Computers & Industrial Engineering 61 (2011) 760–777

Contents lists available at ScienceDirect

Computers & Industrial Engineering

journal homepage: www.elsevier.com/locate/caie

Kriging metamodel with modified nugget-effect: The heteroscedastic variance case ☆

J. Yin, S.H. Ng *, K.M. Ng

Department of Industrial and System Engineering, National University of Singapore, 10 Kent Ridge Crescent, Singapore 119260, Singapore

ARTICLE INFO

Article history:
Received 18 October 2008
Received in revised form 15 April 2011
Accepted 19 May 2011
Available online 27 May 2011

Keywords:
Simulation
Heterogeneous variance
Kriging
Nugget-effect
Metamodel

ABSTRACT

Metamodels are commonly used to approximate and analyze simulation models. However, in cases where the simulation output variances are non-zero and not constant, many of the current metamodels which assume homogeneity, fail to provide satisfactory estimation. In this paper, we present a kriging model with modified nugget-effect adapted for simulations with heterogeneous variances. The new model improves the estimations of the sensitivity parameters by explicitly accounting for location dependent non-constant variances and smoothes the kriging predictor's output accordingly. We look into the effects of stochastic noise on the parameter estimation for the classic kriging model that assumes deterministic outputs and note that the stochastic noise increases the variability of the classic parameter estimation. The nugget-effect and proposed modified nugget-effect stabilize the estimated parameters and decrease the erratic behavior of the predictor by penalizing the likelihood function affected by stochastic noise. Several numerical examples suggest that the kriging model with modified nugget-effect outperforms the kriging model with nugget-effect and the classic kriging model in heteroscedastic cases.

1. Introduction

Computer simulation is commonly used in industry as a tool to aid in studying the system's characteristics and behaviors. It is especially useful in system optimization problems, where the costs can be greatly reduced by running experiments on the simulation models instead of the real systems. As the complexity of the simulation model increases, the computing cost of running experiments on the simulation model becomes much higher. Metamodels have been applied as simplified approximations to the complex simulation model; see Kleijnen (1987, 1998). Replacing the simulation model with a metamodel in expensive experiments can increase the efficiency and lower the computing costs. A review of metamodel applications in engineering can be found in Simpson, Peplinski, Koch, and Allen (2001). Among the different types of metamodels available, the spatial correlation model, also known as the kriging model, is one of the more promising metamodels as it is more flexible than regression models and not as complicated and time consuming as artificial intelligence (AI) techniques; see Li, Ng, Xie, and Goh (2010) for a comparative study. The kriging model was originally developed in the field of geo-statistics; see Matheron (1963). It was first introduced into Design and Analysis of Computer Experiment (DACE) by Sacks, Welch, Mitchell, and Wynn (1989) and Sacks, Schiller, and Welch (1989). Recently, there is an increasing interest in adopting kriging metamodels in industrial engineering problems and applications (e.g. Ankenman, Nelson, & Staum, 2010; Huang, Allen, Notz, & Zeng, 2006; Sakata, Ashida, & Zako, 2007; Wang, Li, Li, & Zhong, 2008).

The kriging model is very suitable for deterministic simulation problems. It is attractive for its interpolating characteristic, providing predictions with the same values as the observations. For example, in Gupta, Yu, Xu, and Reinikainen (2006), the kriging metamodel is adopted for its interpolating characteristic. For stochastic simulations where the responses at the same location vary (for example in a simulation of a queueing system), the interpolation characteristic of kriging models becomes less desirable. In order to model the random fluctuations in stochastic situations, the nugget-effect is introduced. The term "nugget" is borrowed from geo-statistics, referring to the unexpected nugget of gold found in a mining process. According to Cressie (1993, p. 127), the nugget-effect in geo-statistics is caused by two factors: micro-scale variation and measurement error. In this article, we assume that the system studied can be modeled as an L_2-continuous random process (see Cressie, 1993, p. 112), and hence the nugget-effect studied here is purely caused by the random measurement error (or random noise).

The nugget-effect in kriging assumes second-order stationarity and is typically used to model white noise effect. Most kriging publications assume that the variance of the random error is homogeneous and the kriging model with nugget-effect is sufficient to

☆ This manuscript was processed by Area Editor Paul Savory.
* Corresponding author. Tel.: +65 6516 3095; fax: +65 6777 1434.
E-mail address: isensh@nus.edu.sg (S.H. Ng).

0360-8352/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cie.2011.05.008

solve the problem. However, there are many real world situations where the homoscedastic assumption does not hold. These include queueing systems and networks which can be found in many industrial engineering problems. When applying the homoscedastic kriging model in a heteroscedastic case, the fit can be poor, especially when the sample size is small. We illustrate the noisy applications with the simple function displayed in Fig. 1.

The test function consists of a second-order signal function and a noisy function with step variance.

$$y = S(x) + \varepsilon = x^2 + \varepsilon \tag{1}$$

where ε indicates the random noise component, with variance $\sigma_\varepsilon^2 = 0.083$ when $x \in [-5, 2)$, and $\sigma_\varepsilon^2 = 8.3$ when $x \in [2, 5]$. In Fig. 1, the solid line indicates the signal function $y = x^2$, and the dots are the noisy observations of the signal function $y = x^2 + \varepsilon$.

In the traditional application of kriging in stochastic simulations, replications are taken at each observation point and the averages of the replicates at each point are used as the inputs to the model. Kleijnen (2008, p. 92) recommends at least $n \geqslant 2$ replications to be taken equally at each observation point when no prior knowledge on the variance forms is available, otherwise, the simulation exercise may be meaningless due to the variability in the data. In this test function example, we assume that a budget for only 76 runs is available. Based on this, we spread 19 points from -5 to 5, taking four replicates at each point. The averages of the four replicates at each of the 19 points are used as the inputs of the model. The solid line in Fig. 2 plots the fit of the traditional deterministic ordinary kriging (OK) model.

With limited replications and input points, the ordinary kriging model's predictor output is poor with obvious fluctuations away from the true function when $x < -4$ and $x > 1$. Because the traditional ordinary kriging model is designed under deterministic assumptions, random noise can cause an ill fit and result in disappointing predictions. We note that the predictor output will improve as more replications and observation points are taken. However, in many practical applications of simulation, the computer model can be complicated and time consuming to run (see Gramacy and Lee, 2009; Gupta et al., 2006), limiting the number of observation points and replications that can be taken.

Considering the kriging model with nugget-effect which has a homogenous variance assumption, we pool the sample variances at the 19 observation points to estimate the nugget-effect. The predictor output adopting this model is plotted as the dashed line in Fig. 2.

As seen in Fig. 2, the nugget-effect predictor's output is smoother than the OK predictor. However, in the region $x \in [2, 5]$ where the variance is higher, the fit is poor compared with the fit in the region $x \in [-5, 2)$. This indicates that the nugget-effect model can still be inadequate as the heterogeneous variance can have an impact on local predictions. Moreover, due to the homogeneous noise assumptions of this model, there is no clear method to estimate the nugget-effect under these heterogeneous conditions.

This same phenomenon occurs in the simulation of the M/M/1 queue, one of the most basic queueing models. Van Beers and Kleijnen (2003) proposed a detrending approach to model out the trend in the data using least squares methods and then apply the deterministic ordinary kriging model to the detrended data. Two alternative methods were later proposed by Kleijnen and Van Beers (2005) to improve the application of kriging in stochastic problems: the replication method and the studentization method. The replication method proposes that the heteroscedastic problem can be converted into a homoscedastic problem by taking appropriate replications at all the observation locations. This method requires a sequential design with sufficient computing resources to run all the replications. For example, in Fig. 1, the number of replicates needed in the region with higher variance should be 100 times larger than the number of replicates in the region with lower variance (because the variance is 100 times bigger in the former region) in order to convert the heteroscedastic case into a homoscedastic case. In the study of the M/M/1 queue system for the case where the computing budget is limited, both the OK and nugget-effect model with the application of this replication method can still be inadequate. The studentization method is developed on the basis of the detrended kriging approach. The main idea is to

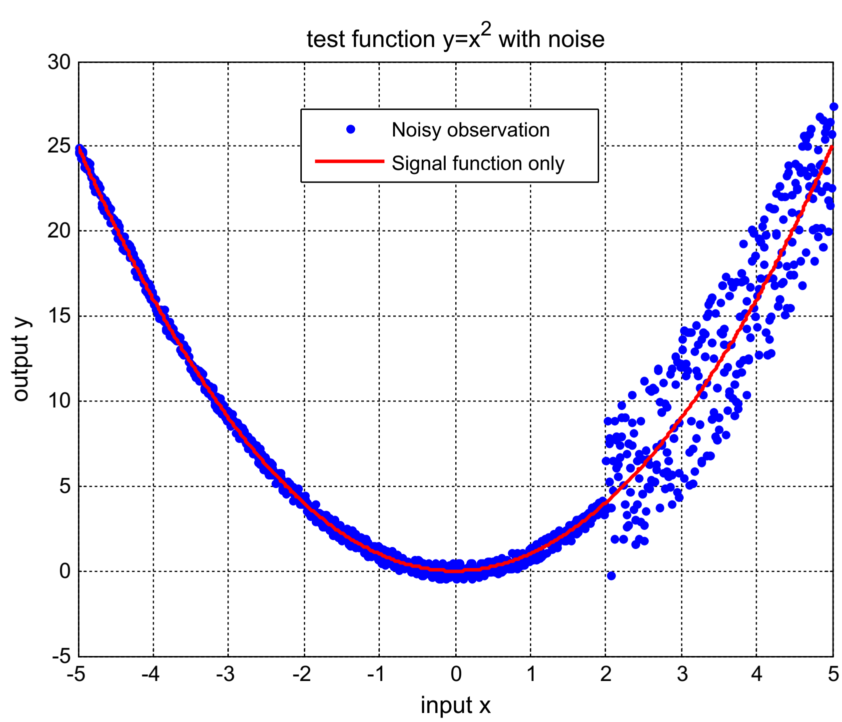

Fig. 1. Test function with step variance function.

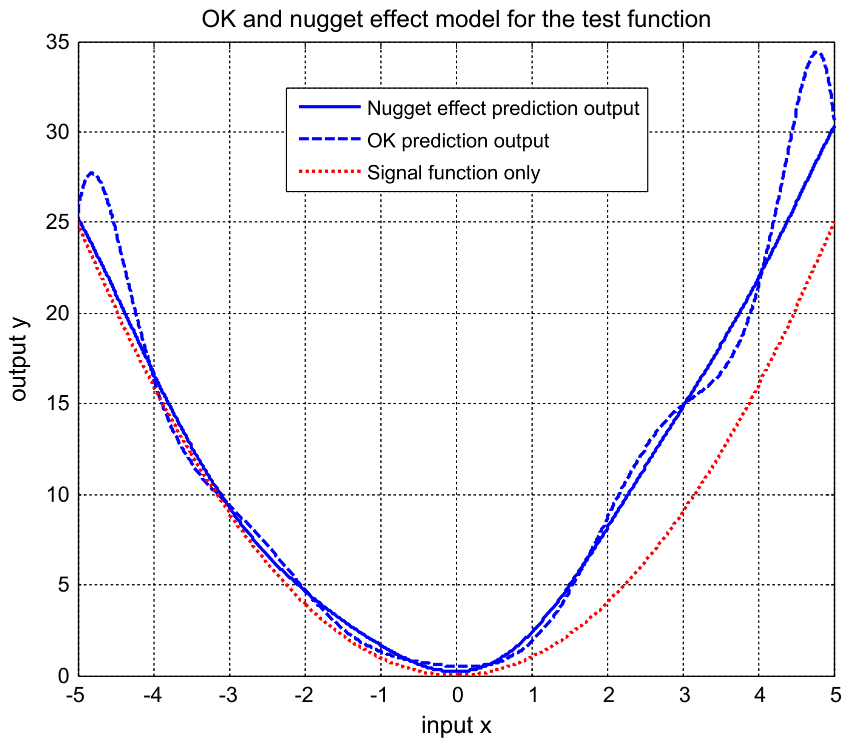

Fig. 2. Ordinary kriging and nugget-effect model for the test function.

model the trend in the data and then standardize the detrended data. It is an intuitive method to handle inputs with different variances. However, in their numerical examples, this method did not improve much over the OK model. This is due to the amplification of the uncertainty in the estimation of the signal function and variance in the transformation of the predictor, especially when the sample size is small.

As seen in Fig. 2, both the OK model and the nugget-effect model perform poorly when dealing with heteroscedastic data. In this paper, we relax the stationarity assumption on the covariance process and propose the kriging model with modified nugget-effect to model heteroscedastic observations. This model follows the basic framework of the kriging model with nugget-effect, but extends it by taking the sample variance as an additional input to provide variance information. This method has two main benefits: first, the new model retains the original simple structure of the kriging model with nugget-effect, and second, the computing resources needed for computing the sample variance can be significantly lower than the requirement of the replication method. The sample variance is used as an additional input variable and it can reduce the impact of the heterogeneous variance on the local prediction by penalizing the data with higher variance. In the numerical experiments shown in this paper, the modified nugget-effect model's performance in the heteroscedastic case is consistently better than the OK model and nugget-effect model.

