

04/23/2003

DETC2003/CIE-91

GENERIC SEQUENTIAL SAMPLING FOR METAMODEL APPROXIMATIONS

Cameron J. Turner, Matthew I. Campbell, Richard H. Crawford
Department of Mechanical Engineering
The University of Texas at Austin

ABSTRACT

Metamodels approximate complex multivariate data sets from simulations and experiments. These data sets often are not based on an explicitly defined function. The resulting metamodel represents a complex system's behavior for subsequent analysis or optimization. Often an exhaustive data search to obtain the data for the metamodel is impossible, so an intelligent sampling strategy is necessary. While multiple approaches have been advocated, the majority of these approaches were developed in support of a particular class of metamodel, known as a Kriging. A more generic, commonsense approach to this problem allows sequential sampling techniques to be applied to other types of metamodels. This research compares recent search techniques for Kriging metamodels with a generic, multi-criteria approach combined with a new type of B-spline metamodel. This B-spline metamodel is competitive with prior results obtained with a Kriging metamodel. Furthermore, the results of this research highlight several important features necessary for these techniques to be extended to more complex domains.

INTRODUCTION

Collecting data for use in defining a metamodel is a daunting task. High dimensionalities, expensive computational simulations, finite experimental trials, and limited resources with which to define a metamodel are only some of the major issues to be overcome. While an exhaustive sampling strategy might be attractive from a modeling standpoint, exhaustive sampling is generally infeasible for most applications. Instead, an intelligent method of sampling data is needed.

While classical design of experiments techniques offer methods by which to systematically sample the unknown space [Montgomery, 1997, Wu, 2000], these approaches do not efficiently search with respect to particular goals that may be important in the model. For example, it may be the goal of the metamodel to characterize the performance of a system (such as the limit of stable operation), instead of optimizing performance. Characterization of the system depends on the overall accuracy of the metamodel.

Existing techniques can be classified as either global search techniques, which explore unsampled regions of the unknown space, or local search techniques, which explore "interesting" regions previously sampled in greater detail. Sequential sampling techniques (SST) attempt to balance these competing goals.

Sasena's work [2002a, 2002b] pursued a search strategy that was initially global, and then explored local optima until convergence, when the search would revert to a global search. This paper uses a multi-objective combination of global and local criteria, and provides mechanisms by which to examine other local phenomena of interest in addition to regions surrounding optima. Periodic restarts were also considered as a complimentary modification. Furthermore, this approach is developed for use with alternative metamodel forms. Preliminary results indicate that the performance of this technique is at least as good as that achieved with prior techniques aimed at Kriging metamodels.

The remainder of this paper will briefly review metamodeling and an existing metamodeling technique, Kriging, as well as a B-Spline technique proposed by the authors. The issues associated with data acquisition for large and small data sets are also considered, along with potential criteria for sequential sampling techniques. Potential variations for the multi-criteria objective function are considered, prior to comparing the results of multiple trials obtained from the approach developed in this paper with B-spline models to prior results obtained for Kriging models. On the basis of the trials, it was concluded that the multi-criteria optimization problem is competitive with prior techniques, but that further work is necessary to extend the capabilities of this approach.

1. METAMODELING

Fundamentally, a metamodel is a model of a model. Metamodels are commonly employed to produce accurate but computationally efficient representations of an actual system, using data derived from multiple experimental and simulation data sources [Shyy, 2001] as shown in Figure 1. The goal of the metamodel is to replace these "expensive" simulations, direct measurements and experiments with a single "efficient" model.



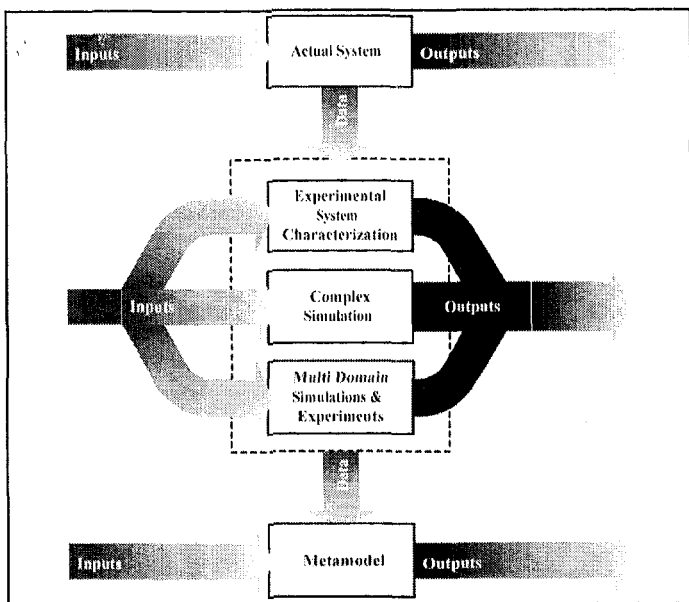


Figure 1. System, Model and Metamodel Relationships.

