Proceedings of DETC'02
ASME 2002 Design Engineering Technical Conferences And Computers and Information in Engineering Conference
MONTREAL, Canada, September 29–October 2, 2002

DETC2002/DAC-34092

ON SEQUENTIAL SAMPLING FOR GLOBAL METAMODELING IN ENGINEERING DESIGN

Ruichen Jin
Wei Chen*
Dept. of Mechanical and Industrial Engineering
University of Illinois at Chicago

Agus Sudjianto
V-Engine Engineering Analytical Powertrain
Ford Motor Company

*Corresponding Author
Mechanical and Industrial Engineering (M/C 251)
842 W. Taylor St., University of Illinois at Chicago, Chicago IL 60607-7022
Phone: (312) 996-6072; Fax: (312) 413-0447; E-mail: weichen1@uic.edu

ABSTRACT

Approximation models (also known as metamodels) have been widely used in engineering design to facilitate analysis and optimization of complex systems that involve computationally expensive simulation programs. The accuracy of metamodels is directly related to the sampling strategies used. Our goal in this paper is to investigate the general applicability of sequential sampling for creating global metamodels. Various sequential sampling approaches are reviewed and new approaches are proposed. The performances of these approaches are investigated against that of the one-stage approach using a set of test problems with a variety of features. The potential usages of sequential sampling strategies are also discussed.

NOMENCLATURE

d	Distance between two sample points
d_s	Scaled distance between two sample points
k	Number of input variables
n	Number of sample points
l	Number of sample points generated at all the previous sampling stages
m	Number of sample points generated at the new sampling stage
$\mathbf{X}_D$	A sample set with n sample points $\mathbf{x}_{D1}, \mathbf{x}_{D2}, \ldots, \mathbf{x}_{Dn}$
$\mathbf{X}_P$	A sample set with all l previous sample points $\mathbf{x}_{P1}, \mathbf{x}_{P2}, \ldots, \mathbf{x}_{Pl}$
$\mathbf{X}_C$	A sample set with m new sample points $\mathbf{x}_{C1}, \mathbf{x}_{C2}, \ldots, \mathbf{x}_{Cm}$
$\mathbf{R}$	Correlation matrix

INTRODUCTION

Mathematical models have been widely used to simulate and analyze complex real world systems in the area of engineering design. These mathematical models, often implemented by computer codes (e.g., Computational Fluid Dynamics and Finite Element Analysis), could be computationally expensive. For example, one run of a finite element model for vehicle crashworthiness can take several hours. While the capacity of computer keeps increasing, to capture the real world systems more accurately, today's simulation codes are even getting much more complex and unavoidably more expensive. The multidisciplinary nature of design and the need for incorporating uncertainty in design optimization have posed additional challenges. A widely used strategy is to utilize approximation models which are often referred to as metamodels as they provide a model of the model [1], replacing the expensive simulation model during the process. Recent studies on using metamodels in design applications include [2, 3, 4, 5, 6], etc. For dealing with multidisciplinary systems, Meckesheimer, et al. [7] presented a generic integration framework to integrate metamodels from multiple subsystems.

An important research issue related to metamodeling is how to achieve a good accuracy of a metamodel with a reasonable number of sample points. While the accuracy of a metamodel is directly related to the metamodeling technique used and the properties of a problem itself, the types of sampling approaches also have direct influences on the performance of a metamodel. Koehler and Owen [8] provided a good review on various sampling approaches for computer experiments. Simpson, et al. [9] compared five sampling strategies and four metamodeling approaches in terms of their

capability to generate accurate approximations for two engineering applications. It is generally believed that space-filling designs, e.g., Latin Hypercube design [10], Orthogonal Array [11], Minimax and Maximin design [12], and Entropy design [13], etc., are preferable for metamodeling. In addition to the types of sampling approaches, equally important is the sample size and how the sample points are generated. Typically, sample points are generated all at once, or in other words, at one-stage. However, when a limited sample size is affordable, sequentially selecting a (or several) sample point(s) at a time allows the sample size to be determined as data accumulate, or in other words, it allows the sampling process to be stopped as soon as there is sufficient information [14]. Based on a sequential sampling strategy, a metamodel is updated sequentially with new sample points and the associated response evaluations. Sequential sampling strategies that can take the advantage of information gathered from the existing (earlier created) metamodel could be very helpful. As Sacks, et al. [14] pointed out: Fully sequential design is, therefore, the most natural for computer experiments.

Sequential sampling can be used for both global metamodeling and metamodel-based design optimization. Sequential sampling for global metamodeling focuses on sequentially improving the accuracy of a metamodel over the entire design space of interest; Sequential sampling for metamodel-based design optimization emphasizes finding the global optimum by balancing a search for the optimal point of a metamodel and a search for unexplored regions to avoid missing promising areas due to the inaccuracy of the metamodel. For global metamodeling, Currin, et al [15] applied two stages optimal design based on the maximum entropy criterion to gain insight into a circuit-simulator example; Sacks [14] used two stages optimal design based on Integrated Mean Squared Error (IMSE) criterion for the same example; Osio and Amon [16] proposed Bayesian surrogates and a sequential-adaptive optimal sampling method based on the A-optimality criterion and applied the method for a case study on thermal properties of a embedded chip. For metamodel-based design optimization, based on different criteria for a balanced search (sequential sampling), various optimization approaches have been developed, e.g., Efficient Global Optimization (EGO) [17], Surrogate Management Framework (SMF) [5], Sequential Design Optimization (SDO) [18], and Model-Assisted Grid Search (MAGS) [19], etc. Jones [20] provided a comprehensive investigation on several criteria; Sasena et al. [21] compared several sampling criteria based on a few test problems.

This paper focuses on sequential sampling for global metamodeling. Without additional notes in the rest of this paper, *sequential sampling* is referred for global metamodeling. *The major issue in sequential sampling is how to choose new sample points based on the locations of available sample points and the information gathered from a created (updated) metamodel*. To select new sample points, optimality criteria, such as the maximum entropy criterion and the maxmin distance criterion are often used. Regarding the utilization of information from an earlier created model, Sacks, et al. [14] classified sequential sampling approaches as sequential approaches with adaptation to the existing metamodel and those without adaptation. One immediate question is, how effective are the sequential sampling approaches compared to a one-stage approach? Clearly, any sequential approaches without adaptation will generally be less efficient than a one-stage optimal sampling approach, since these approaches simply split the optimization in the one-stage approach into a sequence of small optimizations and their results (sample points) are only sub-optimums of the one-stage approach. On the other hand, if the information gathered from an existing metamodel is dependable and usable, sequential approaches with adaptation may be superior to one-stage approaches.

Even though most of the existing works provide good details about a particular type of sequential sampling strategy, it is not clear the effectiveness of a sequential sampling strategy when used for a variety of applications. There is also a lack of comparison between various sequential sampling approaches. Further, in the literature of computer experiments, sequential sampling approaches with adaptation are mainly developed for the Kriging [14] method. They typically make use of the model structure information available from a Kriging model. These techniques are no longer applicable when metamodeling techniques other than Kriging are used. *Our goal in this paper is to investigate the general applicability of sequential sampling for creating global metamodels and to develop new sequential sampling approaches, which are not limited to Kriging metamodel.* Using a set of test problems with a variety of features, we will test and compare the performance of several sequential sampling approaches that exist in the literature as well as the two approaches proposed by us. Through this study, we seek the answers to the following questions: (1) How effective are the sequential sampling approaches compared to the traditional one-step approach? (2) Among the various techniques investigated, is there any technique superior to the others? (3) What are the general benefits one can expect when applying the sequential sampling approaches?

