{x}_i) := \{\mathscr{Y}_j(\mathbf{x}_i)\}_{j=1}^{n_i}$. We draw B bootstrap samples each of size n_i from $\mathbf{S}(\mathbf{x}_i)$, resulting in B bootstrap samples $\{\mathbf{S}_b^*(\mathbf{x}_i)\}_{b=1}^B$, where $\mathbf{S}_b^*(\mathbf{x}_i) = \{\mathscr{Y}_{j,b}^*(\mathbf{x}_i)\}_{j=1}^{n_i}$, for $b = 1, 2, \ldots, B$. Then the bth sample variance, $s_{i,b}^2$, can be calculated using $\mathbf{S}_b^*(\mathbf{x}_i)$, $b = 1, 2, \ldots, B$. As a result, a variance estimate of the sample variance can be given as

$$\widehat{\mathsf{V}}_{s^2}(\mathbf{x}_i) = \frac{1}{B-1} \sum_{b=1}^{B} \left(s_{i,b}^2 - \overline{s_{i,\cdot}^2}\right)^2, \quad \text{with } \overline{s_{i,\cdot}^2} = \frac{1}{B} \sum_{b=1}^{B} s_{i,b}^2. \tag{8}$$

With the sample variance s_i^2 and its corresponding variance estimate $\widehat{\mathsf{V}}_{s^2}(\mathbf{x}_i)$ for $i = 1, 2, \ldots, k$, a variance SK metamodel can be built, and a pointwise predictor of the simulation variance can be given as described in Subsection 3.1.

3.4 Variance Estimation via SK with Bootstrap Log-Transformed Sample Variances

We consider building an SK metamodel for the log-transformed variance function and estimate the variance function based on the metamodel constructed. Following the steps described in Subsection 3.3, we obtain B bootstrap samples $\{\mathbf{S}_b^*(\mathbf{x}_i)\}_{b=1}^B$. Then the bth log-transformed sample variance can be calculated as $\ln(s_{i,b}^2)$, where $s_{i,b}^2$ is the sample variance obtained using the bth bootstrap sample, $b = 1, 2, \ldots, B$. A variance estimate of the log-transformed sample variance, $\widehat{\mathsf{V}}_{\ln s^2}(\mathbf{x}_i)$, can be obtained by replacing $s_{i,b}^2$ with $\ln(s_{i,b}^2)$ in (8).

Following the discussion given in Subsection 3.2, a bias-corrected point estimate of $\ln \mathrm{V}(\mathbf{x}_i)$ follows as $\ln(s_i^2) - \psi\left(\frac{(n_i-1)}{2}\right) + \ln\left(\frac{(n_i-1)}{2}\right)$; and its corresponding variance estimate can be given as $\widehat{\mathrm{V}}_{\ln s^2}(\mathbf{x}_i)$ at design point $\mathbf{x}_i$ for $i = 1, 2, \ldots, k$. As a result, an SK metamodel for the log-variance function can be built, and a pointwise predictor $\widehat{\mathrm{V}}(\mathbf{x}_0)$ of the simulation variance at $\mathbf{x}_0$ can be obtained in a similar fashion as described in Subsection 3.2.

3.5 Variance Estimation via Kernel Regression with Sample Variances

We consider fitting the variance function using the Nadaraya-Watson (NW) kernel regression estimator which is independently proposed by Nadaraya (1964) and Watson (1964). The estimator for the variance function value at $\mathbf{x}_0 \in \mathscr{X} \subseteq \mathbb{R}^d$ is given by

$$\widehat{\mathrm{V}}(\mathbf{x}_0) = \sum_{i=1}^{k} w_i s_i^2, \text{ with } w_i = \frac{K_{\mathbf{H}}(\mathbf{x}_0 - \mathbf{x}_i)}{\sum_{j=1}^{k} K_{\mathbf{H}}(\mathbf{x}_0 - \mathbf{x}_j)}, \tag{9}$$

