10th World Congress on Structural and Multidisciplinary Optimization
May 19 -24, 2013, Orlando, Florida, USA

A Study on Improving the Accuracy of Kriging Models by Using Correlation Model/Mean Structure Selection and Penalized Log-Likelihood Function

Hyeongjin Song[1], KK. Choi[1] and David Lamb[2]

[1] The University of Iowa, Iowa City, Iowa, USA, kkchoi@engineering.uiowa.edu
[2] US Army RDECOM/TARDEC, Warren, Michigan, USA

1. Abstract
In this paper, four schemes are presented to improve the accuracy of the Kriging model. The sampling-based methods are used in reliability-based design optimization (RBDO) if the sensitivity information is not available. In sampling-based RBDO, true models are approximated by using surrogates, since direct reliability estimation requires evaluations at a very large number of sample locations and true samples are computationally expensive in most practical design applications. Since the Kriging model is accurate compared to other surrogate methods, it is widely used not only in geostatistics but also in engineering applications. The Kriging model is based on a Gaussian random process, which includes a spatial correlation function. To identify the best correlation function parameter for the given correlation function, maximum likelihood estimation (MLE) is widely applied. However, MLE can be inaccurate when the sample size is small and the selected spatial correlation length (Kriging parameter) is smaller than distances between samples. In such situations, the estimated Kriging parameters may yield either misleading results or flat predictions. Therefore, in the literature, a penalized likelihood function is introduced to avoid such a wrong result by using a penalty function and cross-validation error. The authors of the original penalized likelihood function recommended using the smoothly clipped absolute deviation (SCAD) penalty function along with grid samples for the parameter λ of the penalty function. The appropriate grid size for the parameter λ is dependent on the problem at hand. A small grid size is usually more stable and gives better performance, but it increases the number of cross-validation (CV) errors to estimate. On the other hand, when the sample size is relatively large, the effect of the penalty function is limited. Therefore, a generalized pattern search (GPS) algorithm is applied to find better λ starting from $\lambda = 0$, where the penalty function is zero. At the same time, some correlation function types are better than other correlation function types for describing the correlation structure of given data. In engineering applications, the Gaussian correlation function is often used, since they are infinitely differentiable and provide a relatively smooth surface. However, examples show that other correlation function types could yield better results. To select the best correlation function, MLE can be applied. In the dynamic Kriging method, better mean structure is selected based on the process variance of the Kriging model. However, a wrong correlation model could be selected. If CV is used instead of the process variance to select a better mean structure, the performance is improved. The proposed methods are applied to mathematical and engineering examples to demonstrate effectiveness.
2. Keywords: Kriging, Correlation Model, Penalized MLE, Mean Structure, Dynamic Kriging

3. Introduction
Deterministic computer simulations of physical phenomena have been widely used in science and engineering for design guidance; and as more sophisticated and larger size models are used, they become computationally expensive. To design engineering products, accurate reliability analysis is necessary. Otherwise, an unreliable or overly conservative design can be obtained. When accurate sensitivity information of performance functions is available, reliability-based design optimization (RBDO) using the most probable point (MPP) has been applied to many engineering problems [1-11]. However, the sensitivity is often not available or is difficult to obtain accurately in complex problems such as multi-physics or multidisciplinary design/structural optimization. Without sensitivity, the MPP-based reliability analysis cannot be used and, alternatively, direct numerical probability integration can be performed using Monte Carlo simulation (MCS). However, this method is impractical when computer experiments are computationally expensive because it requires a large number of performance function evaluations, especially for small probability of failures.
Therefore, surrogate models are introduced to reduce the computational burden for solving problems without sensitivity. The Kriging method is one of the widely used surrogate modeling methods [12-18]. One advantage of the Kriging method is that it is an interpolation method and not a regression method. Thus, this method reproduces the same responses at given sample locations, and it is appropriate to approximate deterministic computer experiments. Another advantage is that it provides uncertainty information at the prediction point, which has motivated a number of adaptive sampling methods [12, 13]. To construct an accurate Kriging model, an

appropriate form of the Kriging model should be selected and the parameters should be estimated accurately. In geostatistics, a sample variogram is usually plotted first from given data [19-21]. Based on the variogram, an appropriate correlation model is selected and parameters are estimated. However, this process requires users' knowledge on the Kriging method. Therefore, the correlation model is usually fixed, and the maximum likelihood estimation (MLE) approach is applied to estimate parameters in many engineering applications [22-26]. Cross-validation can be an alternative to MLE, but MLE usually outperforms CV [22, 27].

Zhao et al. showed the importance of parameter estimation in the Kriging model, and the generalized pattern search (GPS) algorithm was used to find the optimal parameter in MLE [26, 28]. They found that GPS performed better than Hooke and Jeeves (H-J) [29], Levenberg-Marquardt (L-M) [24], or genetic algorithm (GA) [26]. However, performances of GPS are also influenced by initial parameter values. Therefore, GA is introduced to find better initial parameters for GPS.

