Proceedings of the 2009 Winter Simulation Conference
M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin and R. G. Ingalls, eds.

A STUDY ON THE EFFECTS OF PARAMETER ESTIMATION ON KRIGING MODEL'S PREDICTION ERROR IN STOCHASTIC SIMULATIONS

Jun Yin
Szu Hui Ng
Kien Ming Ng

Department of Industrial & Systems Engineering
10 Kent Ridge Crescent, Singapore
National University of Singapore
Singapore 119260,SINGAPORE

ABSTRACT

In the application of kriging model in the field of simulation, the parameters of the model are likely to be estimated from the simulated data. This introduces parameter estimation uncertainties into the overall prediction error, and this uncertainty can be further aggravated by random noise in stochastic simulations. In this paper, we study the effects of stochastic noise on parameter estimation and the overall prediction error. A two-point tractable problem and three numerical experiments are provided to show that the random noise in stochastic simulations can increase the parameter estimation uncertainties and the overall prediction error. Among the three kriging model forms studied in this paper, the modified nugget effect model captures well the various components of uncertainty and has the best performance in terms of the overall prediction error.

1 INTRODUCTION

The kriging model has been used as a meta-model in the design and analysis of computer experiments (DACE) since Sacks *et al.* (1989). Although kriging originated in geostatistics, see Matheron (1963), it has been successfully applied in many deterministic computer experiments (Pham and Wagner 1999; Roshan 2006; Gupta *et al*. 2006; Wu and Sun 2007). The general kriging approach assumes that the sample observations are realizations of a random process given as:

$$Y(x) = \mu(x) + \delta(x) + \epsilon(x) = S(x) + \varepsilon(x) \tag{1}$$

where μ is the mean of the process, also known as the large-scale variation; δ is the zero-mean L_2-continuous, second order stationary process used to model the difference between single observation and process mean, also known as the small-scale variation; and ε represents the random measure error (random noise). In the deterministic (noiseless) simulation context of DACE, ε is not present and $\mu(x)+\delta(x)$ is denoted as the deterministic signal function S. $\mu(x)$ is typically modeled with a constant term or a polynomial function, and δ a Gaussian random process with a mean of 0 and a spatial correlation function.

In building the kriging model and its predictor, in addition to the sample observations y, the best linear unbiased predictor depends on the parameters in $\mu(x)$ and the covariance parameters in $\delta(x)$. In an ideal situation, these parameters are assumed known. In practice however, these parameters can only be estimated from sample data, making them random variables dependent on the experimental design and sample observations. Moreover, the prediction error of the kriging model, which is a commonly used quality measure of the fit and accuracy of the model, is a function of sample data and model parameters. In the ideal case when the parameters are known, the prediction error measures the "true" prediction error. Typically however, the parameters are unknown and the prediction error is estimated by replacing the unknown parameters with point estimates. This 'plug-in' estimator will however underestimate the true prediction error as it does not take into account the uncertainty of the model parameters. In some cases, it can cause overconfidence in the predictors. This additional error is also noted in Cressie (1993).

Similar problems in parameter uncertainty have been studied in time series models, heteroscedastic regression models, mixed linear models (Khatri and Shah 1981, Kackar and Harville 1984, Reinsel 1984) and general linear models (Toyooka 1982, Harville 1985, Harville and Jeske 1992). Zimmerman and Cressie (1992) extended this research in parameter estimation uncertainty to the mean squared prediction error (MSPE) of spatial linear models, and examined the appropriateness of the 'plug-in' estimator of the MSPE with alternative approximations. Hertog, Kleijnen and Siem (2006) studied the similar problem with the sensitivity parameter θ of the deterministic kriging model with a boot-strapping approach, and showed that the traditional kriging variance underestimates the true kriging variance as noted in Cressie (1993). These studies demonstrate that the problem of parameter uncertainty in spatially correlated models can have a large influence on the prediction error in the deterministic (noiseless) case. In stochastic simulations, this problem can be amplified by the random noise ε of the system as it increases the variability of the parameter estimates (see illustrative example below). In this paper, we extend previous studies to look at the effects of stochastic noise on parameter estimation and the overall prediction error. We further decompose the mean squared error into individual components reflecting parameter uncertainty, response model misspecification and the stochastic noise.

To illustrate the influence of random noise in stochastic simulation on parameter estimation and prediction error, consider the example where the response is $Y(x) = sin(x) + \varepsilon$, and $\varepsilon \sim N(0, \sigma_\varepsilon^{\ 2})$. Observations are obtained at seven equally spaced input locations from $[0, 2\pi]$ for various levels of $\sigma_\varepsilon^{\ 2}$. Suppose a deterministic kriging model with an exponential correlation function is used to fit the data. For each level of $\sigma_\varepsilon^{\ 2}$, 1000 replications of observations are taken. For each replication, we estimate the sensitivity parameter θ of the exponential correlation function and compute the overall prediction error at the point x_0=3π/4, where the overall prediction error is the squared differences between the estimated kriging predictor and the signal function $E[Y(x_0)] = S(x_0) = sin(x_0)$. Table 1 summarizes the results.

Table 1: Numerical results on the parameter estimation and prediction error.

Variance of the ε, $\sigma_\varepsilon^{\ 2}$	*Variance of the θ estimator*	*Averaged Overall Prediction Error*
0.1	9.161*E-12	0.0159
0.2	5.044*E-11	0.0268
0.5	1.181*E-8	0.1335
1	1.894*E-7	0.5923

From Table 1, we see that both of the variance of the estimated θ and averaged overall prediction error increase as the variance of the random noise ε increases. As the input design for x is fixed for all four noise levels, the inherent model misspecification error is the same throughout, and hence, the increase in the overall prediction error can be attributed to the noisy sample data.

