Kriging Model with Modified Nugget Effect for Random Simulation with Heterogeneous Variances

J. Yin, S.H. Ng, K.M. Ng
Department of Industrial and Systems Engineering, National University of Singapore, Singapore

Abstract – **Random simulation with heterogeneous variance is a common phenomenon in real-world simulation cases. Most of the current meta-models assume homogeneous variances and hence are unable to provide satisfactory estimations of these models. This article presents a new kriging model with modified nugget effect, which accounts for heterogeneous variances in simulations. The new model can reduce the influence of noise with heterogeneous variance by incorporating local variance information into the model. A theoretical explanation and numerical experiment will be shown to provide clear illustration of this new modified model.**

Keywords - **heterogeneous variance, simulation, kriging, nugget effect**

I. INTRODUCTION

Simulation is commonly used in industrial scenarios as a more convenient way to study the system's characteristics and behaviors. As the complexity of simulation model increases, its inner structure is unknown and the behavior is unpredictable. The meta-modeling method can be used to approximate the simulation model's behavior. The spatial correlation model, also known as the kriging model, was first introduced into Design and Analysis of Computer Experiment (DACE) by [1-2]. It has been widely used in a variety of fields since then.

For deterministic problems, the kriging model is attractive for its interpolating characteristic, providing the predictions of the same values as the observations. For stochastic cases however, responses at the same location might be varied. The interpolation characteristic of kriging model in this case seems to be less desirable. In order to model the random fluctuation in stochastic cases, the nugget effect is introduced into the kriging model. According to [3], the traditional nugget effect in geo-statistic is caused by two factors: micro-scale variation and measurement error. However in this article, the random process studied is assumed to be L_2 –continuous, so it is purely caused by measurement error.

Most of the literatures applying the nugget effect assume that variance of random error is homogeneous only, [4-6]. However, there are many real world cases where the homoscedastic assumptions do not hold, like in queuing systems and networks etc. To handle heteroscedastic case with homogeneous model, certain preliminary adjustments are taken before the data is entered to build the model in [7]. This transforms the noise with heterogeneous variance into a noise with homogeneous variance. This preliminary process requires information on the signal function, which is the noiseless target function. In many real world simulations however, the signal function is unknown.

The kriging model with modified nugget effect proposed in this article is developed based on the kriging model with nugget effect to handle the heterogeneous variance situation. Initial kriging model assumptions need to be changed to incorporate the non-stationary behavior into the stationary model. Other aspects such as the MLE function and correlation matrix is also modified. A theoretical explanation and a numerical experiment will be provided to give a better view of this new modified model.

II. KRIGING MODEL REVISIT

The random simulation output can be treated as a random process Z(x), a function of the simulation input x. Typically, the mean response S(x) of the random process is of interest. According to reference [3], the random process can be decomposed as:

$$Z(x) = S(x) + \varepsilon(x) = \mu(x) + \delta(x) + \varepsilon(x) \qquad (1)$$

where S is the deterministic signal function; μ is the mean of the process, also known as the large-scale variation; δ is the bias between the signal function and mean, also known as the small-scale variation; ε stands for the random measure error (random noise).

There are several forms of kriging model. In this paper, we adopt the commonly used form, the ordinary kriging model. The ordinary kriging predictor at point x_0 is a linear combination of all the observation values:

$$P(z(x_0)) = \sum_{k=1}^{m} \lambda_k Z(x_k) \; with \; \sum_{i=1}^{m} \lambda_i = 1 \qquad (2)$$

and

$$\lambda_k = r^T R^{-1} e_k + 1^T R^{-1} e_k \frac{(1 - 1^T R^{-1} r)^T}{1^T R^{-1} 1} \qquad (3)$$

where $r = (corr(d_{01}), corr(d_{02}), ..., corr(d_{0m}))$, is the vector of the correlations between the point to be estimated and observed points; 1 stands for the vector of one with length of m; R is the matrix of all the correlations between any two observation points. The general form of the covariance function is given below:

$$C(d_{ij}) = \text{cov}(Z(x_i), Z(x_j)) = \begin{cases} c_0 + c_1 & d_{ij} = 0 \\ c_1 corr(d_{ij}) & d_{ij} \neq 0 \end{cases} \qquad (4)$$

