

Blind Kriging: A New Method for Developing Metamodels

V. Roshan Joseph¹

School of Industrial and Systems Engineering,
Georgia Institute of Technology, Atlanta, GA 30332

Ying Hung

School of Industrial and Systems Engineering,
Georgia Institute of Technology, Atlanta, GA 30332

Agus Sudjianto

Bank of America, Charlotte, NC 28255

Abstract

Kriging is a useful method for developing metamodels for product design optimization. The most popular kriging method, known as ordinary kriging, uses a constant mean in the model. In this article, a modified kriging method is proposed, which has an unknown mean model. Therefore it is called blind kriging. The unknown mean model is identified from experimental data using a Bayesian variable selection technique. Many examples are presented which show remarkable improvement in prediction using blind kriging over ordinary kriging. Moreover, blind kriging predictor is easier to interpret and seems to be more robust to misspecification in the correlation parameters.

KEY WORDS: Computer experiments, Design optimization, Cross validation, Finite element models, Kriging, Metamodels, Variable selection.

To appear in the *ASME Journal of Mechanical Design*

¹Corresponding Author. roshan@isye.gatech.edu. School of Industrial and Systems Engineering, Georgia Institute of Technology, 755 Ferst Drive NW, Atlanta, GA 30332, phone: 404 894 0056, fax: 404 894 2301.

1 Introduction

The use of computer modeling and experiments is becoming more and more popular for product design optimization [1]. Based on the physical knowledge of the product, models such as finite element models can be formulated and solved on computers. Although cheaper than experimenting on products or prototypes, computer experiments can still be time consuming and expensive. An approach to reduce the computational time and cost is to perform optimization on a metamodel that approximates the original computer model. The metamodel can be estimated from data by running the computer experiment on a sample of points in the region of interest.

Kriging is widely used for obtaining the metamodels [2, 3, 4]. For examples, [5] uses kriging for the thermal design of wearable computers and [6] uses kriging for the design of a variable thickness piezoelectric bimorph actuator. See [7, 8, 9, 10] for more examples. The popularity of kriging is due to the fact that computer models are often deterministic (i.e., no random error in the output) and thus interpolating metamodels are desirable. Kriging gives an interpolating metamodel and is therefore more suitable than the other common alternatives such as quadratic response surface model.

A kriging model, known as *universal kriging*, can be stated as follows [11]. Assume that the true function $y(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^p$, is a realization from a stochastic process

$$Y(\mathbf{x}) = \mu(\mathbf{x}) + Z(\mathbf{x}), \tag{1}$$

where $\mu(\mathbf{x}) = \sum_{i=0}^m \mu_i v_i(\mathbf{x})$ and $Z(\mathbf{x})$ is a weak stationary stochastic process with mean 0 and covariance function $\sigma^2 \psi$. The v_i 's are some known functions and μ_i 's are unknown parameters. Usually $v_0(\mathbf{x}) = 1$. The covariance function is defined as $cov\{Y(\mathbf{x}+\mathbf{h}), Y(\mathbf{x})\} = \sigma^2 \psi(\mathbf{h})$, where the correlation function $\psi(\mathbf{h})$ is a positive semidefinite function with $\psi(\mathbf{0}) = 1$ and $\psi(-\mathbf{h}) = \psi(\mathbf{h})$. In this formulation $\mu(\mathbf{x})$ is used to capture the known trends, so that $Z(\mathbf{x})$ will be a stationary process. But, in reality, rarely will those trends be known and thus the following special case, known as *ordinary kriging*, is commonly used [11, 12, 13],

$$Y(\mathbf{x}) = \mu_0 + Z(\mathbf{x}). \tag{2}$$

The metamodel (or the predictor) can be obtained as follows. Suppose we evaluated the function at n points $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ and let $\mathbf{y} = (y_1, \dots, y_n)'$ be the corresponding function values. Then ordinary kriging predictor is given by

$$\hat{y}(\mathbf{x}) = \hat{\mu}_0 + \boldsymbol{\psi}(\mathbf{x})' \boldsymbol{\Psi}^{-1}(\mathbf{y} - \hat{\mu}_0 \mathbf{1}), \quad (3)$$

where $\mathbf{1}$ is a column of 1's having length n , $\boldsymbol{\psi}(\mathbf{x})' = (\psi(\mathbf{x} - \mathbf{x}_1), \dots, \psi(\mathbf{x} - \mathbf{x}_n))$, $\boldsymbol{\Psi}$ is an $n \times n$ matrix with elements $\psi(\mathbf{x}_i - \mathbf{x}_j)$, and $\hat{\mu}_0 = \mathbf{1}' \boldsymbol{\Psi}^{-1} \mathbf{y} / \mathbf{1}' \boldsymbol{\Psi}^{-1} \mathbf{1}$. It is the best linear unbiased predictor, which minimizes the mean squared prediction error $E\{\hat{Y}(\mathbf{x}) - Y(\mathbf{x})\}^2$ under the model in Eq. (2).

The predictor in Eq. (3) is an interpolating predictor and is easy to evaluate. However, it has some problems. First, the prediction can be poor if there are some strong trends (see the simulation results in [14]). Second, it is not easy to understand the effects of the factors by just looking at the predictor. Of course, sensitivity analysis techniques such as the functional analysis of variance can be used for understanding and quantifying their effects [1], but the proposed predictor in this article is a much simpler alternative. Third, the predictor is not robust to the misspecification in the correlation parameters (see [15] for examples). In this article, we propose a modification of universal kriging predictor that overcomes the foregoing problems of ordinary kriging predictor.

2 Blind Kriging

We propose a simple modification to universal kriging model in Eq. (1). We do not assume the functions v_i 's to be known. Instead, they are identified through some data-analytic procedures. Because v_i 's are unknown in our model, we name it *blind kriging*. Thus, the blind kriging model is given by

$$Y(\mathbf{x}) = \mathbf{v}(\mathbf{x})' \boldsymbol{\mu}_m + Z(\mathbf{x}), \quad (4)$$

where $\mathbf{v}(\mathbf{x})' = (1, v_1, \dots, v_m)$, $\boldsymbol{\mu}_m = (\mu_0, \mu_1, \dots, \mu_m)'$, and m are unknown. Here $Z(\mathbf{x})$ is assumed to be a weak stationary stochastic process with mean 0 and covariance function

$\sigma_m^2 \psi$. The correlation function ψ can also depend on m , but for the moment assume it to be independent. The blind kriging predictor, which has the same form as that of universal kriging predictor, is given by

$$\hat{y}(\mathbf{x}) = \mathbf{v}(\mathbf{x})' \hat{\boldsymbol{\mu}}_m + \psi(\mathbf{x})' \boldsymbol{\Psi}^{-1}(\mathbf{y} - \mathbf{V}_m \hat{\boldsymbol{\mu}}_m), \quad (5)$$

where $\mathbf{V}_m = (\mathbf{v}(\mathbf{x}_1), \dots, \mathbf{v}(\mathbf{x}_n))'$ and $\hat{\boldsymbol{\mu}}_m = (\mathbf{V}_m' \boldsymbol{\Psi}^{-1} \mathbf{V}_m)^{-1} \mathbf{V}_m' \boldsymbol{\Psi}^{-1} \mathbf{y}$. Note that \mathbf{V}_m is an $n \times (m + 1)$ matrix.