The results from this example show that the performance of deterministic kriging model worsens as the noise level increases. This indicates that a more appropriate model reflecting the stochastic inputs is required when the noise levels are high. In such situations, the kriging model has been improved. Under homoscedastic assumptions on the noise, the kriging model with the nugget effect (nugget effect model) has been proposed and applied (Cressie 1993, Huang *et al.* 2006). Yin, Ng, and Ng (2008) and Nelson, Staum, and Ankenman (2008) propose the modified nugget effect model and the stochastic kriging model respectively to address the more general heteroscedastic case. Although the mathematical predictor forms of both models are equivalent, their initial assumptions differ in that the modified nugget effect model treats the additional noise component ε as a non stationary component of the random process; and the stochastic kriging model considers the additional noise component ε as the intrinsic uncertainty of the simulation itself and use it to model the effect of common random number (CRN).

In this paper, we will look more closely at the influence of random noise in stochastic simulations on three different model types: traditional deterministic kriging model (Cressie, 1993), nugget effect kriging model (Cressie, 1993), and the modified nugget effect kriging model (Yin, Ng and Ng 2008). We first decompose the prediction error into components reflecting the model misspecification error, parameter estimation error and the stochastic error. We will then examine a simple two-point tractable problem and provide some insights on the effects of the random noise in stochastic simulations on the individual components of the three different model types. Finally, we provide a numerical study on three additional examples. In the next section, we first review the three different forms of the kriging model considered.

2 KRIGING MODELS FOR STOCHASTIC SIMULATION

Deterministic kriging model, nugget effect model and modified nugget effect model are the three kriging model forms we will study in this paper. The nugget effect model and modified nugget effect model are developed on the basis of the deterministic kriging model. In this section, we will review the basic theory of kriging model and summarize the structures and characteristics of these three models.

2.1 Kriging model basis

The basic model form and structure for the kriging model is given in (1). Based on the model assumption, we can derive the general kriging predictor. The kriging predictor is a combination of all the observations $y(x_k)$ and the "kriging weights" λ, which is also a function of the observed data:

$$P(Y(x_0)) = \sum_{k=1}^{m}(\lambda_k Y(x_k))$$

where

$$\lambda_k = r^T R^{-1} e_k + 1^T R^{-1} e_k \frac{(1 - 1^T R^{-1} r)^T}{1^T R^{-1} 1} \tag{2}$$

and r=($corr(d_{01})$, $corr(d_{02})$,…, $corr(d_{0m})$), the vector of the correlations between the point to be estimated and all the m observations. It is a function of the distances d_{0i} of the observations to the prediction point x_0. R is the correlation matrix of all the observation points, and e_k is the vector of ones of length m.

Under the second-order stationarity assumption (Cressie 1993), the covariance between any two points is only dependent on the distance between these two points:

$$C(d_{ij}) = cov(Y(x_i), Y(x_j)) = \begin{cases} c_0 + c_1 & d_{ij} = 0 \\ c_1 corr(d_{ij}) & d_{ij} \neq 0 \end{cases} \tag{3}$$

where c_0 is the nugget effect representing the random noise component ε in the model; c_1 is the "partial sill" which models the variance of the random process without the random noise. The correlation function $corr(\sim)$ is a function of distance only. In this paper, we focus on the exponential family correlation functions which are most commonly applied for its smooth output and fast convergence. The form of the exponential family correlation functions are given as follows:

$$corr(d_{ij}) = \exp(-\theta |d_{ij}|^q) \tag{4}$$

Equation (4) becomes the exponential correlation function when q=1, and Gaussian correlation function when q=2. It represents the relationship between the points in the sample space. The sensitivity parameter θ is typically estimated by the maximum likelihood estimation method. Given the functional forms of (3) and (4), the kriging predictor in (2) is not a linear function of θ.

2.2 Kriging models for deterministic case and stochastic case

The three different forms of the kriging model applied in this research can be distinguished from each other by the random noise component ε in (1). For the predictor and kriging weight, all three models have the same structure in (2). The only difference is in the correlation matrix R. For the traditional deterministic model, the random error ε is assumed to be 0. Combining the covariance function in (3), with the nugget effect term $c_0 = 0$, the correlation matrix R of the deterministic kriging model is:

$$R = \begin{bmatrix} 1 & corr(d_{12}) & \dots & corr(d_{1(m-1)}) & corr(d_{1m}) \\ corr(d_{21}) & 1 & \dots & corr(d_{2(m-1)}) & corr(d_{2m}) \\ \dots & \dots & \ddots & \dots & \dots \\ corr(d_{(m-1)1}) & corr(d_{(m-1)2}) & \dots & 1 & corr(d_{(m-1)m}) \\ corr(d_{m1}) & corr(d_{m2}) & \dots & corr(d_{m(m-1)}) & 1 \end{bmatrix} \tag{5}$$

Under the stochastic assumptions, we further divide it into two sub-cases: homoscedastic case and heteroscedastic case. With the homoscedastic assumption, the variance of the random noise component ε is a constant throughout the whole sample space. From the definition in (3), the correlation matrix R of the nugget effect model can be written as:

$$R = \begin{bmatrix} 1+\frac{c_0}{c_1} & corr(d_{12}) & \dots & corr(d_{1(m-1)}) & corr(d_{1m}) \\ corr(d_{21}) & 1+\frac{c_0}{c_1} & \dots & corr(d_{2(m-1)}) & corr(d_{2m}) \\ \dots & \dots & \ddots & \dots & \dots \\ corr(d_{(m-1)1}) & corr(d_{(m-1)2}) & \dots & 1+\frac{c_0}{c_1} & corr(d_{(m-1)m}) \\ corr(d_{m1}) & corr(d_{m2}) & \dots & corr(d_{m(m-1)}) & 1+\frac{c_0}{c_1} \end{bmatrix} \tag{6}$$