where the NW kernel regression estimator is just a weighted sum of the sample variances s_i^2, $i = 1, 2, \ldots, k$. In (9), $\mathbf{H}$ represents a $d \times d$ bandwidth matrix that is symmetric and positive definite, $K(\mathbf{u}) = K(u_1, u_2, \ldots, u_d)$ represents a d-variate kernel function such that $\int K(\mathbf{u}) d\mathbf{u} = 1$, and for $\mathbf{x} \in \mathbb{R}^d$, $K_{\mathbf{H}}(\mathbf{x}) := |\mathbf{H}|^{-1/2} K(\mathbf{H}^{-1/2}\mathbf{x})$. Popular kernel functions include the Gaussian kernel, the cubic kernel, and the Epanechnikov kernel, among others (Härdle 1990). The inclusion of off-diagonal elements in $\mathbf{H}$ allows the data to be arbitrarily rotated; however, the estimation of a full bandwidth matrix is difficult when d is not small. Wand and Jones (1993) argue that typically it suffices to smooth different directions independently by using a diagonal bandwidth matrix $\mathbf{H} = \mathrm{diag}(h_1, h_2 \ldots, h_d)$ with $h_i \neq h_j$ for $i \neq j$.

The choice of kernel function is not critical to the performance of the kernel regression estimator whereas the choice of the bandwidths h_i's is (Härdle 1990, Wand and Jones 1993). There exists a rich literature on bandwidth selection, and numerous methods have been developed (Härdle and Müller 2000), which can be classified into two types: (1) plug-in approaches which minimize some criterion (such as the mean integrated squared error) theoretically and estimate the minimizers directly; and (2) resampling methods such as cross-validation (CV) and bootstrap techniques that estimate the integrated squared error function and locate the minimizers correspondingly. In this paper we adopt the Gaussian kernel function $K(\mathbf{x}) = (2\pi)^{-d/2} \exp\{-\frac{1}{2}\mathbf{x}^\top \mathbf{x}\}$ and a diagonal bandwidth matrix $\mathbf{H}$ with the rule-of-thumb bandwidths given by Härdle and Müller (2000). Besides the NW estimator, we note that a local polynomial estimator can also be constructed for variance function estimation (Fan and Gijbels 1992, Fan and Gijbels 1995), which may enjoy a better behavior near the edges of the design region. We will only consider the NW estimator in this paper.

4 NUMERICAL EXAMPLES

In this section, we present two numerical experiments to show the predictive performance of SK with smoothed variance estimates obtained by methods described in Section 3. The examples considered are an M/M/1 queue example and a three-dimensional synthetic example.

We give a brief description of the common experiment setup used in both examples. A simulation experiment is performed with a total simulation budget of C replications to expend at k distinct design points, with n_i simulation replications assigned at design point $\mathbf{x}_i$, for $i = 1, 2, \ldots, k$. Two budget allocation schemes, i.e., the equal and unequal allocations, are considered. Specifically, the unequal budget allocation scheme sets $n_i = \left\lceil \frac{\sqrt{\mathrm{V}(\mathbf{x}_i)}}{\sum_{i=1}^{k} \sqrt{\mathrm{V}(\mathbf{x}_i)}} C \right\rceil$, whereas $n_i = \lceil C/k \rceil$ is specified under the equal allocation scheme. We note that the unequal allocation scheme assigns the simulation budget in accordance with the scale function $\sigma(\mathbf{x}) := \sqrt{V(\mathbf{x})}$ for $\mathbf{x} \in \mathscr{X}$; it is proposed based on the optimal design density $h^*(\cdot)$ established

for controlled nonparametric regression experiments. Specifically, $h^*(x) = \sigma(x)/\int_{\mathscr{X}} \sigma(x)dx$ achieves the asymptotic sharp minimax lower bound for the mean integrated squared error (MISE) when $x \in \mathscr{X} = [0,1]$; see Efromovich (2007) and references therein for more details.

For both examples, we repeat the simulation experiment for 100 independent macro-replications and calculate the empirical root mean squared error (ERMSE) achieved by SK with different variance estimation methods applied on each macro-replication. Specifically, a check-point set X^K consists of K points in $\mathscr{X}$ is generated by a space-filling design, and the predictive accuracy measure ERMSE is defined by

$$\text{ERMSE}_\ell = \sqrt{\frac{1}{K}\sum_{i=1}^{K}\left(\widehat{\text{Y}}_\ell(\mathbf{x}_i) - m(\mathbf{x}_i)\right)^2}, \quad \ell = 1,2,\ldots,100, \tag{10}$$

where $\widehat{\text{Y}}_\ell(\cdot)$ represents the prediction delivered by SK with a given variance estimation method on the ℓth macro-replication; and $m(\cdot)$ denotes the underlying mean function of interest.