Metamodels are commonly used to facilitate optimization of a design. For instance, in order to optimize the design of a gas turbine engine, it is necessary to combine expensive computational models of fluid flow, with relatively inexpensive thermodynamic cycle calculations, and experimentally obtained aerodynamic models of the engine nacelle. Each iteration of the optimization process requires a new set of computational and experimental data. The time required to obtain this information is prohibitive and costly. Alternatively, a smaller data set obtained from these sources can be used to create a metamodel, which can be used to find an optimal solution in less time and at a reduced cost. As a design tool, improved metamodels can be used to map and search the design space for better designs.

Many types of metamodel formulations have been offered. Fundamentally, these types can be classified as geometric (including polynomial and spline based techniques), as stochastic (including Kriging and Radial Basis Function techniques), or as heuristic metamodels (including neural network models). This research focuses on comparing Kriging and a variant of spline based metamodels derived from the B-spline basis function (BSMs) developed by the authors and previously compared with another geometric metamodel, called Response Surface Models (RSMs). [Turner, 2002] In this work BSMs were shown to be more promising than RSMs in nonlinear problems.

Several surveys of metamodeling approaches have been conducted. [Barton, 1994 and 1998 and Simpson, 2001] In addition to Turner [2002], several papers have attempted to compare different metamodeling techniques for a variety of criteria including Sasena [1998], Hussain [2002], and Wang [1999] who surveyed metamodels with respect to their suitability for mechanical engineering applications.

Despite these many studies, few definitive conclusions have been drawn. RSMs, as introduced in most Design of Experiments (DOE) texts, [Montgomery, 1997, and Wu, 2000] are perfectly suitable for problems with few variables, and well behaved, near quadratic responses. Smoothing cubic splines (often referred to as smoothing splines) are an improvement over RSMs, but only at an increased computational cost.

Smoothing splines are distinct from B-spline models (BSMs), offer a promising geometric method and are used in this work.

The most popular stochastic or heuristic technique is currently Kriging. A competing technique, Radial Basis Functions (RBFs) are closely related mathematically to Kriging, but are somewhat more limited in their capabilities and therefore, Kriging models are currently more practical and because of their origins, have a number of SST methods developed specifically for this metamodel type. Neural Networks also have promise, but have seen limited applications in metamodeling. Thus, Kriging was selected as an appropriate technique for comparison with BSMs.

1.1. B-SPLINE METAMODELS

The main distinction between RSMs and BSMs is that RSMs are defined with polynomials, while BSMs are defined using parametric variables in a parametric space. A parametric space allows locally valid polynomials to be seamlessly blended together into a continuous function to form a BSM. The BSMs developed for this work are based on Non Uniform Rational B-Splines (NURBS), defined in 1-D by Equation 1.

$$P(u) = \frac{\sum_{i=1}^{n_c} B_i w_i N_{i,k}(u)}{\sum_{i=1}^{n_c} w_i N_{i,k}(u)} \quad \text{for } a \leq u \leq b \quad (1)$$

where $\{B\}$ is a vector defining the location of the n_c control points in \mathbf{R}^{n+1} space, w defines the weight of a particular control point, and $N_{i,k}(u)$ is the B-spline basis function in terms of the parameter u , as defined by Equations 2 and 3 [Rogers, 1990].

$$N_{i,k}(u) = \left(\frac{u - x_i}{x_{i+k-1} - x_i} \right) N_{i,k-1}(u) + \left(\frac{x_{i+k} - u}{x_{i+k} - x_{i+1}} \right) N_{i+1,k-1}(u) \quad (2)$$

$$N_{i,1}(u) = \begin{cases} 1 & \text{if } x_i \leq u < x_{i+1} \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

where $\{x\}$ is the knot vector, a sequence of parameter values defining the region of control point influence within the NURBS. Equations 2 and 3 are subject to the conditions given by Equation 4 [Rogers, 1990].

$$\frac{0}{0} = 0, \text{ and} \quad \sum_{i=1}^{n+1} N_{i,k}(u) \equiv 1 \quad \forall k \text{ and } u. \quad (4)$$

Higher order objects, such as surfaces, are produced with a tensor product. [Cohen, 2001] Control points, associated weights, and one or more parameters and knot vectors lends additional flexibility to NURBS, producing a highly flexible curve definition [Gopi, 1997].