In this paper, the proposed new kriging model form is based on an extension of the nugget-effect model. Ankenman, Nelson and Staum (2010) recently proposed an alternative stochastic kriging model for stochastic simulations. Although the mathematical predictor forms of both models are equivalent (as will be seen in Section 2), our initial assumptions differ in that our proposed modified nugget-effect model is developed from the traditional nugget-effect model, extending it to treat the additional noise component ε as a non stationary component of the random process. The stochastic kriging model is developed based on the deterministic kriging model, and considers the additional noise component ε as the intrinsic uncertainty of the simulation itself. Ankenman et al. (2010) and Chen et al. (2010) go onto look at the effects of common random numbers on the model and describe experimental design strategies under the stochastic model. In this paper, our focus differs in that we study in detail the effects and influence of stochastic noise on the traditional deterministic ordinary kriging model and nugget-effect model, looking more deeply into the effects on parameter estimation and characteristics of the likelihood function. We also compare in detail the prediction performances of the three models, providing insights on when each model form is sufficient and adequate.

This article is organized as follows: In Section 2, we develop the proposed modified nugget-effect model. We then address the issues of parameter estimation and error measurement and further study the effects of stochastic noise on the traditional models as well as illustrate how the modified nugget-effect model mitigates this problem. In Section 3, we study the prediction performance and characteristics of the proposed model. Then in Section 4, the performance of the modified nugget-effect model is illustrated with several numerical experiments and a case study. Comparisons with the traditional kriging model and the nugget-effect model are given, and finally, comparisons with the studentization method are also made.

2. Kriging model with modified nugget-effect

In order to introduce the modified nugget-effect model, the details of the kriging model are first discussed. The differences between the modified nugget-effect model, the classic kriging model, and the nugget-effect model will be discussed in three aspects: the development of the modified nugget-effect model, parameter estimation, and error measurement of the model.

2.1. Classic kriging (deterministic and nugget-effect models)

In kriging metamodeling, the response of the simulation is treated as a random process $Z(\boldsymbol{x})$ where $\boldsymbol{x}$ stands for the simulation's

p-dimensional ($p \geqslant 1$) input. Typically, the mean response $S(\boldsymbol{x})$ of the random process is of interest. According to Cressie (1993), the general form of the random process can be decomposed as:

$$Z(\boldsymbol{x}) = S(\boldsymbol{x}) + \varepsilon(\boldsymbol{x}) = \mu(\boldsymbol{x}) + \delta(\boldsymbol{x}) + \varepsilon(\boldsymbol{x}) \tag{2}$$

where $S(\boldsymbol{x})$ is the deterministic signal function; $\mu(\boldsymbol{x})$ is the mean of the process, also known as the large-scale variation; $\delta(\boldsymbol{x})$ is the bias between the signal function and mean, also known as the small-scale variation; $\varepsilon(\boldsymbol{x})$ represents the random measure error (or random noise). In the application of the kriging model to deterministic simulations, the response takes the above form without the random noise component $\varepsilon(\boldsymbol{x})$. For stochastic simulations with homogenous variances throughout, the nugget-effect kriging model takes the form with $\varepsilon(\boldsymbol{x}) = \varepsilon$.

As with most applications of response metamodeling in stochastic simulations, when replicates at each observation point are observed, the sample means of the replicates are typically used as the input for the metamodel estimation. We denote the sample mean and sample variance as:

$$\overline{Z}(\boldsymbol{x}_i) = \sum_{j=1}^{n} \frac{Z_j(\boldsymbol{x}_i)}{n} \tag{3}$$

$$s^2(\boldsymbol{x}_i) = \sum_{j=1}^{n} \frac{(Z_j(\boldsymbol{x}_i) - \overline{Z}(\boldsymbol{x}_i))^2}{n-1} \tag{4}$$

where $Z_j(\boldsymbol{x}_i)$ denotes the jth replicate at location $\boldsymbol{x}_i$ and n is the constant number of replications.

Among several kinds of original deterministic kriging models, the one used in this article is the ordinary kriging (OK) model. The OK predictor for point $\boldsymbol{x}_0$, $P(Z(\boldsymbol{x}_0))$ is a linear combination of all m observation values:

$$P(Z(\boldsymbol{x}_0)) = \sum\nolimits_{i=1}^{m} \lambda_i \overline{Z}(\boldsymbol{x}_i) \quad \text{with} \quad \sum_{i=1}^{m} \lambda_i = 1 \tag{5}$$

where λ_i is the MSE optimal kriging weight:

$$\lambda_i = r^T R^{-1} e_i + 1^T R^{-1} \frac{[1 - 1^T R^{-1} r]^T}{1^T R^{-1} 1} e_i \tag{6}$$

$r = (corr(d_{01}), corr(d_{02}), \ldots, corr(d_{0m}))$ is the correlation between the point to be estimated and the m observed points, and d_{0i} is the Euclidean distance between point $\boldsymbol{x}_0$ and $\boldsymbol{x}_i$; R is the matrix of all the correlations between any two observation points;$e_i = [0, 0, \ldots, \underbrace{1}_{\text{the } i\text{th element}}, \ldots, 0, 0]$; and 1 is the vector of ones with the length of m. It is clear that the kriging weight λ_i is a function of the correlations $corr(\sim)$.

The stationary assumption of the kriging metamodel assumes that the correlation between any two points in the sample space depends only on the distance between the two points. As a result, the covariance function (or its corresponding correlation function) becomes the key component in the model. The general form of the covariance function is given below:

$$C(d_{ij}) = cov(\overline{Z}(\boldsymbol{x}_i), \quad \overline{Z}(\boldsymbol{x}_j)) = \begin{cases} c_0 + c_1 & d_{ij} = 0 \\ c_1 corr(d_{ij}) & d_{ij} \neq 0 \end{cases} \tag{7}$$

where c_0 is the nugget-effect value usually can be estimated from the sample variance as $\hat{c}_0 = s^2/n$; c_1 is called the partial sill which is the term borrowed from the geo-statistics. For applications in stochastic simulation, we treat the simulation output $Z(\boldsymbol{x})$ as a random variable which can be decomposed as the summation of the deterministic signal function $S(\boldsymbol{x})$ and random noise function $\varepsilon(\boldsymbol{x})$. c_0 and c_1 accordingly represent the variance of $S(\boldsymbol{x})$ and $\varepsilon(\boldsymbol{x})$, and hence also indicate the influences of the signal function and noise function on the simulation outputs; $corr(d_{ij})$ is the correlation function based on d_{ij}. So it is clear that the kriging weight λ_i in Eq. (6) is dependent only on the Euclidean distances between all the observation locations. Further discussion on the covariance function will be given in the following subsections.

For the ordinary kriging predictor, the weights are selected by minimizing the mean squared error defined as:

$$\text{MSE} = E[P(Z(\boldsymbol{x}_0)) - Z(\boldsymbol{x}_0)]^2 \tag{8}$$

The minimization result gives the optimal predictor

$$P(Z(\boldsymbol{x}_0)) = \vec{\lambda}\overline{Z} \tag{9}$$

where $\vec{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_m)$ is the vector of the kriging weights given in Eq. (6) and $\overline{Z} = (\overline{Z}(\boldsymbol{x}_1), \overline{Z}(\boldsymbol{x}_2), \ldots, \overline{Z}(\boldsymbol{x}_m))$ is the observation vector of sample means (Cressie, 1993, p. 123). The minimal mean squared prediction error (also known as the kriging variance) is then given by (see Cressie, 1993, p. 123)

$$\text{MSE}(\boldsymbol{x}_0) = c_1 \left(1 - \left[r + 1 \frac{(1 - 1^T R^{-1} r)}{1^T R^{-1} 1} \right]^T R^{-1} r + \frac{1 - 1^T R^{-1} r}{1^T R^{-1} 1} \right) \tag{10}$$

As the weights in the kriging predictor are dependent only on the Euclidean distances, this can be inadequate in many heteroscedastic cases where the randomness of the system is also dependent on the location. To solve this problem, the kriging model with modified nugget-effect proposes to relax the stationarity assumption and use the local variance information as an additional input variable. As a result, the predictor can penalize at locations where the variance is high.

2.2. Modified nugget-effect kriging model

From Section 2.1, we see that the kriging predictor is a function of the observations $Z(\boldsymbol{x}_i)$, $i = 1, 2, \ldots, m$, and the covariance function C. The general form of the covariance function was given in Eq. (7). Under different underlying assumptions, c_0 in Eq. (7) has different forms. We consider two underlying cases: the deterministic case and the stochastic case.

In the deterministic case, the same input at a given location gives the same output. The traditional deterministic kriging model can be used to model this case. In the deterministic kriging model, the random noise is assumed to be 0, so the nugget value c_0 is 0. The correlation matrix R is given as

$$R = \begin{bmatrix} 1 & corr(d_{12}) & \cdots & corr(d_{1(m-1)}) & corr(d_{1m}) \\ corr(d_{21}) & 1 & \cdots & corr(d_{2(m-1)}) & corr(d_{2m}) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ corr(d_{(m-1)1}) & corr(d_{(m-1)2}) & \ldots & 1 & corr(d_{(m-1)m}) \\ corr(d_{m1}) & corr(d_{m2}) & \ldots & corr(d_{m(m-1)}) & 1 \end{bmatrix} \tag{11}$$

and the kriging variance is given in Eq. (10).

The stochastic case can be further divided into two sub-cases: homoscedastic and heteroscedastic. The nugget-effect model is developed under the constant variance assumption to handle the homoscedastic case. As the random noise in this model is assumed to be a constant, the nugget-effect c_0 is a constant which equals the constant variance. The correlation matrix for this model is then given by

$$R' = \begin{bmatrix} 1+\frac{c_0}{c_1} & corr(d_{12}) & \cdots & corr(d_{1(m-1)}) & corr(d_{1m}) \\ corr(d_{21}) & 1+\frac{c_0}{c_1} & \cdots & corr(d_{2(m-1)}) & corr(d_{2m}) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ corr(d_{(m-1)1}) & corr(d_{(m-1)2}) & \ldots & 1+\frac{c_0}{c_1} & corr(d_{(m-1)m}) \\ corr(d_{m1}) & corr(d_{m2}) & \ldots & corr(d_{m(m-1)}) & 1+\frac{c_0}{c_1} \end{bmatrix} \tag{12}$$

where c_0/c_1represents the ratio of the variance of the input noise to the process variance. The kriging variance is (see Cressie, 1993, p. 123)

$$\mathrm{MSE}(\boldsymbol{x}_0) = c_0 + c_1\left(1 - \left[r + 1\frac{(1-1^T R'^{-1} r)}{1^T R'^{-1} 1}\right]^T R'^{-1} r + \frac{1-1^T R'^{-1} r}{1^T R'^{-1} 1}\right) \tag{13}$$

In the heteroscedastic case, the variances of the random noise are different at different locations. Alternative approaches like the replication method have been proposed to modify the heteroscedastic outputs directly to homoscedastic ones in order to apply the nugget-effect model. These methods, however, have limited applicability under tight computing budget constraints as extra replications are typically needed to drive down the random variability.

In this paper, we propose the modified nugget-effect model to address the heteroscedastic case. We relax the stationarity assumption in the homoscedastic model, and assume that the random noise is independent (no CRN used) but not identical. The only difference between Eqs. (7) and (14) is the constant nugget-effect c_0 is relaxed to become a variable $c(\boldsymbol{x}_i)^*$, which is dependent on location. As a result, the covariance function is given as:

$$C(d_{ij}) = cov(Z(\boldsymbol{x}_i), Z(\boldsymbol{x}_j)) = \begin{cases} c(\boldsymbol{x}_i)^* + c_1 & d_{ij} = 0 \\ c_1 corr(d_{ij}) & d_{ij} \neq 0 \end{cases} \tag{14}$$

where $c(\boldsymbol{x}_i)^*$ represents the variance of the input random error at location $\boldsymbol{x}_i$, and can be estimated from the sample variance and number of replications at location $\boldsymbol{x}_i$ as $\hat{c}(\boldsymbol{x}_i)^* = s^2(\boldsymbol{x}_i)/n_i$. Accordingly, the correlation matrix becomes

$$R' = \begin{bmatrix} 1+\frac{c_1^*}{c_1} & corr(d_{12}) & \cdots & corr(d_{1(m-1)}) & corr(d_{1m}) \\ corr(d_{21}) & 1+\frac{c_2^*}{c_1} & \cdots & corr(d_{2(m-1)}) & corr(d_{2m}) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ corr(d_{(m-1)1}) & corr(d_{(m-1)2}) & \cdots & 1+\frac{c_{m-1}^*}{c_1} & corr(d_{(m-1)m}) \\ corr(d_{m1}) & corr(d_{m2}) & \cdots & corr(d_{m(m-1)}) & 1+\frac{c_m^*}{c_1} \end{bmatrix} \tag{15}$$

As shown in Appendix A, the kriging predictor for this heteroscedastic model is

$$P(Z(\boldsymbol{x}_0)) = \sum_{i=1}^{m} \lambda_i' \overline{Z}(\boldsymbol{x}_i) \tag{16}$$

where λ_i' equals to the λ_i defined in Eq. (6) with the correlation matrix R replaced by the R' defined in Eq. (15).