The variance estimation methods considered are labeled as s^2, $\ln s^2$, s^2-bs, $\ln s^2$-bs, and kr, respectively, in the order of which they are introduced in Section 3. We note that the parameters and their respective values considered in the numerical examples are summarized in Table 1.

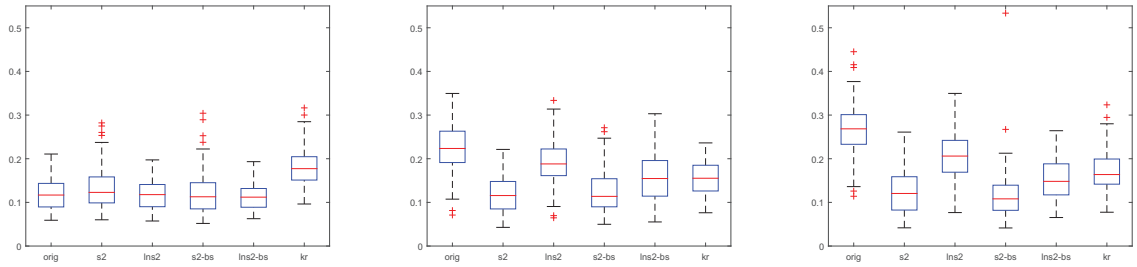
Table 1: Parameters for numerical experiments.

Parameter	Interpretation	Value
K	number of prediction points	1000
C	total computation budget	2000
k	number of design points	$\{10d$, 50, 80, 100, 125, 200$\}$ for $d = 1,3$
n_i	sample size at design point $\mathbf{x}_i$	$\lceil C/k \rceil$ for an equal allocation, $\left\lceil \frac{\sqrt{\text{V}(\mathbf{x}_i)}}{\sum_{i=1}^{k}\sqrt{\text{V}(\mathbf{x}_i)}} C \right\rceil$ for an unequal allocation
B	bootstrap sample size for variance estimation	100

4.1 An M/M/1 Queue Example

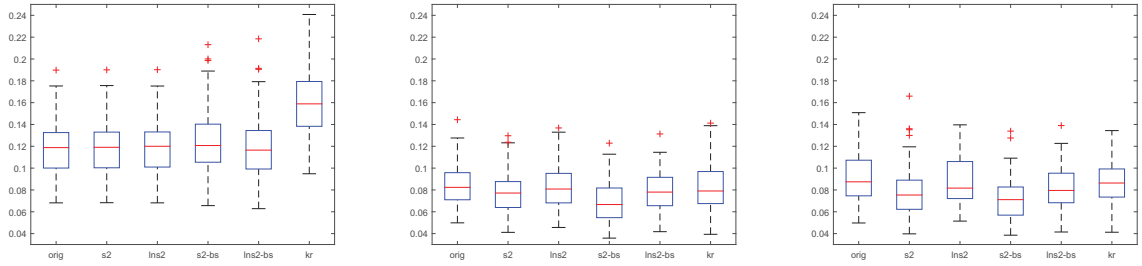
Consider simulating an M/M/1 queue with arrival rate 1 and service rate x with $x \in \mathscr{X} = [1.1,2]$. It is well known from queueing theory that the mean steady-state waiting time in the queue is $m(x) = 1/\big(x(x-1)\big)$; this is the function we intend to estimate. The general experiment setup is as given at the beginning of Section 4, and the example specific setup is given as follows. For each simulation experiment, a set of k equispaced design points are chosen from $\mathscr{X}$, among which $x_1 = 1.1$ and $x_k = 2$. Each simulation replication (run) is initialized in steady-state, and the run length T is determined by the number of simulated customers, with $T \in \{500, 1000, 3000\}$. The simulation output on the jth replication, $\mathscr{Y}_j(\mathbf{x})$, is the average waiting time of the T customers. For large T, the variance function $\text{V}(x) = \text{Var}[\mathscr{Y}_j(\mathbf{x})] \approx 4/\left(x(1-1/x)^4 T\right)$ (Whitt 1989). A grid of $K = 1000$ equispaced check-points are chosen from $\mathscr{X}$ to calculate the ERMSEs.