In many engineering applications, the Gaussian correlation function is the most commonly used spatial correlation function (SCF) since it provides a relatively smooth and infinitely differentiable surface, which can be beneficial for gradient-based optimization algorithms [22-26]. However, there are many different data structures, and thus the fixed correlation model may not be able to describe the given data well [19, 21, 29, 30]. MLE is used for identifying the best correlation model among seven different SCF types in the literature [29].

When enough samples are not provided, performance of MLE is often not satisfactory due to inaccurate log-likelihood function. To overcome this problem, penalized MLE (PMLE) is introduced, which includes the penalty function [31]. However, PMLE is computationally more expensive due to CV estimations and is not always more accurate than MLE. Thus, an appropriate condition should be applied for the usage of PMLE, and PMLE is only applied for problems with a small sample size.

Zhao et al. also showed that the accuracy of the Kriging method can be enhanced by selecting appropriate basis functions instead of using all possible basis functions [26]. Even though their original idea is noble, the process variance is not appropriate for selecting better mean structures. In this paper, leave-one-out cross-validation (LOOCV) is used instead of the process variance. In this paper these proposed improvements are combined and implemented to test.

Section 4 briefly reviews the formulation of the Kriging method, and Section 5 explains different parameter estimation methods. In Section 6, correlation models and mean structures in the Kriging model are selected, and a combined method, which includes all improvement schemes, is proposed. Performances of proposed methods for numerical and engineering examples are shown in Section 7. Section 8 concludes the current study and discusses its limitations and future research.

4. Kriging Model

In a Kriging model, there are two parts, and the mathematical form can be expressed as [14],

$$\begin{aligned} y(\mathbf{x}) &= \sum_{i=1}^{k} \beta_i f_i(\mathbf{x}) + Z(\mathbf{x}) \\ &= \mathbf{f}^T \boldsymbol{\beta} + Z \end{aligned} \tag{1}$$

where k is the number of basis functions, f_i is the i^{th} basis function, β_i is the i^{th} coefficient, and $\mathbf{x}$ is the prediction location. Z is a model of a Gaussian random process with zero mean and covariance:

$$V(\mathbf{x}_1, \mathbf{x}_2) = \sigma^2 \mathbf{R}(\boldsymbol{\theta}, \mathbf{x}_1, \mathbf{x}_2) \tag{2}$$

where σ^2 is the process variance, $\mathbf{R}$ is the SCF, $\boldsymbol{\theta}$ is the correlation function parameter, and $\mathbf{x}_1$ and $\mathbf{x}_2$ are two sample sites.

The universal Kriging (UKG) method is defined with a set of basis functions:

$$\mathbf{f}(\mathbf{x}) = \{f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_k(\mathbf{x})\}^T \tag{3}$$

The ordinary Kriging (OKG) method is a special case of universal Kriging when its basis function is expressed as

$$\mathbf{f}(\mathbf{x}) = \{1\}^T \tag{4}$$

The Kriging prediction is expressed as

$$\hat{y}(\mathbf{x}) = \mathbf{f}^T(\mathbf{x})\boldsymbol{\beta} + \mathbf{r}^T(\mathbf{x})\mathbf{R}^{-1}(\mathbf{Y} - \mathbf{F}\boldsymbol{\beta}) \tag{5}$$

where $\mathbf{Y}$ is true responses at known samples, $\mathbf{r}(\mathbf{x})=\{\mathbf{R}(\boldsymbol{\theta}, \mathbf{x}, \mathbf{x}_1), \mathbf{R}(\boldsymbol{\theta}, \mathbf{x}, \mathbf{x}_2), \ldots, \mathbf{R}(\boldsymbol{\theta}, \mathbf{x}, \mathbf{x}_n)\}^T$; n is the number of samples; $\mathbf{F}= \{\mathbf{f}(\mathbf{x}_1), \mathbf{f}(\mathbf{x}_2), \ldots, \mathbf{f}(\mathbf{x}_n)\}^T$; and the least-squares estimate of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}}(\mathbf{x}) = (\mathbf{F}^T\mathbf{R}^{-1}\mathbf{F})^{-1}\mathbf{F}^T\mathbf{R}^{-1}\mathbf{Y} \tag{6}$$

The SCF can have different forms, and seven different correlation functions in the literature [29] are considered, as shown in Table 1.

Table 1: Correlation functions

Name	$R_j(\theta_j, x_1, x_2)$
Exponential	$\exp(-\theta_j \lvert x_2 - x_1 \rvert)$
General Exponential	$\exp(-\theta_j \lvert x_2 - x_1 \rvert^{\theta_{n+1}}),\ 0 < \theta_{n+1} \le 2$
Gaussian	$\exp(-\theta_j \lvert x_2 - x_1 \rvert^2)$
Linear	$\max\{0, 1-\theta_j \lvert x_2 - x_1 \rvert\}$
Spherical	$1-1.5\xi_j + 0.5\xi_j^3,\ \xi_j = \min\{1,\ \theta_j \lvert x_2 - x_1 \rvert\}$
Cubic	$1-3\xi_j^2 + 2\xi_j^3, \qquad \xi_j = \min\{1,\ \theta_j \lvert x_2 - x_1 \rvert\}$
Spline	$\begin{cases} 1-15\xi_j^2 + 30\xi_j^3, & \text{for } 0 \le \xi_j \le 0.2 \\ 1.25(1-\xi_j)^3, & \text{for } 0 < \xi_j < 1 \\ 0, & \text{for } \xi_j \ge 1, \end{cases}$ where $\xi_j = \theta_j \lvert x_2 - x_1 \rvert$