The kriging variance can be obtained by substituting $c(\boldsymbol{x}_0)^*$ for the constant term c_0 in Eq. (13) where $c(\boldsymbol{x}_0)^*$ denotes the variance of the input random error at location $\boldsymbol{x}_0$.

As discussed, the kriging model with nugget-effect is used to handle the homoscedastic case by adding a constant onto the diagonal of the correlation matrix of the deterministic kriging model. The model proposed for the heteroscedastic case has a form similar to the kriging model with nugget-effect, with a variable term added onto the diagonal instead. We can rewrite the correlation matrix as the summation of the correlation matrix R for the deterministic kriging model and an additional term η. For the homoscedastic case, η is a matrix with a constant $\eta_c = c_0/c_1$ on the diagonal:

$$R' = R + \eta \quad \text{with} \quad \eta = \begin{bmatrix} \eta_c & 0 & \cdots & 0 & 0 \\ 0 & \eta_c & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \eta_c & 0 \\ 0 & 0 & \cdots & 0 & \eta_c \end{bmatrix} \tag{17}$$

For the heteroscedastic case, η is a matrix with a variable $\eta_i = c(\boldsymbol{x}_i)^*/c_1$ on the diagonal:

$$R' = R + \eta \quad \text{with} \quad \eta = \begin{bmatrix} \eta_1 & 0 & \cdots & 0 & 0 \\ 0 & \eta_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \eta_{m-1} & 0 \\ 0 & 0 & \cdots & 0 & \eta_m \end{bmatrix} \tag{18}$$

Hence, we call this model the kriging model with modified nugget-effect. Note that substituting R' defined in Eq. (15) as R in the deterministic predictor in Eq. (9) gives the same mathematical predictor form as Ankenman et al. (2010). In the next section, we will discuss in detail the nugget-effect's influence on parameter estimation.

2.3. Parameter estimation and characteristics of likelihood function with noisy data

The correlation function $corr(\sim)$ in R'can have different forms. The Gaussian correlation function has been widely applied. It has the following form:

$$corr(d_{ij}) = \prod_{k=1}^{p} exp(-\theta_k d_{ij}^2), \qquad \theta_k > 0 \tag{19}$$

where θ_k is the sensitivity parameter for the kth input, which is usually estimated by the maximum likelihood (ML) approach. The DACE toolbox is a widely used free Matlab kriging toolbox that is well documented in Lophaven, Nielsen, and Songdergaard (2002).

For stochastic responses, especially in the heteroscedastic case, the likelihood function is likely to have an erratic behavior caused by the high variability of the data. In this case, the estimation of $\theta = (\theta_1, \theta_1, \ldots, \theta_p\})$ may have a large variance because the likelihood function can be flat near the optimum. Because θ is the sensitivity parameter of the correlation function, inaccurate estimation may cause oscillations and fluctuations in the kriging predictor's output. The test function in Eq. (1) will be used as an example to demonstrate this phenomenon.

First, we will show how the likelihood function for the deterministic model will change with noisy observations. According to Cressie (1993, p. 92), the likelihood function for the deterministic kriging model is:

$$\ell(\theta) = \frac{1}{2}\ln\det(R) + \frac{1}{2}(\overline{Z} - 1\mu)^T R^{-1}(\overline{Z} - 1\mu) \tag{20}$$

where R is the correlation matrix, which is a function of θ. The correlation function used here is the Gaussian correlation function, given in Eq. (19). Obviously, if any $\theta_k \to \infty$ (for $k = 1, 2, \ldots, p$), $R \to I$ (identity matrix), and so, $\lim_{\theta_k \to \infty} \ell(\theta) \to \frac{m}{2}\mathrm{Var}(\overline{Z})$. For the one dimensional test function of the form in Eq. (1) where the variability in the observations is high, the likelihood function will increase in the region where $\theta \to \infty$. This phenomenon will cause a bad estimate of θ. The plot of the likelihood function for both the signal function and the noisy observations on the signal function is shown in Fig. 3

In Fig. 3, considering the likelihood function for the signal (solid line), the maximum likelihood estimator (MLE) θ^* is very close to 0. θ^* can be viewed as the best θ possible without the influence of noise. From Eq. (19), we see that θ is the sensitivity parameter of the correlation function. Intuitively, a small θ implies that the correlation is not very sensitive to distance, providing a smooth predictor. However, with noisy observations (see the dotted line in Fig. 3), much of the likelihood function has been lifted up in the right side of the plot when θ gets large. This causes an ML estimated θ to be much larger than θ^*, making the correlation very sensitive to distance. As a result, the output of the kriging predictor will oscillate. Moreover, as noise gets larger, the variability of the MLE increases.

When assuming the kriging model with nugget-effect or modified nugget-effect, this situation is improved. As seen in Fig. 4, the lifted up portion is pulled back and the likelihood function is corrected with a much smaller ML estimate of θ.

We can see this from the likelihood function of the nugget-effect and modified nugget-effect model:

$$\ell'(\theta) = \frac{1}{2}\ln\det(R') + \frac{1}{2}(\bar{Z} - 1\mu)^T R'^{-1}(\bar{Z} - 1\mu) \tag{21}$$

where $R' = R + \eta$ and η is given in Eqs. (17) and (18) for the nugget-effect and modified nugget-effect models. Because R is positive definite, by the Woodbury identity (Woodbury, 1950), we have

$$R'^{-1} = R^{-1} - R^{-1}(R^{-1} + \eta^{-1})^{-1}R^{-1}$$

So, the likelihood function becomes

$$\ell'(\theta) = \frac{1}{2}\ln\det(R') + \frac{1}{2}(\bar{Z} - 1\mu)^T R^{-1}(\bar{Z} - 1\mu) - \frac{1}{2}(\bar{Z} - 1\mu)^T R^{-1}(R^{-1} + \eta^{-1})^{-1}R^{-1}(\bar{Z} - 1\mu)$$

The last term on the right hand side

$$p_\eta(\theta) = -\frac{1}{2}(\bar{Z} - 1\mu)^T R^{-1}(R^{-1} + \eta^{-1})^{-1}R^{-1}(\bar{Z} - 1\mu) \tag{22}$$

behaves like a penalty function when θ gets large. The plot of $p_\eta(\theta)$ when η defined by (18) is given in Fig. 5. We see that $p_\eta(\theta)$ behaves much like the penalty function approach proposed in Li and Sudjianto (2005) to correct highly variable likelihood estimates due to the lack of data in computationally intensive deterministic simulation models. From the deterministic perspective, $p_\eta(\theta)$ serves as a natural penalty function to the log likelihood function to reduce the variability of the estimated θ when "noisy" observations are obtained, hence reducing the erratic behavior of the predictor.

When applying the kriging model with nugget-effect in a heteroscedastic situation, there is no straightforward interpretation for selecting an appropriate nugget value. Kleijnen and Van Beers (2005) noted that variogram results are meaningless for heteroscedastic data. An *ad hoc* approach is to pool sample variances as done in the test function example. Here, we provide a more intuitive argument for selecting the nugget-effect value for this model. For cases like the test function in Eq. (1) with a step variance function, the variance of the random noise is considerably large in some regions and cannot be ignored, increasing the variance of all the observations, $\mathrm{Var}(\bar{Z})$. For the likelihood function of the nugget-effect model, when any $\theta_k \to \infty$ (for $k = 1, 2, \ldots, p$), $R' \to I$, the function will become:

$$\lim_{\theta_k \to \infty} \ell'(\theta) \to \frac{m}{2}\ln(1 + \eta_0) + \frac{m}{2(1 + \eta_0)\mathrm{Var}(\bar{Z})} \tag{23}$$

In order to fix the abnormal likelihood function caused by the high variability in the observations, the influence of $\mathrm{Var}(\bar{Z})$ can be reduced by increasing η_0. To do so, the largest of the heterogeneous variances should be used as the nugget-effect value for kriging metamodel with nugget-effect. We adopt this value in the examples given in Section 4.

The likelihood function for the nugget-effect model in Fig. 4 has a similar profile as the plot of the likelihood function with only the signal function in Fig. 3 (the solid line). As a result, the estimation of θ by maximizing the likelihood function with nugget-effect will be closer to θ^* than the estimation given by maximizing the

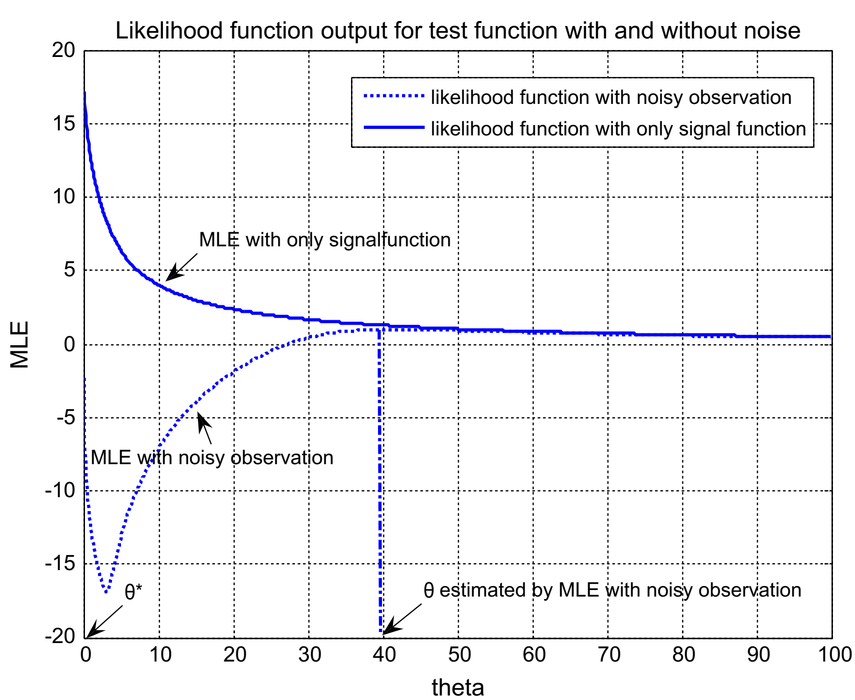

Fig. 3. Likelihood function for θ (signal function only and noisy observation in Eq. (1)).

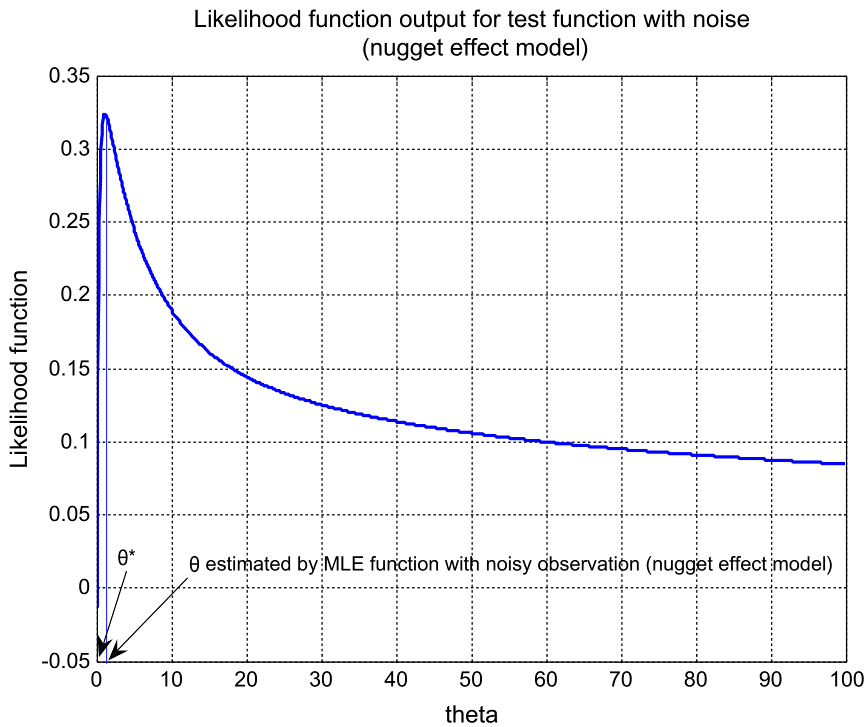

Fig. 4. Likelihood function for θ with nugget effect model (noisy observation of the signal function).