The only difference between (5) and (6) is the nugget effect term added onto the diagonal. For the heteroscedastic assumption, the variance of the random noise ε is a variable. Assuming ε to be independent but not necessarily identical, the nugget effect terms added onto the diagonal will also be a variable. The correlation matrix R for the modified nugget effect model can be given as:

$$R = \begin{bmatrix} 1+\frac{c_1^*}{c_1} & corr(d_{12}) & \dots & corr(d_{1(m-1)}) & corr(d_{1m}) \\ corr(d_{21}) & 1+\frac{c_2^*}{c_1} & \dots & corr(d_{2(m-1)}) & corr(d_{2m}) \\ \dots & \dots & \ddots & \dots & \dots \\ corr(d_{(m-1)1}) & corr(d_{(m-1)2}) & \dots & 1+\frac{c_{m-1}^*}{c_1} & corr(d_{(m-1)m}) \\ corr(d_{m1}) & corr(d_{m2}) & \dots & corr(d_{m(m-1)}) & 1+\frac{c_m^*}{c_1} \end{bmatrix} \tag{7}$$

where c_i^* stands for the local variance for the *i*th location observation, or $\sigma_{\varepsilon i}^2$. From (5), (6) and (7), we see that the nugget effect model is a special case of modified nugget effect model when all the c_i^* are the same and deterministic kriging model is a special case of nugget effect model when c_0=0.

3 DECOMPOSITION OF THE OVERALL PREDICTION ERROR

The additional error caused by parameter estimation uncertainty is reported in several different research papers. For the kriging model, the sensitivity parameter estimation uncertainty's influence on prediction error has been discussed in Hertog, Kleijnen and Siem (2006). Bootstrapping numerical experiments show that the actual kriging model prediction error which correctly accounts for parameter estimation uncertainty is larger than the traditional kriging variance given the known parameter.

Assuming the predictor $P[S(x_0)]_Y$ is a linear function of the parameter θ and the estimator for θ is unbiased, from the results of Kacker and Harville (1984), the prediction error with unknown parameter can be approximated by:

$$MSE[P[S(x_0)]_Y] = tr[A(\hat{\theta})B(\hat{\theta})] + MSE[P[S(x_0)]_Y|\hat{\theta}] \tag{8}$$

where $tr[A(\theta)B(\theta)]$ is an approximation of the additional error introduced when the estimator $\hat{\theta}$ is used, $A(\theta) = var(\frac{\partial P[S(x_0)]_Y}{\partial \theta})$, and $B(\theta) = E[(\hat{(\theta)} - \theta)(\hat{(\theta)} - \theta)^T]$. The second component of the right-hand side of (8) is the traditional mean squared error (MSE) when $\hat{\theta}$ is used, and this can be further decomposed as:

$$\begin{aligned} MSE[P[S(x_0)]_Y|\hat{\theta}] = E[(\sum_{k=1}^{m}\lambda_k Y_k - S(x_0))^2|\hat{\theta}] &= E[(\sum_{k=1}^{m}\lambda_k(S(x_k)+\varepsilon_k) - S(x_0))^2|\hat{\theta}] \\ &= E[(\sum_{k=1}^{m}\lambda_k S(x_k) - S(x_0) + \sum_{k=1}^{m}\lambda_k\varepsilon_k)^2|\hat{\theta}] \\ &= E[(\sum_{k=1}^{m}\lambda_k S(x_k) - S(x_0))^2|\hat{\theta}] + \sum_{k=1}^{m}E[\lambda_k^2|\hat{\theta}]\sigma^2_{\varepsilon_k^2} \end{aligned} \tag{9}$$

The first term on the right hand side of (9) is the prediction error caused by model misspecification. The second term is the direct effect of the stochastic noise ε on the prediction error, and is a function of the variance of ε. Combining (8) and (9) together, we find that the overall prediction error can be decomposed into the following three error components:

$$MSE[P[S(x_0)]_Y] = MSE[P[S(x_0)]_S|\hat{\theta}] + \sum_{k=1}^{m} E[\lambda_k^2|\hat{\theta}]\sigma_{\varepsilon_k}^2 + tr[A(\hat{\theta})B(\hat{\theta})] \tag{10}$$

The prediction error caused by model misspecification is inherent in the metamodel selection and will not be the focus in this research. In the next section, we analyze how the random noise ε affects the parameter estimation of θ and the additional error caused by parameter estimation uncertainty (the last term in (10)).

4 TWO-POINT PROBLEM

We design a simple two-point problem to provide some theoretical insights to the parameter estimation problem for the kriging model:

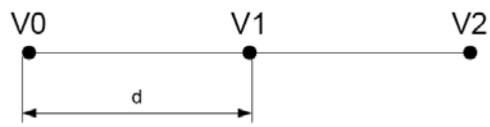

Figure 1. Design for two-point problem

Points V_0, V_1 and V_2 are evenly spaced. Take points V_1 and V_2 to be the observation points, and point V_0 to be the prediction point. Suppose that the signal function s is an unknown Gaussian random process with the mean function μ and bias function δ. The additional random noise function ε follows the normal distribution with zero mean and unknown heterogeneous variance function $\sigma_\varepsilon^2(V_i)$. Then at points:

$$V_1(x_1) : y_1 = s_1 + \varepsilon_1, V_2(x_2) : y_2 = s_2 + \varepsilon_2 \tag{11}$$

In the next subsection, we describe the parameter estimation techniques for the unknown parameter before discussing its effects on the overall prediction error.