The BSMs used here have all weights set to 1, the typical method of weight calculation [Piegl, 1997a, Rogers 1990], and use maximum error reduction spline-fitting techniques developed by Legault [2000] and Turner [2000]. This approach to fitting BSMs to data is distinctly different from traditional BSM metamodel approaches that match control points to data points in a one-to-one relationship [Simpson, 2001].

1.2. KRIGING METAMODELS

A South African Geologist, D.G. Krige, originally developed Kriging in the late 1950s and early 1960s. [Isaaks, 1989] Krige's original techniques were further developed during the 1970s, and by the late 1980s engineers interested in improving upon response surface models began to adapt his techniques. [Sasena, 1998 and Sasena, 2002b]

A Kriging metamodel superimposes a stochastic process model, upon a Response Surface Model with a Spatial Correlation Function (SCF), $R(u, x)$, defining the influence of u on x . with Equation 5 in terms of the covariance of the two points. Different choices for the functional form of the SCF can lead to different data "fits" of various qualities. Fortunately, many good choices seem to exist. Geostatistics tends to favor SCF's similar to that shown in Equation 5.

$$R(u, x) = \begin{cases} C_0 + C_1 \sqrt{(u-x)^2} = 0 \\ C_1 e^{\left(-3\sqrt{(u-x)^2}/a\right)} \sqrt{(u-x)^2 \neq 0} \end{cases} \quad (5)$$

where C_0 is referred to as the "nugget" effect (you either have a mineral or you don't), $C_1 + C_0$ defines the "sill" or a value at large distances from a sample point, and a defines a range of influence of the data. [Isaaks, 1989] In contrast, engineering metamodels (called Design and Analysis of Computer Experiments, DACE, Models) commonly employ a SCF similar to Equation 6. [Sasena, 2002a, and Sasena, 2002b]

$$R(u, x) = e^{-\theta|u-x|^p} \quad (6)$$

where θ defines the range of influence of the data ($\theta > 0$), and p defines the smoothness of the model ($0 < p < 2$) where increasing values of p lead to a smoother model. Both models interpolate given data points. [Sasena, 2002b]

The key difference between Kriging models used in Geostatistics and DACE metamodels is in the determination of the SCF parameters. Geostatistics utilizes a wealth of experimental data from centuries of geological exploration to provide experimentally valid parameters for C_0 , C_1 , and a from Equation 6. Kriging models used for engineering metamodels tend to lack this wealth of data, and therefore resort to an optimization approach where a maximum likelihood estimation (MLE) is maximized by finding optimal values of p and θ . [Sasena, 2002b] For large data sets, this is cumbersome and due to the nonlinearity of the problem, an optimal solution is not guaranteed. [Sasena, 2002b]

For interpolation, the Gaussian process modeled in Kriging is incredibly robust. It can represent virtually all the variations that might be represented with linear or quadratic response surface models. Since these models provide little benefit to the Kriging model, it is common to assume an "ordinary" Kriging model where the underlying polynomial RSM is simply a constant offset term. [Isaaks, 1989 and Sasena, 2002b]

In addition, Kriging allows for the estimation of the variance throughout the modeled space, through Equation 7.

$$\hat{\sigma}^2(x) = \sigma_z^2 \left(1 - \{r_x\}^T [R]^{-1} \{r_x\}\right) \quad (7)$$

where $\{r_x\}$ represents the affect of each of the n data points on the point of interest, x , and $[R]$ is the SCF "distance" between each data point, and σ_z is the variance of the data set. [Sasena, 2002b] This measure is small near data points, and large far

away from the data points. Thus the model tends to deviate from the underlying polynomial near data points, but follow a polynomial trend further away. The resulting Kriging model for the response \hat{y} at an unknown point x , denoted $\hat{y}(x)$, is defined with Equations 8 and 9.

$$\hat{y}(x) = \{f_x\}^T \{\beta\} + \{r_x\}^T [R]^{-1} (\{y\} - [F]\{\beta\}) \quad (8)$$

$$\{\beta\} = ([F]^T [R]^{-1} [F])^{-1} [F]^T [R]^{-1} \{y\} \quad (9)$$

where $\{f_x\}$ is a vector representing the power series basis of the underlying polynomial at the point x , $\{y\}$ is a vector of corresponding data outputs to the input data locations $\{x\}$, and $[F]$ is a matrix representing the power series basis for the polynomial at the corresponding data points in $\{x\}$.

2. METAMODEL DATA ACQUISITION

Metamodel data is generally available in one of two forms. There is either an excess of data, possibly derived from exhaustive search techniques, or there is a lack of data with corresponding limitations upon how much data can be acquired. There does not seem to be any well-defined middle ground where just the right amount of data exists. Therefore, we will examine the ramifications of these two conditions.