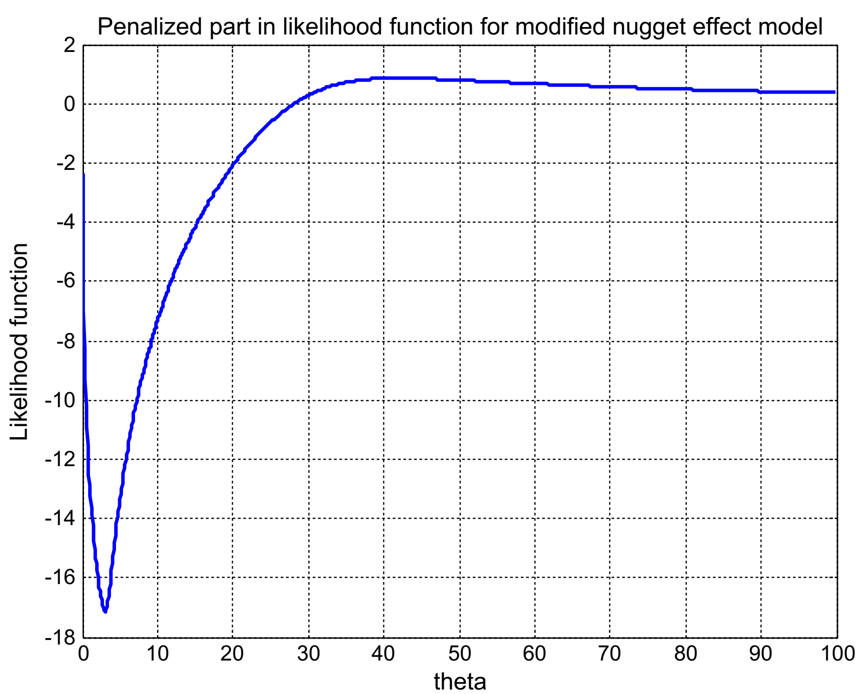

Fig. 5. Profile of the penalized portion of the likelihood function for modified nugget effect model.

likelihood function without the nugget-effect. The likelihood function for the modified nugget-effect model has similar behavior and characteristics.

Thus, in the heteroscedastic case, both the kriging model with nugget-effect and the kriging model with modified nugget-effect can provide a better estimator of θ than the traditional deterministic kriging model. Therefore, the performance of either kriging model with nugget-effect or kriging model with modified nugget-effect will be much better than the original kriging model, especially when the variability of the data is high, and the number of available observations is small.

2.4. Error measurement

In order to compare the performance of the nugget-effect model and the modified nugget-effect model in the heteroscedastic case, an error measurement standard is needed. We use the Mean Squared Error (MSE),

$$\mathrm{MSE}(\boldsymbol{x}_0) = \sigma^2_{kg}(\boldsymbol{x}_0) = E[P(Z(\boldsymbol{x}_0)) - Z(\boldsymbol{x}_0)]^2 \tag{24}$$

The squared error has several benefits: it simplifies the Bayes risk, it has symmetric confidence intervals, and it is related to the variance terms, which is especially suitable for kriging. As can be seen, the MSE is used to measure the difference between the predictor and the observation. In the deterministic case, because the observation strictly equals the signal function, MSE can offer a clear view of the prediction accuracy with respect to the signal function. For the stochastic case however, when predicting at a new location $\boldsymbol{x}_0$, our interest is in the actual signal function, $S(\boldsymbol{x}_0)$, and not the observation $Z(\boldsymbol{x}_0)$ which is distorted by random noise. As a result, the kriging predictor will be:

$$P(S(\boldsymbol{x}_0)) = \sum_{i=1}^{m} \xi_i \overline{Z}(\boldsymbol{x}_i) \tag{25}$$

with the kriging weight ξ_i, which is generated by minimizing the mean squared error with respect to signal function $S(\boldsymbol{x})$ (also see Eq. (5)). The predictor is still based on all the observations, $\overline{Z}(\boldsymbol{x}_i),\ i = 1, 2, \ldots, m$.

The notation MSE_S is used for the mean squared error computed with respect to the signal function $S(\boldsymbol{x})$, which according to Cressie (1993, p. 128) is

$$\begin{aligned}\mathrm{MSE}_s(\boldsymbol{x}_0) &= \tau^2_{kg}(\boldsymbol{x}_0) = E[P(S(\boldsymbol{x}_0)) - S(\boldsymbol{x}_0)]^2 \\ &= c_1\left(1 - \left(r + 1\frac{1 - 1^T R'^{-1} r}{1^T R'^{-1} 1}\right)^T R'^{-1} r + \frac{1 - 1^T R'^{-1} r}{1^T R'^{-1} 1}\right) - \sigma^2_{\varepsilon(\boldsymbol{x}_0)}\end{aligned} \tag{26}$$

where for the nugget-effect model, $\sigma^2_{\varepsilon(\boldsymbol{x}_0)} = c_0$; for the modified nugget-effect model, $\sigma^2_{\varepsilon(\boldsymbol{x}_0)} = c(\boldsymbol{x}_0)^*$.

Clearly, the MSE_S can offer the evaluation of the predictor's performance with respect to the signal function in the presence of random noise. In the heteroscedastic case, however, one limitation is that the variance of the random error at every prediction point is needed. As it is assumed that the random error is independent but not identical, the variance information is not available unless the location is observed. Hence, comparing the performances of the kriging metamodel with nugget-effect and the kriging metamodel with modified nugget-effect is not straightforward.

As an alternative, we decompose the sample space into two parts: observation points and the areas in between the observations. The MSE_S at all the observation points can be compared, which will be given in the following section. For the areas in between the observations, the prediction is dependent on the prediction at the observations and the sensitivity parameter θ. For comparison purposes, we fix the sensitivity parameter θ in both models to be the MLE estimated with nugget-effect.

3. Prediction performance of the kriging model with modified nugget-effect

The prediction performance of a kriging model with modified nugget-effect can be divided into two parts: predictor's output and variance of the predictor.

3.1. Comparison through MSE_S

From Eq. (26), it can be shown that the MSE_S at the ith observation point for the kriging model with modified nugget-effect is

$$\mathrm{MSE}_s(\boldsymbol{x}_i) = c_1\left(\eta_i^2 \frac{\Delta^2_{mi}}{\Delta_m} - \eta_i^2 \Delta_{mi}\right) \tag{27}$$

where $\Delta_m = 1^T R'^{-1} 1$ indicate the summation of all the elements in the inverse correlation matrix, Δ_{mi} represent the summation of the ith column or row, and $\eta_i = c_i^*/c_1$. The details of this derivation are provided in Appendix B.

Similarly, for the kriging metamodel with nugget-effect, the MSE_S at the ith observation point is

$$\mathrm{MSE}_s(\boldsymbol{x}_i) = c_1\left(\eta_c^2 \frac{\Delta^2_{ni}}{\Delta_n} - \eta_c^2 \Delta_{ni}\right) \tag{28}$$

Because the difference between Δ_m and Δ_n is typically not significant, assuming $\Delta_m \approx \Delta_n = \Delta$ and $\Delta_{mi} \approx \Delta_{ni} = \Delta_i$, when the nugget value c_0 exceeds $l_0 = \sum_{i=1}^m \left[c_i^{*2}\left(\frac{\Delta_i^2}{\Delta} - \Delta_i\right)\right] \Big/ \sum_{i=1}^m \left(\frac{\Delta_i^2}{\Delta} - \Delta_i\right)$, the modified nugget-effect model outperforms the nugget-effect model in terms of MSE. For most situations, in order to have a better estimation of the θ, a larger nugget-effect is desired, as seen in Eq. (23). Hence, it is often likely that the selected c_0 will be larger than the l_0.

3.2. Estimating predictor's variance

The covariance function of the proposed modified nugget-effect model is given in Eq. (14). Based on the model assumptions, the random noise in the modified nugget-effect model is independent but not identical. Hence the closed form formulation of the predictor's variance at un-sampled points is not available. Numerical approaches can be used to approximate of the variance of the predictor. Here we propose two different methods to estimate the variance of the predictor.

The first method is the nonparametric bootstrap approach. Bootstrapping is a re-sampling method developed by Efron (1979), and has been widely used in estimating the distribution and the properties of estimators. Nonparametric bootstrapping draws samples from the original observed data with replacement to form the bootstrap samples. The bootstrapping statistic is then estimated from the bootstrap samples. Parametric bootstrapping can also be applied as an alternative approach to estimate the variance of the kriging predictor. In the parametric approach, based on the model assumptions in Eq. (2), bootstrapped samples can be drawn from the normal distribution with plug-in parameter estimators of $\hat{\beta},\ \hat{\sigma}^2_z,\ \hat{\theta}$; see Den Hertog, Kleijnen, and Siem (2006) Kleijnen, Van Beers, and Van Nieuwenhuyse (2010). Both these bootstrapping approaches can provide estimators of the predictor variance. However, to avoid the assumption of the plug-in estimator and its additional variability due to the random noise (as discussed in Section 2.3) in the parametric approach, here we apply the nonparametric approach to estimate the variance of the predictor.

Applying the nonparametric bootstrap approach, the bootstrap estimator of the predictor's variance can be given as:

$$\begin{aligned}\widehat{var}(P(Z(\boldsymbol{x}_0))) &= \widehat{var}^B(P(Z(\boldsymbol{x}_0))) \\ &= \sum_{j=1}^{b} \frac{\left(P_j^B(Z(\boldsymbol{x}_0)) - \overline{P}^B(Z(\boldsymbol{x}_0))\right)^2}{b-1}\end{aligned} \tag{29}$$

where $P_j^B((Z(\boldsymbol{x}_0))$ denotes the bootstrapped predictor's output at location $\boldsymbol{x}_0$ with the jth set of bootstrap samples, $\overline{P}^B(Z(\boldsymbol{x}_0))$ is the average of all the bootstrapped predictor's outputs at location $\boldsymbol{x}_0$ given b sets of bootstrap samples, where b is the total number of the bootstrap sample sets. Given the simulation observations $Z(\boldsymbol{x}_i) = \{Z_1(\boldsymbol{x}_i), Z_2(\boldsymbol{x}_i), \ldots, Z_n(\boldsymbol{x}_i)\}_{i=1,2,\ldots,m}$, the general bootstrap procedure to estimate the predictor's variance at a new point is given as follows:

Step 1: For each observation location $\boldsymbol{x}_i$, re-sample n observations (with replacement) from the n original simulation observations obtained at that location. Repeat this b times to obtain b bootstrap samples of n re-sampled observations at each location $\boldsymbol{x}_i$, $\vec{Z}^B(\boldsymbol{x}_i)_j = \left\{Z_1^B(\boldsymbol{x}_i)_j, Z_2^B(\boldsymbol{x}_i)_j, \ldots, Z_n^B(\boldsymbol{x}_i)_j\right\}_{i=1,2,\ldots,m;\, j=1,2,\ldots,b}$;

Step 2: For each bootstrap sample j, $j = 1, \ldots, b$, compute the sample mean $\overline{Z}^B(\boldsymbol{x}_i)_j = \frac{\sum_{k=1}^{n} Z_k^B(\boldsymbol{x}_i)_j}{n}$ and sample variance $\widehat{var}(\overline{Z}^B(\boldsymbol{x}_i)_j) = \frac{\sum_{k=1}^{B}\left(Z_k^B(\boldsymbol{x}_i)_j - \overline{Z}^B(\boldsymbol{x}_i)_j\right)^2}{n-1}$ at each location $\boldsymbol{x}_i$. Using these as inputs to the kriging model, we can obtain the bootstrap predictor at location $\boldsymbol{x}_0$ for the jth bootstrap sample, $P_j^B(Z(\boldsymbol{x}_0))$.

Step 3: With b bootstrap predictors at location $\boldsymbol{x}_0$, $P_j^B(Z(\boldsymbol{x}_0))_{j=1,2,\ldots,b}$, the non-parametric bootstrapping predictor variance at $\boldsymbol{x}_0$ can be computed with Eq. (28).

For an alternative method, Kleijnen & Van Beers (2005) recommended that the variance of the predictor can be estimated by interpolation. Because the variances at all the observation points are available, piecewise linear interpolation can be used to cover the space between any two observation points.

In the following examples, we compare the bootstrapping variance and interpolating variance with the empirical brute-force predictor variance generated through 10,000 replications of the independent simulation results.

From Fig. 6, we can see that the bootstrapping estimated variance is close to the empirical variance of the predictor, as well as the interpolating estimated variance. Fig. 7 plots the 95% prediction intervals obtained with the estimated variances.