4.1 Maximum Likelihood Estimation and Restricted Maximum Likelihood Estimation

The Maximum Likelihood Estimation (MLE) method is commonly used in kriging model's estimation. Assuming a Gaussian random process, the log-likelihood function is given by:

$$\ell(\mu, \theta, \sigma_S^2) = -\frac{1}{2}\log|\sigma_S^2 R(\theta)| - \frac{1}{2}(Y - F\beta)^T(\sigma_S^2 R(\theta))^{-1}(Y - F\beta) \tag{12}$$

where F is the design matrix for the ordinary least squares model, β is the regression parameters and $F\beta$ represents the mean function μ; σ_S^2 is the variance of the Gaussian random process S, which indicates the variability of an unknown point in S. Here we write the correlation matrix as $R(\theta)$ to denote that it is a function of parameter θ, as seen in equations (4) – (7). We can find the estimators for μ, θ and σ_S^2 by taking the first order derivatives:

$$\frac{\partial \ell(\mu, \theta, \sigma_S^2)}{\partial \mu} = 0, \frac{\partial \ell(\mu, \theta, \sigma_S^2)}{\partial \theta} = 0, \frac{\partial \ell(\mu, \theta, \sigma_S^2)}{\partial \sigma_S^2} = 0$$

Solving the above three equations, the MLE estimators for μ and σ_S^2 result as functions of θ. For simplification purposes, μ and σ^2 are typically assumed fixed or known in order to estimate the sensitivity parameter θ. This simplifies the likelihood function to a function of θ only.

However, in this simplification, the MLE estimator for θ is biased as the estimation of θ depends also on β, which is usually unknown (see Cressie 1993). In order to simplify the approximation in (8), we use instead the restricted maximum likelihood (REML) proposed by Patterson and Thompson (1971, 1974) which provides an unbiased estimator of θ for this Gaussian random process. With this unbiased estimator, $B(\theta)$ will equal to the variance of the θ estimator. The log-likelihood function for the REML is given by:

$$\ell(\theta,\sigma_S^2) = \frac{1}{2}\log|F^TF| - \frac{1}{2}\log|\sigma_S^2 R(\theta)| - \frac{1}{2}\log|F^T(\sigma_S^2 R(\theta))^{-1}F| - \frac{1}{2}Y^T((\sigma_S^2 R(\theta))^{-1} - \frac{(\sigma_S^2 R(\theta))^{-1}FF^T(\sigma_S^2 R(\theta))^{-1}}{F^T(\sigma_S^2 R(\theta))^{-1}F})Y \quad (13)$$

which is independent of β. The difference between (12) and (13) is especially significant in the small sample case, like the two-point problem here (see Cressie, 1993).

For this two-point problem, the terms in (13) are given as:

$$\sigma_S^2 = 1, F = \begin{bmatrix}1\\1\end{bmatrix}, Y = \begin{bmatrix}y_1\\y_1\end{bmatrix}, R(\theta) = \begin{bmatrix}1 & \exp(-d\theta)\\ \exp(-d\theta) & 1\end{bmatrix}$$

As mentioned in section 2.1, the kriging predictor is not a linear function of the parameter θ in this two-point case. In order to make the predictor a linear function of the estimated parameter to apply the approximation in (8), a re-parameterization is made as follows:

$$\rho = \exp(-d\theta), R(\rho) = \begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}$$

With this reparameterization, (13) becomes a function of ρ instead of θ:

$$L_W(\rho) = \frac{1}{2}\log|F^TF| - \frac{1}{2}\log|R(\rho)| - \frac{1}{2}\log|F^T(R(\rho))^{-1}F| - \frac{1}{2}Y^T((R(\rho))^{-1} - \frac{(R(\rho))^{-1}FF^T(R(\rho))^{-1}}{F^T(R(\rho))^{-1}F})Y \quad (14)$$

4.2 Analytical Results

For the three different model forms described in (5), (6) and (7), maximizing (14), we obtain have the following results:
Deterministic kriging model:

$$\hat{\rho} = 1 - \frac{(y_1 - y_2)^2}{2}$$

Nugget effect model:

$$\hat{\rho}_N = \frac{2 + 2c_0 - (y_1 - y_2)^2}{2}$$

Modified nugget effect model:

$$\hat{\rho}_M = \frac{2 + c_1^* + c_2^* - (y_1 - y_2)^2}{2}$$

Following, the expectation and variance of the parameter estimators are given as:

$$E(\hat{\rho}) = 1 - \frac{1}{2}E((y_1 - y_2)^2)$$
$$var(\hat{\rho}) = \frac{1}{4}var((y_1 - y_2)^2) \quad (15)$$

From (11), it is straightforward to see that y_1 follows normal distribution with mean s_1 and variance $\sigma_{\varepsilon 1}^2$, and y_2 follows normal distribution with mean s_2 and variance $\sigma_{\varepsilon 2}^2$. Separating the signal and pure noise components, we get:

$$E((y_1 - y_2)^2) = \sigma_{\varepsilon_1}^2 + \sigma_{\varepsilon_2}^2 + (s_1 - s_2)^2$$
$$var((y_1 - y_2)^2) = 2(\sigma_{\varepsilon_1}^2 + \sigma_{\varepsilon_2}^2)^2 + 4(\sigma_{\varepsilon_1}^2 + \sigma_{\varepsilon_2}^2)(s_1 - s_2)^2 \quad (16)$$