The most important step in blind kriging is to identify the unknown functions v_i 's. They can be chosen from a set of candidate functions (or variables) using variable selection techniques. If some simple functions are used in the candidate set, then the predictor can be easily interpreted using the first part $\mathbf{v}(\mathbf{x})' \hat{\boldsymbol{\mu}}_m$. The second part of the predictor $\psi(\mathbf{x})' \boldsymbol{\Psi}^{-1}(\mathbf{y} - \mathbf{V}_m \hat{\boldsymbol{\mu}}_m)$ helps to achieve interpolation.

2.1 Variable Selection

There are many variable selection techniques that are popular in regression analysis such as forward selection, backward elimination, and step-wise regression [16]. Recently, many other techniques have also been proposed [17, 18, 19, 20]. All of these techniques have a drawback for using in the analysis of experiments and in particular for blind kriging. They do not lead to models that satisfy the well known principles of effect hierarchy and effect heredity [21]. The *effect hierarchy principle* states that lower order effects (such as main effects) are more important than higher order effects (such as two-factor interactions) and the *effect heredity principle* states that in order for an interaction effect to be significant, at least one of its parent factors should be significant. These principles are useful for identifying models that are simple and interpretable. Ref [22] introduced a Bayesian variable selection technique that incorporates these two principles. Another Bayesian variable selection technique introduced in [23, 24] seems to be more useful for our purpose because of its connections with kriging. It can be considered as a Bayesian version of the forward selection strategy. Below we explain this technique briefly. Additional details of the technique are included in the Appendix. We

note that the work in [23, 24] focus on physical experiments and therefore, the Bayesian variable selection technique was applied only to linear models and not kriging models.

The candidate variables are selected as the linear effects, quadratic effects, and two-factor interactions. Here the two-factor interactions include the linear-by-linear, linear-by-quadratic, quadratic-by-linear, and quadratic-by-quadratic interactions. There are a total of $t = 2p^2$ candidate variables (excluding the constant term). We note that this Bayesian variable selection technique can easily handle three and higher order effects, but in this article we focus on the lower order effects for the simplicity of exposition and interpretation. Following [24], first scale the factors in $[1.0, 3.0]$. Other ranges such as $[0, 1]$ or $[-1, 1]$ maybe used, however, Eqs (6) and (7) should be changed accordingly (see the Appendix). The linear and quadratic effects can be defined using the orthogonal polynomial coding [25]

$$x_{jl} = \frac{\sqrt{3}}{\sqrt{2}}(x_j - 2) \quad \text{and} \quad x_{jq} = \frac{1}{\sqrt{2}}(3(x_j - 2)^2 - 2), \quad (6)$$

for $j = 1, 2, \dots, p$. The variables x_{jl} and x_{jq} are scaled so that they have the same length $\sqrt{3}$ when x_j takes the values 1, 2, and 3. The two-factor interaction terms can be defined as the products of these variables. For example, the linear-by-quadratic interaction term between x_1 and x_3 can be defined as $x_{1l}x_{3q}$.

Denote the candidate variables by u_1, \dots, u_t . Consider approximating $y(\mathbf{x})$ by the linear model $\sum_{i=0}^m \mu_i v_i + \sum_{i=0}^t \beta_i u_i$, where $u_0 = 1$. As an example, for two factors x_1 and x_2 , the linear model is $\sum_{i=0}^m \mu_i v_i + \sum_{i=0}^8 \beta_i u_i$, where $u_0 = 1$, $u_1 = x_{1l}$, $u_2 = x_{1q}$, $u_3 = x_{2l}$, $u_4 = x_{2q}$, $u_5 = x_{1l}x_{2l}$, $u_6 = x_{1l}x_{2q}$, $u_7 = x_{1q}x_{2l}$, and $u_8 = x_{1q}x_{2q}$. Note that when $t > n - 1$, a frequentist estimation of the β_i 's is not possible. However, all of the t effects can be simultaneously estimated using a Bayesian approach. For doing this, we need to postulate a prior distribution for $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_t)'$. Let

$$\boldsymbol{\beta} \sim \mathcal{N}(\mathbf{0}, \tau_m^2 \mathbf{R}),$$

where $\mathbf{0}$ is a vector of 0's having length $t + 1$ and \mathbf{R} is a $(t + 1) \times (t + 1)$ diagonal matrix.

The matrix \mathbf{R} can be constructed as follows. Assume that the correlation function in ordinary kriging model has a product correlation structure given by $\psi(\mathbf{h}) = \prod_{j=1}^p \psi_j(h_j)$.

Let $l_{ij} = 1$ if β_i includes the linear effect of factor j and 0 otherwise. Similarly, $q_{ij} = 1$ if β_i includes the quadratic effect of factor j and 0 otherwise. Then the i th diagonal element of \mathbf{R} is given by $\prod_{j=1}^p r_{jl}^{l_{ij}} r_{jq}^{q_{ij}}$, where

$$r_{jl} = \frac{3 - 3\psi_j(2)}{3 + 4\psi_j(1) + 2\psi_j(2)} \quad \text{and} \quad r_{jq} = \frac{3 - 4\psi_j(1) + \psi_j(2)}{3 + 4\psi_j(1) + 2\psi_j(2)}. \quad (7)$$

The foregoing connection with kriging makes this Bayesian variable selection technique the most suitable among its competitors. As shown in [23, 24] the effect hierarchy and effect heredity principles are embedded in the prior.

Assume that $Z(\mathbf{x})$ in Eq. (4) follows a Gaussian process. Then the posterior mean of β can be approximated by [24]

$$\hat{\beta} = \frac{\tau_m^2}{\sigma_m^2} \mathbf{R}\mathbf{U}'\Psi^{-1}(\mathbf{y} - \mathbf{V}_m\hat{\boldsymbol{\mu}}_m), \quad (8)$$

where \mathbf{U} is the model matrix corresponding to the experimental design. A variable can be declared important if its absolute coefficient is large. Thus the variable to enter at each step $m = 0, 1, 2, \dots$ can be selected as the variable with the largest $|\hat{\beta}_i|$. We note that [23, 24] instead uses the standardized coefficient for variable selection. Both produce similar results, but the computation of the former is easier. For maximizing $|\hat{\beta}_i|$, without loss of generality we can set $\tau_m^2/\sigma_m^2 = 1$ in Eq. (8), which further simplifies the computations.

There remains an important issue to address in this Bayesian forward variable selection strategy. When should we stop adding terms to the mean part? In other words, what is the best value for m ? The difficulty in choosing m is that, irrespective of its value, kriging predictor interpolates the data and thus gives a perfect fit. Therefore, the prediction errors are all 0. This prevents us from using the standard model selection criteria in regression analysis such as C_p -statistic and Akaike information criterion [16]. We overcome this problem by using cross validation errors.