5. Parameter Estimation

Parameter estimation is the process of selecting the regression function coefficient, process variance, and correlation function parameters $\boldsymbol{\gamma} = \{\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}\}$. To estimate accurate parameters, various statistical techniques are available, such as the variographic analysis (VA), Bayesian estimation (BE), CV, or MLE [18, 22]. In geostatistics, the empirical variogram is usually obtained first from the autocovariance structure of the data [19-21]. However, this methodology requires users' interaction, so it is very sophisticated and not appropriate for our purpose. With the BE technique, prior information is needed, which might not be available [18]. In computer experiments, the MLE technique is most widely used, and Martin and Simpson [22] showed that the MLE method outperformed the CV technique. Therefore, the MLE technique will be mainly used.

5.1. Maximum Likelihood Estimation (MLE)

Under the Gaussian assumption, the log-likelihood of the model parameters is defined as

$$l = -\frac{n}{2}\ln[2\pi\sigma^2] - \frac{1}{2}\ln[|\mathbf{R}|] - \frac{1}{2\sigma^2}(\mathbf{Y}-\mathbf{F}\boldsymbol{\beta})^T\mathbf{R}^{-1}(\mathbf{Y}-\mathbf{F}\boldsymbol{\beta}) \tag{7}$$

By taking the derivative of Eq.(7) with respect to $\boldsymbol{\beta}$ and σ^2 and using Eq.(6), σ^2 is estimated as

$$\hat{\sigma}^2 = \frac{1}{n}(\mathbf{Y}-\mathbf{F}\hat{\boldsymbol{\beta}})^T\mathbf{R}^{-1}(\mathbf{Y}-\mathbf{F}\hat{\boldsymbol{\beta}}). \tag{8}$$

Then the log-likelihood function l also can be estimated. The goal of the MLE method is to find optimal $\boldsymbol{\theta}$ that maximizes the likelihood function based on all observations.

It is a global optimization problem, and many different optimization algorithms have been applied, such as the downhill simplex method, the Newton-Raphson method, the quasi-Newton method, the Fisher scoring algorithm, the adaptive simulated annealing, the GA, and the GPS method [23-26]. The first four gradient-based methods are local optimization methods, so they are not appropriate for highly nonlinear problems. Adaptive simulated annealing is much more computationally intensive because it is a Monte Carlo method [23, 25]. Zhao et al. showed that GPS performed better than H-J, L-M, or GA [17, 24, 26]. However, it is observed that the performance of GPS is influenced by its initial design point, and thus if the initial design is far from true optimum, GPS is computationally expensive and sometimes inaccurate. Therefore, GA can be used to find a better initial point for GPS.

5.2. Penalized MLE (PMLE)

The log-likelihood function near the optimum may be flat in some situations, or it can give wrong information. One way to solve this problem is to add a constraint $\theta < \theta_{max}$ in MLE, but it is another problem to choose the threshold θ_{max}. Therefore, PMLE is introduced for parameter estimation instead of MLE [21, 31-33]. In PMLE, a penalty function is added to the log-likelihood function, such as

$$Q = -\frac{n}{2}\ln[2\pi\sigma^2] - \frac{1}{2}\ln[|\mathbf{R}|] - \frac{1}{2\sigma^2}(\mathbf{Y}-\mathbf{F}\boldsymbol{\beta})^T\mathbf{R}^{-1}(\mathbf{Y}-\mathbf{F}\boldsymbol{\beta}) - n\sum_{i=1}^{d} p_\lambda(\gamma_i) \tag{9}$$

where n is the number of samples, d is the number of variables, p_λ is the penalty function, and λ is the parameter in the penalty function. Three penalty functions are introduced; they are the L_1 penalty, $p_\lambda(|\theta|)=\lambda|\theta|$; the L_2 penalty, $p_\lambda(|\theta|)=0.5\lambda|\theta|^2$; and the smoothly clipped absolute deviation (SCAD) penalty. The first derivative of the SCAD penalty is defined as

$$p'_\lambda(\theta) = \lambda\left\{I(\theta \le \lambda) + \frac{(a\lambda-\theta)_+}{(a-1)\lambda}I(\theta > \lambda)\right\} \tag{10}$$

where $a = 3.7$, $\theta > 0$, with $p_\lambda(0) = 0$ [31]. The SCAD penalty function is recommended in the literature [31], but other penalty functions showed similar performances with global optimization algorithms, so the L_1 penalty is applied in this paper. The amount of penalty is decided by the penalty function parameter λ. In the literature [31], a set of grid samples for λ are chosen, and the best λ is selected based on CV error. However, when exhaustive grid samples are used, it is not easy to decide the grid size and the parameter range for λ. If the grid size is too small, the computational cost becomes unnecessarily high. On the other hand, if the grid size is too large, the optimum may not be found. With grid samples, a large parameter range also increases the computational cost. Therefore, an optimization algorithm needs to be applied to enhance the efficiency. Local optimization methods such as a golden section method [34] or a gradient-based method are tested, but they often fail to find the optimum. Cross-validation error is not a smooth function and may have multiple local optima, so local optimization algorithms often fall into local optima. Therefore, GPS is applied to find optimal λ in PMLE.