The paper is organized as the following. In the next section, the technological base of our study is provided. After it, the test problems, test scheme, and test results will be provided. The last section is the closure of the paper where we summarize the important observations made from our study.

TECHNOLOGICAL BASE - SEQUENTIAL SAMPLING FOR GLOBAL METAMODELING

In this section, we first introduce two metamodeling techniques used in this study. And then a review of the existing sequential sampling approaches is provided, followed by the introduction of our proposed approaches.

Metamodeling Techniques

A variety of metamodeling techniques exists in the literature, e.g., Response surface methodology (see, e.g., [22]), artificial neural network methods (see, e.g., [23]), Kriging method (see, e.g., [14]), Multivariate adaptive regression splines [15], and radial basis function methods [25,26]. In this paper, we will focus on two interpolation metamodels, i.e., Kriging method and Radial Basis Function (RBF) methods. We are interested in these two representative techniques, because both can provide good accuracy for fitting highly nonlinear behaviors. The major difference of these two techniques is that the former treats the output of the unknown function as a realization of a stochastic process and can provide an estimation of the prediction error while the later cannot.

The Kriging Method

The Kriging model used here treats the deterministic output $y(\mathbf{x})$ as a realization of a stochastic process $Y(\mathbf{x})$:

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

where μ is a constant; $Z(\mathbf{x})$ is assumed to a stochastic process with zero mean and the Gaussian spatial correlation function is given by:

$$Cov(Z(\mathbf{v}), Z(\mathbf{w})) = \sigma^2 \rho(\mathbf{v}, \mathbf{w}) = \sigma^2 \exp\left(-d_s^2(\mathbf{v}, \mathbf{w})\right), \quad (2)$$

where σ^2 is the process variance, ρ stands for correlation functions, and $d_s(\mathbf{v}, \mathbf{w})$ is the scaled distance between two points $\mathbf{v}$ and $\mathbf{w}$:

$$d_s(\mathbf{v}, \mathbf{w}) = \sqrt{\sum_{h=1}^{k} \theta_h (v_h - w_h)^2} \ . \quad (3)$$

The weight parameters θ_h (h = 1,2,…,k), also called correlation parameters, can be considered as the importance of variable h: the larger is the value of θ_h, the larger influence $\mathbf{v}$ could have on $\mathbf{w}$ in the h direction. θ_h, together with μ and σ^2, determines the stochastic process $Y(\mathbf{x})$.

Given a sample set $\mathbf{X}_D$ with n observation points $\mathbf{x}_{D1}, \mathbf{x}_{D2}, ..., \mathbf{x}_{Dn}$ and their outputs (responses) $\mathbf{y}_D = [y_{D1}, y_{D2}, ..., y_{Dn}]^T$, the best linear unbiased predictor (BLUP), i.e., the Kriging metamodel, can be expressed by:

$$\hat{y}(\mathbf{x}) = \hat{\mu} + \mathbf{r}_D(\mathbf{x})^T \mathbf{R}_D^{-1} (\mathbf{y}_D - \mathbf{J}_n \hat{\mu}), \quad (4)$$

where, $\mathbf{J}_n$ is an n-vector of 1's; $\mathbf{R}_D$ is a n*n matrix, whose elements are the correlation between two sample points: $R_{D\,ij} = \rho(\mathbf{x}_{Di}, \mathbf{x}_{Dj}), \quad 1 \le i, j \le n$; the elements of an n-vector $\mathbf{r}_D(\mathbf{x})$ are the correlation between a new point $\mathbf{x}$ and sample points: $r_{Di}(\mathbf{x}) = \rho(\mathbf{x}, \mathbf{x}_{Di}), \quad 1 \le i \le n$; $\hat{\mu} = \dfrac{\mathbf{J}_n^T \mathbf{R}_D^{-1} \mathbf{y}_D}{\mathbf{J}_n^T \mathbf{R}_D^{-1} \mathbf{J}_n}$ is the generalized least-squares estimate of μ. The unknown parameters $\boldsymbol{\theta} = [\theta_1, \theta_2, ..., \theta_k]^T$ and σ^2 are obtained by maximizing the likelihood function.

Kriging provides an estimation of the prediction error on an unobserved point, which is also called as the mean squared error (MSE):

$$s^2(\mathbf{x}) = \sigma^2 \left[1 - \mathbf{r}_D(\mathbf{x})^T \mathbf{R}_D^{-1} \mathbf{r}_D(\mathbf{x}) + \frac{(1 - \mathbf{J}_n^T \mathbf{R}_D^{-1} \mathbf{r}_D(\mathbf{x}))^2}{\mathbf{J}_n^T \mathbf{R}_D^{-1} \mathbf{J}_n} \right]. \quad (5)$$

More detailed introduction of Kriging is provided by [14, 2, 17], etc. Kriging method is very flexible in capturing nonlinear behaviors because the correlation functions can be statistically tuned by the sample data. Another good feature of Kriging is its ability to provide the estimation of the prediction error. This feature is the basis of several sequential approaches examined in this paper. Kriging metamodels have been used in design and analysis of computer experiments in engineering applications. For example, Sacks, et al. [14] used Kriging to create a metamodel for a circuit-simulator code; Currin, et al. [15] applied Kriging in a Thermal Energy Storage System problem; Jones, et al. [17] used it to investigate how the performance of an integrated circuit depended on 36 input variables. Recently, Kriging has also been used in design optimization. For example, Simpson et al. [27] used the Kriging metamodel in structural design optimization; Stinstra et al. [28] used the Kriging metamodel and a multi-start optimization approach for design optimization of electron gun in a color TV.

The Radial Basis Function Method

Given n sample points $\mathbf{x}_{D1}, \mathbf{x}_{D2}, ..., \mathbf{x}_{Dn}$ and their outputs (responses) $\mathbf{y}_D = [y_{D1}, y_{D2}, ..., y_{Dn}]^T$, the radial basis function interpolation is expressed as follows:

$$\hat{y}(\mathbf{x}) = p(\mathbf{x}) + \sum_{i=1}^{n} \lambda_i \phi(d(\mathbf{x}, \mathbf{x}_{Di})), \quad (6)$$

where $p(\mathbf{x})$ is a polynomial model; d is the Euclidean distance; ϕ is a basis function, for which we have many different choices, such as linear, cubic, thin plate spline, multiquadric, Gaussian, etc. Detailed review on RBF methods can be found in [29, 30]. This approach is suitable for highly nonlinear problems. Yang et al. [31] used the RBF method with multiquadric basis functions for a vehicle frontal impact problem.