Let $\hat{y}_{(i)}(\mathbf{x})$ be the predictor after removing the i th data point. Then the leave-one-out cross validation error is defined as

$$cv_i = y_i - \hat{y}_{(i)}(\mathbf{x}_i),$$

for $i = 1, 2, \dots, n$. Define the cross validation prediction error (CVPE) by

$$CVPE(m) = \sqrt{\frac{1}{n} \sum_{i=1}^n cv_i^2}.$$

Now we can choose the value of m that minimizes $CVPE(m)$. We should point out that the foregoing approach of using cross validation errors works well only if the experimental data points are able to capture the trends in the true function.

The cross validation errors can be computed only after estimating the unknown parameters from the data, which is discussed in the next section. Among the parameters, those associated with the correlation function are computationally difficult to estimate. Clearly, the computations will become even more difficult if we need to estimate those parameters after removing each data point. Therefore, we recommend keeping the correlation parameters the same when computing cross validation errors.

2.2 Estimation

We choose the following Gaussian product correlation function

$$\psi(\mathbf{h}) = \exp\left(-\sum_{j=1}^p \theta_j h_j^2\right),$$

which is the most popular correlation function used in computer experiments. Other correlation functions such as cubic correlation function and Matérn correlation function could also be used [3]. Let $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)'$. The parameters $\boldsymbol{\mu}_m$, σ_m^2 , and $\boldsymbol{\theta}$ can be estimated by maximizing the likelihood. Because the model is selected based on a cross validation criterion, it may seem more appropriate to use the same criterion for estimation. However, many empirical studies have shown that the maximum likelihood estimates perform better than the estimates based on cross validation [3, 14].

Under the assumption that $Z(\mathbf{x})$ in Eq. (4) follows a Gaussian process, the negative of the log-likelihood is given by

$$NL = \frac{n}{2} \log(2\pi) + \frac{n}{2} \log \sigma_m^2 + \frac{1}{2} \log |\boldsymbol{\Psi}| + \frac{1}{2\sigma_m^2} (\mathbf{y} - \mathbf{V}_m \boldsymbol{\mu}_m)' \boldsymbol{\Psi}^{-1} (\mathbf{y} - \mathbf{V}_m \boldsymbol{\mu}_m).$$

For the moment assume that $\boldsymbol{\theta}$ is known. Minimizing NL with respect to $\boldsymbol{\mu}_m$ and σ_m^2 , we obtain [3]

$$\hat{\boldsymbol{\mu}}_m = (\mathbf{V}'_m \boldsymbol{\Psi}^{-1} \mathbf{V}_m)^{-1} \mathbf{V}'_m \boldsymbol{\Psi}^{-1} \mathbf{y}, \quad (9)$$

$$\hat{\sigma}_m^2 = \frac{1}{n} (\mathbf{y} - \mathbf{V}_m \hat{\boldsymbol{\mu}}_m)' \boldsymbol{\Psi}^{-1} (\mathbf{y} - \mathbf{V}_m \hat{\boldsymbol{\mu}}_m). \quad (10)$$

Thus, the minimum value of NL is

$$NL = \frac{n}{2} (1 + \log(2\pi)) + \frac{1}{2} (n \log \hat{\sigma}_m^2 + \log |\boldsymbol{\Psi}|). \quad (11)$$

Now consider the case with unknown $\boldsymbol{\theta}$. It can also be estimated by minimizing NL in Eq. (11). However, the minimization is not a trivial task. We have encountered multiple local minima in many examples and thus, finding the global minimum is difficult. Therefore, we propose to estimate $\boldsymbol{\theta}$ only at $m = 0$. Thus

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} n \log \hat{\sigma}_0^2 + \log |\boldsymbol{\Psi}|. \quad (12)$$

Keeping the correlation parameters the same at each step also helps in identifying a mean model that satisfies effect heredity [24]. At the final step, that is after choosing m , the correlation parameters can be again estimated (i.e., by minimizing NL in Eq. (11)), which can give a better prediction. Because $\boldsymbol{\theta}$ is estimated two times, the computational complexity in fitting a blind kriging model is roughly twice as that of an ordinary kriging model. The approach is explained with examples in the next section.

3 Examples

3.1 Example 1: Engine block and head joint sealing Experiment

The engine block and head joint sealing assembly is one of the most crucial and fundamental structural design in the automotive internal combustion engine. Design decisions must be made upfront, prior to the availability of a physical prototype, because it affects downstream design decisions for other engine components as well as significantly impacts the long lead

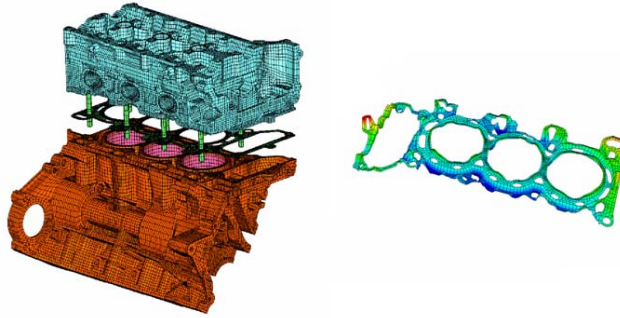


Figure 1: Finite element model of engine head and block joint sealing assembly

time tooling and machining facility setup. Reversing a decision about this assembly at a later time has very expensive consequences. Thus, the use of a computer simulation model is indispensable. The design of the engine block and head joint sealing assembly is very complex due to multiple functional requirements (e.g., combustion gas, high pressure oil, oil drain, and coolant sealing) and complicated geometry; thus, the interactions among design parameters in this assembly (block and head structures, gasket, and fasteners) have significant effects. To best simulate the engine assembly process and operating conditions, a finite element model was developed to capture the complexity of part geometry, the compliance in the components, non-linear material properties, and contact interface between the parts (see Fig. 1). To address performance robustness of the joint sealing, manufacturing variability of the mating surfaces and head bolt tensional load are included in the analysis for which design parameters are optimized. Because the assembly model is computationally expensive, the availability of a computationally efficient and accurate metamodel is important for optimizing the design.

Eight factors are selected for experimentation: gasket thickness (x_1), number of contour zones (x_2), zone-to-zone transition (x_3), bead profile (x_4), coining depth (x_5), deck face surface flatness (x_6), load/deflection variation (x_7), and head bolt force variation (x_8). Because of the complexity in the simulation setup and the excessive computing requirements, only 27 runs are used for the experiment. The experimental design, which is a 27-run orthogonal array [25], is given in Table 1. In this example, we analyze only the gap lift (y).