As seen in Fig. 7, the signal function $S(x) = \frac{x}{1-x}$ lies within the prediction interval and the differences between these two figures are not very big. The prediction interval is especially large when the input x is close to 0.9. For the M/M/1 model, the variance of the output goes to infinity when the input traffic rate x goes to 1. This phenomenon is also known as the "variance explosion". This sudden increase of input variance results in the unusually large prediction interval when x is close to 0.9.

The interpolation method is much easier and faster than the bootstrapping method, and is preferred for its simplicity. However its performance is comparable only when there are sufficient observation points. If the observations are limited and the variance function is complicated, then the precision of result with the interpolation method may be low.

4. Examples

In this section, two numerical examples and a short case study will be presented. All the examples used here have heteroscedastic variances: a test function with step variance function, the M/M/1 queueing system and a queueing network problem. The quadratic test function provides insight into the heteroscedastic case, focusing on the comparison between a low and high step variance. The M/M/1 queue is a classic queueing system where analytical results are available for comparison. It has a continuous variance function and is a basic component of many more complex queueing systems. The final case study is of a comprehensive queueing network system where the underlying functional forms are unknown.

4.1. Test function

The test function here is given in Eq. (1) with a step variance function. In order to illustrate the influence of the heterogeneous variance, we set 3 different ratio levels for the step variance function: $\sigma_\varepsilon^2 = 0.083$ when $x \in [-5, 2)$, and $\sigma_\varepsilon^2 = 0.83,\ 8.3,\ 83$ when $x \in [2, 5]$. Figs. 8–10 illustrate the predictors' outputs (ordinary kriging, kriging with nugget-effect and kriging with modified nugget-effect) for r_{var} (ratio of max variance to min variance) = 10, 100, 1000 respectively.

From these three figures, we can see the erratic behavior in the output of the ordinary kriging predictor. This is the result of the

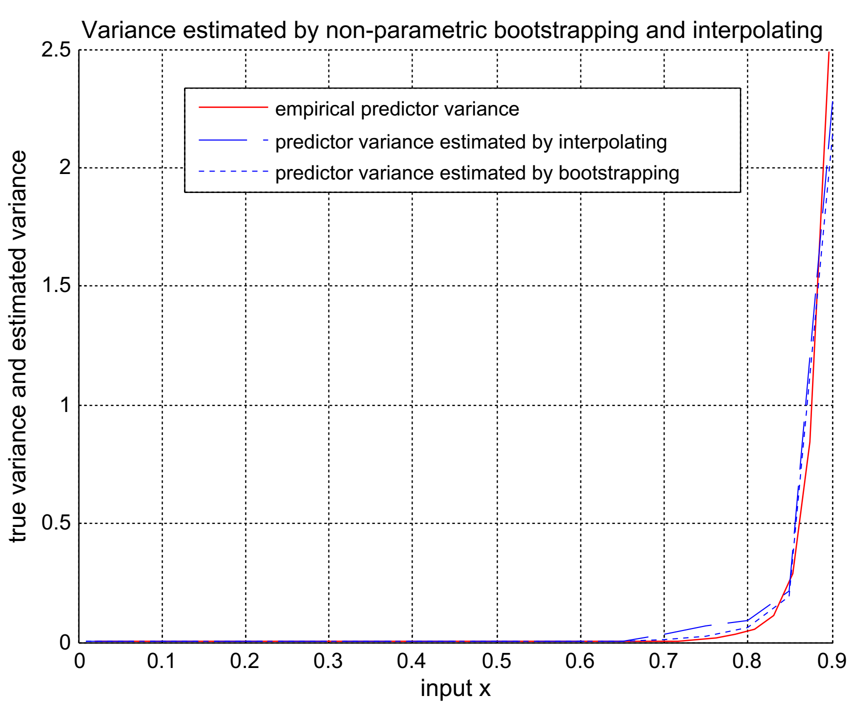

Fig. 6. Estimated variance for modified nugget effect predictor (M/M/1).

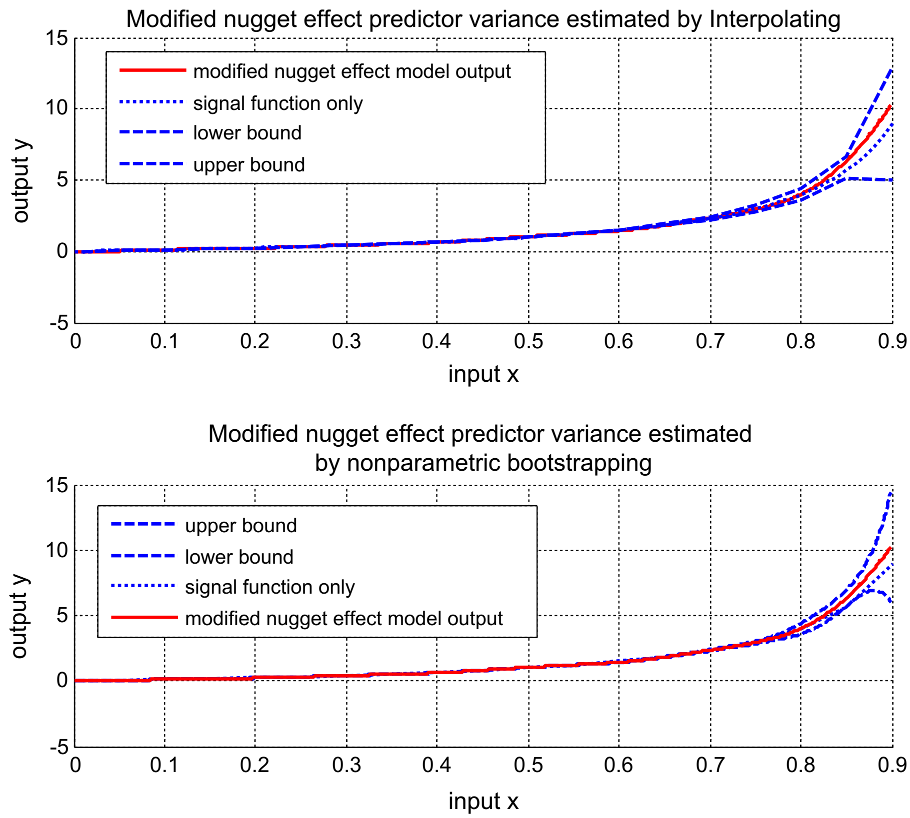

Fig. 7. Prediction interval for modified nugget effect model in M/M/1 example (bootstrapping).

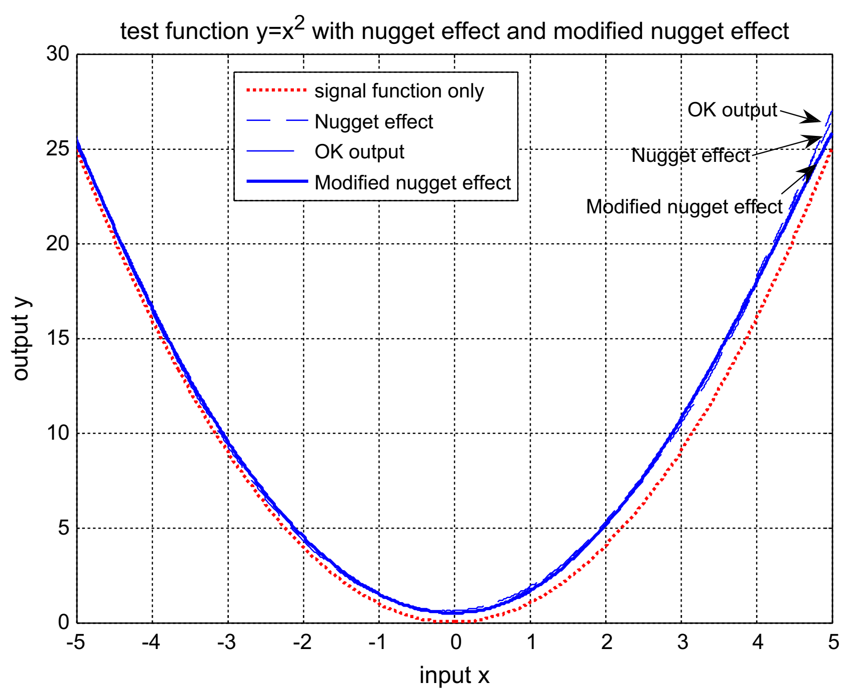

Fig. 8. Different predictors' output for test function ($r_{var} = 10$).

high variance in parameter estimation. The output of the kriging predictor with nugget-effect is much smoother, but it is still far from the signal function in the region with higher variance, especially for the case when $r_{\text{var}} = 1000$. The modified nugget-effect predictor lies closest to the signal function. Selecting a second-order polynomial function ("universal kriging") instead of a constant as the process mean μ in Eq. (2) can improve the prediction accuracy. Fig. 11 illustrates the predictor output when a second-order model is assumed in place of the constant model for the three kriging model forms. As seen in this figure, the

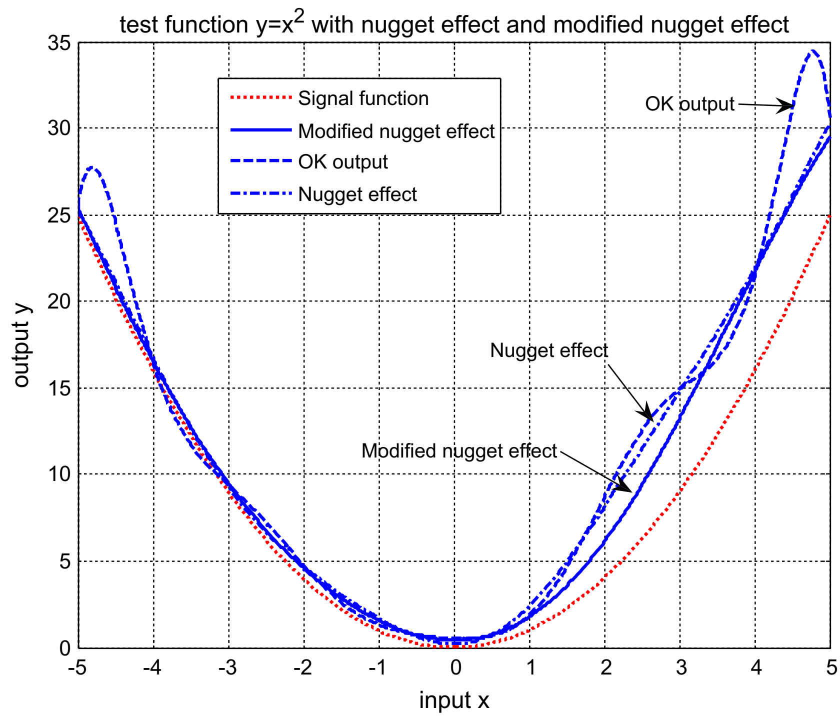

Fig. 9. Different predictors' output for test function (r_{var} = 100).

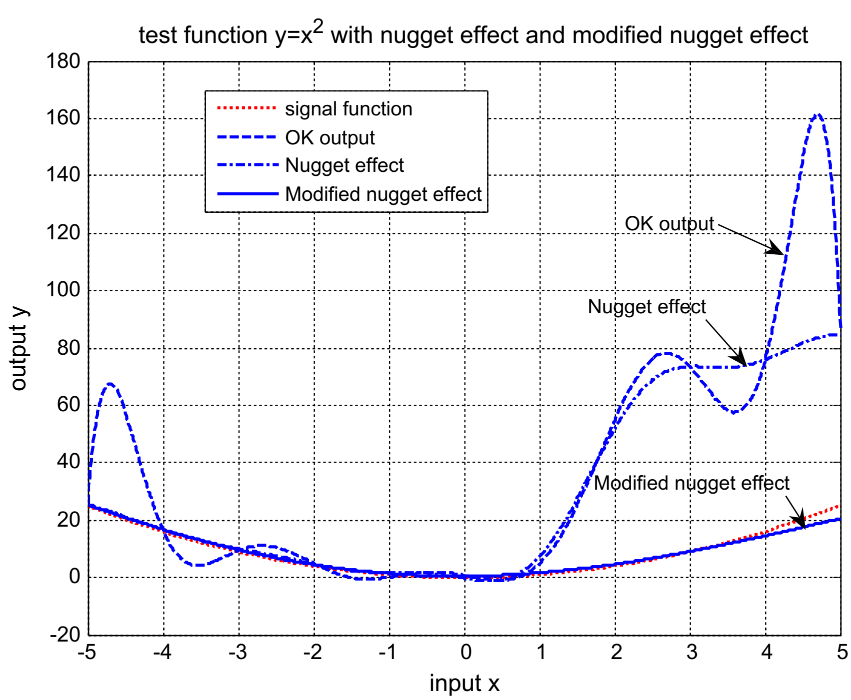

Fig. 10. Different predictors' output for test function (r_{var} = 1000).

modified nugget-effect model with a second-order polynomial regression model provides a smoother output over the entire region.