6. Correlation Models and Mean Structures

6.1. Correlation Models

The choice of the correlation model is crucial to the Kriging model. Gaussian correlation models are widely applied for many engineering problems, since corresponding Kriging models are infinitely differentiable and smooth. However, the best correlation model can be different for different data or problems. If mean structures are the same for given data, the residual information is also the same. Then MLE can be used to select the best correlation function, since the correlation model, which maximizes the log-likelihood function, is more likely to describe the residual information well compared to other correlation models. Seven different correlation models are considered in Table 1.

6.2. Mean Structures

In DKG, the best mean structure is selected based on the process variance σ^2 in Eq. (8) [26]. However, DKG tends to choose the model with full basis functions for its mean structure, and chosen models are often not the best in terms of accuracy. This means that σ^2 may not be a good measure to select the better mean structure. Cross-validation is applied instead of σ^2 to find better mean structures. However, CV is computationally expensive, and it may not be sensitive enough to identify the difference among all basis function combinations. Therefore, the best mean structure is chosen among only OKG and first- and second-order UKG methods.

7. Numerical Experiments

7.1. Analytical Examples

To test the effectiveness of the proposed methods, a set of analytical functions are employed [5, 7, 35-37]. These are:

Iowa G2 function (2 variables)

$$\begin{aligned} y(\mathbf{x}) = {} & 1 + (0.9063\cdot x_1 + 0.4226\cdot x_2)^2 + (0.9063\cdot x_1 + 0.4226\cdot x_2 - 6)^3 \\ & -0.6(0.9063\cdot x_1 + 0.4226\cdot x_2)^4 - (-0.4226\cdot x_1 + 0.9063\cdot x_2) \\ & 0.01 \le x_1, x_2 \le 10 \end{aligned} \tag{11}$$

Branin-Hoo (2 variables)

$$y(\mathbf{x}) = \left(x_2 - \frac{5.1x_1^2}{4\pi^2} + \frac{5x_1}{\pi} - 6 \right)^2 + 10\left(1 - \frac{1}{8\pi}\right)\cos(x_1) + 10, \tag{12}$$
$$-5 \le x_1 \le 10, 0 \le x_2 \le 15.$$

Camelback (2 variables)

$$y(\mathbf{x}) = \left(\frac{x_1^4}{3} - 2.1x_1^2 + 4 \right)x_1^2 + x_1x_2 + (4x_2^2 - 4)x_2^2, \tag{13}$$
$$-3 \le x_1 \le 3, -2 \le x_2 \le 2.$$

Hartman 3 (3 variables)

$$y(\mathbf{x}) = -\sum_{i=1}^{q} a_i \exp\left(-\sum_{j=1}^{m} b_{ij}(x_j - d_{ij})2 \right),$$
$$0 \le x_j \le 1, m = 3, q = 4, \mathbf{a} = [1, 1.2, 3.0, 3.2],$$
$$\mathbf{b} = \begin{bmatrix} 3.0 & 10.0 & 30.0 \\ 0.1 & 10.0 & 35.0 \\ 3.0 & 10.0 & 30.0 \\ 0.1 & 10.0 & 35.0 \end{bmatrix}, \tag{14}$$
$$\mathbf{d} = \begin{bmatrix} 0.3689 & 0.1170 & 0.2673 \\ 0.4699 & 0.4387 & 0.7470 \\ 0.1091 & 0.8732 & 0.5547 \\ 0.03815 & 0.5743 & 0.8828 \end{bmatrix}$$

Hartman 6 (6 variables)

$$y(\mathbf{x}) = -\sum_{i=1}^{q} a_i \exp\left(-\sum_{j=1}^{m} b_{ij}(x_j - d_{ij})2 \right),$$
$$0 \le x_j \le 1, m = 6, q = 4, \mathbf{a} = [1, 1.2, 3.0, 3.2],$$
$$\mathbf{b} = \begin{bmatrix} 10.0 & 3.0 & 17.0 & 3.5 & 1.7 & 8.0 \\ 0.05 & 10.0 & 17.0 & 0.1 & 8.0 & 14.0 \\ 3.0 & 3.5 & 1.7 & 10.0 & 17.0 & 8.0 \\ 17.0 & 8.0 & 0.05 & 10.0 & 0.1 & 14.0 \end{bmatrix}, \tag{15}$$
$$\mathbf{d} = \begin{bmatrix} 0.1312 & 0.1696 & 0.5569 & 0.0124 & 0.8283 & 0.5886 \\ 0.2329 & 0.4135 & 0.8307 & 0.3736 & 0.1004 & 0.9991 \\ 0.2348 & 0.1451 & 0.3522 & 0.2883 & 0.3047 & 0.6650 \\ 0.4047 & 0.8828 & 0.8732 & 0.5743 & 0.1091 & 0.0381 \end{bmatrix}$$