Table 1: Example 1, Data for the engine head and block joint sealing experiment

Run	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	y
1	2	2	3	2	2	1	2	3	1.53
2	3	3	3	2	3	1	3	1	2.21
3	1	1	2	3	2	1	3	3	1.69
4	3	1	2	1	2	2	3	1	1.92
5	1	1	2	2	3	1	1	2	1.42
6	1	3	2	3	3	3	2	2	5.33
7	1	3	1	2	1	2	3	3	2.00
8	2	3	2	1	1	1	1	1	2.13
9	3	2	1	3	3	2	1	2	1.77
10	2	1	1	2	1	3	1	3	1.89
11	1	3	3	1	3	2	1	3	2.17
12	3	2	2	3	1	2	1	3	2.00
13	3	3	1	3	2	1	2	3	1.66
14	2	1	1	3	3	2	3	1	2.54
15	1	2	1	1	3	1	2	1	1.64
16	3	1	3	2	3	3	2	3	2.14
17	1	2	3	1	1	3	3	2	4.20
18	3	2	2	2	1	3	2	1	1.69
19	1	2	1	2	2	3	1	1	3.74
20	2	2	2	1	3	3	3	3	2.07
21	2	3	3	3	2	3	1	1	1.87
22	2	3	2	2	2	2	2	2	1.19
23	3	3	1	1	2	3	3	2	1.70
24	2	2	3	3	1	1	3	2	1.29
25	2	1	1	1	1	1	2	2	1.82
26	1	1	3	3	1	2	2	1	3.43
27	3	1	3	1	2	2	1	2	1.91

First consider ordinary kriging. The maximum likelihood estimate of $\boldsymbol{\theta}$ is given by

$$\hat{\boldsymbol{\theta}} = (2.75, .26, .02, .01, .01, 4.00, .01, .01)'$$

To avoid numerical problems, we have constrained each θ_i in $[.01, 4]$ in the optimization of the likelihood. We obtain $CVPE(0) = .5784$. The ordinary kriging predictor is given by

$$\hat{y}(\mathbf{x}) = 2.27 + \hat{\boldsymbol{\psi}}(\mathbf{x})' \hat{\boldsymbol{\Psi}}^{-1}(\mathbf{y} - 2.27 \mathbf{1}),$$

where $\hat{\boldsymbol{\psi}}(\mathbf{x})$ is a vector of length 27 with i th element $\psi(\mathbf{x} - \mathbf{x}_i) = \exp(-\sum_{k=1}^8 \hat{\theta}_k (x_k - x_{ik})^2)$ and $\hat{\boldsymbol{\Psi}}$ is a 27×27 matrix whose ij th element is $\psi(\mathbf{x}_i - \mathbf{x}_j) = \exp(-\sum_{k=1}^8 \hat{\theta}_k (x_{ik} - x_{jk})^2)$.

Now consider blind kriging. To apply the Bayesian forward selection technique in [23, 24], we first need to construct the \mathbf{R} matrix. It is a 129×129 diagonal matrix given by

$$\hat{\mathbf{R}} = \text{diag}(1, \hat{r}_{1l}, \hat{r}_{1q}, \hat{r}_{2l}, \dots, \hat{r}_{7q}, \hat{r}_{8q}),$$

where

$$\hat{r}_{jl} = \frac{3 - 3e^{-4\hat{\theta}_j}}{3 + 4e^{-\hat{\theta}_j} + 2e^{-4\hat{\theta}_j}} \quad \text{and} \quad \hat{r}_{jq} = \frac{3 - 4e^{-\hat{\theta}_j} + e^{-4\hat{\theta}_j}}{3 + 4e^{-\hat{\theta}_j} + 2e^{-4\hat{\theta}_j}}.$$

Now compute

$$\hat{\boldsymbol{\beta}} = \hat{\mathbf{R}}\mathbf{U}'\hat{\boldsymbol{\Psi}}^{-1}(\mathbf{y} - 2.27 \mathbf{1}),$$

where \mathbf{U} is a 27×129 matrix whose first column is $\mathbf{1}$ and the other columns correspond to the values of $x_{1l}, x_{1q}, x_{2l}, \dots, x_{7q}, x_{8q}$. Note that because we are only interested in finding the maximum value of $|\hat{\beta}_i|$, we have set $\tau_0^2/\sigma_0^2 = 1$ in Eq. (8). A half-normal plot [25] of the absolute values of $\hat{\beta}_i$'s is shown in Fig. 2. We can see that the maximum value of $|\hat{\beta}_i|$ occurs for the coefficient of the linear-by-linear interaction term of x_1 and x_6 . This could have been easily identified without using a half-normal plot; it is given here only for illustration.

Thus, take $v_1 = x_{1l}x_{6l}$. Again estimate the coefficients using

$$\hat{\boldsymbol{\beta}} = \hat{\mathbf{R}}\mathbf{U}'\hat{\boldsymbol{\Psi}}^{-1}(\mathbf{y} - \mathbf{V}_1\hat{\boldsymbol{\mu}}_1),$$

where $\hat{\boldsymbol{\mu}}_1$ is obtained from Eq. (9) and \mathbf{V}_1 is a 27×2 matrix whose first column is $\mathbf{1}$ and the second column is the values of v_1 . Note that in this computation, the matrices $\hat{\mathbf{R}}$, \mathbf{U} , and

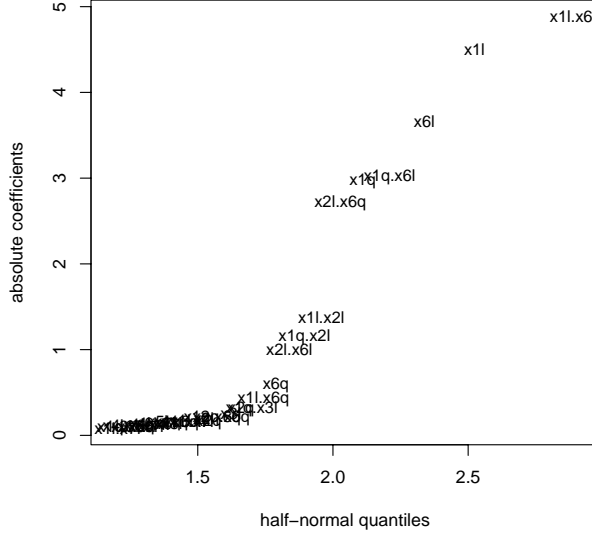


Figure 2: Half-normal plot of $|\hat{\beta}_i|$'s at $m = 0$

$\hat{\Psi}$ remain the same as before. At this step, we identify x_{1l} as the most significant among the remaining variables, because it has the largest $|\hat{\beta}_i|$. Thus, take $v_2 = x_{1l}$ and continue the forward selection procedure. In the next four steps, the variables x_{6l} , $x_{1q}x_{6l}$, x_{1q} , and $x_{2l}x_{6q}$ are selected. The $CVPE(m)$ decrease as shown in Fig. 3 (in the figure ordinary kriging is denoted by OK). The next variable to enter is x_{6q} , but it increases the $CVPE(m)$. We checked a few more steps and found that $CVPE(m)$ is continued to increase and thus we choose $m = 6$. We obtain $CVPE(6) = .4243$. It is also informative to calculate the usual R^2 value used in regression analysis. For our problem, we can define it by [23]

$$R^2(m) = 1 - \frac{\sum_{j=1}^n (y_j - \sum_{i=0}^m \hat{\mu}_i v_{ij})^2}{\sum_{j=1}^n (y_j - \hat{\mu}_0)^2}.$$

It is also plotted in Fig. 3. We can see that the six variables in the mean part explains about 86% of the variation in the data. The kriging part captures the remaining 14%.