Combining (15) and (16),

$$E(\hat{\rho}) = 1 - \frac{1}{2}(\sigma_{\varepsilon_1}^2 + \sigma_{\varepsilon_2}^2 + (s_1 - s_2)^2)$$
$$var(\hat{\rho}) = \frac{1}{2}(\sigma_{\varepsilon_1}^2 + \sigma_{\varepsilon_2}^2)^2 + \frac{1}{4}(\sigma_{\varepsilon_1}^2 + \sigma_{\varepsilon_2}^2)(s_1 - s_2)^2 \quad (17)$$

Similarly, for the nugget effect model and modified nugget effect model, we obtain the following:

$$E(\hat{\rho}) = 1 + c_0 - \frac{1}{2}(\sigma_{\varepsilon_1}^2 + \sigma_{\varepsilon_2}^2 + (s_1 - s_2)^2)$$
$$var(\hat{\rho}) = \frac{1}{2}(\sigma_{\varepsilon_1}^2 + \sigma_{\varepsilon_2}^2)^2 + (\sigma_{\varepsilon_1}^2 + \sigma_{\varepsilon_2}^2)(s_1 - s_2)^2 \quad (18)$$

$$E(\hat{\rho}) = 1 + \frac{c_1^*}{2} + \frac{c_2^*}{2} - \frac{1}{2}(\sigma_{\varepsilon_1}^2 + \sigma_{\varepsilon_2}^2 + (s_1 - s_2)^2)$$
$$var(\hat{\rho}) = \frac{1}{2}(\sigma_{\varepsilon_1}^2 + \sigma_{\varepsilon_2}^2)^2 + (\sigma_{\varepsilon_1}^2 + \sigma_{\varepsilon_2}^2)(s_1 - s_2)^2 \quad (19)$$

For deterministic model, we see from (17) that the expectation and variance of the estimated parameter are functions of the input variance $\sigma_{\varepsilon 1}^2$ and $\sigma_{\varepsilon 2}^2$. If the input variances increase, the variance of the estimated parameter will also increase and its mean will decrease, indicating a weaker correlation-ship between the points. From (17), the expectation of the estimated parameter can be negative if the variance $\sigma_{\varepsilon 1}^2$ and $\sigma_{\varepsilon 2}^2$ are high enough. However as we assume the exponential correlation function in this two-point problem, we consider only non-negative correlations. For the cases when estimated ρ is negative, the restricted likelihood function is monotonically decreasing, indicating that extra sample data is needed. For the modified nugget effect model, from (19), we see that the influence of the input variance can cancel out if c_1^* and c_2^* are the exact estimators of $\sigma_{\varepsilon 1}^2$ and $\sigma_{\varepsilon 2}^2$. Similarly, from (18), we see that the nugget effect model can have the same results when the nugget value c_0 equals to the average of $\sigma_{\varepsilon 1}^2$ and $\sigma_{\varepsilon 2}^2$.

This partially explains what was observed in (Yin, Ng, and Ng 2008), where it was observed that the estimated θ's for the modified nugget effect model and nugget effect model is closer to the optimal value than the deterministic model.

4.3 Influence of Parameter Estimation on Overall Prediction Error

From the approximation in (8), the additional prediction error caused by parameter estimation uncertainty can be approximated as $tr[A(\hat{\rho})B(\hat{\rho})]$, where $tr[Q]$ stands for trace of matrix Q. Based on the REML, $A(\hat{\rho})$ can be computed (Harville and Jeske 1992) and $B(\hat{\rho})$ is the variance of $\hat{\rho}$. As a result, we can formulate the approximation as a function of ρ:

$$tr[A(\hat{\rho})B(\hat{\rho})] = \frac{1-\hat{\rho}}{2} var(\hat{\rho})$$

The detailed derivation is given in the appendix. Since the variance of the estimator is the same for all three models, an estimator $\hat{\rho}$ closer to 1 is favorable. Comparing (17) and (19), the modified nugget effect estimator $\hat{\rho}_M$ is closer to one in expectation than the estimated parameter given with deterministic kriging model. With careful selection of the nugget value c_0 in (18), the additional error incurred by the nugget effect model can be as small as the modified nugget effect model.

In this simple example, we see that in stochastic situations where random noise is present, selection of the appropriate stochastic model can reduce the additional error introduced by parameter estimation in θ. Furthermore, although not addressed in this paper, good knowledge or accurate estimation of σ_ε^2's can also improve the estimation error.

5 NUMERICAL EXPERIMENTS

The two-point problem given in the section 4 illustrates how the random noise ε affects the parameter estimation uncertainty and increases the additional prediction error in the end. In order to simultaneously study the effects of the random noise on the three individual error components, three numerical examples are further studied in this section. The first example is a one-dimension problem with a step variance function studied in Yin, Ng and Ng (2008). In this example, the correlation is strong in the noiseless case. The second example extends to a two-dimension problem with a step variance function. Several different noise level scenarios are studied in this example. The third example is more complex functional form taken from Hussain et al. (2002) with a continuous heterogeneous variance function.

For each numerical example, 1000 replications are applied. The mean and variance of the θ estimator, prediction error caused by model misspecification, additional prediction error caused by noisy data and additional prediction error caused by parameter estimation uncertainty are provided based on all the 1000 replications. For notation convenience, we use the following abbreviations in the numerical experiments:

Table 2. Abbreviations for notations

DK	Deterministic kriging model
NK	Nugget effect kriging model
MK	Modified nugget kriging model
ErrA	Overall prediction error
ErrM	Prediction error caused by model misspecification
ErrN	Additional prediction error caused by noisy data
ErrP	Additional prediction error caused by parameter estimation uncertainty (approximation)
θ^*	The estimated θ based on noiseless observations

The ErrA, ErrM, ErrN and ErrP refer to the individual components of the prediction error decomposition in (10). θ^* is the θ estimated based on noiseless observations. The comparison between θ^* and $\hat{\theta}$ estimated based on the noisy sample data can provide an insight of parameter estimation uncertainty in the stochastic case.