In this paper, we focus on the RBF method with linear basis functions $\phi(d(\mathbf{v}, \mathbf{w})) = d(\mathbf{v}, \mathbf{w})$ and a constant value for the polynomial part, i.e.,

$$\hat{y}(\mathbf{x}) = \mu + \sum_{i=1}^{n} \lambda_i \left\| \mathbf{x} - \mathbf{x}_{Di} \right\|, \quad \textit{where} \quad \mu = \frac{1}{n} \sum_{i=1}^{n} y_{Di} \ . \quad (7)$$

The n unknown parameters λ_i (i = 1, ..n) can be obtained by solving n linear equations and the prediction equation is expressed as:

$$\hat{y}(\mathbf{x}) = \mu + \mathbf{b}_D(\mathbf{x})^T \mathbf{B}_D^{-1} (\mathbf{y}_D - \mathbf{J}_n \mu), \quad (8)$$

where $\mathbf{J}_n$ is a n-vector of 1's, $\mathbf{B}_D$ is a n*n matrix, and $\mathbf{b}_D$ is a n*1 vector with elements, respectively:

$$B_{Dij} = \left\| \mathbf{x}_{Di} - \mathbf{x}_{Dj} \right\|, 1 \le i, j \le n \ \text{ and } \ b_{Di} = \left\| \mathbf{x} - \mathbf{x}_{Di} \right\|, 1 \le i \le n. \quad (9)$$

It can be noted that the function form of the predictor in Eq. (4) for Kriging is very similar to that of the predictor in Eq. (8) for RBF. In fact, Kriging is a RBF interpolation with Gaussian basis functions, except that the scaled distance is used and statistical approaches are applied to tune the correlation parameters θ_h in Kriging. The RBF method, unlike Kriging, cannot provide the information of prediction error. Therefore, some of the sequential sampling approaches introduced below are not applicable when the RBF is used as the metamodeling technique.

Existing Sequential Sampling Approaches

As mentioned before, sequential sampling approaches are typically based on optimality criteria for experimental design. Here, we introduce three sequential sampling approaches, i.e., the Entropy approach, the IMSE approach, and the Maximin Distance approach. We also review the MSE approach, which is a variant of the Entropy approach. The Entropy approach, the IMSE approach, and the MSE approach are sequential approaches with adaptation to existing Kriging metamodels, and hence they are not suitable for other types of metamodels. The Maximin Distance approach is a sequential approach without adaptation and it is not limited to Kriging.

Entropy Approach

Shannon [32] used entropy to quantify the "amount of information". Currin, et al [13,15] introduced the entropy criterion for design of computer experiments under the

Bayesian framework. Given the existing sample set $\mathbf{X}_P$ (with all l previous sample points $\mathbf{x}_{P1}, \mathbf{x}_{P2}, ..., \mathbf{x}_{Pl}$) and the existing Kriging metamodel (created based on $\mathbf{X}_P$), the Entropy approach is to select a new sample set $\mathbf{X}_C$ (with m sample points $\mathbf{x}_{C1}, \mathbf{x}_{C2}, ..., \mathbf{x}_{Cm}$) to maximize the amount of information obtainable from the new sample set. It can be shown [8] that this criterion is equivalent to:

$$\max_{\mathbf{X}_C} \left|\mathbf{R}_A\right| * \left|\mathbf{J}_{l+m}{}^T \mathbf{R}_A{}^{-1} \mathbf{J}_{l+m}\right|, \tag{10}$$

where $\mathbf{R}_A$ is the correlation matrix of all the $l + m$ sample points in $\mathbf{X}_A = \mathbf{X}_C \bigcup \mathbf{X}_P$; $\mathbf{J}_{n+m}$ is an $(n+m)$-vector of 1's. The correlation parameters θ_h (h=1,..,k) in the correlation matrix $\mathbf{R}_A$ are set to the same as those in the existing Kriging metamodel, which makes the new sample points adapted to the existing metamodel.

MSE Approach

The MSE approach is to select a new sample point $\mathbf{x}_C$ with the largest estimation of prediction error (Eq. (5)) in the existing Kriging metamodel (created based on the existing sample set $\mathbf{X}_P$), i.e.,

$$\max_{\mathbf{x}_C} s^2(\mathbf{x}_C), \tag{11}$$

This approach is in fact a special case of the Entropy approach if only one new sample point is selected at each stage. The proof is provided in Appendix A.

IMSE Approach

This integrated mean squared error (IMSE) criterion is introduced by Sacks, et al [14] for design of computer experiments. Given the existing sample set $\mathbf{X}_P$ and the existing Kriging metamodel (created based on $\mathbf{X}_P$), the IMSE approach is to select a new sample set $\mathbf{X}_C$ to minimize the integrated mean square error, i.e.,

$$\min_{\mathbf{X}_C} \int s^2(\mathbf{x}) d\mathbf{x}, \tag{12}$$

where

$$s^2(\mathbf{x}) = \sigma^2 \left[1 - \mathbf{r}_A(\mathbf{x})^T \mathbf{R}_A{}^{-1} \mathbf{r}_A(\mathbf{x}) + \frac{(1 - \mathbf{J}_{l+m}{}^T \mathbf{R}_A{}^{-1} \mathbf{r}_A(\mathbf{x}))^2}{\mathbf{J}_{l+m}{}^T \mathbf{R}_A{}^{-1} \mathbf{J}_{l+m}}\right]. \tag{13}$$

Refer to Eqs. (5) and (10) for the notations of parameters. The correlation parameters used in $\mathbf{R}_A$ and $\mathbf{r}_A(\mathbf{x})$ are the same as those used in the existing metamodel.

The differences between the IMSE approach (when used to generate one sample point at a stage) and the MSE approach are: 1) the IMSE approach uses an averaged MSE in the whole design space; 2) in the IMSE approach, the mean squared error is not only based on existing sample set $\mathbf{X}_P$, but also based on new sample set $\mathbf{X}_C$, while in the MSE approach, the mean squared error is only based on the previous sample points $\mathbf{X}_P$.

Maximin Distance Approach

The maximin distance criterion was proposed by Johnson, et al. [12] for computer experiments. Given the existing sample set $\mathbf{X}_P$, the Maximin Distance approach is to select a new sample set $\mathbf{X}_C$ to maximize the minimum distance between any two sample points in the sample set $\mathbf{X}_A = \mathbf{X}_C \bigcup \mathbf{X}_P$ (with all l + m sample points), i.e.,

$$\max_{\mathbf{X}_C} \left[\min_{\substack{\mathbf{x}_{Ci} \neq \mathbf{x}_{Aj} \\ 1 \le i \le m, 1 \le j \le l+m}} \left(d(\mathbf{x}_{Ci}, \mathbf{x}_{Aj}) \right) \right], \tag{14}$$

where $\mathbf{x}_{Ci} \in \mathbf{X}_C$ (i =1,..,m) and $\mathbf{x}_{Aj} \in \mathbf{X}_A = \mathbf{X}_C \bigcup \mathbf{X}_P$ (j = 1,…,l + m) . This approach can be used for both Kriging and RBF methods. However, one potential limitation is that unlike other approaches reviewed here it cannot adapt new sample points to the information obtained from the existing metamodel (created based on $\mathbf{X}_P$).

Proposed Sequential Sampling Approaches

In this work, two sequential sampling approaches, i.e., the Maximin Scaled Distance approach and the cross-validation approach, are proposed. Both approaches take use of the information from the existing metamodels and are not limited to the Kriging method.