The correlation parameters θ can again be estimated by minimizing NL in Eq. (11). The new $\hat{\theta}$ is obtained as

$$\hat{\theta} = (.01, .01, .01, .01, 4, .24, 4, .14)'$$

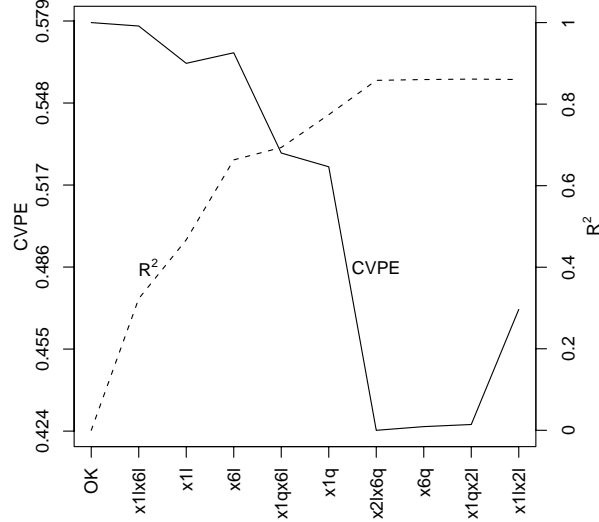


Figure 3: Plots of $CVPE(m)$ and $R^2(m)$ in Example 1

We obtain $CVPE(6) = .2702$, which is much smaller than using the θ estimated at the beginning. The CVPE shows about 53% improvement in prediction using blind kriging over ordinary kriging ($CVPE(0) = .5784$).

The blind kriging predictor is given by

$$\hat{y}(\mathbf{x}) = 2.18 - .44x_{1l}x_{6l} - .48x_{1l} + .39x_{6l} + .21x_{1q}x_{6l} + .19x_{1q} + .30x_{2l}x_{6q} + \hat{\psi}(\mathbf{x})' \hat{\Psi}^{-1}(\mathbf{y} - \mathbf{V}_6 \hat{\mu}_6).$$

It is clear from the mean model that x_1 and x_2 have interactions with x_6 . Because x_6 (the deck face surface flatness) is a noise factor, robustness against it can be achieved by adjusting the two control factors x_1 and x_2 . This cannot be understood from ordinary kriging predictor without performing additional sensitivity analysis [26].

3.2 Example 2: Piston Slap Noise Experiment

Piston slap is an unwanted engine noise resulting from piston secondary motion. A computer experiment was performed by varying six factors to minimize the noise. The factors were set clearance between the piston and the cylinder liner (x_1), location of peak pressure (x_2),

skirt length (x_3), skirt profile (x_4), skirt ovality (x_5), and pin offset (x_6). The experimental design and the data are given in Table 2. More details of the experiment can be found in [27, 28].

Table 2: Example 2, Data for the piston slap noise experiment

Run	x_1	x_2	x_3	x_4	x_5	x_6	y
1	71	16.8	21	2	1	0.98	56.75
2	15	15.6	21.8	1	2	1.3	57.65
3	29	14.4	25	2	1	1.14	53.97
4	85	14.4	21.8	2	3	0.66	58.77
5	29	12	21	3	2	0.82	56.34
6	57	12	23.4	1	3	0.98	56.85
7	85	13.2	24.2	3	2	1.3	56.68
8	71	18	25	1	2	0.82	58.45
9	43	18	22.6	3	3	1.14	55.5
10	15	16.8	24.2	2	3	0.5	52.77
11	43	13.2	22.6	1	1	0.5	57.36
12	57	15.6	23.4	3	1	0.66	59.64

To apply the Bayesian forward selection, first we scale x_1 , x_2 , x_3 , and x_6 to [1.0, 3.0]. For ordinary kriging, we obtain $\hat{\boldsymbol{\theta}} = (1.17, .01, .23, .01, .01, .71)$ and $CVPE(0) = 1.4511$. The $CVPE(m)$ and $R^2(m)$ are plotted in Fig. 4 based on the variables identified by the Bayesian variable selection technique. We see that the three variables x_{1l} , $x_{1l}x_{6l}$, and $x_{1q}x_{6l}$ give the minimum $CVPE(3) = 1.2777$. The corresponding $R^2(3) = .79$, which shows that the three variables alone explain about 79% of the variability in the data. Estimating $\boldsymbol{\theta}$ again, we obtain $\hat{\boldsymbol{\theta}} = (.01, .01, .09, 1.32, .01, .46)'$ and $CVPE(3) = 1.1168$. Thus, we can expect about a 23% improvement in prediction using blind kriging over ordinary kriging. The blind kriging predictor is given by

$$\hat{y}(\mathbf{x}) = 56.6 + 1.40x_{1l} - 1.12x_{1l}x_{6l} + .93x_{1q}x_{6l} + \hat{\boldsymbol{\psi}}(\mathbf{x})' \hat{\boldsymbol{\Psi}}^{-1}(\mathbf{y} - \mathbf{V}_3 \hat{\boldsymbol{\mu}}_3).$$

We can see that in this example the $CVPE$ increased after the first step but then came down significantly after two more steps. This shows that we should not stop the procedure immediately when we observe an increase in $CVPE$. The procedure should be continued

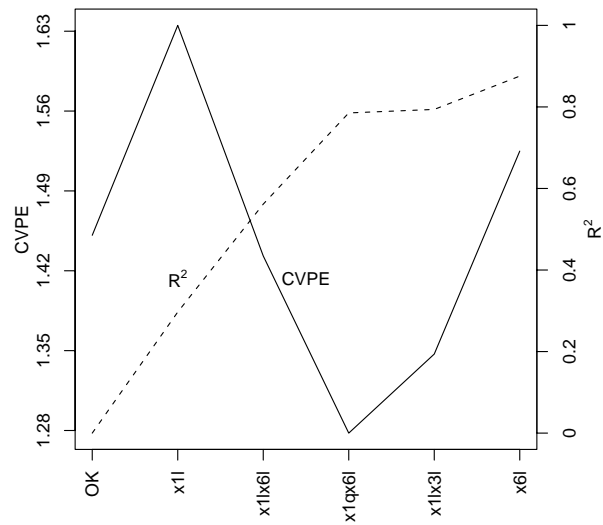


Figure 4: Plots of $CVPE(m)$ and $R^2(m)$ in Example 2

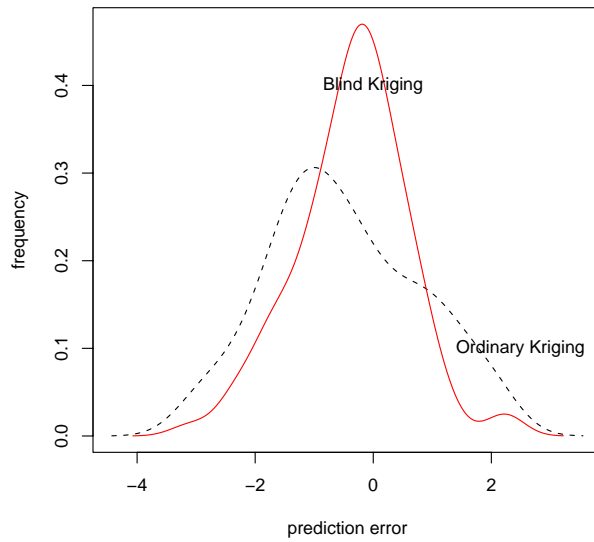


Figure 5: Density plot for the prediction errors in Example 2

for a few more steps before choosing the value of m . Note that the R^2 plot is used only for interpretation and not for selecting the best m .