Extended Rosenbrock (9 variables)

$$y(\mathbf{x}) = \sum_{i=1}^{m-1} \left[(1 - x_i)^2 + 100(x_{i+1} - x_i^2)^2 \right], \tag{16}$$
$$-5 \le x_i \le 10, i = 1, 2, \cdots, m = 9.$$

Dixon-Price (12 variables)

$$y(\mathbf{x}) = (x_1 - 1)^2 + \sum_{i=2}^{m} i\left[2x_i^2 - x_{i-1}\right]^2,$$
$$-10 \le x_i \le 10, i = 1, 2, \cdots, m = 12. \tag{17}$$

Different numbers of samples are used depending on the number of variables in the examples. Sample locations are generated using Latin Centroidal Voronoi Tessellations (LCVT), which has better uniformity compared with Latin hypercube sampling (LHS) [38]. For the CV error, LOOCV is applied. For the Kriging model, DACE MATLAB toolbox is used [29].

7.2. Engineering Example

As described by He et al. [39], statistical analysis in irregular wave and uncertainty quantification in variable regular wave are presented for resistance and motions. There are two variables and twelve performance functions. One hundred twenty-nine true samples are evaluated using CFD simulations. Among the 129 samples, 9, 17, 33, and 65 samples are selected and used to construct Kriging models. The unused 120, 112, 96, and 64 samples, respectively, are used to evaluate the accuracy of surrogates.

Normalized root mean square error (NRMSE) is used to compare the accuracies of different surrogates and is defined as

$$NRMSE = \frac{\sqrt{\frac{1}{N_t}\sum_{i=1}^{N_t}\left[y_i - \hat{y}_i\right]^2}}{\max\left\{y_i\right\} - \min\left\{y_i\right\}} \tag{18}$$

where N_t is the number of test points, y_i is the i^{th} true response or the i^{th} CFD result, and $\hat{y}_i$ is the i^{th} prediction using Kriging models.

7.3 Results and Discussions

7.3.1. Parameter Estimation in MLE

The genetic algorithm is proposed to find the initial point for GPS in Section 5.1. Using mathematical examples, three optimization methods are compared. The mathematical examples are Iowa G2, Branin-Hoo, Camelback, Rosenbrock, and Dixon-Price. The correlation function is fixed as Gaussian.

Table 2: NRMSE with different parameter search algorithms

Examples	Iowa G2	Branin-Hoo	Camelback	Rosenbrock	Dixon-Price
No. of Samples	10	10	50	50	60
H-J	10.57%	9.95%	0.04%	12.37%	12.35%
GPS	10.55%	9.80%	0.04%	12.37%	12.35%
GPS using GA	5.85%	9.80%	0.04%	0.70%	1.13%

According to Table 2, GPS using GA performs better than H-J or GPS, especially for high-dimensional problems, even though the computational cost is about twice that of GPS.

7.3.2. PMLE

When MLE is inaccurate, PMLE can be used as described in Section 5.2. Five mathematical examples are used to compare the performances of MLE and PMLE.

Table 3: NRMSE of MLE and PMLE

Examples	Iowa G2	Branin-Hoo	Camelback	Rosenbrock	Dixon-Price
No. of Samples	10	10	10	20	30
MLE	12.08%	9.35%	15.83%	15.21%	13.02%
PMLE	9.13%	9.21%	11.56%	10.24%	10.59%

According to Table 3, PMLE performs better than MLE for a relatively small sample size.

Next, to find the best λ in PMLE, GPS is proposed instead of the grid sampling method. The grid sampling method, golden section search, and GPS are compared in five mathematical examples. The range of λ is [0, 10]. For the grid

sampling method, 100 grid samples are used, and thus 100 LOOCV evaluations are required. For golden section search and GPS, stopping tolerance is set to 10^{-4}, which is a lot smaller resolution than the grid sampling method.

Table 4: Parameter λ from different search algorithms

Examples	Iowa G2	Branin-Hoo	Camelback	Rosenbrock	Dixon-Price
No. of Samples	10	10	10	20	30
Grid Sampling	0.01	3.64	0.01	0.01	0.01
Golden Section	1.62	3.82	1.46	10.00	10.00
GPS	0.01	9.00	0.01	0.00	0.00

Golden section search required 19~25 LOOCV evaluations, and GPS required 23~42. Therefore, they are more efficient than the grid sampling method. However, golden section search often falls into local optimum in Table 4. Therefore, GPS is more efficient and accurate than other methods.
When the log-likelihood function is flat near the optimum or is misleading, PMLE performs better than MLE. However, PMLE performs worse than MLE for relatively large samples since the performance of PMLE is based on the CV error. Furthermore, PMLE is much more expensive than MLE, so application of PMLE should be limited to cases with small sample size. Maximum likelihood estimation has difficulty when the log-likelihood function is flat near the optimum or the optimum is on the upper bound. Therefore, PMLE is only applied when the optimum is located at the upper bound or the log-likelihood function value at the optimum is the same as that at the upper bound.