978-1-4244-2630-0/08/$25.00 ©2008 IEEE

where d_{ij} is the Euclidean distance between point x_i and point x_j; c_0 is called nugget effect used to describe the random noise; c_1 is called partial sill; $corr(d_{ij})$ is the correlation function based on d_{ij}. As a result, the kriging weights are only dependent on distances.

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

$$MSE = E\{[P(Z(x_0)) - Z(x_0)]^2\} \quad (5)$$

The minimization result gives the optimal predictor:

$$P(Z(x_0)) = r^T R^{-1} Z + 1^T R^{-1} Z \frac{[1 - 1^T R^{-1} r]^T}{1^T R^{-1} 1} \quad (6)$$

where $Z=(Z(x_1),Z(x_2),\ldots,Z(x_m))$ is the observation vector. The minimal mean squared error (also known as kriging variance) is:

$$MSE(x_0) = c_1(1 - (r + 1\frac{1 - 1^T R^{-1} r}{1^T R^{-1} 1})^T R^{-1} r + \frac{1 - 1^T R^{-1} r}{1^T R^{-1} 1}) \quad (7)$$

III. PROPOSED KRIGING MODEL WITH MODIFIED NUGGET EFFECT

The weights of the kriging predictor are only dependent on the distances. It 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 uses the local variance information as an additional input variable. This penalizes the predictor at locations where the variance is high.

A. Kriging model for deterministic and stochastic case

From (2) and (3), we see that the kriging predictor is a function of observations $Z(x_k)$ and covariance function C. The general form of covariance function is given in (4). Under different circumstances, the covariance function may have different behaviors as the distance approaches zero, we consider three different models: deterministic model, nugget effect model (a homoscedastic model) and modified nugget effect model (a heteroscedastic model).

In the deterministic kriging model, the random noise is assumed to be 0, hence the nugget effect $c_0=0$. The correlation matrix R is given as:

$$R = \begin{pmatrix} 1 & corr(d_{12}) & \ldots & corr(d_{1(m-1)}) & corr(d_{1m}) \\ corr(d_{21}) & 1 & \ldots & corr(d_{2(m-1)}) & corr(d_{2m}) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ corr(d_{(m-1)1}) & corr(d_{(m-1)2}) & \ldots & 1 & corr(d_{(m-1)m}) \\ corr(d_{m1}) & corr(d_{m2}) & \ldots & corr(d_{m(m-1)}) & 1 \end{pmatrix} \quad (8)$$

The kriging variance is given in (7). This deterministic model is suitable for deterministic simulation case.

In stochastic case, the variance can be further divided into two categories: homoscedastic and heteroscedastic. The random noise in the nugget effect model is assumed to be a constant, hence the nugget effect c_0 is a constant which equals to the constant variance. The correlation matrix R for this model is:

$$R' = \begin{pmatrix} 1+\frac{c_0}{c_1} & corr(d_{12}) & \ldots & corr(d_{1(m-1)}) & corr(d_{1m}) \\ corr(d_{21}) & 1+\frac{c_0}{c_1} & \ldots & corr(d_{2(m-1)}) & corr(d_{2m}) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ corr(d_{(m-1)1}) & corr(d_{(m-1)2}) & \ldots & 1+\frac{c_0}{c_1} & corr(d_{(m-1)m}) \\ corr(d_{m1}) & corr(d_{m2}) & \ldots & corr(d_{m(m-1)}) & 1+\frac{c_0}{c_1} \end{pmatrix} \quad (9)$$

The kriging variance is then given as follows:

$$MSE(x_0) = c_0 + c_1(1 - (r + 1\frac{(1 - 1^T R^{-1} r)}{1^T R^{-1} 1})^T R^{-1} r + \frac{1 - 1^T R^{-1} r}{1^T R^{-1} 1}) \quad (10)$$

As a result, the nugget effect model can be used in homoscedastic case in which the random noise function is assumed to be a white-noise process.