An additional 100 runs were performed for validating the results. The two densities of the prediction errors for ordinary kriging and blind kriging are shown in Fig. 5. It clearly shows that blind kriging gives a much better prediction. We can also calculate the root-mean squared prediction error (RMSPE) using

$$RMSPE = \sqrt{\frac{1}{100} \sum_{i=1}^{100} (y(\mathbf{x}_i) - \hat{y}(\mathbf{x}_i))^2}.$$

For ordinary kriging $RMSPE = 1.3626$ and for blind kriging $RMSPE = 1.0038$, which shows that the prediction error of blind kriging is smaller than that of ordinary kriging by about 26%.

There are several case studies reported in the literature where universal kriging is applied instead of ordinary kriging. Ref [30] used universal kriging with all linear effects in the mean part of the model for the optimization in a material cellular design problem; see [31] for other examples. In this example, we fitted a universal kriging model with linear effects for all of the factors. The universal kriging predictor is given by

$$\hat{y}(\mathbf{x}) = 55.3 + 1.02x_{1l} - .15x_{2l} - .96x_{3l} + .01x_{4l} - .45x_{5l} - .31x_{6l} + \hat{\boldsymbol{\psi}}(\mathbf{x})' \hat{\boldsymbol{\Psi}}^{-1}(\mathbf{y} - \mathbf{V}\hat{\boldsymbol{\mu}}),$$

with $\hat{\boldsymbol{\theta}} = (0.14, 0.01, 0.17, 0.01, 0.01, 0.09)$. The $RMSPE$ for the 100 validation runs is obtained as 1.5109, which is larger than both ordinary and blind kriging. The reason for this poor performance is that the mean part of the universal kriging model contains some unimportant effects (R^2 is only 25.4%). This shows the danger of using a universal kriging model without proper variable selection.

3.3 Example 3: Borehole Model

The following simple function for the flow rate through a borehole is used by many authors to compare different methods in computer experiments (see e.g., [29]):

$$y = \frac{2\pi T_u(H_u - H_l)}{\ln(r/r_w) \left[1 + \frac{2LT_u}{\ln(r/r_w)r_w^2 K_w} + \frac{T_u}{T_l} \right]},$$

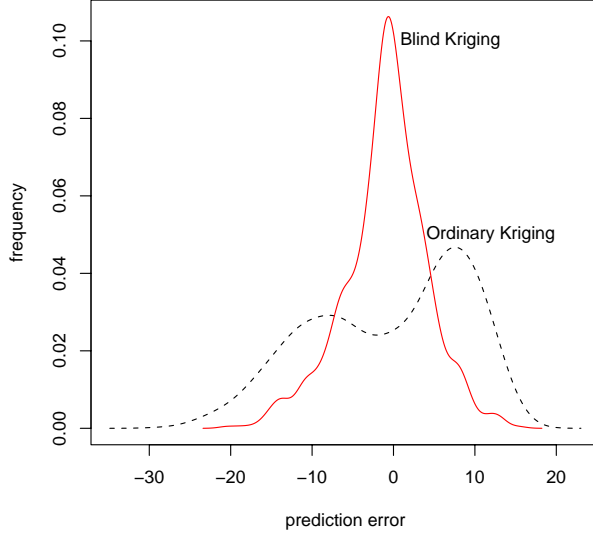


Figure 6: Density plot for the prediction errors in Example 3

where the ranges of interest for the eight variables are $r_w : (0.05, 0.15)$, $r = (100, 50000)$, $T_u = (63070, 115600)$, $H_u = (990, 1110)$, $T_l = (63.1, 116)$, $H_l = (700, 820)$, $L = (1120, 1680)$, and $K_w = (9855, 12045)$. We re-scale the variables in $[1.0, 3.0]$ and denote them as x_1, x_2, \dots, x_8 . For convenience, we use the same 27-run experimental design in Table 1.

Using the Bayesian variable selection technique, we identified the linear effect of x_1 as the only important variable. The blind kriging predictor is given by

$$\hat{y}(\mathbf{x}) = 93.4 + 60.1x_{1l} + \hat{\boldsymbol{\psi}}(\mathbf{x})' \hat{\boldsymbol{\Psi}}^{-1}(\mathbf{y} - \mathbf{V}_1 \hat{\boldsymbol{\mu}}_1),$$

with $\hat{\boldsymbol{\theta}} = (.31, .01, .01, .09, .01, .08, .07, .02)'$. We randomly generated 1,000 values within the experimental region and the prediction errors are plotted in Fig. 6. It shows remarkable improvement in prediction for blind kriging over ordinary kriging.

To check for the robustness against misspecification of correlation parameters, we repeated the calculations by varying $\boldsymbol{\theta}$. Let $\theta_1 = \dots = \theta_8 = \theta$. Fig. 7 shows the plot of RMSPE values for different values of θ . We can see that the RMSPE values of blind kriging are almost half of those of ordinary kriging and have much less variation. This shows that

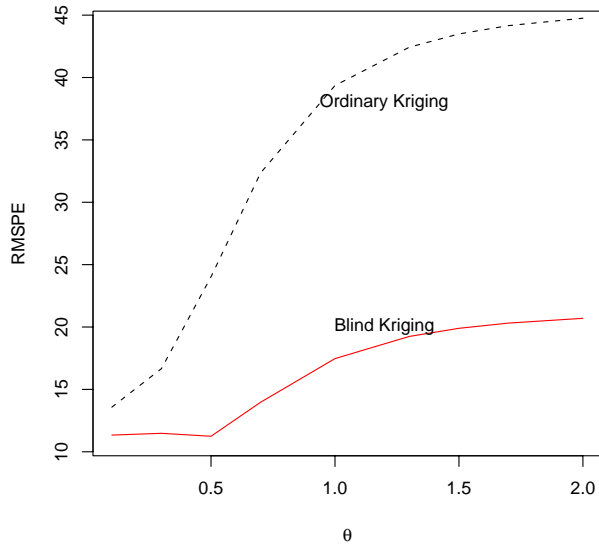


Figure 7: RMSPE values for different θ in Example 3

blind kriging is more robust to misspecification in the correlation parameters than is ordinary kriging. This is a great advantage, because in practice it is difficult to obtain precise estimates of the correlation parameters.

We also tried universal kriging method for the borehole example. Two models are fitted, one with all linear terms and the other with all linear and quadratic terms. The RMSPE values for the 1,000 runs are given in Table 3. We can see that they are much higher than that of ordinary kriging and blind kriging. Thus, including unimportant variables in the mean part can actually deteriorate the performance. This clearly shows the importance of selecting variables carefully and the superiority of blind kriging over universal kriging.

4 Conclusions

It is a common practice in the literature to use a constant mean for the kriging model. Although some recent studies point out the benefits of using more complex models for the mean [14, 30], none of them have proposed a systematic methodology to obtain such models.