7.3.3. Correlation Function Selection
In Section 6.1, correlation function selection by MLE is proposed. Five mathematical examples are used for comparison tests.

Table 5: NRMSE for different correlation functions

	Gaussian	Cubic	Exponential	Spline	Linear	Spherical	General Exponential
Iowa G2	**0.04%**	10.49%	6.16%	1.75%	5.85%	5.91%	0.69%
Branin-Hoo	**0.58%**	8.90%	4.54%	1.28%	3.87%	4.31%	0.53%
Camelback	**0.04%**	10.55%	2.26%	0.76%	3.33%	2.29%	1.55%
Rosenbrock	12.35%	12.35%	12.35%	12.35%	12.35%	12.35%	**0.44%**
Dixon-Price	12.36%	12.36%	12.36%	12.36%	12.36%	12.36%	**0.73%**

- Bold numbers are the correlation functions chosen by MLE.

As shown in Table 5, performances are different with different correlation functions, and correct correlation functions are identified by MLE except for the Branin-Hoo example. However, the difference in NRMSE is very small (0.58% vs. 0.53%), which means that those two correlation functions are similar in their shapes. Overall, better correlation models are identified by MLE, and the performance is improved.

7.3.4. Mean Structure Selection
In Section 6.2, a CV-based mean structure selection method is proposed instead of σ^2-based DKG. To compare the two dynamic Kriging methods, Iowa G2, Branin-Hoo, Camelback, Rosenbrock, Dixon-Price, and CFD engineering problems are used. Two different sample sets are applied to mathematical examples, and four different samples sets are applied to engineering examples. Therefore, 58 cases are tested overall.

Table 6: Performances of σ^2-based and CV-based DKG

	Average NRMSE	Total Elapsed Time (Sec)
σ^2-based DKG	7.91%	784
CV-based DKG	6.85%	1,537

In Table 6, CV-based DKG is more accurate than σ^2-based DKG. Therefore, the process variance σ^2 does not seem to be a good measure for mean structure selection. However, CV-based DKG is very expensive since many different basis function combinations are examined. Furthermore, CV may not be accurate enough to identify the differences between all basis function combinations, especially with large sample size. With the new dynamic

Kriging method, the best model is chosen among the zeroth-, first- and second-order UKG based on CV error. The Hartman3 and Hartman6 functions are included in this test, so 62 cases are tested.

Table 7: Performances of different Kriging methods

Methods	0^{th} UKG	1^{st} UKG	2^{nd} UKG	CV-based DKG	New DKG
Average NRMSE	6.28%	6.55%	8.42%	6.45%	5.92%
Average Time (sec)	0.71	0.71	0.84	48.49	6.54

According to Table 7, the accuracy of CV-based DKG is similar to that of the zeroth- and first-order UKG, but it is far more expensive than the other methods. New DKG is the most accurate, and the computational cost is much less than CV-based DKG. Thus, even though CV-based DKG uses more combinations, CV does not seem to be accurate enough to identify all small differences between tested mean structures.

7.3.5. Combined Scheme

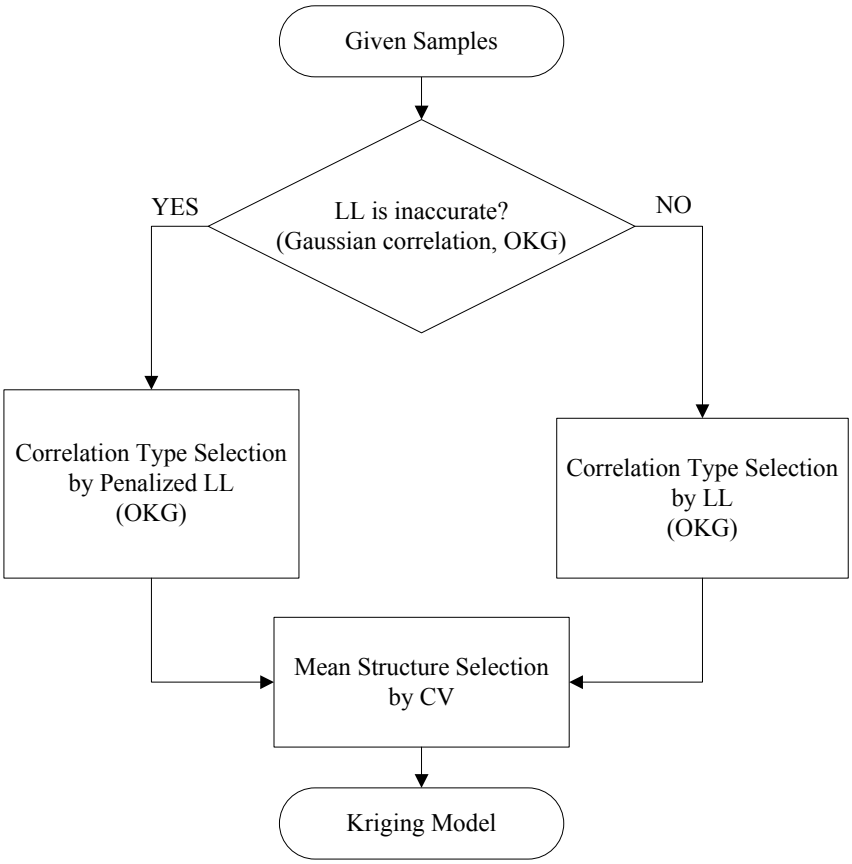

Figure 1: Flowchart of new dynamic Kriging (LL means log-likelihood function)