Summary. The ERMSEs obtained for simulation experiments with $T = 1000$ and $k \in \{10, 50, 100\}$ are shown in Figures 1 and 2, but those for the cases with $T \in \{500, 3000\}$ and $k \in \{80, 125, 200\}$ are left out, based on which similar conclusions can be reached. First, a comparison of boxplots shown in Figures 1 and 2 indicates that all the methods perform better under the unequal allocation scheme for the M/M/1 example. Second, using smoothed variance estimates does help improve the predictive accuracy of SK as compared to standard SK (using the sample variances directly); this is particularly true when the number of design points k is not small (e.g., $k > 20$). We note without showing the details that this benefit is more evident as the impact of the simulation variance becomes more pronounced, an observation made when examining the results obtained with $T = 500, 1000$, and 3000. Third, it seems that each method achieves



(a) $k = 10$ (b) $k = 50$ (c) $k = 100$

Figure 1: M/M/1: ERMSEs obtained with an equal allocation and the run length of 1000.



(a) $k = 10$ (b) $k = 50$ (c) $k = 100$

Figure 2: M/M/1: ERMSEs obtained with an unequal allocation and the run length of 1000.

its best performance under the unequal allocation scheme with a moderately large number of design points (e.g., $k = 50$ or 80).

Some observations can be made regarding the performance of SK with each variance estimation method applied as well. (i) Variance estimation using the sample variances with their respective bootstrap variance estimates seems to help SK achieve the best predictive performance when compared to the other variance estimation methods. (ii) Variance estimation using the sample variances with their respective theoretical variance approximates provides a comparable predictive performance relative to that of its bootstrap variance estimation counterpart. (iii) Variance estimates generated by smoothing the log-transformed sample variances (resp. the bootstrap log-transformed sample variances) do not seem to provide a noticeable improvement on the predictive accuracy of SK as compared to the variance estimates obtained by smoothing the sample variances (resp. bootstrap sample variances) without applying the logarithm transformation. (iv) The performance of SK with variance estimates obtained by kernel regression is not competitive when the number of design points k is small (say, $k \leq 20$). However, the performance of SK with this variance method applied is relatively robust to the experimental design used, as long as the number of design points k is moderately large (say, $k \geq 40$).

4.2 The Three-Dimensional Hartmann Example

Consider the following 3-dimensional Hartmann function:

$$m(\mathbf{x}) = -\sum_{i=1}^{4} c_i \exp\left(-\sum_{j=1}^{3} a_{ij}(x_j - p_{ij})^2\right), \tag{11}$$

where $x_i \in [0,1], i = 1,2,3$; the parameter vector is $c = (1, 1.2, 3, 3.2)^\top$, and a_{ij}, p_{ij} are respectively defined in the 4×3 matrices $A = \{a_{i,j}\}$ and $P = \{p_{i,j}\}$ specified in Table 2.

Table 2: Parameters for the 3D-Hartmann example.

A			P		
3	10	30	0.3689	0.117	0.2673
0.1	10	35	0.4699	0.4387	0.747
3	10	30	0.1091	0.8732	0.5547
0.1	10	35	0.03815	0.5743	0.8828

The simulation output on each replication at a point $\mathbf{x}$ is generated as $\mathscr{Y}_j(\mathbf{x}) = m(\mathbf{x}) + \varepsilon_j(\mathbf{x})$, where the $\varepsilon_j(\mathbf{x})$'s are i.i.d. $\mathscr{N}(0, \mathsf{V}(\mathbf{x}))$ random variables. The simulation variance is set as $\mathsf{V}(\mathbf{x}) = \delta|m(\mathbf{x})|$, where the parameter δ controls the magnitude of the simulation variance. The general experiment setup is as described at the beginning of Section 4, and the example specific setup is given as follows. For each simulation experiment, a Latin hypercube sample of k design points in $\mathscr{X} = [0,1]^3$ are generated. The impact of the simulation variance is adjusted by varying δ in $\{0.1, 1, 10\}$. A Latin hypercube sample of $K = 1000$ check-points in $\mathscr{X}$ is generated to calculate the predictive accuracy measure ERMSE.