Table 1 summarizes four error measures MSE_{SO}, MSE_S, AAE_S and MAE_S for the different models and variance ratios used. All the results are averaged based on 1000 macro-replications. For the nugget-effect model in Table 1, the min variance refers to the nugget value c_0 equals to the minimum value for the step variance function, and the max variance refers to the nugget value c_0 equals to the maximum value for the step variance function. The modified nugget-effect (optimal) refers to the modified nugget effect model with the optimal sensitivity parameter θ^*. The first measure MSE_{SO} refers to the mean squared error between the predictor's output and signal function at all the observed points:

$$MSE_{SO} = \frac{1}{m}\sum_{i=1}^{m}(P(S(\boldsymbol{x}_i)) - S(\boldsymbol{x}_i))^2 \quad (30)$$

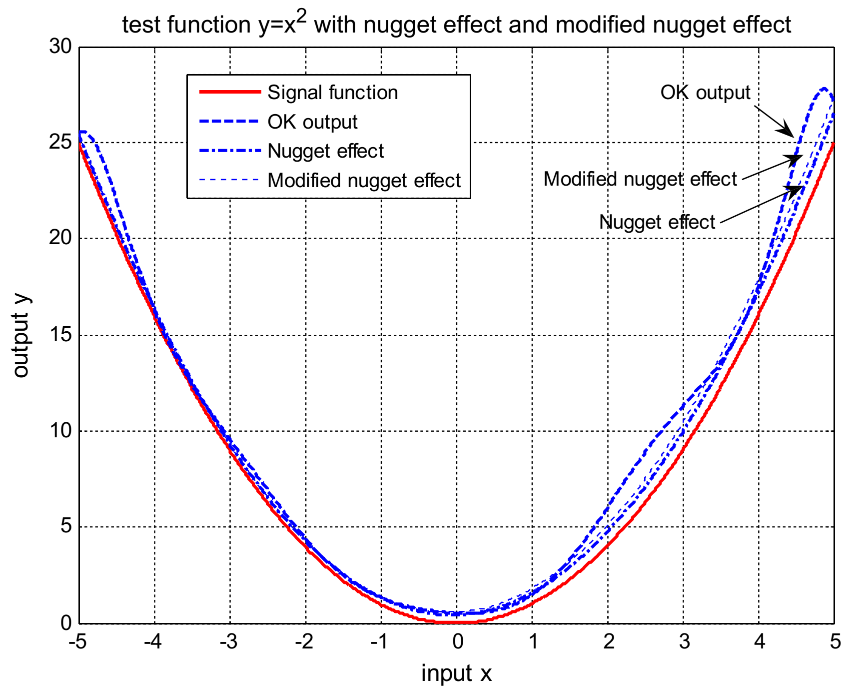

Fig. 11. Different predictors' output for test function (r_{var} = 100, 2^{nd}-order polynomial regression model).

Table 1
Different error measures of different metamodels for the test function example.

Ratio (r_{var})	10				100				100(2^{nd} poly)				1000			
Metamodels	MSE_{SO}	MSE_S	AAE_S	MAE_S	MSE_{SO}	MSE_S	AAE_S	MAE_S	MSE_{SO}	MSE_S	AAE_S	MAE_S	MSE_{SO}	MSE_S	AAE_S	MAE_S
OK	0.990	2.572	0.780	3.315	8.474	17.16	2.234	14.11	8.462	5.492	1.954	4.963	83.04	241.9	10.10	29.49
Nugget-effect (min variance)	0.967	1.588	0.762	3.212	7.836	11.11	1.836	9.743	7.402	0.513	0.738	2.854	66.13	108.3	9.469	14.40
Nugget-effect (max variance)	0.927	1.426	0.664	2.758	7.222	9.490	1.473	8.165	6.926	0.506	0.539	2.201	62.47	91.16	8.653	12.85
Modified nugget-effect	0.646	0.971	0.513	1.992	4.030	6.957	1.181	5.489	3.698	0.133	0.351	1.201	16.44	21.78	7.394	9.565
Modified nugget-effect (optimal)	0.618	0.835	0.444	1.815	3.340	5.725	0.930	5.153	2.902	0.092	0.198	1.065	12.39	17.15	5.535	7.921

The second MSE_S refers to the differences at all the points (including the m = 19 observed points and k = 982 unobserved test points):

$$MSE_S = \frac{1}{m+k}\sum_{i=1}^{m+k}(P(S(\boldsymbol{x}_i)) - S(\boldsymbol{x}_i))^2 \tag{31}$$

The Average Absolute Error (AAE_S) is

$$AAE_S = \frac{1}{m+k}\sum_{i=1}^{m+k}|P(S(\boldsymbol{x}_i)) - S(\boldsymbol{x}_i)| \tag{32}$$

and also assesses the overall performance, and the Maximum Absolute Error (MAE_S)

$$MAE_S = \max|P(S(\boldsymbol{x}_i)) - S(\boldsymbol{x}_i)|_{i=1,2,\ldots m+k} \tag{33}$$

reflects the presence of the poor prediction in local areas.

From this table, we note that the modified nugget-effect model is better than the nugget-effect model and the traditional ordinary kriging model. For all the models listed in Table 1, the error increases as the r_{var} increases; the second-order polynomial regression can greatly reduce the error by providing a better estimator for the process mean.

Given the prediction results over 1000 macro-replications, paired t-test results suggest that the differences between the ordinary kriging model and the modified nugget-effect model, the differences between the nugget-effect model (max variance) and the modified nugget-effect model for all the measurements are statistically significant at alpha level of 0.05. The difference between the ordinary kriging model and the nugget-effect model (max variance) is statistically significant only for the high variance scenario (r_{var} = 100, 1000).

As mentioned in Section 3.1, for most cases, we tend to select a higher nugget-effect value in order to have a better estimator of θ. For the nugget-effect model, c_0 is set at the largest variance observed. To understand this selection, we study the impact of c_0 on the prediction of the nugget-effect model. Fig. 12 plots the influence of the nugget-effect value c_0 on the estimation error.

As can be seen, the larger nugget-effect value, the lower the MSE. This empirically verifies our observation in Section 3.1.

4.2. M/M/1 queueing system

The M/M/1 queue system is a typical stochastic system which is widely studied in the literature. Because the expected waiting time

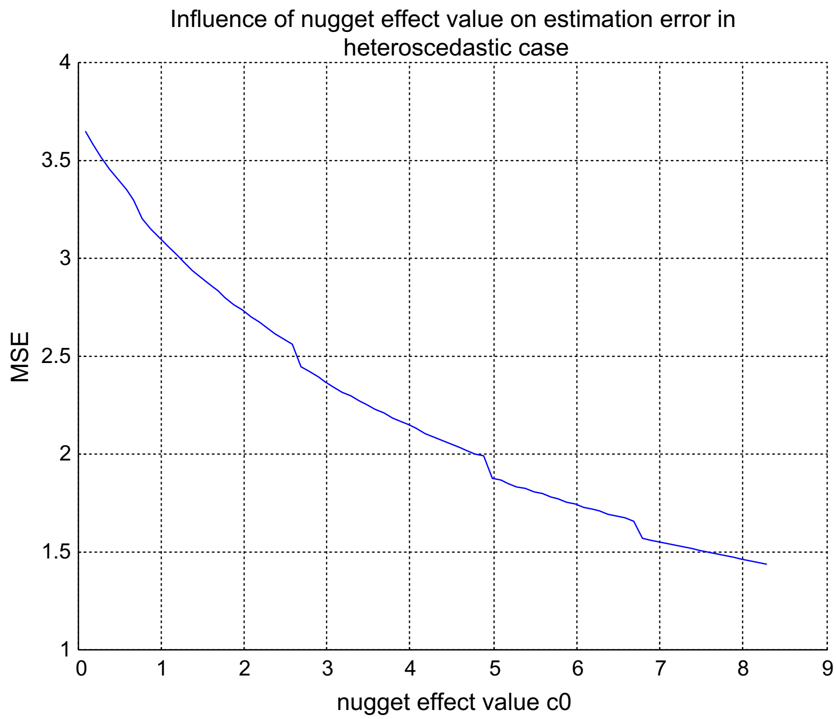

Fig. 12. Influence of nugget value on MSE (test function).

for the M/M/1 system has a closed form (see Hillier & Lieberman, 2001), it is easy to compare the model's performance with the true performance.

The data for this experiment is based on the M/M/1 queueing simulation system, which displays the heterogeneous variance characteristics. For the simulation system design, the input is the traffic load of the queueing system, and output is the expected waiting time over 10,000 customer arrivals with a warm up period of 2000 customer arrivals. The input–output combinations of the M/M/1 system are used to build the model, with input points located at $x = 0.01, 0.05, 0.10, \ldots, 0.80, 0.85, 0.90$, totaling 19 points. To show the metamodel's performance, 1000 macro-replications are taken. For each replication, one kriging metamodel is built based on the $m = 19$ observed input–output combinations. The output of the corresponding kriging predictor is generated at $k = 40$ evenly distributed points, $x = 0.02, 0.0425, 0.065, \ldots, 0.875, 0.8975$ (no extrapolations). Both the nugget-effect model and the modified nugget-effect model are tested and the comparisons of these two models are given in Table 2.

Paired *t*-test results suggest that the differences between the nugget-effect model and the modified nugget-effect model are statistically significant at the alpha level of 0.05. This indicates that the modified nugget-effect model outperforms the nugget-effect model in all three measures for this M/M/1 queue example.

The detrended studentization method proposed in Kleijnen and Van Beers (2005) takes replications at each location to obtain an average value, and standardizes it as the input. In order to make a comparison with the studentization method, we use the expected average waiting time instead of the average waiting time as the output of interest. We repeat this for 100 sub-groups, where in each sub-group, we obtain a sample size of 10 replications at each of the input points and calculate the sample means and the sample variances. For the studentization method, the sample mean and the sample variance are used to standardize the input. For the modified nugget-effect, both the sample mean and the sample variance are used as the inputs. Figs. 13 and 14 illustrate the studentization method and the modified nugget-effect model's performance on 100 sub-groups with the sample size of 10 per sub-group:

From Figs. 13 and 14, we see that the modified nugget-effect predictor's outputs are closer to the signal function than the studentization method. The performance measures MSE_S, AAE_S and MAE_S for the two methods are summarized in Table 3 for a variety of sample sizes.

From Table 3, paired *t*-test results suggest that the differences between the studentization method and the modified nugget-effect model are statistically significant at the alpha level of 0.05 for sample size equals to 10 and 100. The modified nugget-effect model performs better in terms of the three measures. This is because the variability in the predictor is higher in the studentization method from the standardization and transformation of the predictor by its estimated signal function and variance. We note that the differences between the two methods decrease as the sample size for each subgroup increases. The improvement of the studentization method as the sample size increases and estimators improve is also noted in Kleijnen and Van Beers (2005). Hence, the modified nugget-effect model is a better choice when the total samples available are limited.

Table 2
Error measurements of the nugget-effect model and the modified nugget-effect model for the M/M/1 example.

	Nugget-effect model		Modified nugget-effect model	
	Mean	Var	Mean	Var
MSE_S	11.72	49.82	9.04	32.38
AAE_S	0.85	0.21	0.69	0.12
MAE_S	8.56	32.83	6.08	18.35

4.3. PAD system

In this example, we adopt the complex queueing network model studied in Cheng and Kleijnen (1999). This packet assembly/disassembly device (PAD) is illustrated in Fig. 15 below:

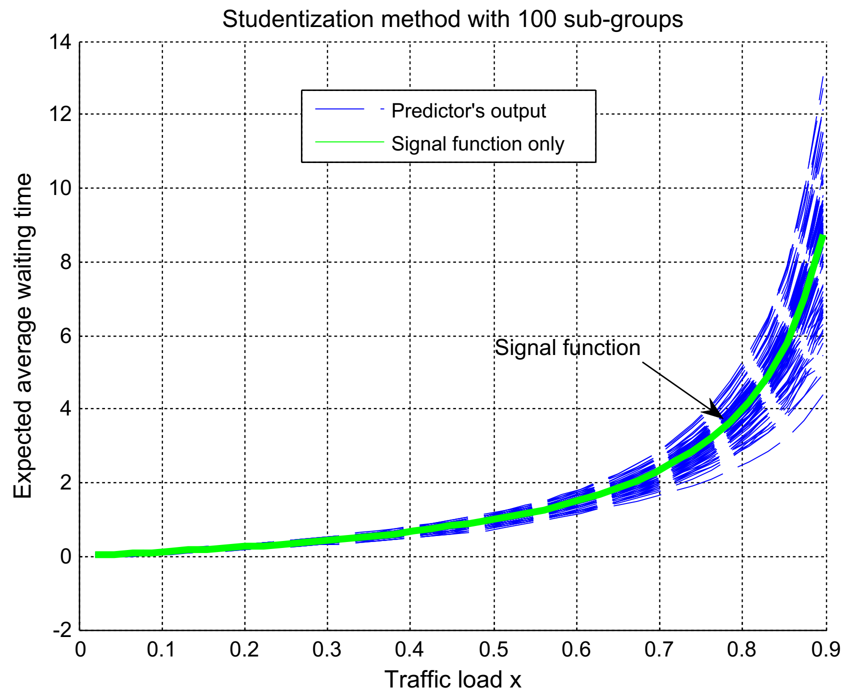

Fig. 13. Studentization method with 100 sub-groups (sample size per sub-group = 10).