Four different schemes are introduced to improve the accuracy of Kriging. They are better parameter search in MLE, PMLE for small sample size, better correlation model selection, and better mean structure selection. They can be combined into a new dynamic Kriging process as shown in Fig. 1. For given samples, the shape of the log-likelihood function is examined to see if it has a flat region near the optimum. If the log-likelihood function is inaccurate, PMLE is used in correlation function selection. Otherwise, the best correlation function is identified by MLE. Once a correlation model is selected, the best mean structure is chosen based on CV error. Parameters are estimated at each stage, and all parameter estimations are done by GPS using GA.
With 62 tests, average NRMSE is 5.71%, but average elapsed time is 62.83 sec. It is quite expensive, so GPS is

applied instead of GPS using GA since multiple parameter estimations are performed within the new dynamic Kriging process. With the same 62 tests, average NRMSE is 5.77%, and average elapsed time is 24.88 sec. The accuracy became a bit worse, but it is more efficient.

8. Conclusions

The accuracy of the Kriging method is improved by employing four different methods – GPS using GA for parameter search in MLE, PMLE with small sample size, correlation model selection by MLE, and mean structure selection by CV error. Parameter estimation in PMLE is also improved. Each method is shown to improve the accuracy of the Kriging model. They are combined into one process, and the new dynamic Kriging method shows improved accuracy with mathematical and engineering examples.

However, there are several issues to be resolved. In new dynamic Kriging, PMLE is very conservatively applied. If the threshold condition when MLE is bad can be identified accurately, PMLE can contribute more. Only seven correlation function types are considered in this paper, but any other correlation functions can be considered too. However, more correlation function types mean more computational cost, so only distinctive correlation types are used. In mean structure selection, a more accurate identifier should be investigated. If the differences between different mean structures can be identified correctly, dynamic Kriging will become more effective. New DKG is computationally expensive, so efficiency schemes need to be investigated.

9. Acknowledgements

The research is supported by the Automotive Research Center, which is sponsored by the U.S. Army TARDEC. The research is also partially supported by the World Class University Program through the National Research Foundation of Korea (NRF) grant funded by the Ministry of Education, Science and Technology (Grant Number R32–2008–000–10161–0 in 2009). Professor Frederick Stern's group at The University of Iowa provided CFD sample information for numerical tests. These supports are greatly appreciated.

10. References

[1] A.M. Hasofer and N.C. Lind, Exact and invariant second-moment code format, *Journal of the Engineering Mechanics Division*, 100(1), 111-121, 1974.

[2] K. Breitung, Asymptotic approximations for multinormal integrals, *Journal of Engineering Mechanics*, 110(3), 357-366, 1984.

[3] M. Hohenbichler and R. Rackwitz, Improvement of second-order reliability estimates by importance sampling, *Journal of Engineering Mechanics*, 114(12), 2195-2199, 1988.

[4] O. Ditlevsen and H.O. Madsen, *Structural reliability methods*, Citeseer, 1996.

[5] J. Tu, K.K. Choi and Y.H. Park, A new study on reliability-based design optimization, *Journal of Mechanical Design*, 121(4), 557-564, 1999.

[6] A. Haldar and S. Mahadevan, *Probability, Reliability and Statistical Methods in Engineering Design*, John Wiley & Sons, New York, 2000.

[7] B. Youn, K.K. Choi and L. Du, Enriched performance measure approach for reliability-based design optimization, *AIAA Journal*, 43(4), 2005.

[8] H.O. Madsen, S. Krenk and N.C. Lind, *Methods of structural safety*, Courier Dover Publications, 2006.

[9] S. Rahman and D. Wei, A univariate approximation at most probable point for higher-order reliability analysis, *International Journal of Solids and Structures*, 43(9), 2820-2839, 2006.

[10] D. Wei, A univariate decomposition method for higher-order reliability analysis and design optimization, Ph. D. Thesis, The University of Iowa, 2006.

[11] I. Lee, K.K. Choi, L. Du and D. Gorsich, Inverse analysis method using MPP-based dimension reduction for reliability-based design optimization of nonlinear and multi-dimensional systems, *Computer Methods in Applied Mechanics and Engineering*, 198(1), 14-27, 2008.

[12] J. Sacks, W.J. Welch, T.J. Mitchell and H.P. Wynn, Design and analysis of computer experiments, *Statistical Science*, 4(4), 409-423, 1989.