Maximin Scaled Distance Approach

The Maximin Scaled Distance approach is a modification to the Maximin Distance approach (Eq. (14)). The original Maximin Distance approach cannot utilize the information obtained from the existing metamodel. To make new sample points adapted to the information from the existing metamodel, we introduce a scaled distance approach:

$$\max_{\mathbf{X}_C} \left[\min_{\substack{\mathbf{x}_{Ci} \neq \mathbf{x}_{Aj} \\ 1 \le i \le m, 1 \le j \le l+m}} \left(d_s(\mathbf{x}_{Ci}, \mathbf{x}_{Aj}) \right) \right], \tag{15}$$

where,

$$d_s(\mathbf{v}, \mathbf{w}) = \sqrt{\sum_{h=1}^{k} a_h (v_h - w_h)^2}\ . \tag{16}$$

The weight parameters a_h reflect the importance of different variables identified from the existing metamodel. If one variable is more important than the other variables, more weights should be assigned to that variable. This approach is expected to lead to a better uniformity of the projection of sample points into the space made of those important variables and therefore to improve the quality of the information obtained. For example, consider an extreme case in which some input variables are totally irrelevant (a_h = 0), this approach should select sample points that uniformly fill the space made of the rest of variables.

When Kriging is used for metamodeling, the correlation coefficients θ_h could be used as a_h. In fact, the correlation coefficients to some extent are indictors of the smoothness or predictability along the x_h coordinate (e.g., the smaller the θ_h is, the smoother along x_h coordinate; if θ_h is close to zero, the influence of x_h is close to linear). In this case, the scaled distance d_s is the same distance used in the correlation function of Kriging (Eq. (3)). For other metamodeling techniques, the relative importance of variables can be obtained by global sensitivity analysis approaches, e.g., variance-based approaches by Sobol' [33].

Cross-validation Approach

For those metamodeling techniques with which an estimation of prediction errors are not provided, such as the RBF method, we propose a cross validation approach to estimate the prediction error. The basic idea of this method is to leave out one or several sample points each time, to fit a metamodel based on the rest of sample points, to use the

metamodel to predict the response on those leave-out points, and then to calculate the difference between the prediction and the real response. With this approach, no new sample points are needed for assessing the accuracy of a metamodel. Instead of using cross-validation for accuracy assessment, here we use it to estimate the prediction error for sequential sampling. The prediction error is estimated by averaging the cross-validation errors. Based on the existing sample set $\mathbf{X}_P$ with l points $\mathbf{x}_{P1}, \mathbf{x}_{P2}, \ldots, \mathbf{x}_{Pl}$, the prediction error on point $\mathbf{x}$ is estimated by:

$$e(\mathbf{x}) = \sqrt{\frac{1}{l}\sum_{i=1}^{l}(\hat{y}_{-i}(\mathbf{x}) - \hat{y}(\mathbf{x}))^2}\,, \qquad (17)$$

where $\hat{y}(\mathbf{x})$ denotes the prediction of the response on $\mathbf{x}$ based on the metamodel created based on all l existing sample points; $\hat{y}_{-i}(\mathbf{x})$ denotes the prediction of the response on $\mathbf{x}$ using the metamodel created based on l-1 existing sample points with point i moved out (i =1,2,…, l).

With the cross validation approach, a point with the largest prediction error (Eq. (17)) is selected as the new sample point. The idea is similar to the MSE Approach. It has been found from our preliminary study that in some cases the point with the largest estimated prediction error tends to be close to one of the existing sample points. To avoid that new sample points clustering around the existing sample points, we further modify Eq. (17) to take into consideration the distance between points:

$$\max_{\mathbf{x}_c}[e(\mathbf{x}_C) * \min_i(d(\mathbf{x}_C, \mathbf{x}_{Pi}))]. \qquad (18)$$

TEST SCHEME AND RESULTS

Test Examples

Table 1. Summary of Test Examples

Problem NO.	Number of Variables (k)	Features
1	2	Highly nonlinear
2	2	Waving
3	2	Waving in x_1 direction, Linear in x_2 direction
4	2	Highly nonlinear, Sharp change in some region
5	6	Highly nonlinear
6	12	Nonlinear

In this study, six examples are used to test sequential sampling approaches. Table 1 shows a summary of those examples including the variety of features they have. The first five examples are mathematical functions: function 1 (the Branin function), function 4 (the Goldstein-Price function), and function 5 (the Hartman 6 function) come from [34]; function 2 comes from [35]; and function 3 comes from [16]. Function 6, a shaft press fit problem for a V6 Engine, is provided by Ford Motor Company. The response (output) of this problem is the press fit pressure, while the inputs include shape parameters (housing outer diameter, housing inner diameter, bearing outer diameter and bearing inner diameter), material parameters (housing Young's modulus, bearing Young's modulus, housing Poisson's ratio, bearing Poisson's ratio), thermal parameters (housing thermal expansion, bearing thermal expansion), and other parameters (interference, temperature). The equations for the mathematical functions and 3-D figures for all the test examples are provided in Appendix B.

Test Scheme

The test scheme is shown in Table 2. The IMSE approach is not tested because without applying special techniques sample points tend to pile up [14]. We do not apply the Maximin Scaled Distance approach to RBF because the relative importance of each variable needed in this approach cannot be obtained directly as with Kriging. However, once the relative importance information is available, e.g., through methods such as sensitivity analysis discussed by Sobol (1993), this approach can be easily used for RBF or other metamodeling techniques.

As shown before, the Entropy approach, the Maximin Distance approach, and the Maximin Scaled Distance approach are capable to select one points or several sample points at each stage; while the MSE approach and the Cross Validation approach are designed to select one sample point at each stage. The multi-point per stage scheme needs much more computational efforts due to the high-dimensional optimization involved in searching for new sample set and therefore its implementation is much difficult. In this study, we test the Entropy approach by using the multi-point per stage scheme. We test all the other approaches by the one point per stage scheme (note: the MSE approach is equivalent to the Entropy approach in this case).

For the purpose of comparison, one-stage (os) sampling strategy is tested with the Optimal Latin Hypercubes Sampling (OLHS) algorithm [36]. The OLHS algorithm can efficiently select an optimal sample set from Latin Hypercube designs based on the entropy criterion. The Entropy approach (ep) is tested with the OLHS (the algorithm can also select sample points sequentially). Because the good space filling property of OLHS, it is also used to generate initial sample points for all other sequential approaches. Since the accuracy of metamodels is very sensitive to the selection of sample points and the OLHS is not unique, we use 10 different initial sample sets obtained from OLHS for each test. For the one-point per stage scheme, at each sequential stage, we choose points from a large set of random sample points (50000 points) based on the optimality criterion of a particular sequential approach. Table 3 shows the sample size for sequential sampling approaches (based on the number of variables (k) of the test examples). The sample size for the one-stage approach is the same as the total sample size for sequential sampling approaches. Here, the number of stages for multi-point per stage scheme (the Entropy approach) is three, while the number of stages for one-point per stage scheme is $8k$+1 if k<6 and $6k$+1 if k≥6.

Table 2. Test Scheme for Kriging and RBF

	Sequential Approaches		One-stage (os) Approach
	Multi-point Per Stage	**One-point Per Stage**	
Kriging	Entropy (ep)	MSE (mse), Maximin distance (md), Maximin scaled distance(msd)	OLHS
RBF	N/A	Cross validation (cv), Maximin distance (md)	OLHS

Table 3. Sample Size for Sequential Sampling Approaches

Number of Variables (*k*)	Total Sample Size	Initial Sample Size (for the First Stage)	Sample Size for Each Later Stage
<6	12*k*	4 *k*	4 *k* (1)*
≥6	9k	3 *k*	3 *k* (1)*

* For one-point per stage scheme, only one point is added at each stage.