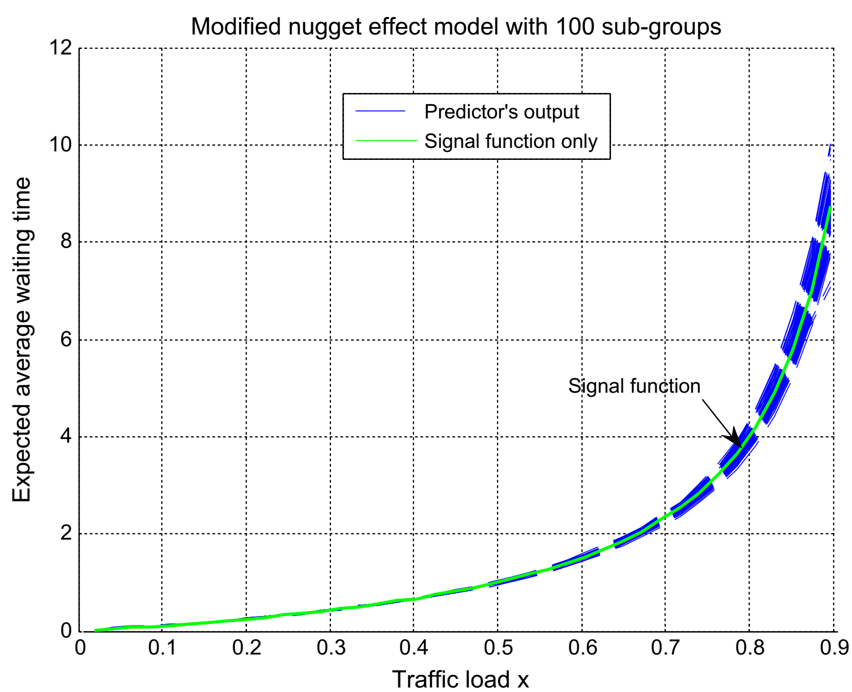

Fig. 14. Modified nugget effect model with 100 sub-groups (sample size per sub-group = 10).

The PAD system receives characters from several terminals. All the character arriving rates of the terminals are the same, and we denote this as x. The number of terminals is $N = 10$. The buffer size of each terminal is set at 32. Once the buffer is full, the characters in the buffer will be packed into a packet and sent to the output queue. If a special character is sent to the buffer, then all the characters in the buffer will be packed and sent to the output queue. The probability of a special character's arrival is 0.02. The packets at the output queue will be served in a FIFO rule and sent to the network at the speed of C char/s.

In this study, the input variable is the character arrival rate $\boldsymbol{x}$ while the output of interest is the average character delay T. This is the time interval from the moment the character arrives at the terminal to the moment the character leaves the output queue. Obviously, when the arrival rate x is low, the delay T will be high due to the long inter-arrival interval. In this study, we are

Table 3
Error measurement of the studentization method and the modified nugget-effect model for the M/M/1 example.

Sample size per sub-group	Studentization method		Modified nugget-effect model	
	Mean	Var	Mean	Var
10				
MSE_S	2.2040	2.6041	0.6525	0.2162
AAE_S	0.1795	0.0149	0.0454	0.0009
MAE_S	1.4373	1.2573	0.4584	0.1164
100				
MSE_S	1.5485	0.9168	0.4677	0.1110
AAE_S	0.1695	0.0103	0.0223	0.00007
MAE_S	0.7417	0.2646	0.4299	0.1047
1000				
MSE_S	0.5444	0.1167	0.3966	0.0406
AAE_S	0.0568	0.0013	0.0216	0.00002
MAE_S	0.2793	0.0343	0.3377	0.0390

Table 4
Error measurement of the nugget-effect model and the modified nugget-effect model for the PAD system example

	Nugget-effect model		Modified nugget-effect model	
	Mean	Var	Mean	Var
MSE_S	0.72	3.26	0.65	2.93
AAE_S	0.82	0.15	0.76	0.11
MAE_S	1.21	2.81	0.98	2.56

interested in the higher saturation regions, so the focus region will be in the region [0.5, 0.9]. 9 observed points are evenly distributed in this region and 100 replications are taken for each observation location. Based on the estimated mean and variance of the 100 replications, we plot the prediction intervals below:

In Table 4, we use 41 evenly distributed unobserved test points to determine the performance of both the nugget-effect and the

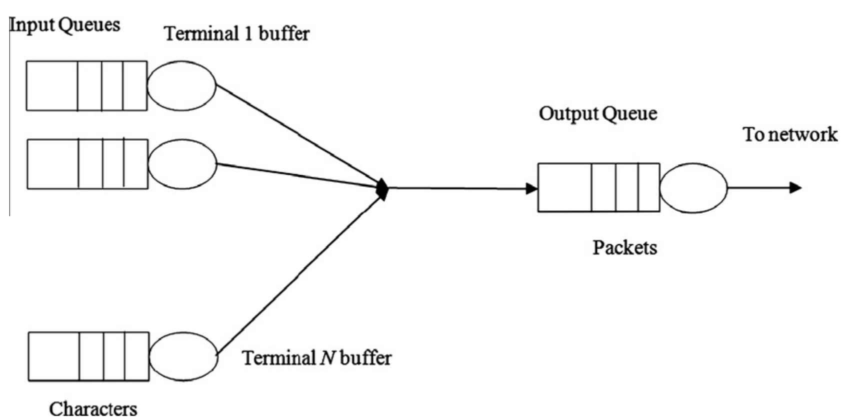

Fig. 15. Queueing model for computer PAD system.

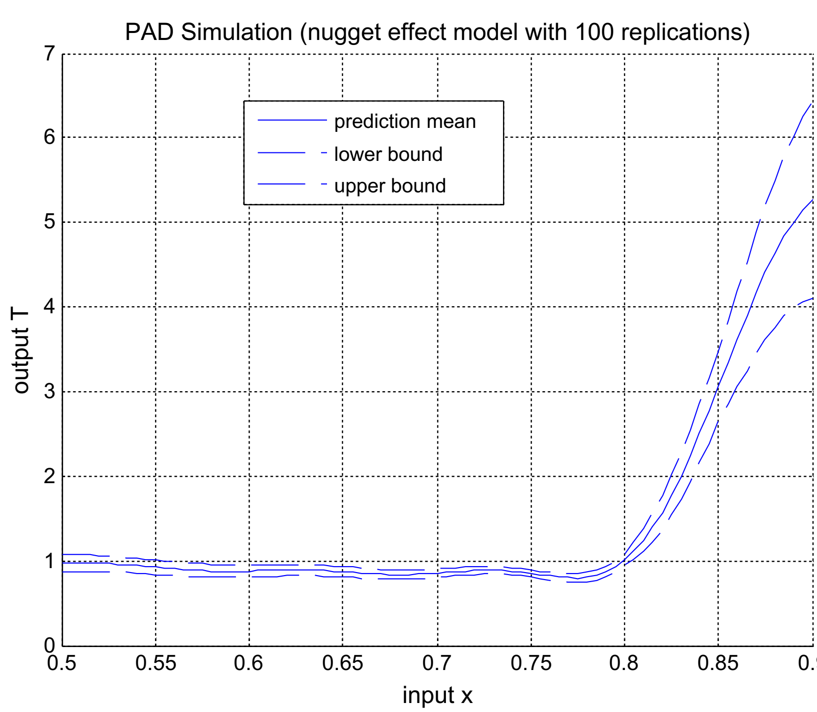

Fig. 16. Prediction interval for nugget effect predictor (PAD system).

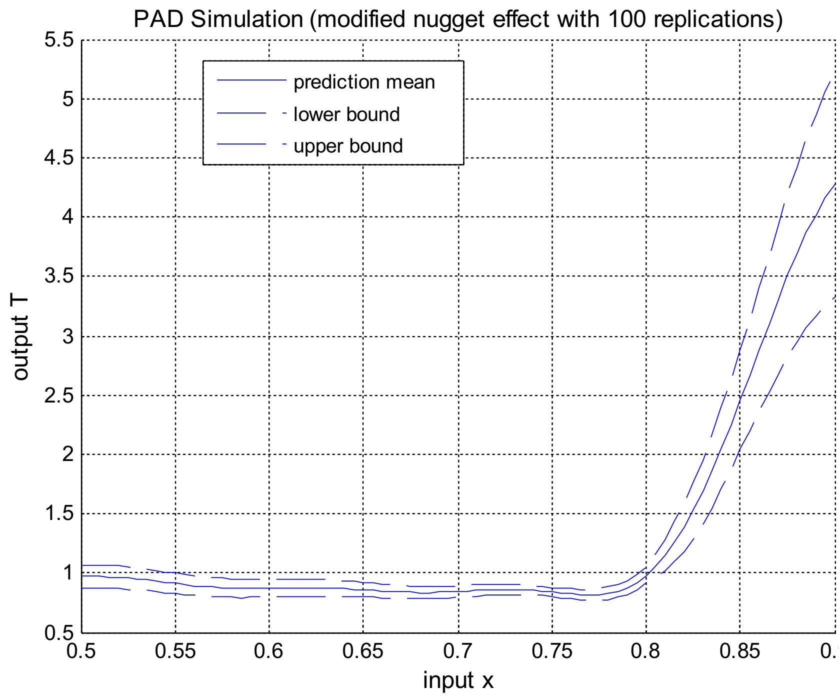

Fig. 17. Prediction interval for modified nugget effect predictor (PAD system).

modified nugget-effect models. Paired t-test results show that the differences between nugget-effect model and the modified nugget-effect model are not statistically significant at the alpha level of 0.05. As the closed-form signal function is not known for this PAD system, the three measurements evaluate the differences between the single replication and the overall mean.

From Figs. 16, 17 and Table 4, we see that the differences between the modified nugget-effect model and the nugget-effect model are not obvious. Although the modified nugget-effect model has a tighter prediction interval and is better able to reduce the influence of noise with high variance in the region between 0.85 and 0.9. Based on the test data collected at test points $x = 0.85$, 0.86, 0.87, 0.88, 0.89, 0.90. The differences between the modified nugget-effect model and the nugget-effect model are significant at alpha level of 0.05.

The numerical experiments in this section provide insights into the application of modified nugget-effect model in stochastic simulation. Although we focus on the development of our proposed model in this paper, in practice it can be applied within sequential experimental design or optimization frameworks such as those proposed in Kleijnen (2009) and Huang et al. (2006). Because the proposed modified nugget-effect model requires prior information of the target simulation model, these stage-wise sequential approaches are useful where an initial stage can be used to estimate variances and structure, and subsequent follow up stages can be used to focus on refining the model in more interesting or promising regions. Hence the proposed model also can work in the real world scenarios with these stage-wise sequential approaches.

5. Conclusion

In this research, we investigate the kriging model's application in stochastic simulation, especially in heteroscedastic situations. We discuss the kriging model's behavior in this situation, and propose a modified nugget-effect model by relaxing the stationarity assumptions of the nugget-effect model. We then look into the behavior of the likelihood function in the parameter estimation for the traditional OK model given the stochastic inputs, and note that the nugget-effect model and the modified nugget-effect model can reduce the erratic behavior in the parameter estimation by penalizing the likelihood function affected by noisy input. This modified nugget-effect model is also compared with the stationary kriging model in terms of the mean squared error. We propose two methods to easily obtain the prediction variance of our new predictor; namely, the non-parametric bootstrapping method and the interpolation method. We illustrate our model and approach with several numerical examples. Our results indicate that the modified nugget-effect model outperforms the stationary kriging model in heteroscedastic cases. It is a promising method for applications in simulation systems with heterogeneous variances. Future research includes generalizing the results to a larger class of correlation functions, and extending its application to examples with more than one input.

Acknowledgements

This research was partially supported by a research project grant (R-266-000-051-720) funded by the NOL (Neptune Orient Lines) Fellowship Programme in Singapore. The authors also thank the associate editor and three anonymous referees for their valuable comments and suggestions that have led to improvements in the article. Parts of this paper were published in Yin, Ng, and Ng (2008).