[13] J. Sacks, S.B. Schiller and W.J. Welch, Designs for computer experiments, *Technometrics*, 31(1), 41-47, 1989.

[14] N.A.C. Cressie, *Statistics for Spatial Data*, John Wiley & Sons, New York, 1991.

[15] R.R. Barton, Metamodeling: a State of the Art Review, *WSC '94: Proceedings of the 26th conference on Winter Simulation*, Society for Computer Simulation International, San Diego, CA, USA, 1994 (0-7803-2109-X).

[16] T.W. Simpson, J.D. Poplinski, P.N. Koch and J.K. Allen, Metamodels for computer-based engineering design: Survey and recommendations, *Engineering with Computers*, 17 (2), 129-150, 2001.

[17] A.I.J. Forrester and A.J. Keane, Recent advances in surrogate-based optimization, *Progress in Aerospace Sciences*, 45(1-3), 50-79, 2009.

[18] V. Dubourg, B. Sudret and J. Bourinet, Reliability-based design optimization using kriging surrogates and subset simulation, *Structural and Multidisciplinary Optimization*, 44(5), 673-690, 2011.
[19] G. Bohling, Introduction to geostatistics and variogram analysis, *Kansas Geological Survey*, 2005.
[20] T. Hengl, *A practical guide to geostatistical mapping of environmental variables*, Joint Research Centre, Luxembourg, 2007 (978-92-79-06904-8).
[21] O. Roustant, D. Ginsbourger and Y. Deville, DiceKriging, DiceOptim: Two R packages for the analysis of computer experiments by kriging-based metamodeling and optimization, *Journal of Statistical Software*, 51(1), 2012.
[22] J.D. Martin and T.W. Simpson, Use of kriging models to approximate deterministic computer models, *AIAA Journal*, 43(4), 853-863, 2005.
[23] S.E. Gano, J.E. Renaud, J.D. Martin and T.W. Simpson, Update strategies for kriging models used in variable fidelity optimization, *Structural and Multidisciplinary Optimization*, 32(4), 287-298, 2006.
[24] J.D. Martin, Computational improvements to estimating kriging metamodel parameters, *Journal of Mechanical Design*, 131(8), 084501, 2009.
[25] H. Deng, W. Shao, Y. Ma and Z. Wei, Bayesian metamodeling for computer experiments using the Gaussian kriging models, *Quality and Reliability Engineering International*, 2011.
[26] L. Zhao, K.K. Choi and I. Lee, Metamodeling method using dynamic kriging for design optimization, *AIAA Journal*, 49(9), 2034-2046, 2011.
[27] P. Refaeilzadeh, L. Tang and H. Liu, Cross-validation, *Encyclopedia of Database Systems*, 532-538, 2009.
[28] R.M. Lewis and V. Torczon, Pattern search algorithms for bound constrained minimization, *SIAM Journal on Optimization*, 9(4), 1082-1099, 1999.
[29] S.N. Lophaven, H.B. Nielsen and J. Søndergaard, *DACE: A Matlab kriging toolbox*, Citeseer, 2002.
[30] J. Koehler and A. Owen, Computer experiments, *Handbook of Statistics*, 13(13), 261-308, 1996.
[31] R. Li and A. Sudjianto, Analysis of computer experiments using penalized likelihood in gaussian kriging models, *Technometrics*, 47(2), 111-120, 2005.
[32] J. Fan and R. Li, Variable selection via nonconcave penalized likelihood and its oracle properties, *Journal of the American Statistical Association*, 96(456), 1348-1360, 2001.
[33] S. Kok, The asymptotic behaviour of the maximum likelihood function of Kriging approximations using the Gaussian correlation function, *EngOpt 2012 - International Conference on Engineering Optimization*, Anonymous Rio de Janeiro, Brazil, 1-5 July 2012 .
[34] J.S. Arora, *Introduction to optimum design*, Elsevier/Academic Press, 2004 (9780120641550).
[35] L.C.W. Dixon and G.P. Szego, *Towards global optimisation 2*, North-Holland Pub. Co, 1978 (9780444851710).
[36] J. Lee, A novel three-phase trajectory informed search methodology for global optimization, *Journal of Global Optimization*, 38(1), 61-77, 2007.
[37] F.A. Viana, R.T. Haftka, V. Steffen Jr, S. Butkewitsch and M. F. Leal, Ensemble of surrogates: A framework based on minimization of integrated square error, *49th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, American Institute of Aeronautics and Astronautics, Schaumburg, IL, April 2008.
[38] Y. Saka, M. Gunzburger and J. Burkardt, Latinized, improved LHS, and CVT point sets in hypercubes, *International Journal of Numerical Analysis and Modeling*, 4(3-4), 729-743, 2007.
[39] W. He, M. Diez, D. Peri, E.F. Campana, Y. Tahara and F. Stern, Uncertainty quantification of delft catamaran total added resistance and motions in irregular and variable regular head wave, *29th Symposium on Naval Hydrodynamics*, Gothenburg, Sweden, August 2012 .