Table 3: A summary of the ERMSEs obtained with an equal allocation for the 3-D example with $\mathsf{V}(\mathbf{x}) = |m(\mathbf{x})|$.

k	orig	s^2	$\ln s^2$	s^2-bs	$\ln s^2$-bs	kr
	0.197	0.155	0.221	0.164	0.229	0.195
200	(0.003)	(0.002)	(0.004)	(0.002)	(0.004)	(0.003)
	0.209	0.194	0.225	0.203	0.220	0.215
125	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
	0.213	0.172	0.221	0.176	0.218	0.193
100	(0.004)	(0.003)	(0.004)	(0.003)	(0.004)	(0.003)
	0.214	0.198	0.227	0.200	0.233	0.214
80	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)
	0.242	0.227	0.242	0.227	0.244	0.240
50	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
	0.325	0.319	0.326	0.299	0.308	0.322
30	(0.006)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)

Table 4: A summary of the ERMSEs obtained with an unequal allocation for the 3-D example with $\mathsf{V}(\mathbf{x}) = |m(\mathbf{x})|$.

k	orig	s^2	$\ln s^2$	s^2-bs	$\ln s^2$-bs	kr
	0.152	0.170	0.147	0.153	0.144	0.136
200	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.001)
	0.160	0.165	0.170	0.158	0.161	0.163
125	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
	0.150	0.145	0.152	0.143	0.152	0.139
100	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
	0.161	0.162	0.159	0.162	0.160	0.159
80	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
	0.206	0.202	0.205	0.201	0.203	0.201
50	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)
	0.290	0.284	0.289	0.289	0.294	0.287
30	(0.004)	(0.004)	(0.004)	(0.006)	(0.006)	(0.004)

Table 5: A summary of the ERMSEs obtained with an equal allocation for the 3-D example with $\mathsf{V}(\mathbf{x}) = 10|m(\mathbf{x})|$.

k	orig	s^2	$\ln s^2$	s^2-bs	$\ln s^2$-bs	kr
	0.446	0.350	0.467	0.366	0.478	0.331
200	(0.009)	(0.005)	(0.009)	(0.005)	(0.010)	(0.005)
	0.442	0.364	0.459	0.370	0.464	0.346
125	(0.009)	(0.005)	(0.009)	(0.005)	(0.009)	(0.006)
	0.491	0.388	0.503	0.395	0.500	0.333
100	(0.009)	(0.007)	(0.009)	(0.007)	(0.009)	(0.006)
	0.441	0.379	0.438	0.367	0.425	0.355
80	(0.008)	(0.006)	(0.008)	(0.004)	(0.007)	(0.006)
	0.473	0.390	0.470	0.396	0.473	0.352
50	(0.009)	(0.006)	(0.009)	(0.006)	(0.008)	(0.006)
	0.409	0.386	0.407	0.374	0.400	0.405
30	(0.008)	(0.007)	(0.008)	(0.005)	(0.006)	(0.008)

Table 6: A summary of the ERMSEs obtained with an unequal allocation for the 3-D example with $\mathsf{V}(\mathbf{x}) = 10|m(\mathbf{x})|$.

k	orig	s^2	$\ln s^2$	s^2-bs	$\ln s^2$-bs	kr
	0.376	0.430	0.356	0.380	0.324	0.291
200	(0.006)	(0.007)	(0.004)	(0.007)	(0.005)	(0.005)
	0.302	0.318	0.325	0.313	0.321	0.317
125	(0.005)	(0.006)	(0.004)	(0.005)	(0.004)	(0.005)
	0.324	0.320	0.341	0.314	0.339	0.299
100	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)
	0.327	0.310	0.335	0.325	0.336	0.309
80	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)	(0.005)
	0.332	0.318	0.327	0.316	0.324	0.333
50	(0.005)	(0.004)	(0.005)	(0.004)	(0.004)	(0.006)
	0.359	0.358	0.358	0.355	0.356	0.395
30	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.008)

Summary. A summary of the average ERMSEs obtained for the 100 macro-replications with $\delta = 1$ and 10 are given in Tables 3 to 6; the values in parentheses are the corresponding standard errors. The results for the cases with $\delta = 0.1$ are omitted to economize on space. First, a comparison of Tables 3 and 5 with Tables 4 and 6 indicates that all the methods perform better under the unequal allocation scheme. Second, using smoothed variance estimates helps improve the predictive accuracy of SK as compared to standard SK, especially when the number of design points k is not small (e.g., $k > 30$). It becomes more beneficial as the simulation variances increase, as observed in the M/M/1 example. Third, regardless of the level of the simulation variances, it seems that each method achieves its best performance under the unequal allocation scheme with a moderately large number of design points (e.g., $k \geq 100$).