Table 3: Comparison of different methods in Example 3

Method	m	RMSPE
Ordinary Kriging	0	9.7
Blind Kriging	1	5.5
Universal Kriging (linear)	8	11.3
Universal Kriging (linear and quadratic)	16	18.0

In fact, the problem is much more complicated than merely using a complex model for the mean. Unnecessary variables in the mean model can deteriorate the performance. Therefore only those variables that have a significant effect on the response should be used for the mean model. We showed that they can be identified using a Bayesian forward selection technique proposed in [23, 24].

The Bayesian forward selection technique is directly related to kriging, which makes it attractive to use in blind kriging method. The most difficult step in this Bayesian technique is the estimation of correlation parameters. However, the estimates are readily available from ordinary kriging model and thus, the technique can be applied with no additional difficulty. Another advantage of the technique is that it incorporates the effect hierarchy and heredity principles through prior specification and thus, produces interpretable models.

We also note that a naive strategy of identifying important variables using a variable selection technique and then fitting the kriging part, in general will not work. This is because the performance of blind kriging is quite sensitive to the number of variables used in the mean part. Our approach computes the cross validation errors at each step of the Bayesian forward selection technique and selects the model with minimum error. It may happen that ordinary kriging itself is the optimal predictor, which cannot be detected in the naive strategy that applies a variable selection technique without considering the kriging part. Thus, we believe that the use of cross validation errors along with the Bayesian forward selection technique is critical for obtaining a good blind kriging predictor.

Several examples presented in the article demonstrate that substantial improvement in prediction can be achieved by using blind kriging. It is also shown that blind kriging predictor is simpler to interpret and is more robust to the misspecification in the correlation parameters than ordinary kriging predictor.

Acknowledgments

The research of Joseph was supported by the U.S. National Science Foundation grants CMMI-0448774 and CMMI-0654369.

Appendix: Bayesian Variable Selection Technique

Here we provide some additional details for the Bayesian variable selection technique. The computer model can be represented as $Y = f(\mathbf{x})$, where the transfer function f can be highly nonlinear. First assume that each x_i takes only three values 1, 2, and 3. Later we will explain how to generalize this. Define the linear and quadratic effects for each x_i as in Eq. (6). Now consider approximating $f(\mathbf{x})$ by a linear model containing all of the interaction terms (up to the p th order interaction). The linear model can be written as $\sum_{i=0}^{3^p-1} \beta_i u_i$, where $u_0 = 1$, $u_1 = x_{1l}$, \dots , and $u_{3^p-1} = x_{1q} \cdots x_{pq}$.

A major step in the Bayesian variable selection technique is to postulate a prior distribution for $\boldsymbol{\beta} = (\beta_0, \dots, \beta_{3^p-1})'$. This is a difficult task because of the huge number of parameters. To simplify this task, Refs [23, 24] proposed an interesting idea. Instead of directly postulating a prior for $\boldsymbol{\beta}$, postulate a functional prior for $f(\mathbf{x})$ and use it to induce a prior for $\boldsymbol{\beta}$. Assume that

$$f(\mathbf{x}) \sim GP(\mu_0, \sigma_0^2 \psi),$$

where μ_0 is the mean and $\sigma_0^2 \psi$ is the covariance function of the Gaussian process (GP). Because there are 3^p parameters in the linear model, their distribution can be obtained based on 3^p function values. One simple choice is to evaluate the function at the full factorial design for the p factors (which contains 3^p points). To simplify the results further, write

the linear model as $\mu_0 + \sum_{i=0}^{3^p-1} \beta_i u_i$ and assume a product correlation structure given by $\psi(\mathbf{h}) = \prod_{j=1}^p \psi_j(h_j)$. Then, it can be shown that [24]

$$\begin{aligned}\beta_0 &\sim \mathcal{N}(0, \tau_0^2), \\ \beta_1 &\sim \mathcal{N}(0, \tau_0^2 r_{1l}), \\ \beta_2 &\sim \mathcal{N}(0, \tau_0^2 r_{1q}), \\ &\vdots \\ \beta_{3^p-1} &\sim \mathcal{N}(0, \tau_0^2 r_{1q} r_{2q} \cdots r_{pq}),\end{aligned}$$

where r_{jl} and r_{jq} for $j = 1, \dots, p$ are calculated using Eq. (7). Further, [24] shows that β_i 's are approximately independent. Thus, the prior distribution for $\boldsymbol{\beta}$ is a multivariate normal distribution with mean $\mathbf{0}$ and variance-covariance matrix $\tau_0^2 \mathbf{R}$, where $\mathbf{R} = \text{diag}\{1, r_{1l}, r_{1q}, \dots, r_{1q} \cdots r_{pq}\}$.

Let the experiment has n runs and let \mathbf{y} be the data. We have $\mathbf{y} = \mu_0 \mathbf{1} + \mathbf{U}\boldsymbol{\beta}$, where \mathbf{U} is the model matrix with dimension $n \times 3^p$. Using Bayes theorem, the posterior distribution of $\boldsymbol{\beta}$ is given by

$$\boldsymbol{\beta}|\mathbf{y} \sim \mathcal{N}\left(\frac{\tau_0^2}{\sigma_0^2} \mathbf{R}\mathbf{U}'\boldsymbol{\Psi}^{-1}(\mathbf{y} - \mu_0 \mathbf{1}), \tau_0^2 \mathbf{R} - \frac{\tau_0^4}{\sigma_0^2} \mathbf{R}\mathbf{U}\boldsymbol{\Psi}^{-1}\mathbf{U}\mathbf{R}\right).$$

The posterior mean can be used as an estimate of $\boldsymbol{\beta}$. This forms the basis for the forward selection technique discussed in Section 2.1.

Note that if we are interested only up to the two-factor interactions, then approximate results can be obtained by replacing \mathbf{R} and \mathbf{U} by their appropriate sub-matrices. Moreover, if a factor takes values in a continuous interval, then it should be scaled in the interval [1.0, 3.0]. Other ranges such as [0, 1] or [-1, 1] may also be used. However, the formulas for the linear-quadratic effects and r 's should be changed accordingly. For example, if the factors are scaled in [0, 1], then the linear-quadratic effects should be calculated using Eq. (6) after replacing $x_j - 2$ with $2(x_j - .5)$ and the r 's using Eq. (7) after replacing the arguments of ψ_j by .5 and 1 instead of 1 and 2.