In the heteroscedastic variance case, the variance of the random noise is different at different locations. In order to incorporate the heterogeneous variance into kriging model, we assume that the random noise is independent but not identical. The covariance function is then given as:

$$C(d_{ij}) = \mathrm{cov}(Z(x_i), Z(x_j)) = \begin{cases} c_i^* + c_1 & d_{ij} = 0 \\ c_1 corr(d_{ij}) & d_{ij} \neq 0 \end{cases} \quad (11)$$

where the c_i^* stands for the variance of the random error at location x_i. The only difference between (4) and (11) is the constant nugget effect c_0 becomes a variable c_i^*, which is dependent on location. Accordingly, the correlation matrix will become:

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

The kriging variance will be given as follows:

$$MSE(x_0) = c_0^* + c_1(1 - (r + 1\frac{(1 - 1^T R^{-1} r)}{1^T R^{-1} 1})^T R^{-1} r + \frac{1 - 1^T R^{-1} r}{1^T R^{-1} 1}) \quad (13)$$

As can be seen, the nugget effect model is used to handle homoscedastic case by adding a constant onto the diagonal of the correlation matrix of the deterministic model. While the model used for heteroscedastic case has a similar form to the kriging model with nugget effect, the term added onto the diagonal is a variable. We can rewrite the correlation matrix as:

$$R' = \begin{cases} R + \eta_c & \eta_c \textit{ is a diagonal matrix with constant } \eta_0 = c_0 / c_1 \\ R + \eta_v & \eta_c \textit{ is a diagonal matrix with variable } \eta_i = c_i^* / c_1 \end{cases}$$

Hence, we call this model the kriging model with modified nugget effect. In the next subsection, we will discuss the nugget effect's influence on parameter estimation.

B. Parameter Estimation

The correlation function corr(~) can have several different forms, see [8]. In this article, we use the Gaussian correlation function:

$$corr(d_{ij}) = \exp(-\theta d_{ij}^2) \quad (14)$$

where θ is the sensitivity parameter, which is usually estimated by MLE function in Design and Analysis of Computer Experiment (DACE).

According to [3], the MLE function for θ is:

$$\ell(\theta) = \frac{1}{2}\ln\det(R) + \frac{1}{2}(y - 1\mu)^T R^{-1}(y - 1\mu) \quad (15)$$

where R is the correlation matrix, which is a function of θ; y is the vector of all observations; and μ is the mean function.

Maximizing the MLE can provide a good estimator of θ, but in the stochastic case, the random noise can cause ill estimation. For example, in (15), when θ→∞, R→I and $\lim_{\theta\to\infty}\ell(\theta) \to \frac{m}{2}Var(y)$. For stochastic cases with high Var(y), the MLE function at the region where θ→∞ will also increase. This phenomenon will cause an ill estimation of θ. A test function is shown below to illustrate it better:

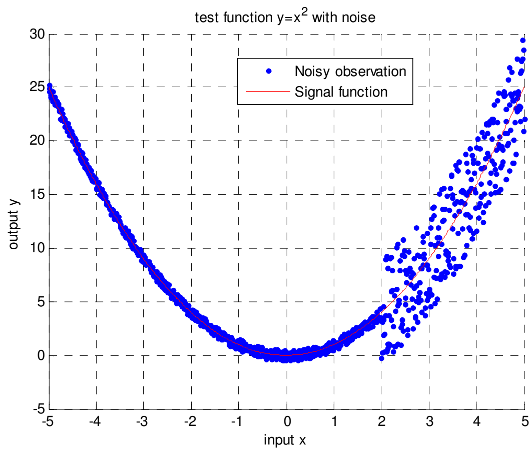

Fig. 1. Test function with step variance function

The test function used in Fig. 1 is a noisy second-order power function:

$$y = x^2 + \varepsilon \quad (16)$$

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

The plot of the MLE function for both of signal function only and noisy observation of signal function is shown below:

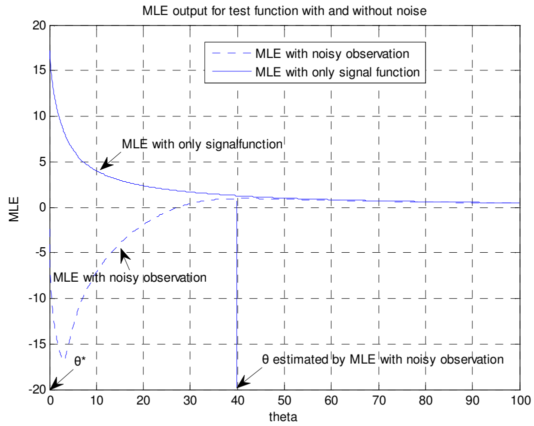

Fig. 2. MLE function for θ (signal function only and noisy observation)

We can see from Fig. 2, for the MLE function with signal function, the optimal θ is very close to 0, and we denote this optimal value as θ^*. For the MLE function with noisy observation, much of the function has been lifted as θ gets large. As a result, the MLE estimated θ is much larger than θ^*, and the output of the kriging predictor will oscillate.

When assuming the Kriging model with nugget model or modified nugget effect, this situation is improved:

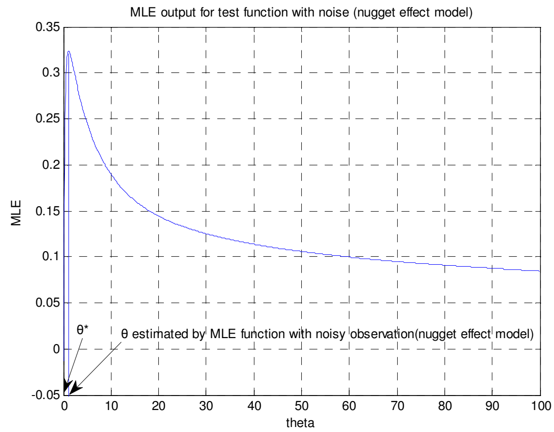

Fig. 3. MLE function for θ with nugget effect (noisy observation of the signal function)

As seen in Fig. 3, the lifted up portion is pulled back and the MLE function is corrected with a much smaller MLE estimate of θ. When θ→∞, R→I+η, the MLE function for θ with nugget effect will become:

$$\lim_{\theta\to\infty}\ell'(\theta) = \frac{m}{2}\ln(1+\eta_0) + \frac{m}{2(1+\eta_0)}Var(y) \quad (17)$$

In (17), the first component lifts the whole function because adding a constant onto the diagonal will increase the value of the determinant of R. The second component reduces the influence of Var(y). As a result, the abnormal MLE output in Fig. 2 (dash line) is corrected, resulting in a MLE output shown in Fig. 3. This form is similar to the MLE output with only signal function in Fig.2 (solid line).

When applying the kriging model with nugget effect in a heteroscedastic situation, there is no interpretation for selecting an appropriate nugget value. Other methods such as using the intercept of the variogram plot as an estimation cannot work when the signal function is not a constant, see [8]. As seen in (17), in order to fix the abnormal MLE function by reducing the influence of Var(y), the largest variance of the heterogeneous variance will be used as the nugget effect value for kriging meta-model with nugget effect.

Thus, in heteroscedastic example, both of the kriging model with nugget effect and kriging model with modified nugget effect can have a better estimation of **θ** than traditional 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 traditional kriging model.