Some observations follow regarding the performance of SK with each variance estimation method applied. (i) The two variance estimation methods that use the sample variances without applying the logarithm transformation and the kernel regression estimation method lead to the most accurate prediction results, when applied with SK. (ii) Smoothing the log-transformed sample variances (resp. the bootstrap log-transformed sample variances) does not seem to provide a noticeable improvement in the predictive accuracy of SK as compared to smoothing the sample variances (resp. bootstrap sample variances) without applying the logarithm transformation; this is particularly true under the equal allocation scheme. (iii) SK with variance estimates due to kernel regression works best with a moderately large number of design points (say, $k \geq 50$).

5 STOCHASTIC KRIGING-BASED DUAL METAMODELING

In this section, we propose an SK-based dual metamodeling approach for estimating the mean and variance functions simultaneously, inspired by the literature on residual-based nonparametric variance estimation (Fan and Yao 1998, Carroll and Ruppert 1988) and Gaussian process based prediction (Boukouvalas and Cornford 2009, Kersting et al. 2007, Opsomer et al. 1999). One distinctive feature of this approach is that it intends to make an efficient budget allocation of C simulation replications based on the initial estimation of the variance function by smoothing the squared residuals resulting from an earlier fit of the mean function. The approach is briefly summarized by Algorithm 1.

Algorithm 1 SK-based dual metamodeling

Step 1 Given an initial space-filling design consisting of k design points, perform a single simulation replication at each one of them to obtain the simulation output $\mathscr{Y}_0(\mathbf{x}_i), i = 1,2,\ldots,k$. Build a kriging model with nugget effect using the dataset $\{(\mathbf{x}_i, \mathscr{Y}_0(\mathbf{x}_i))\}_{i=1}^k$, and make an initial prediction at the k design points, $\widehat{\mathrm{Y}}_0(\mathbf{x}_i), i = 1,2,\ldots,k$.

Step 2 Calculate the residual $r_0(\mathbf{x}_i)$ resulting from this initial prediction at $\mathbf{x}_i$ using $r_0(\mathbf{x}_i) = \mathscr{Y}_0(\mathbf{x}_i) - \widehat{\mathrm{Y}}_0(\mathbf{x}_i)$, for $i = 1,2,\ldots,k$. Construct a metamodel for variance function estimation by smoothing the squared residuals $r_0^2(\mathbf{x}_i)$'s using kernel regression and obtain the variance estimate $\widehat{\mathrm{V}}_1(\mathbf{x}_i)$ at design point $\mathbf{x}_i$ for $i = 1,2,\ldots,k$.

Step 3 Allocate the remaining budget of $C-k$ simulation replications to the k design points. Specifically, run n_i replications at $\mathbf{x}_i$, where $n_i = \left\lceil \frac{\sqrt{\widehat{\mathrm{V}}_1(\mathbf{x}_i)}}{\sum_{i=1}^k \sqrt{\widehat{\mathrm{V}}_1(\mathbf{x}_i)}}(C-k) \right\rceil$ and obtain the outputs $\{\mathscr{Y}_j(\mathbf{x}_i)\}_{j=1}^{n_i}$. At $\mathbf{x}_i$, based on the simulation outputs obtained $\mathscr{S}(\mathbf{x}_i) = \{\mathscr{Y}_j(\mathbf{x}_i)\}_{j=0}^{n_i}$, calculate the sample variance and sample mean s_i^2 and $\bar{\mathscr{Y}}(\mathbf{x}_i)$, for $i = 1,2,\ldots,k$.

Step 4 Build a new metamodel for variance function estimation based on $\{(\mathbf{x}_i, s_i^2)\}_{i=1}^k$ using a variance estimation method (e.g., those given in Section 3). Obtain the updated variance estimates $\widehat{\mathrm{V}}_2(\mathbf{x}_i)$ at $\mathbf{x}_i$, for $i = 1,2,\ldots,k$.