Nomenclature

- cv_i = leave-one-out cross validation error
- $\mathbf{v}(\mathbf{x})$ = Set of functions in the mean model
- \mathbf{V}_m = $n \times (m + 1)$ model matrix for the mean
- l_{ij}, q_{ij} = Indicator variables for linear and quadratic terms
- m = Number of variables in the mean model
- n = Number of design points
- p = Number of factors
- t = Number of candidate variables
- u_i = i th candidate variable
- \mathbf{x} = p – dimensional vector of factors
- \mathbf{x}_i = i th design point
- x_i = i th factor
- x_{il}, x_{iq} = Linear and quadratic components of the i th factor
- \mathbf{y} = Data vector
- $\hat{y}(\mathbf{x})$ = Predictor at \mathbf{x}
- $\hat{y}_{(i)}(\mathbf{x})$ = Predictor at \mathbf{x} after removing i th data point
- $Z(\mathbf{x})$ = Stochastic process
- β_i = Coefficient of u_i
- $\boldsymbol{\beta}$ = A $(t + 1)$ – dimensional vector containing β_0, \dots, β_t
- $\boldsymbol{\theta}$ = Coefficients in the Gaussian correlation function
- $\tau_m^2 \mathbf{R}$ = Prior variance-covariance matrix of $\boldsymbol{\beta}$
- σ_m^2 = Variance of $Z(\mathbf{x})$
- $\psi(\mathbf{h})$ = Correlation function
- $\boldsymbol{\psi}(\mathbf{x})$ = An n – dimensional vector with i th element $\psi(\mathbf{x} - \mathbf{x}_i)$

Ψ	=	Correlation matrix
$\mathbf{1}$	=	Vector of 1's having length n
$\mu(\mathbf{x})$	=	Mean function
μ_i	=	Coefficient of v_i
$\boldsymbol{\mu}_m$	=	An $(m + 1)$ – dimensional vector containing μ_0, \dots, μ_m
CVPE	=	Cross validation prediction error
NL	=	Negative of log-likelihood
OK	=	Ordinary kriging
RMSPE	=	Root mean squared prediction error

References

- [1] Fang, K. T., Li, R., and Sudjianto, A., 2006, *Design and Modeling for Computer Experiments*, CRC Press, New York.
- [2] Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P., 1989, “Design and Analysis of Computer Experiments,” *Statistical Science*, **4**, 409-423.
- [3] Santner, T. J., Williams, B. J., and Notz, W. I., 2003, *The Design and Analysis of Computer Experiments*, Springer, New York.
- [4] Jin, R., Chen, W., and Simpson, T., 2001, “Comparative Studies of Metamodeling Techniques under Multiple Modeling Criteria,” *Journal of Structural & Multidisciplinary Optimization*, **23**, 1-13.
- [5] Pacheco, J. E., Amon, C. H., and Finger, S., 2003, “Bayesian Surrogates Applied to Conceptual Stages of the Engineering Design Process,” *ASME J. Mech. Des.*, **125**, 664-672.

- [6] Cappelleri, D. J., Frecker, M. I., Simpson, T. W., and Snyder, A., 2002, “Design of a PZT Bimorph Actuator Using a Metamodel-Based Approach,” *ASME J. Mech. Des.*, **124**, 354-357.
- [7] Sasena, M. J., Parkinson, M., Reed, M. P., Paplambros, P. Y., Goovaerts, P., 2005, “Improving an Ergonomics Testing Procedure via Approximation-based Adaptive Experimental Design,” *ASME J. Mech. Des.*, **127**, 1006-1013.
- [8] Yang, R. J., Wang, N., Tho, C. H., Bobineau, J. P., and Wang, B. P., 2005, “Metamodeling Development for Vehicle Frontal Impact Simulation,” *ASME J. Mech. Des.*, **127**, 1014-1020.
- [9] Apley, D. W., Lin, J. L., and Chen, W., 2006, “Understanding the Effects of Model Uncertainty in Robust Design with Computer Experiments,” *ASME J. Mech. Des.*, **128**, 945-958.
- [10] Martin, J. D. and Simpson, T. W., 2006, “A Methodology to Manage System-level Uncertainty During Conceptual Design,” *ASME J. Mech. Des.*, **128**, 959-968.
- [11] Wackernagel, H., 2002, *Multivariate Geostatistics*, Springer, New York.
- [12] Currin, C., Mitchell, T. J., Morris, M. D. and Ylvisaker, D., 1991, “Bayesian Prediction of Deterministic Functions, with Applications to the Design and Analysis of Computer Experiments,” *J. Amer. Statist. Assoc.*, **86**, 953-963.
- [13] Welch, W. J., Buck, R. J., Sacks, J., Wynn, H. P., Mitchell, T. J., and Morris, M. D., 1992, “Screening, Predicting, and Computer Experiments,” *Technometrics*, **34**, 15-25.
- [14] Martin, J. D. and Simpson, T. W., 2005, “On the Use of Kriging Models to Approximate Deterministic Computer Models,” *AIAA Journal*, **43**, 853-863.
- [15] Joseph, V. R., 2006, “Limit kriging,” *Technometrics*, **48**, 458-466.
- [16] Miller, A., 2002, *Subset Selection in Regression*, CRC Press, New York.

- [17] George, E. I. and McCulloch, R. E., 1993, "Variable Selection via Gibbs Sampling," J. Amer. Statist. Assoc., **88**, 881-889.
- [18] Breiman, L., 1995, "Better Subset Regression using the Nonnegative Garrote," Technometrics, **37**, 373-384.
- [19] Tibshirani, R., 1996, "Regression Shrinkage and Selection via the Lasso," J. Royal. Statist. Soc. B., **58**, 267-288.
- [20] Efron, B., Johnstone, I., Hastie, T. and Tibshirani, R., 2004, "Least Angle Regression," Annals of Statistics, **32**, 407-499.
- [21] Hamada, M. and Wu, C. F. J., 1992, "Analysis of Designed Experiments with Complex Aliasing," J. Qual. Tech., **24**, 130-137.
- [22] Chipman, H., Hamada, M. and Wu, C. F. J., 1997, "A Bayesian Variable Selection Approach for Analyzing Designed Experiments with Complex Aliasing," Technometrics, **39**, 372-381.
- [23] Joseph, V. R., 2006, "A Bayesian Approach to the Design and Analysis of Fractionated eExperiments," Technometrics, **48**, 219-229.
- [24] Joseph, V. R. and Delaney, J. D., 2007, "Functionally Induced Priors for the Analysis of Experiments," Technometrics, **49**, 1-11.
- [25] Wu, C. F. J., and Hamada, M., 2000, *Experiments: Planning, Analysis, and Parameter Design Optimization*, Wiley, New York.
- [26] Chen, W., Jin, R., and Sudjianto, A., 2005, "Analytical Variance-Based Global Sensitivity Analysis in Simulation-Based Design Under Uncertainty," ASME J. Mech. Des., **127**, 875-886.
- [27] Hoffman, R. M., Sudjianto, A., Du, X., Stout, J., 2003, "Robust Piston Design and Optimization Using Piston Secondary Motion Analysis," SAE Paper 2003-01-0148, SAE Transactions.

- [28] Li, R. and Sudjianto, A., 2005, "Analysis of Computer Experiments Using Penalized Likelihood in Gaussian Kriging Models," *Technometrics*, **47**, 111-120.
- [29] Morris, M. D., Mitchell, T. J., and Ylvisaker, D., 1993, "Bayesian Design and Analysis of Computer Experiments: Use of Derivatives in Surface Prediction," *Technometrics*, **35**, 243-255.
- [30] Qian, Z., Seepersad, C. C., Joseph, V. R., Allen, J. K. and Wu, C. F. J., 2006, "Building surrogate models based on detailed and approximate simulations," *ASME J. Mech. Des.*, **128**, 668-677.
- [31] Sacks, J., Schiller, S. B., and Welch, W. J., 1989, "Design of Computer Experiments," *Technometrics*, **31**, 41-47.