5.1 One Dimension Quadratic Test Function

The quadratic test function $y=x^2$ was used in Yin, Ng, and Ng (2008). The function has the following shape:

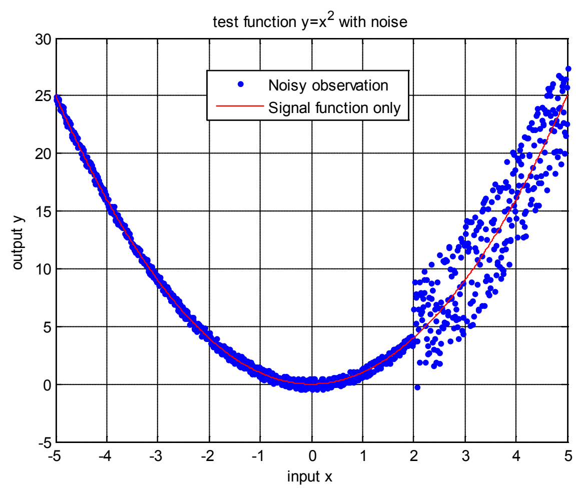

Figure 2. Quadratic test function, $y = x^2 + \varepsilon$

where ε is the additional noise function with step variance $\sigma_\varepsilon^2=0.083$ when $x \in [-5,2)$, $\sigma_\varepsilon^2=8.3$ when $x \in [2,5]$. Observation points are located at $x=[-5,-4,-3,-2,-1,0,1,2,3,4,5]$, and the prediction point is located at x=[-0.5]. The results are given in Table 3:

Table 3. Results for the quadratic test function

Estimation based on noisy data															*Estimation based on noiseless data*	
DK					*NK*					*MK*					θ^*	*ErrM*
$\hat{\theta}$		ErrA	ErrN	ErrP	$\hat{\theta}$		ErrA	ErrN	ErrP	$\hat{\theta}$		ErrA	ErrN	ErrP		
Mean	Var				Mean	Var				Mean	Var					
0.832	5.261	1.375	1.102	0.123	0.245	0.625	0.573	0.427	0.003	0.007	0.001	0.294	0.205	2E-7	0.002	5E-13

From Table 3, we see that the modified nugget effect model helps in reducing the overall prediction error ErrA. Both the prediction error caused by the noisy data and parameter estimation uncertainty for the modified nugget effect model are lower than the other two models. Considering the ErrP to ErrA ratio, an indicator of the fraction of the overall error attributed to the parameter estimation uncertainty, we see that the modified nugget effect model has a lower ratio than the other two models. As for the estimator, we see from the last two columns that the θ^* and prediction error caused by model misspecification are very small. This indicates that there is sufficient samples in the sample space and the correlation is strong in the noiseless case. The estimated θ is higher than the θ^* for all three models, which indicates the correlation is weakened by the random noise. Comparing all the three estimated θ's, the modified nugget effect model has the estimator the closest to θ^*.

5.2 Two Dimension Linear Function

The two dimension linear function is given as follows: $S=x_1+x_2$

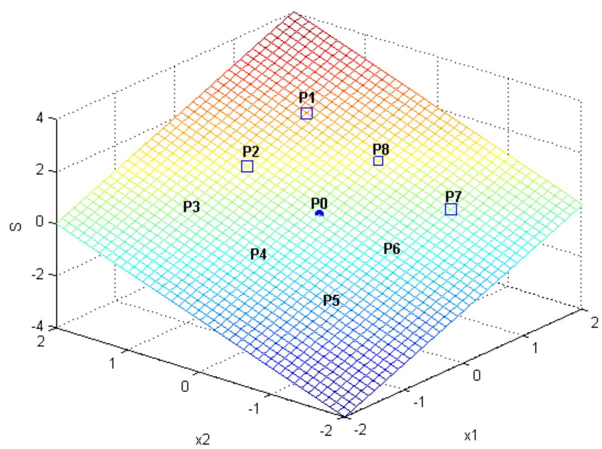

Figure 3. Two dimension linear test function

Observation points are located at $P_i(x_1,x_2)$=[P1*(1,1)*,P2*(0,1)*, P3*(-1,1)*, P4*(-1,0)*, P5*(-1,-1)*, P6*(0,-1)*, P7*(1,-1)*, P8*(1,0)*], and the prediction point is located at $P_0(x_1,x_2)$=[*(0,0)*]. To find out how the three models perform as the variance levels at the different locations increase, several different noise level scenarios for $\sigma_{\varepsilon i}^2$, i=1,…,8 (shown in Table 4) are tested. The results are given in Table 5.

Table 4. Noise level scenarios for two dimension linear test function

Scenarios	$\sigma_{\varepsilon 1}^2$	$\sigma_{\varepsilon 2}^2$	$\sigma_{\varepsilon 3}^2$	$\sigma_{\varepsilon 4}^2$	$\sigma_{\varepsilon 5}^2$	$\sigma_{\varepsilon 6}^2$	$\sigma_{\varepsilon 7}^2$	$\sigma_{\varepsilon 8}^2$
L1	0.5	0.5	0.1	0.1	0.1	0.1	0.1	0.1
L2	0.5	0.5	0.5	0.5	0.1	0.1	0.1	0.1
L3	0.5	0.5	0.5	0.5	0.5	0.5	0.1	0.1