C. Error Measurement

The error measurement used for kriging meta-model is MSE (Mean Squared Error), which is given as follows:

$$MSE(x_0) = \sigma_k^2(x_0) = E[(P(Z(x_0)) - Z(x_0))^2]$$

As can be seen, the MSE is used to measure the difference between the predictor and observation. However for stochastic case, when predicting at an unknown location like x_0, we want to know the actual signal function, say $S(x_0)$, but not the observation $Z(x_0)$ which is distorted by random noise.

According to [3], the kriging predictor will be:

$$P(S(x_0)) = \sum_{i=1}^{m} \xi_i Z(x_i) \tag{18}$$

Compared with (2), the only difference is the kriging weight. The predictor is still based on all the observations, say $Z(x_i)$, i=1,2,…,m.

The notation MSE_S is used for the MSE with respect to signal function S(x). According to [3], it can be given as:

$$\begin{aligned} MSE_S &= \tau_k^2(x_0) = E[(P(S(x_0)) - S(x_0))^2] \\ &= c_1(1-(r+1\frac{(1-1^T R'^{-1} r)}{1^T R'^{-1} 1})^T R'^{-1} r + \frac{1-1^T R'^{-1} r}{1^T R'^{-1} 1}) - \sigma^2_{\varepsilon(x_0)} \end{aligned} \tag{19}$$

where for homoscedastic model, $R' = R + \eta_c$, $\sigma^2_{\varepsilon(x_0)} = c_0$; for heteroscedastic model, $R' = R + \eta_v$, $\sigma^2_{\varepsilon(x_0)} = c_0^*$.

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. But in heteroscedastic case, one limitation is that the variance of the random error is needed. Since we assumed that the random error at every prediction point is independent but not identical, the variance information is not available unless the location is observed. Then comparing the performances of the kriging meta-model with nugget effect and kriging meta-model 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 parts. For the areas in between the observations, the prediction is dependent on the prediction at the observations and the sensitivity parameter θ. For comparison purpose, we fix the sensitivity parameter θ in both models to be the θ estimated by MLE function with nugget effect.

D. MSE_S Comparison for Observed Points

From (19), the MSE_S at the k*th* observation point for the kriging model with modified nugget effect is given as:

$$\begin{aligned} MSPE_S(x_k) &= c_1[1-(r+1\frac{(1-1^T R'^{-1} r)}{1^T R'^{-1} 1})^T R'^{-1} r + \frac{(1-1^T R'^{-1} r)}{1^T R'^{-1} 1}] - c_k^* \\ &= c_1[1-(r+1\frac{1^T R^{-1}(R^{-1}+\eta^{-1})^{-1} R^{-1} r}{1^T R'^{-1} 1})^T R'^{-1} r + \frac{1^T R^{-1}(R^{-1}+\eta^{-1})^{-1} R^{-1} r}{1^T R'^{-1} 1}] - c_k^* \\ &= c_1[1-(r+1\frac{1^T R'^{-1} \eta e_k}{1^T R'^{-1} 1})^T R'^{-1} r + \frac{1^T R'^{-1} \eta e_k}{1^T R'^{-1} 1}] - c_k^* \end{aligned}$$

where $e_k = [0,0,\ldots,\underbrace{1}_{the\ kth\ element},\ldots,0,0]$

If we set $\Delta_n = 1^T R'^{-1} 1$, which indicates the summation of all the elements in the inverse correlation matrix. And Δ_{nk} represents the summation of the *kth* column or row. Then we have:

$$\begin{aligned} MSPE_S(x_k) &= c_1[1-(r+1\frac{\eta_k \Delta_{mk}}{\Delta_m})^T R'^{-1} r + \frac{\eta_k \Delta_{mk}}{\Delta_m}] - c_k^* \\ &= c_1[1-1-\sum_{i=1}^{m} r_i \eta_k (\frac{\Delta_{mk}\Delta_{mi}}{\Delta_m} - \Delta_{mi}) + \frac{\eta_k \Delta_{mk}}{\Delta_m}] - c_k^* \\ &= c_1[-\eta_k(\frac{\Delta_{mk}}{\Delta_m} - 1)\sum_{i=1}^{m} r_i \Delta_{mi} + \frac{\eta_k \Delta_{mk}}{\Delta_m}] - c_k^* \\ &= c_1[-\eta_k(\frac{\Delta_{mk}}{\Delta_m} - 1)(1-\eta_k \Delta_{mk}) + \frac{\eta_k \Delta_{mk}}{\Delta_m}] - c_k^* \\ &= c_1[-\eta_k(\frac{\Delta_{mk}}{\Delta_m} - 1 - \eta_k \frac{\Delta_{mk}^2}{\Delta_m} + \eta_k \Delta_k) + \frac{\eta_k \Delta_{mk}}{\Delta_m}] - c_k^* \\ &= c_1[\eta_k + \eta_k^2 \frac{\Delta_{mk}^2}{\Delta_m} - \eta_k^2 \Delta_{mk}] - c_k^* \end{aligned}$$

For modified nugget effect model, $\eta_k = c_k^* / c_1$, so:

$$MSE_S(x_k) = c_1(\eta_k^2 \frac{\Delta_{mk}^2}{\Delta_m} - \eta_k^2 \Delta_{mk}) \tag{20}$$

Similarly, for kriging meta-model with nugget effect, the MSE_S at the k*th* observation point:

$$MSE_S(x_k) = c_1(\eta_0^2 \frac{\Delta_{nk}^2}{\Delta_n} - \eta_0^2 \Delta_{nk}) \tag{21}$$

Compared to the difference between η_k^2 and η_0^2, the difference between Δ_m and Δ_n and the difference between Δ_{mk} and Δ_{nk} are not very significant. Hence, we assume that the Δ stays approximately the same for both nugget effect model and modified nugget effect model. Under this assumption, when the nugget value c_0 exceeds $l_0 = \sum_{k=1}^{m}[c_k^{*2}(\frac{\Delta_k^2}{\Delta} - \Delta_k)] / \sum_{k=1}^{m}(\frac{\Delta_k^2}{\Delta} - \Delta_k)$, the modified nugget effect model outperforms the nugget effect model.

For most of the situations, in order to have a better estimation of the θ, a larger nugget effect is desired, as seen in (17). Hence, it is often likely the selected c_0 will be larger than the l_0.

E. Variance of the New Predictor

As the variances at unobserved new predictor points are unknown, the sub-sampling technique can be used to generate an estimate of the predictor's variance:

$$\mathrm{var}(P(Z(x_0))) = \sum_{i=1}^{s} \frac{(P_i(Z(x_0)) - \bar{P}(Z(x_0)))^2}{s-1} \quad (22)$$

According to [9], a reasonable design is: n-b+1 sub-samples of the form $\{X_i, X_{i+1}, \ldots, X_{i+b-1}\}$, and n=100 and b=10 in our experiment. Then s=100-10+1=91 groups of sub-samples will be generated. The M/M/1 signal function y=x/(1-x) is used as the test function, and the result is given as follows:

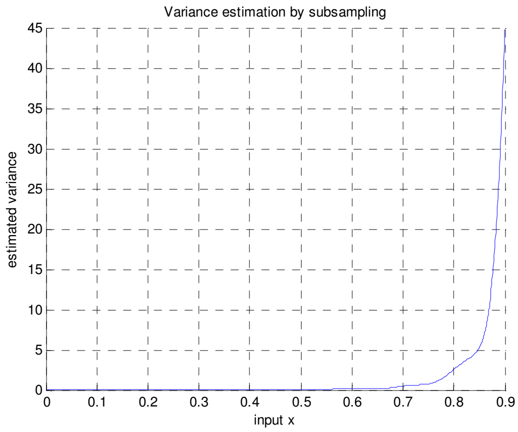

Fig. 4 Subsampling variance for kriging model with modified nugget effect

This will then give us an estimate of the predictor variance at every input level.