The RMSE (root mean squared error) based on the confirmation tests is used to assess the accuracy of metamodels, i.e.,

$$RMSE = \sqrt{\frac{\sum_{i=1}^{nt} (\hat{y}(\mathbf{x}_{Ti}) - y(\mathbf{x}_{Ti}))^2}{nt}}, \qquad (19)$$

where $\mathbf{x}_{Ti}$ (i =1,2,..,*nt*) is a set of random sample points used for assessing the metamodel accuracy (here, *nt* = 50000); $\hat{y}(\mathbf{x}_{Ti})$ are the predicted responses and $y(\mathbf{x}_{Ti})$ are the real responses at $\mathbf{x}_{Ti}$, respectively;.

Comparisons of Sequential Approaches to One-Stage Approach

As mentioned in the last subsection, 10 different initial sample sets are used for each sampling approach and therefore there are ten sets of accuracy results for each sampling approach. Box plots are used to show the deviations of the accuracy (RMSE) of the final metamodels when 10 different initial sample sets are used (see, e.g. Figs. 1 and 2). Lower values of RMSE are preferred. The line in each box is for the median value of the 10 accuracy results. The lines extending from each end of the box are used to show the range of the results. A vertical dash line splits each plot into two parts: the left part is for Kriging and the right part is for the RBF method. For both Kriging and RBF, a dot line extending from the line in the box for the one-stage (os) approach is used as a basis for comparison. When the median line in a box for a sequential approach is below the dot line, it indicates that the median result from the sequential approach outperforms that of the one-stage approach. The results for function 1 and function 6 are shown in Figs. 1 and 2, respectively. Others are shown in Appendix C. For functions 1 and 6, because the performance of Kriging is far better than RBF, different scales are used for RMSE values.

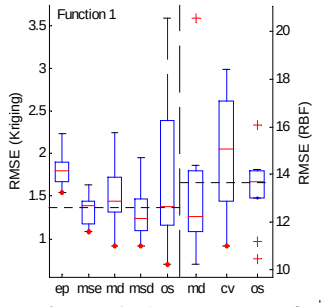

Figure 1. Assessment of Metamodels for Function 1

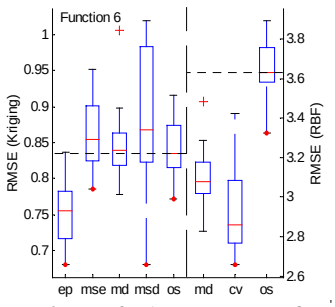

Figure 2. Assessment of Metamodels for Function 6

From the results shown in Figs. 1, 2 and Figs. A7~A10, it is found that in general the performance of sequential sampling approaches is comparable to that of the one-stage approach. It is found that, however, no sequential approaches with adaptation to existing metamodels can consistently (for all test problems) have better performance than approaches without adaptation (i.e., the maximum distance approach). In addition, it is not easy to identify an approach that is generally better than the other approaches.

For Kriging, the entropy (ep) approach outperforms other sequential approaches and the one-stage approach (os) for functions 4 and 6, and outperform other sequential approaches (but not the one-stage approach) for function 5; for functions 1 and 2, the Maximin Scaled Distance approach (msd) outperforms other sequential approaches and the one stage approach; for function 3, the MSE approach (mse) is the best in the sequential approaches. The performance of Maximin Distance approach (md) is never the best but seldom the worst, since the method does not adapt to previous data and tends to select sample points evenly filling the design space. One reason that none of these sequential approaches consistently outperforms the one-stage approach is that the information obtained from early created metamodels might be misleading. As shown in the 3-D plots in the appendix, many test problems are irregular. With a limited sample size, it is difficult for Kriging to catch the correct correlation parameters θ_h (h=1,.., k). This leads to large RMSE deviations for some sequential approaches that directly depend on the correlation parameters, e.g., the Maximin Scaled Distance approach. Further investigation also shows that the multiple variables in most of test problems selected appear to be equally important and none of these variables are ignorable. Due to this problem nature, when applying sequential approaches with adaptation, it is hard to make good use of the information obtained from the existing metamodel.

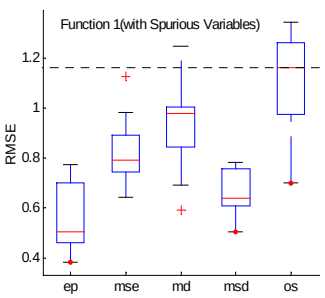

Figure 3. Assessment of Metamodels for Function 1 With Spurious Variables

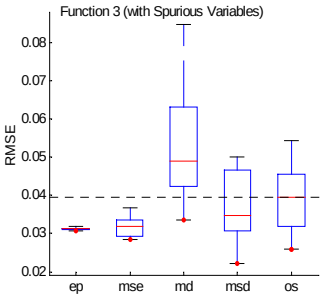

Figure 4. Assessment of Metamodels for Function 3 With Spurious Variables

To further study the adaptation performance of sequential approaches for Kriging, two spurious variables are added to functions 1~4. These two spurious variables have no influence on the response of all testing functions and the correlation parameters for the spurious variables ideally should be 0. It is expected that Kriging may work better to identify such structure and thus lead to better performance of sequential approaches (except Maximin Distance approach, which does not adapt to data). The test results are shown in Figs. 3, 4 and Figs. A8 and A9. It is found that in such cases, sequential approaches with adaptation do outperform the one-stage approach except for function 2, which is highly irregular. It is noted that the performance of the Maximin Distance approach

is always the worst in this test. From this study we note that the sequential approaches with adaptation hold promises for those functions with ignorable variables

For the RBF method, the Cross-validation (cv) approach we proposed significantly outperforms the one-stage approach for function 4 and also outperforms the one-stage approach for function 6. The maximin distance (md) approach outperforms the Cross-validation approach for the remaining functions and outperforms the one-stage approach for function 1, but neither improvement is significant. The mixed performance could be caused by the misleading information obtained from the metamodels created in the early stages for estimating the prediction error in some test problems.

Potential Usages of Sequential Approaches

As mentioned earlier, the major advantage of sequential approaches is that they can be used to monitor the process of metamodeling and to help a user decide when to stop the sampling process. This is desirable for many engineering problems if the simulation or analysis codes are computationally expensive. To save computational costs, designers can stop the sampling process whenever they think the quality or fidelity of metamodels is good enough or the performance of metamodels cannot be further improved (in that case we may change the metamodeling approaches). One issue, however, is how to assess the performance of metamodels without using additional confirmation samples. We propose to use the cross-validation error as an accuracy measure to overcome this difficulty. Fig. 5 shows the sequential metamodeling history for the engine problem (function 6). From the figure, we note that compared to the real RMSE history, cross-validation can roughly provide the general trend of the performance history of a sequential metamodeling process. We also note that after the number of sample points is large than 80, the improvement of the accuracy becomes small. Designers could use this information to decide whether to stop the sampling process.