Table 5. Results for the two dimension linear test function

	Estimation based on noisy data																*Estimation based on noiseless data*	
	DK					*NK*					*MK*					θ^*	*ErrM*
	$\hat{\theta}$ Mean	$\hat{\theta}$ Var	ErrA	ErrN	ErrP	$\hat{\theta}$ Mean	$\hat{\theta}$ Var	ErrA	ErrN	ErrP	$\hat{\theta}$ Mean	$\hat{\theta}$ Var	ErrA	ErrN	ErrP		
L1	0.10	0.01	0.06	0.05	2E-4	0.09	4E-3	0.03	0.02	1E-4	0.04	2E-3	0.01	0.01	6E-5	1E-3	1E-14
L2	0.18	0.02	0.10	0.09	0.01	0.10	0.01	0.06	0.05	3E-3	0.07	0.01	0.04	0.03	1E-3		
L3	0.36	10.0	0.21	0.13	0.05	0.17	4.24	0.11	0.07	0.02	0.17	3.40	0.08	0.06	0.02		

Similar results are observed in Table 5. The modified nugget effect model performs better that the nugget effect model and deterministic kriging model in all the three scenarios. As the variance of the input random noise ε increases from L1 to L3, the overall prediction error and all the additional error components increase for all three models. The ErrP to ErrA ratio also increases from L1 to L3. In L1 where the variance of the random noise is relatively low, the ErrP is not significant considering the scale of ErrN. In L3 the ErrP cannot be ignored as it consists of about 20% of the ErrA.

5.3 Two Dimension Sinusoidal Function

The sinusoidal function is taken from Hussain et al. (2002). The plot of the function and design are given as:

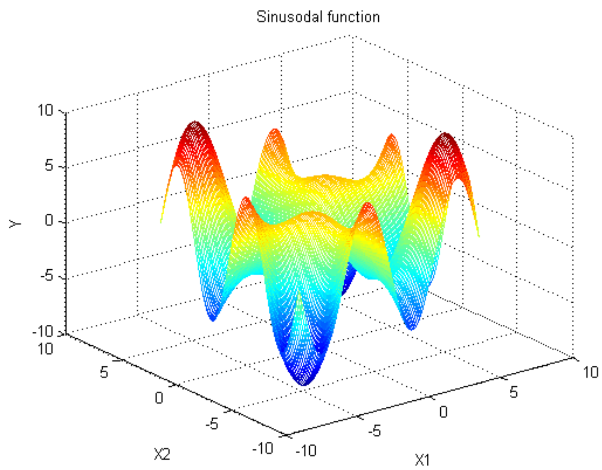

Figure 4. Two dimension sinusoidal test function

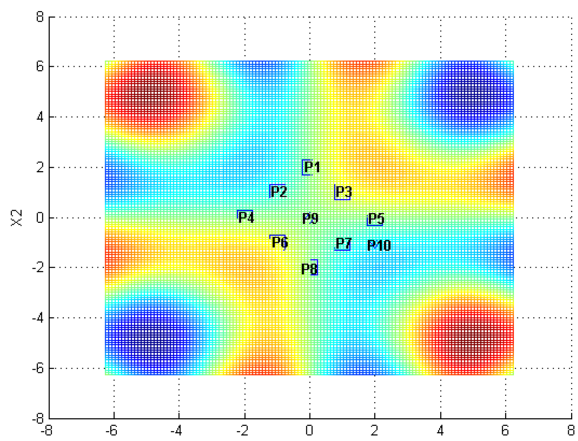

Figure 5. Design of the sinusoidal test function

The mathematical form of sinusoidal function is:

$$Y = x_1 \sin(x_2) + x_2 \sin(x_1) + \varepsilon$$

The additional noise component ε follows normal distribution with zero mean and variance $\sigma_\varepsilon^2 = x_1^2$. Observation points are located at $P_i(x_1,x_2)$=[P1(*0,2*),P2(*-1,1*),P3(*1,1*),P4(*-2,0*),P5(*2,0*),P6(*-1,-1*),P7(*1,-1*),P8(*0,-2*)], and prediction points are located at $P_j(x_1,x_2)$=[P9(*0,0*),P10(*2,-1*)]. The results are given in the following Table 6:

Table 6. Results for the two dimension sinusoidal test function

	Estimation based on noisy data															*Estimation based on noiseless data*	
	DK					*NK*					*MK*					θ^*	ErrM
	$\hat{\theta}$		ErrA	ErrN	ErrP	$\hat{\theta}$		ErrA	ErrN	ErrP	$\hat{\theta}$		ErrA	ErrN	ErrP		
	Mean	Var				Mean	Var				Mean	Var					
P9	118.2	5E3	0.199	0.191	0.005	49.81	4E3	0.192	0.180	0.004	16.26	2E3	0.148	0.129	0.002	14.39	0
P10			7.220	0.317	0.128			7.131	0.242	0.096			7.097	0.224	0.036		6.72

From the results in Table 6, we know that the correlation is rather weak as θ^* is large. For the interpolation at P9, the overall prediction error is low compared with the extrapolation case at P10. The prediction error caused by model misspecification in P10 is considerably larger than the one at P9 due to the lack of information for extrapolation. In this case, considering the more appropriate ErrP to ErrN ratio, we see the additional prediction error caused by parameter estimation uncertainty is significant in the extrapolation case.

Based on the results of the three numerical experiments, some conclusions can be made. The random noise ε inflates the overall prediction error as σ_ε^2 increases. The additional prediction error caused by parameter estimation uncertainty becomes more important when the correlation is weak. This is aligned with the results and conclusions observed in Zimmerman and Cressie (1992). Overall, the modified nugget effect model performs better than the nugget effect model and deterministic kriging model as observed in Yin, Ng and Ng (2008).