IV. NUMERICAL EXPERIMENTS

A. Experiment with noisy function

The noisy test function in part III will be used in this section for comparison purposes.

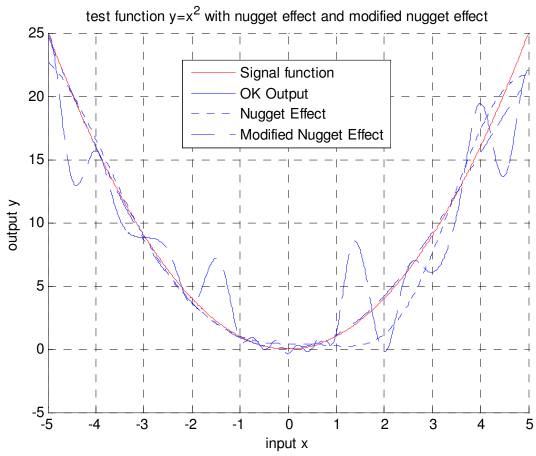

Fig.5 Ordinary Kriging, nugget effect and modified nugget effect output for test function

From Fig.5, we can see that the modified nugget effect model outperforms other models for this test function. Detailed results are given in Table I. There are five different models: ordinary kriging model, nugget effect model with minimum nugget (smallest variance of heterogeneous variance) and maximum nugget (largest variance of heterogeneous variance), modified nugget effect model with fixed θ and optimal θ.

TABLE I
DIFFERENT ERROR MEASURE OF DIFFERENT MODELS

Metamodels	MSE_S	MSE
Ordinary Kriging	2.5531	7.9372
Nugget effect(min variance)	2.4442	3.6441
Nugget effect(max variance)	1.5548	1.4346
Modified nugget effect	0.7409	0.4652
Modified nugget effect(optimal)	0.5983	0.3844

V. CONCLUSION

Finally, the kriging model with modified nugget effect proposed in this paper improves the kriging model's performance in heteroscedastic case by reducing the error between the predictor's output and signal function. It performs well in several experiments, including the M/M/1 queue. For further development, a better variance estimation method and more complete set of experiment is needed. Extension to non-stationary correlation function is also a potential research direction.

REFERENCES

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

[2] Sacks, J., Schiller, S.B., and Welch, W. (1989), Design for Computer Experiments, *Technometrics* 31, 41-47.

[3] Cressie, N. (1993) *Statistics for Spatial Data*, Wiley, New York, NY.

[4] Van Beers, W.C.M. ; Kleijnen, J.P.C. (2003) Kriging for interpolation in random simulation
Source: *Journal of the Operational Research Society*, v 54, n 3, p 255-62

[5] Huang D., Allen T.T., Notz W.I., Zeng N. (2006) Global optimization of stochastic black-box systems via sequential kriging meta-models. *Journal of Global Optimization*, v 34, n 3, p 441-466

[6] MATIAS José M. ; GONZALEZ-MANTEIGA Wenceslao (2003); Regularized kriging: The Support Vectors method applied to kriging; Artificial neural networks and neural information processing - ICANN / ICONIP 2003 Joint international conference, Istanbul,
TURQUIE, vol. 2714, pp. 209-216

[7] Kleijnen JPC and Wim C. M. van Beers. (2004) Robustness of kriging when interpolating in random simulation with heterogeneous variances: Some experiments. Volume 165, Issue 3, Pages 826-834

[8] Van Beers, W.C.M. ; Kleijnen, J.P.C. (2003) Kriging for interpolation in random simulation,Source: *Journal of the Operational Research Society*, v 54, n 3, p 255-62

[9] Dimitris N. Politis, Joseph P.Romano and Michael Wolf (1999) Subsampling, Springer