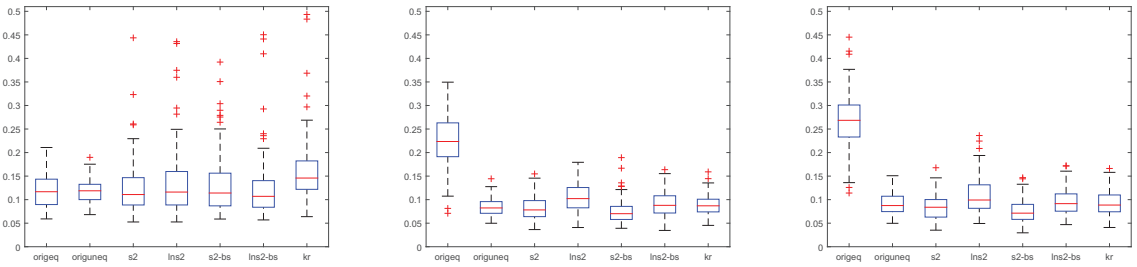
Step 5 Construct an SK metamodel with $\{\mathbf{x}_i, \bar{\mathscr{Y}}(\mathbf{x}_i), \widehat{\mathrm{V}}_2(\mathbf{x}_i)/(n_i+1)\}_{i=1}^k$ and obtain an SK predictor $\widehat{\mathrm{Y}}(\mathbf{x}_i)$ at any $\mathbf{x}_0 \in \mathscr{X}$.

Let us revisit the M/M/1 example given in Subsection 4.1; recall that the unequal allocation scheme is found to be more efficient than the equal budget allocation scheme. However, the unequal allocation given in Section 4 is unavailable due to our lack of knowledge of the variance function. The SK-based dual metamodeling approach can be utilized as a remedy to achieve an efficient budget allocation rule that would otherwise be impossible to apply.

The experimental setup is as described in Subsection 4.1. Figures 3(a)–(c) show the ERMSEs obtained for 100 macro-replications. The boxplots with labels "origeq" and "origuneq" respectively summarize the ERMSEs obtained by SK with the equal and unequal budget allocation schemes given in Subsection 4.1; in particular, "origuneq" corresponds to SK applied with the unequal allocation scheme specified assuming that the variance function is given. The remaining boxplots summarize the ERMSEs obtained by the aforementioned SK-based dual metamodeling approach, with the variance estimation method (specified by the label) applied in Step 4. We observe that when the number of design points k is moderately large, the proposed SK-dual metamodeling approach can achieve a predictive accuracy as high as that of SK with the "ideal" unequal allocation scheme, if not better.

We make a few remarks on Algorithm 1. First, one can use other methods (such as kernel regression, local polynomial regression and least-squares support vector machine, Suykens et al. (2002)) in Steps 1 and 2 for an initial fit of the mean function and the subsequent estimation of the variance function. Second, a relatively dense initial space-filling design is recommended to provide a good coverage of the design

space $\mathscr{X}$, hence k should be reasonably large (e.g., $k \geq 20d$). Third, one can modify Step 3 and iterate the algorithm a few times, until the total simulation budget C is exhausted. In particularly, one can choose to focus on a subset of the k design points for budget allocation, based on some selection criterion. For more details, see Wang and Chen (2016).



(a) $k = 10$ (b) $k = 50$ (c) $k = 100$

Figure 3: ERMSEs obtained by the SK dual metamodeling approach for the M/M/1 example with the simulation run length of 1000.

REFERENCES

Abramowitz, M., and I. A. Stegun. 1964. *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, Volume 55. Courier Corporation.

Aitkin, M. 1987. "Modelling variance heterogeneity in normal regression using GLIM". *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 36:332–339.

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

Boukouvalas, A., and D. Cornford. 2009. "Learning heteroscedastic Gaussian processes for complex datasets". Technical report, Neural Computing Research Group, Aston University.

Brown, L. D., and M. Levine. 2007. "Variance estimation in nonparametric regression via the difference method". *The Annals of Statistics* 35:2219–2232.