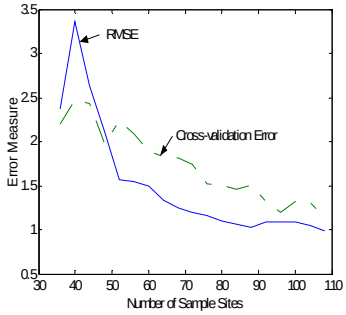

Figure 5. Metamodeling History for Function 6 (Kriging Based on Maximin Scaled Distance, Initial Sample Size: 36, Number of Stages: 73)

Sequential sampling approaches could also be very helpful for identifying an interesting design region and improving the accuracy of a metamodel in a narrowed interesting design region. Consider a situation in which we are only interested in a given range of a response, if we simply use space-filling approaches for the entire design space, many sample points will be wasted. By using sequential sampling approaches, the sample points can be added gradually and used to refine the approximation in the interesting region identified from the early created metamodels.

In addition, sequential approaches require small computational costs compared to one-stage optimal sampling approach. Assuming the number of input variables is k and the number of sample points is n, one-stage optimal sampling approaches involve solving a $k*n$- dimensional optimization problem, which could be computationally prohibitive if k and n are large. By using sequential approaches, the problem is divided into several smaller-dimensional optimizations, which makes the generation of sample points much cheaper.

CLOSURE

By testing various sequential sampling approaches for a set of test problems using different sequential approaches and different metamodeling techniques, we have shown the general applicability of sequential sampling approaches. We find that there is no guarantee that a sequential sampling approach with or without adaptation can improve the accuracy of metamodels compared to one-stage approaches. This is because depending on the samples and the metamodeling techniques used, the information obtained from the early created metamodels (e.g., the correlation parameters in Kriging, the estimation of prediction error, or the importance of variables) may be misleading. However, it has been observed in our study that the sequential approaches have shown promises for functions with variables having small or ignorable influences on a response. This could be helpful to many engineering problems in which only a part of variables are important. By using sequential sampling approaches, engineers can stop the sampling process when the model is accurate enough. Sequential approaches can also be used to identify an interesting design region and further refine the approximation in this region. In addition, sequential approaches require less computational costs compared to one-stage.

It is also observed from the study that if the unknown function is smooth or predictable, such as functions 1 and 6, Kriging can tremendously outperform RBF. Even if the unknown function is smooth or predictable in some coordinates (directions), e.g., function 3, Kriging can still largely outperform RBF. This is due to the use of the scaled distance with the correlation parameters determined by the sample data. However, when the unknown function is very irregular or unpredictable, such as function 2 (waving) and function 4 (sharp change in some region), the correlation parameters in Kriging could be very sensitive to sample points. This may lead to large deviations of the performance of a sequential sampling approach when different initial sample sets are used. It is especially the case for the proposed Maximin Scaled Distance approach, in which the correlation parameters are directly used. The variable importance from the global sensitivity analysis may be a good alternative to the correlation coefficients of Kriging. Therefore, one future work will be to further develop the Maximin Scaled Distance approach to make use of variable importance information. With this development, it is also expected that the approach will also work with other metamodeling techniques such as RBF.

ACKNOWLEDGEMENT

The supports from the National Science Foundation grants DMI-9896300 and CMS-0084477 are gratefully acknowledged.

A part of the research investigations was conducted during the summer interns of the principal author, supported by the Ford Motor Company.

REFERENCES

1. Kleijnen, J.P.C., 1987, *Statistical Tools for Simulation Practitioners*, Marcel Dekker, NY.
2. Simpson, T. W., Peplinski, J., Koch, P. N. and Allen, J. K., 1997, "On the use of statistics in design and the implications for deterministic computer experiments," *Design Theory and Methodology - DTM'97*, Sacramento, CA, ASME, DETC97/DTM-3881.
3. Sobieszczanski-Sobieski, J. and Haftka, R. T., 1997, "Multidisciplinary Aerospace Design Optimization: Survey of Recent Developments," *Structural Optimization*, 14, pp. 1-23.
4. Barton, R. R., 1998, "Simulation metamodels," in Medeiros, D. J., Watson, E. F., et al., eds., *Proceedings of the 1998 Winter Simulation Conference* (*WSC'98*), Washington, DC, IEEE, 167-174.
5. Booker, A. J., Dennis, J. E., Frank, P. D., Serafini, D.B., Torczon, V., and Trosset, M.W., 1999, "A rigorous framework for optimization of expensive functions by surrogates," *Structural Optimization*,17, 1-13.
6. Jin, R., Du, X. and Chen, W., 2001, "The Use of Metamodeling Techniques for Optimization Under Uncertainty," 2001 *ASME Design Automation Conference*, Paper No. DAC21039, Pittsburgh, PA, September 9-12, 2001, accepted by the *Journal of Structural & Multidisciplinary* (Accepted October 2001).
7. Meckesheimer, M., Barton, R. R., Simpson, T. W., Booker, A. J., 2001, "Computationally inexpensive metamodeling assessment strategies," *Proceedings of ASME 2001 Design Engineering Technical Conference and Computers and Information in Engineering Conference*, Pittsburgh, PA, DETC2001/DAC-21028.
8. Koehler, J.R. and Owen, A.B., 1996, "Computer experiments," in Ghosh, S.and Rao, C.R., eds, *Handbook of Statistics*, 13, 261 –308, Elsevier Science, New York.
9. Simpson, T.W., Lin, D.K.J. and Chen, W., 2001, "Sampling strategies for computer experiments: design and analysis," *International Journal of Reliability and Application*, 2(3), 209- 240.
10. McKay, M.D., Beckman, R.J. and Conover, W.J., 1979, "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code," *Technometrics*, 21(2), 239-45.
11. Owen, A., 1992, "Orthogonal arrays for computer experiments, integration, and visualization," *Statistica Sinica*, 2, 439-452.
12. Johnson, M. and Moore, L. and Ylvisaker, D., 1990, "Minimax and maximin distance designs," *Journal of Statistical Planning and Inference*, 26, 131-148.
13. Currin, C., Mitchell, T., Morris, M.D., and Ylvisaker, D., 1991, "Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments," *Journal of the American Statistical Association*, 86 (416), 953~963.
14. Sacks, J., Welch, W. J., Mitchell, T. J. and Wynn, H. P., 1989, "Design and analysis of computer experiments," *Statistical Science*, 4(4), 409-435.
15. Currin,C., Mitchell,T., Morris, M.and Ylvisaker, D.,1988, "A Bayesian approach to the design and analysis of computer experiments," Technical Report ORNL –6498, Oak Ridge National Laboratory.
16. Osio, I.G. and Amon, C.H., 1996, "An engineering design methodology with multistage Bayesian surrogates and optimal sampling," *Research in Engineering Design*, 8, 189-206.
17. Jones, D.R., Schonlau, M., Welch, W.J., 1998, "Efficient global optimization of expensive black-box functions," *Journal of Global Optimization*, 13(4), 455-492.
18. Cox, D. D. and John, S., 1997, "SDO: A statistical method for global optimization, in Alexandrov," N. and M. Y. Hussaini, eds., *Multidisciplinary Design Optimization: State of the Art*, SIAM, Philadelphia, 315-329.
19. Trosset, M.W. and Torczon, V. 1997, "Numerical optimization using computer experiments," Technical Report 97–38, ICASE, NAS, Langley Research Center, Hampton, VA 23681 –2199.
20. Jones, D.R., 2001, "A taxonomy of global optimization methods based on response surfaces," *Journal of Global Optimization*, 21(4), 345-383.
21. Sasena, M.J., Papalambros, P.Y., Goovaerts, P., 2000, "Metamodeling sampling criteria in a global optimization framework," *8th AIAA/USAF/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, Long Beach, CA.
22. Box, G. E. P. and Draper, N. R., 1987, *Empirical Model Building and Response Surfaces*, John Wiley & Sons, New York.
23. Cheng, B. and Titterington, D. M., 1994, "Neural Networks: A Review from a Statistical Perspective," *Statistical Science*, 9(1), 2-54.
24. Friedman, J. H., 1991, "Multivariate Adaptive Regression Splines," *The Annals of Statistics*, 19(1), pp. 1-141.
25. Hardy, R.L., 1971, "Multiquadratic equations of topography and other irregular surfaces," *J. Geophys. Res.*, 76, 1905-1915.
26. Dyn, N., Levin, D. and Rippa, S., 1986, "Numerical procedures for surface fitting of scattered data by radial basis functions," *SIAM Journal of Scientific and Statistical Computing*, 7(2), 639-659.
27. Simpson, T.W., Allen, J.K., and Mistree, F., 1998a, "Spatial correlation metamodels for global approximation in structural design optimization," *Proceedings of 1998 ASME Design Engineering Technical Conference*, Atlanta, GA.
28. Stinstra, E., Driessen, L., and Stehouwer, P., 2001, "Design optimization: some pitfalls and their remedies," *Third ASMO UK/ISSMO Conference on Engineering Design Optimization*.
29. Powell, M. J. D., 1987, "Radial basis functions for multivariable interpolation: a review," in Mason, J.C. and Cox, M.G. eds., *Algorithms for Approximation*, London, Oxford University Press.
30. Gutmann, H.-M., 2001, A radial basis function method for global optimization, *Journal of Global Optimization*, 19, 201-227.
31. Yang, R. J., Wang, N. Tho, C.H., and Bobineau, J. P., Wang, B. P., 2001, "Metamodeling development for vehicle frontal impact simulation," *Proceedings of ASME*