Appendix A. Kriging predictor and kriging variance for heteroscedastic model

According to Cressie (1993), the kriging predictor can be obtained by minimizing the mean squared error in Eq. (8). For the general case (using $Z(\boldsymbol{x})_i$ as the input to the model instead of $\overline{Z}(\boldsymbol{x}_i)$), the MSE for the kriging predictor is

$$\begin{aligned}\mathrm{MSE} &= E[P(Z(\boldsymbol{x}_0)) - Z(\boldsymbol{x}_0)]^2 = E\left[\sum_{i=1}^{m}\lambda_i Z(\boldsymbol{x}_i) - Z(\boldsymbol{x}_0)\right]^2 \\ &= E\left[(Z(\boldsymbol{x}_0)-\mu)^2 + \sum_{i=1}^{m}\sum_{j=1}^{m}\lambda_i\lambda_j(Z(\boldsymbol{x}_i)-\mu)(Z(\boldsymbol{x}_j)-\mu)\right. \\ &\quad \left. -2\sum_{i=1}^{m}\lambda_i(Z(\boldsymbol{x}_0)-\mu)(Z(\boldsymbol{x}_i)-\mu)\right] = var(Z(\boldsymbol{x}_0)) \\ &\quad + \sum_{i=1}^{m}\sum_{j=1}^{m}\lambda_i\lambda_j cov(Z(\boldsymbol{x}_i), Z(\boldsymbol{x}_j)) - 2\sum_{i=1}^{m}\lambda_i cov(Z(\boldsymbol{x}_i), Z(\boldsymbol{x}_0))\end{aligned}$$

given the kriging weight constraint $\sum_{i=1}^{m}\lambda_i = 1$, and from Eq. (14) $var(Z(\boldsymbol{x}_0)) = cov(Z(\boldsymbol{x}_0), Z(\boldsymbol{x}_0)) = c_0^* + c_1$; $cov(Z(\boldsymbol{x})_i, Z(\boldsymbol{x}_j)) = c_1 corr(d_{ij})$.

Then minimizing

$$E\left[\sum_{i=1}^{m}\lambda_i Z(\boldsymbol{x}_i) - Z(\boldsymbol{x}_0)\right]^2 - 2l\left(\sum_{i=1}^{m}\lambda_i - 1\right)$$

with respect to λ_i: $i = 1, 2, \ldots, m$ and the Lagrange multiplier l gives the kriging predictor as follows:

$$P(Z(\boldsymbol{x}_0)) = \sum_{i=1}^{m}\lambda_i\overline{Z}(\boldsymbol{x}_i) \quad \text{with} \quad \lambda_i = r^T R'^{-1} e_i + 1^T R'^{-1}\frac{(1 - 1^T R'^{-1} r)^T}{1^T R'^{-1} 1} e_i$$

where R' is given in Eq. (15).

The kriging variance is given as follows:

$$\begin{aligned}\mathrm{MSE}(\boldsymbol{x}_0) &= c(\boldsymbol{x}_0)^* + c_1 - c_1\left[r + 1\frac{(1 - 1^T R'^{-1} r)}{1^T R'^{-1} 1}\right]^T R'^{-1} r + c_1 \\ &\quad \times \frac{1 - 1^T R'^{-1} r}{1^T R'^{-1} 1}\end{aligned}$$

which is the result in (17).

Appendix B. $\mathrm{MSE_S}$ for the modified nugget-effect model and nugget-effect model

The $\mathrm{MSE_S}$ at the ith observation point for the kriging model with modified nugget-effect is

$$\begin{aligned}\mathrm{MSE}_S(\boldsymbol{x}_i) &= c_1\left[1 - \left(r + 1\frac{(1 - 1^T R'^{-1} r)}{1^T R'^{-1} 1}\right)^T R'^{-1} r + \frac{(1 - 1^T R'^{-1} r)}{1^T R'^{-1} 1}\right] - c(\boldsymbol{x}_i)^* \\ &= c_1\left[1 - \left(r + 1\frac{(1 - 1^T R^{-1} r) + 1^T R^{-1}(R^{-1} + \eta^{-1})^{-1} R^{-1} r}{1^T R'^{-1} 1}\right)^T R'^{-1} r\right] \\ &\quad + \frac{(1 - 1^T R^{-1} r) + 1^T R^{-1}(R^{-1} + \eta^{-1})^{-1} R^{-1} r}{1^T R'^{-1} 1} - c(\boldsymbol{x}_i)^* \\ &= c_1\left[1 - \left(r + 1\frac{1^T R^{-1}(R^{-1} + \eta^{-1})^{-1} R^{-1} r}{1^T R'^{-1} 1}\right)^T R'^{-1} r\right. \\ &\quad \left. + \frac{1^T R^{-1}(R^{-1} + \eta^{-1})^{-1} R^{-1} r}{1^T R'^{-1} 1} - c(\boldsymbol{x}_i)^*\right. \\ &= c_1\left[1 - \left(r + 1\frac{1^T R^{-1}(R^{-1} + \eta^{-1})^{-1}\eta^{-1}\eta R^{-1} r}{1^T R'^{-1} 1}\right)^T R'^{-1} r\right. \\ &\quad \left. + \frac{1^T R^{-1}(R^{-1} + \eta^{-1})^{-1}\eta^{-1}\eta R^{-1} r}{1^T R'^{-1} 1} - c(\boldsymbol{x}_i)^*\right. \\ &= c_1\left[1 - \left(r + 1\frac{1^T R'^{-1}\eta e_i}{1^T R'^{-1} 1}\right)^T R'^{-1} r + \frac{1^T R'^{-1}\eta e_i}{1^T R'^{-1} 1}\right] - c(\boldsymbol{x}_i)^*\end{aligned}$$

Let $\Delta_m = 1^T R'^{-1} 1$ indicate the summation of all the elements in the inverse correlation matrix, and let Δ_{mi} represent the summation of the ith column or row. Then we have

$$\begin{aligned}\mathrm{MSE}_S(\boldsymbol{x}_i) &= c_1\left[1 - \left(r + 1\frac{\eta_i\Delta_{mi}}{\Delta_m}\right)^T R'^{-1} r + \frac{\eta_i\Delta_{mi}}{\Delta_m}\right] - c(\boldsymbol{x}_i)^* \\ &= c_1\left[1 - 1 - \sum_{j=1}^{m} r_j\eta_i\left(\frac{\Delta_{mi}\Delta_{mj}}{\Delta_m} - \Delta_m\right) + \frac{\eta_i\Delta_{mi}}{\Delta_m}\right] - c(\boldsymbol{x}_i)^* \\ &= c_1\left[-\eta_i\left(\frac{\Delta_{mi}}{\Delta_m} - 1\right)\sum_{j=1}^{m} r_j\delta_{mj} + \frac{\eta_i\Delta_{mi}}{\Delta_m}\right] - c(\boldsymbol{x}_i)^* \\ &= c_1\left[-\eta_i\left(\frac{\Delta_{mi}}{\Delta_m} - 1\right)(1 - \eta_i\Delta_{mi}) + \frac{\eta_i\Delta_{mi}}{\Delta_m}\right] - c(\boldsymbol{x}_i)^* \\ &= c_1\left[-\eta_i\left(\frac{\Delta_{mi}}{\Delta_m} - 1 - \eta_i\frac{\Delta_{mi}^2}{\Delta_m} + \eta_i\Delta_i\right) + \frac{\eta_i\Delta_{mi}}{\Delta_m}\right] - c(\boldsymbol{x}_i)^* \\ &= c_1\left[\eta_i + \eta_i^2\frac{\Delta_{mi}^2}{\Delta_m} - \eta_i^2\Delta_{mi}\right] - c(\boldsymbol{x}_i)^*\end{aligned}$$

For the modified nugget-effect model, $\eta_i c(\boldsymbol{x}_i)^*/c_1$, so

$$\mathrm{MSE}_S(\boldsymbol{x}_i) = c_1\left(\eta_i^2\frac{\Delta_{mi}^2}{\Delta_m} - \eta_i^2\Delta_{mi}\right) \tag{34}$$

References

Ankenman, B., Nelson, B. L., & Staum, J. (2010). Stochastic kriging for simulation metamodeling. *Operations Research, 58*(2), 371–382.

Chen, X., Ankenman, B. E., & Nelson, B. L. (2010). *The effects of common random numbers on stochastic kriging. Working paper*. Department of Industrial Engineering & Management Sciences, Northwestern University.

Cheng, R. C. H., & Kleijnen, J. P. C. (1999). Improved design of queueing simulation experiments with highly heteroscedastic responses. *Operations Research, 47*(5), 762–777.

Cressie, N. (1993). *Statistics for spatial data* (revised ed.). New York: Wiley.

Den Hertog, D., Kleijnen, J. P. C., & Siem, A. Y. D. (2006). The correct kriging variance estimated by bootstrapping. *Journal of the Operational Research Society, 57*(4), 400–499.

Efron, B. (1979). Bootstrap methods: Another look at the Jackknife. *The Annals of Statistics, 7*(1), 1–26.

Gramacy, R. B., & Lee, K. H. (2009). Adaptive design and analysis of supercomputer experiments. *Technometrics, 51*(2), 130–145.

Gupta, A., Yu, D., Xu, L., & Reinikainen, T. (2006). Optimal parameter selection for electronic packaging using sequential computer simulations. *Journal of Manufacturing Science and Engineering, Transactions of the ASME, 128*(3), 705–715.

Hillier, F. S., & Lieberman, G. J. (2001). *Introduction to operations research* (7th ed.). Boston: McGraw-Hill.

Huang, D., Allen, T. T., Notz, W. I., & Zeng, N. (2006). Global optimization of stochastic black-nox systems via sequential kriging metamodels. *Journal of Global Optimization, 34*(3), 441–466.

Kleijnen, J. P. C. (1987). *Statistical tools for simulation practitioners*. New York: Marcel Dekker.

Kleijnen, J. P. C. (1998). Experimental design for sensitivity analysis, optimization, and validation of simulation models. In J. Banks (Ed.), *Handbook of simulation* (pp. 173–223). New York: Wiley.

Kleijnen, J. P. C. (2008). *Design and analysis of simulation experiments*. New York: Springer.

Kleijnen, J. P. C. (2009). Kriging metamodeling in simulation: A review. *European Journal of Operational Research, 192*(3), 707–716.

Kleijnen, J. P. C., & Van Beers, W. C. M. (2005). Robustness of kriging when interpolating in random simulation with heterogeneous variances: Some experiments. *European Journal of Operational Research, 165*(3), 826–834.

Kleijnen, J. P. C., Van Beers, W. C. M., & Van Nieuwenhuyse, I. (2010). Constrained optimization in simulation: A novel approach. *European Journal of Operational Research., 202*, 164–174.

Li, Y. F., Ng, S. H., Xie, M., & Goh, T. N. (2010). A systematic comparison of metamodeling techniques for simulation optimization in decision support systems. *Applied Soft Computing, 10*, 1257–1273.

Li, R., & Sudjianto, A. (2005). Analysis of computer experiments using penalized likelihood in gaussian kriging models. *Technometrics, 47*(2), 111–120.

Lophaven, S. N., Nielsen, H. B. & Sondergdaard, J. (2002). *DACE: A Matlab kriging toolbox, version 2.0*. Lyngby: IMM Technical University of Denmark.

Matheron, G. (1963). Principles of geostatistics. *Economic Geology, 58*, 1246–1266.

Sacks, J., Schiller, S. B., & Welch, W. (1989b). Design for computer experiments. *Technometrics, 31*, 41–47.

Sacks, J., Welch, W. J., Mitchell, T. J., & Wynn, H. P. (1989a). Design and analysis of computer experiments (with discussion). *Statistical Science, 4*, 409–435.

Sakata, S., Ashida, F., & Zako, M. (2007). On applying kriging-based approximate optimization to inaccurate data. *Computer Methods in Applied Mechanics and Engineering, 196*(13-16), 2055–2069.

Simpson, T. W., Peplinski, J., Koch, P. N., & Allen, J. K. (2001). Metamodels for computer-based engineering design: Survey and recommendations. *Engineering with Computers, 17*, 129–150.

Van Beers, W. C. M., & Kleijnen, J. P. C. (2003). Kriging for interpolation in random simulation. *Journal of the Operational Research Society, 54*(3), 255–262.

Wang, H., Li, E., Li, G. Y., & Zhong, Z. H. (2008). A metamodel optimization methodology based on multi-level fuzzy clustering space reduction strategy and its applications. *Computers and Industrial Engineering, 55*(2), 503–532.

Woodbury, M. A. (1950). *Inverting modified matrices. Statistical Research Group, Memo. Rep. No. 42*. Princeton, NJ: Princeton University.

Yin, J., Ng, S. H., & Ng, K. M. (2008). Kriging model with modified nugget effect for random simulation with heterogeneous variances. *Proceedings of IEEE International Conference on Industrial Engineering and Engineering Management, 2008*, 1714–1718.