Carroll, R. J., and D. Ruppert. 1988. *Transformation and Weighting in Regression*. New York: Chapman and Hall.

Chen, X., B. E. Ankenman, and B. L. Nelson. 2012. "The effects of common random numbers on stochastic kriging metamodels". *ACM Transactions on Modeling and Computer Simulation* 22:7/1–7/20.

Chen, X., and K.-K. Kim. 2014. "Stochastic kriging with biased sample estimates". *ACM Transactions on Modeling and Computer Simulation* 24:8/1–8/23.

Cheng, R. C. H. 2006. "Resampling methods". In *Elsevier Handbooks in Operations Research and Management Science: Simulation*, edited by S. G. Henderson and B. L. Nelson. New York: Elsevier.

Cox, D. R., and P. J. Solomon. 2003. *Components of Variance*. Boca Raton, Florida: Chapman and Hall CRC.

Efromovich, S. 2007. "Sequential design and estimation in heteroscedastic nonparametric regression". *Sequential Analysis* 26:3–25.

Fan, J., and I. Gijbels. 1992. "Variable bandwidth and local linear regression smoothers". *The Annals of Statistics* 20:2008–2036.

Fan, J., and I. Gijbels. 1995. "Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation". *Journal of the Royal Statistical Society. Series B (Methodological)* 57:371–394.

Fan, J., and Q. Yao. 1998. "Efficient estimation of conditional variance functions in stochastic regression". *Biometrika* 85:645–660.

Gasser, T., L. Sroka, and C. Jennen-Steinmetz. 1986. "Residual variance and residual pattern in nonlinear regression". *Biometrika* 73:625–633.

Hall, P., and R. J. Carroll. 1989. "Variance function estimation in regression: the effect of estimating the mean". *Journal of the Royal Statistical Society. Series B (Methodological)* 51:3–14.

Härdle, W. 1990. *Applied Nonparametric Regression*. Chapman and Hall.

Härdle, W., and M. Müller. 2000. "Multivariate and semiparametric kernel regression". In *Smoothing and Regression: Approaches, Computation, and Application*, edited by M. G. Schimek. New York: John Wiley & Sons, Inc.

Jobson, J. D., and W. A. Fuller. 1980. "Least squares estimation when the covariance matrix and parameter vector are functionally related". *Journal of the American Statistical Association* 75:176–181.

Kamiński, B. 2015. "A method for the updating of stochastic kriging metamodels". *European Journal of Operational Research* 247:859–866.

Kersting, K., C. Plagemann, P. Pfaff, and W. Burgard. 2007. "Most liekly heteroscedastic Gaussian process regression". In *Proceedings of the 24th International Conference on Machine Learning*, ICML '07, 393–400: ACM.

Kleijnen, J. P. C. 2008. *Design and Analysis of Simulation Experiments*. New York: Springer.

Nadaraya, E. A. 1964. "On estimating regression". *Theory of Probability & Its Applications* 9:141–142.

Opsomer, J. D., D. Ruppert, M. P. Wand, U. Holst, and O. Hössjer. 1999. "Kriging with nonparametric variance function estimation". *Biometrics* 55:704–710.

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

Suykens, J., T. V. Gestel, J. D. Brabanter, B. D. Moor, and J. Vandewalle.. 2002. *Least Squares Support Vector Machines*. Singapore: World Scientific.

Wand, M. P., and M. C. Jones. 1993. "Comparison of smoothing parameterizations in bivariate kernel density estimation". *Journal of the American Statistical Association* 88:520–528.

Wang, W., and X. Chen. 2016. "The effects of estimation of heteroscedasticity on stochastic kriging". Technical report, Department of Industiral and Systems Engineering, Virginia Tech.

Watson, G. S. 1964. "Smooth regression analysis". *Sankhyā: The Indian Journal of Statistics, Series A* 26:359–372.

Whitt, W. 1989. "Planning queueing simulations". *Management Science* 35:1341–1366.

AUTHOR BIOGRAPHIES

WENJING WANG is a Ph.D. student in the Grado Department of Industrial and Systems Engineering at Virginia Tech. Her current research interests include stochastic modeling and simulation optimization. Her email address is wenjing@vt.edu.

XI CHEN is an Assistant 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/a/vt.edu/xi-chen/.