2001 Design Engineering Technical Conferences and the Computers and Information in Engineering Conference, Pittsburgh, Pennsylvania, DETC2001/DAC-2101
32. Shannon, C.E., 1948, "A mathematical theory of communication," *Bell System Technical Journal*, 27:379-423, 623-656.
33. Sobol', I.M., 1993, "Sensitivity analysis for nonlinear mathematical models," *Mathematical Modeling & Computational Experiment*, 1, 407-414 [translation of Sobol' (1990), Sensitivity estimates for nonlinear mathematical models, *Matematicheskoe Modelirovanie*, 2, 112-118 (in Russian)].
34. Dixon, L.C.W.and Szego, G.P., 1978, "The global Optimization problem: an introduction," in Dixon, L.C.W.and Szego, G.P. eds., *Toward Global Optimization*, 2, 1-15, North Holland, Amsterdam.
35. Sasena, M.J., 1998, *Optimization of Computer Simulations via Smoothing Splines and Kriging Metamodels*, Master Thesis, University of Michigan, 1998.
36. Sudjianto, A., Juneja, L., Agrawal, H. and Vora, M., 1998, "Computer aided Reliability and robustness assessment," *International Journal of Reliability, Quality and safety Engineering*, 5(2), 181-193.

APPENDIX

A Proof for MSE Approach

If there is only one new point $\mathbf{x}_C$ in $\mathbf{X}_C$, the $(l+1)*(l+1)$ correlation matrix $\mathbf{R}_A$ can be written as (noting $R_{A(l+1)(l+1)} = \rho(\mathbf{x}_C, \mathbf{x}_C) = 1$):

$$\mathbf{R}_A = \begin{bmatrix} \mathbf{R}_P & \mathbf{r}_P(\mathbf{x}_C) \\ \mathbf{r}_P(\mathbf{x}_C)^T & 1 \end{bmatrix}, \quad \text{(A1)}$$

where $\mathbf{R}_P$ stands for the $l*l$ correlation matrix for l previous sample points in $\mathbf{X}_P$ and the elements of the l-vector $\mathbf{r}_P(\mathbf{x}_C)$ are the correlation between $\mathbf{x}_C$ and the sample points in $\mathbf{X}_P$. It can be proved that the determinant of correlation matrix $\mathbf{R}_A$ for all the $l+1$ sample points can be written as:

$$|\mathbf{R}_A| = |\mathbf{R}_P|\,[1 - \mathbf{r}_P(\mathbf{x}_C)^T \mathbf{R}_P^{-1} \mathbf{r}_P(\mathbf{x}_C)]. \quad \text{(A2)}$$

And the inverse matrix for $\mathbf{R}_A$ can be written as:

$$\mathbf{R}_A^{-1} = \begin{bmatrix} \mathbf{R}_P^{-1} + \dfrac{\mathbf{R}_P^{-1}\mathbf{r}_P(\mathbf{x}_C)\mathbf{r}_P(\mathbf{x}_C)^T\mathbf{R}_P^{-1}}{1-\mathbf{r}_P(\mathbf{x}_C)^T\mathbf{R}_P^{-1}\mathbf{r}_P(\mathbf{x}_C)} & \dfrac{-\mathbf{R}_P^{-1}\mathbf{r}_P(\mathbf{x}_C)}{1-\mathbf{r}_P(\mathbf{x}_C)^T\mathbf{R}_P^{-1}\mathbf{r}_P(\mathbf{x}_C)} \\ \dfrac{-\mathbf{r}_P(\mathbf{x}_C)^T\mathbf{R}_P^{-1}}{1-\mathbf{r}_P(\mathbf{x}_C)^T\mathbf{R}_P^{-1}\mathbf{r}_P(\mathbf{x}_C)} & \dfrac{1}{1-\mathbf{r}_P(\mathbf{x}_C)^T\mathbf{R}_P^{-1}\mathbf{r}_P(\mathbf{x}_C)} \end{bmatrix} \quad \text{(A3)}$$

Therefore, $|\mathbf{R}_A| * |\mathbf{J}_{l+1}{}^T \mathbf{R}_A^{-1} \mathbf{J}_{l+1}|$ in Eq. (10) can be expressed by:

$$|\mathbf{R}_A| * |\mathbf{J}_{l+1}{}^T \mathbf{R}_A^{-1} \mathbf{J}_{l+1}| = \mathbf{J}_l{}^T \mathbf{R}_P^{-1} \mathbf{J}_l\, |\mathbf{R}_P| \left[1 - \mathbf{r}_P(\mathbf{x}_C)^T \mathbf{R}_P^{-1} \mathbf{r}_P(\mathbf{x}_C) + \frac{(1 - \mathbf{J}_l{}^T \mathbf{R}_P \mathbf{r}_P(\mathbf{x}_C))^2}{\mathbf{J}_l{}^T \mathbf{R}_P^{-1} \mathbf{J}_l} \right] \quad \text{(A4)}$$

It is easy to find that the only difference between Eq. (A4) and (5) is the const items before square brackets, which means that to maximize $|\mathbf{R}_A| * |\mathbf{J}_{l+1}{}^T \mathbf{R}_A^{-1} \mathbf{J}_{l+1}|$ is actually to maximize the mean squared error as shown in Eq. (11). This makes it clear that MSE approach is really a special case of Entropy approach.