6 CONCLUSION

In this paper, we investigate the impact of parameter estimation uncertainty on three different kriging model forms in stochastic simulations. We analyzed theoretically a simple two-point problem and conducted three numerical studies. Based on the results of these studies, we find that the sensitivity parameter θ estimated by kriging model is affected by the random noise in the stochastic system, and the additional prediction error caused by parameter estimation uncertainty increases as the variance of the random noise ε increases. The proportion of the additional prediction error caused by parameter estimation uncertainty in overall prediction error increases when the variance of the random noise ε increases. In the case when the variability of the noise is low and sufficient sample data is available, this additional error becomes negligible. Among the three kriging model forms studied in this paper, the modified nugget effect model seems to have the best performance in both the overall prediction error and additional prediction error caused by parameter estimation uncertainty. This phenomenon is partially explained in the two-point problem.

In this paper, we assume that the covariance parameters σ_S^2 and σ_ε^2 are known or can be accurately estimated. In practice however, these parameters are likely to be estimated from the data too. In the next step of this research, σ_S^2 and σ_ε^2 should be considered in the parameter estimation problem. In this study also, the approximation used in (8) assumes that the predictor is a linear function of estimated parameters. This however may not hold in many cases and the approximation can be deteriorated. Further studies in alternative approximations can be done.

A APPENDIX

According to Harville and Jeske (1992), the *ij*th element of $A(\theta)$ is:

$$a_{ij}(\rho) = c_i^T Q c_j$$

where in this two-point problem,

$$c_i = \frac{\partial r(\rho)}{\partial \rho} - [rQ + \frac{F^T R(\rho)^{-1}}{F^T R(\rho)^{-1} F}]\frac{\partial R(\rho)}{\partial \rho}$$
$$Q = R^{-1}(\rho)[I - P(\rho)]$$
$$P(\rho) = F\frac{F^T R^{-1}(\rho)}{F^T R^{-1}(\rho) F}$$

As a result, we have:

$$A(\rho) = a_{11}(\rho) = \frac{1-\rho}{2}$$

REFERENCE

Abhishek, G., D. Yu, L. Xu, T. Reinikainen. 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.

Cressie, N. 1993. *Statistics for spatial data*. revised edition. New York, Wiley.

Harville , D. A. 1985. Decomposition of prediction error. *Journal of American Statistical Association*. 80: 132-138.

Harville, D.A. and D. R. Jeske. 1992. Mean squared error of estimation or prediction under a general linear model. *Journal of American Statistical Association*. 87: 724-731.

Hertog, D. d., J.P.C. Kleijnen and A.Y.D. Siem (2006). The correct kriging variance estimated by bootstrapping. *Journal of the Operational Research Society*: 57: 400-409

Hussain M. F., Barton R. R., Joshi S. B. 2002. Metamodeling: radial basis functions, versus polynomials. *European Journal of Operational Research*. 138(1): 142-154.

Kackar, R. N. and D.A. Harville. 1984. Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *Journal of American Statistic Association*. 79: 853-862.

Khatri, C.G. and K. R. Shah .1981. On the unbiased estimation of fixed effects in a mixed model for growth curves. *Communications in Statistics A-Theory and Methods*. 10: 401-406.

Matheron, G. 1963. Principle of geostatistics. *Economic Geology*. 58: 1246-1266.

Nelson, B. L., J. Staum and B. Ankenman. 2008. Stochastic kriging for simulation metamodeling. In *Proceedings of the 2008 Winter Simulation Conference*, ed. S. J. Mason, R.R. Hill, L. Mönch, O.Rose, T. Jefferson, J. W. Fowler, 362-370. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Pham, T. and M. Wagner. 1999. Filtering noisy images using kriging, *Signal Processing and Its Applications. In Proceedings of the Fifth International Symposium* .

Roshan, J. V. 2006. Limit kriging. *Technometrics*. 48(4): 458-466.

Reinsel, G. C. 1984. Estimation and prediction in a multivariate random effects generalized linear model. *Journal of American Statistical Association*. 79: 406-414.

Sacks, J., W.J. Welch, T.J. Mitchell and H.P. Wynn. 1989. Design and Analysis of Computer Experiments (with discussion). *Statistical Science*. 4: 409-435.

Toyooka, Y. 1976. Prediction error in a linear model with estimated parameters. *Biometrika*. 69: 453-459.

Wu, H. and F. Sun. 2007. Adaptive Kriging control of discrete-time nonlinear systems. *Control Theory & Applications, IET* . 1(3): 646-656.

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

Zimmerman, D. L. and N. Cressie. 1992. Mean squared error prediction error in the spatial linear model with estimated covariance parameters. *Annals of the institute of statistical mathematics*. 44(1): 27-43.

AUTHOR BIOGRAPHIES

JUN YIN is a Ph.D candidate in the Department of Industrial and Systems Engineering at the National University of Singapore. He received his B. Eng. in Automation from University of Science and Technology of China in 2005. His research interests include computer simulation, design of experiments and metamodeling. His email address is <yin-jun@nus.edu.sg>.

SZU HUI NG is an Assistant Professor in the Department of Industrial and Systems Engineering at the National University of Singapore. She holds B.S., M.S. and Ph.D. degrees in Industrial and Operations Engineering from the University of Michigan. Her research interests include quality and reliability modeling and analysis, design of experiments and simulation analysis. Her email and web addresses are <isensh@nus.edu.sg> and <http://www.ise.nus.edu.sg/staff/ngsh/index.html>.

KIEN MING NG is an Assistant Professor in the Department of Industrial and Systems Engineering at the National University of Singapore. He obtained his PhD in Management Science and Engineering from Stanford University in 2002. His research interests are in optimization and numerical algorithms, as well as operations research applications in logistics. He can be contacted by email at <isenkm@nus.edu.sg>.