Test Problems

Problem 1 (Branin Function):

$$y = [x_2 - 5.1*(\frac{x_1}{2\pi})^2 + \frac{5x_1}{\pi} - 6]^2 + 10*(1 - \frac{1}{8\pi})*\cos(x_1) + 10 \quad \text{(A5)}$$
$$x_1 \in [-5,10], x_2 \in [0,15]$$

Problem 2:

$$y = 2 + 0.01(x_2 - x_1^2)^2 + (1 - x_1)^2 + 2(2 - x_2)^2 + 7\sin(0.5x_1)\sin(0.7x_1x_2)$$
$$x_1, x_2 \in [0,5] \quad \text{(A6)}$$

Problem 3:

$$y = \cos(6(x_1 - 0.5)) + 3.1(|x_1 - 0.7|) + 2(x_1 - 0.5) + \sin(\frac{1}{|x_1 - 0.5| + 0.31}) + 0.5x_2$$
$$x_1, x_2 \in [0,1] \quad \text{(A7)}$$

Problem 4 (Goldstein-price Function):

$$y = [1 + (x_1 + x_2 + 1)^2(19 - 14x_1 + 3x_1^2 - 14x_2 + 6x_1x_2 + 3x_2^2)]* [30 + (2x_1 - 3x_2)^2 * (18 - 32x_1 + 12x_1^2 + 48x_2 - 36x_1x_2 + 27x_2^2)] \quad \text{(A8)}$$
$$x_1, x_2 \in [-2,2]$$

Problem 5 (Hartman6):

$$y = -\sum_{i=1}^{4} c_i \exp(-\sum_{j=1}^{6} a_{ij}(x_j - p_{ij})^2, \quad \text{(A9)}$$
$$x_j \in [0,1],\ j = 1,2,..,6$$

where, $a = \begin{bmatrix} 10 & 3 & 17 & 3.5 & 1.7 & 8 \\ 0.05 & 10 & 17 & 0.1 & 8 & 14 \\ 3 & 3.5 & 1.7 & 10 & 17 & 8 \\ 17 & 8 & 0.05 & 10 & 0.1 & 14 \end{bmatrix}$, $c = [1\ 1.2\ 3\ 3.2]^T$

and $p = \begin{bmatrix} 0.1312 & 0.1696 & 0.5569 & 0.0124 & 0.8283 & 0.5886 \\ 0.2329 & 0.4135 & 0.8307 & 0.3736 & 0.1004 & 0.9991 \\ 0.2348 & 0.1451 & 0.3522 & 0.2883 & 0.3047 & 0.6650 \\ 0.4047 & 0.8828 & 0.8732 & 0.5743 & 0.1091 & 0.0381 \end{bmatrix}$

3-D Plots for Test Examples Note: 1). All the coordinates are normalized to [0 1]; 2). For problem 5 and 6, all the variables that are not used for plotting are fixed at 0.5.

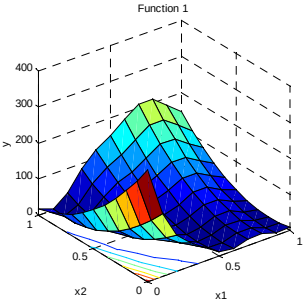

Figure A1. 3-D Plot for Function 1

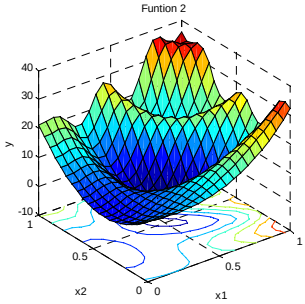

Figure A2. 3-D Plot for Function 2

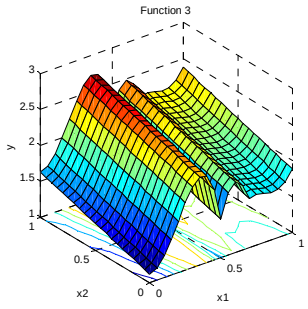

Figure A3. 3-D Plot for Function 3

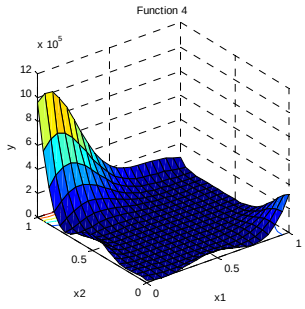

Figure A4. 3-D Plot for Function 4

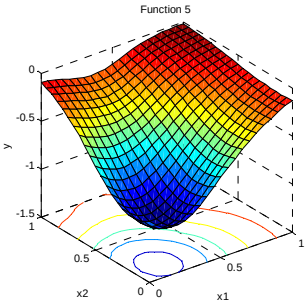

Figure A5. 3-D Plot for Function 5

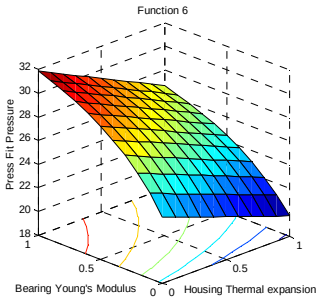

Figure A6. 3-D Plot for Function 6

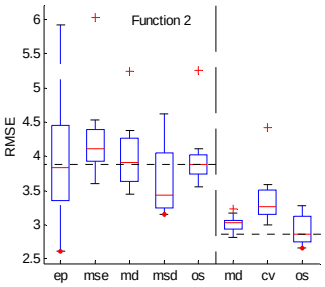

Figure A7. Assessment of Metamodels for Function 2

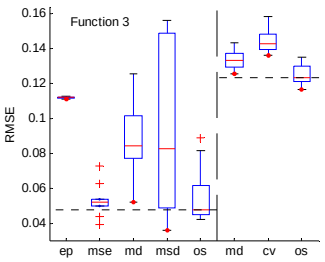

Figure A8. Assessment of Metamodels for Function 3

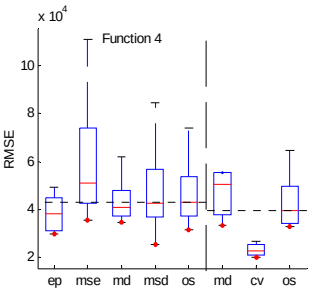

Figure A9. Assessment of Metamodels for Function 4

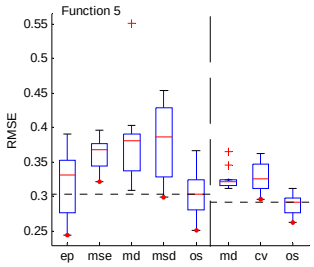

Figure A10. Assessment of Metamodels for Function 5

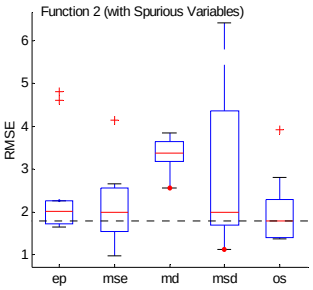

Figure A11. Assessment of Metamodels for Function 2 With Spurious Variables

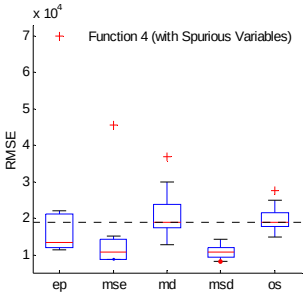

Figure A12. Assessment of Metamodels for Function 4 With Spurious Variables

Test Results

Note: For Figs. A7~A10, the parts on the left of dash lines are for Kriging, while the parts on the right are for RBF. Figs. A11~A12 are only for Kriging.