2.1. LARGE DATA SETS

The problem of having too much data is a much simpler problem to deal with than not having enough data. This problem generally requires some sort of filtering or smoothing so that the resulting model does not interpolate each data point, but instead approximates most data points. This can be achieved with both BSMs and Kriging models.

BSMs approximate data because of the maximum error reduction method used to fit the model. If a data point is well represented by the current model, its error will be low, and it will not be a likely candidate for further fitting of the model, unless the local behavior of the model changes significantly. This phenomenon often occurs in early stages of model fitting.

Kriging approaches the problem through the introduction of a nugget parameter, η . This parameter modifies the SCF used in the Kriging model as shown in Equation 10.

$$R(u, x) = (1 - \eta) e^{-\theta|u-x|^p} \quad (10)$$

where η is less than 1, but greater than or equal to zero. Typically, η is very small, on the order of $1e-5$, and is also determined via optimization. [Sasena, 2002b]

2.2. SMALL DATA SETS

A lack of data is a much more common situation. Obtaining sufficient data for a metamodel is the goal of sequential sampling techniques. A variety of strategies have been experimented with to intelligently sample unknown regions of the modeled space.

Most of these approaches have used a variety of criteria that determine the best point to sample. Sasena [2002a and 2002b] and Jin [2002] include surveys of these criteria. These criteria can be classified as either global or local criteria, based on their search scope. Table 1 is a summary of the criteria reviewed in these surveys, with the search scope of each criteria defined as appropriate.

Table 1. Potential Sequential Sampling Criteria.

Symbol	Name	Formula	Global	Local
EI	Expected Improvement Function	$EI = \begin{cases} (f_{\min} - \hat{y})\Phi(z) + \hat{\sigma}\phi(z) & \text{if } \hat{\sigma} > 0 \\ 0 & \text{if } \hat{\sigma} = 0 \end{cases}$	No	Yes
EI _g	Generalized Improvement Function	$EI_g = \sigma^g \sum_{k=0}^g (-1)^k \left(\frac{g!}{k!(g-k)!} \right) z^{g-k} T_k$ and, $T_0 = \Phi(z)$ $T_1 = -\phi(z)$ $T_k = -\phi(z)z^{k-1} + (k-1)T_{k-2}$	For large g values	For small g values
Cool	EI _g with a Cooling Schedule	Same as above but g decreases with successive iterations	Initially	In final iterations
K	Kushner's Criterion	$K = \Phi\left(\frac{(f_{\min} - \varepsilon) - \hat{y}}{\hat{\sigma}}\right)$	For large ε	For small ε
lcb	Lower Confidence Bounding Function	$lcb = \hat{y} - b\hat{\sigma}$	For large b	For small b
WB ₁	Watson & Barnes Threshold Bounded Criterion	$WB_1 = \Phi\left(\frac{f_{\min} - \hat{y}}{\hat{\sigma}}\right)$	No	Very local (Kushner's criteria with ε=0)
WB ₂	Watson & Barnes Regional Extreme Criterion	$WB_2 = \begin{cases} (f_{\min} - \hat{y})\Phi(z) + \hat{\sigma}\phi(z) + \hat{y} & \text{if } \hat{\sigma} > 0 \\ 0 & \text{if } \hat{\sigma} = 0 \end{cases}$	No	Yes
WB ₃	Watson & Barnes Surprise Minimization	$WB_3 = \min_x \max_v \text{Var}[Y(v) S \text{ and } x]$	Yes	No
Maxvar	Maximum Variance	$\hat{\sigma}^2(x) = \sigma_z^2 \left(\mathbf{1} - \{r_x\}^T [R]^{-1} \{r_x\} \right)$ for the Spatial Correlation Function $R(u, x) = e^{-\theta u-x ^p}$	Yes	No
Switch	Switching Criterion	1. Perform 5 iterations globally (with Maxvar) 2. Explore the minimum point until 3 consecutive iterations result in values within 0.1% of each other 3. Return to global iterations (step 1)	During Step 1	During Step 2

Sasena [2002a, 2002b] found that the most consistently well performing technique was the switching criterion (Switch in Table 1). For specific problems, other criteria might outperform this switching criterion, but only the switching criterion was consistently amongst the best. Fundamentally, this criterion searches globally, and then focuses on a local optimum until the optimum is well modeled before beginning a search for a new optimum.

This method is akin to the cooling schedule, established in the stochastic optimization method known as Simulated Annealing [Kirkpatrick, 1983] that initially searches globally before becoming a local search. However, its local search is focused only on the best-identified local optima. Finding an optimum is not always the focus of metamodeling, and there could be other areas of interest, such as regions where the curve changes rapidly, or regions where a discontinuity has been detected. However, this fundamental idea of the cooling schedule can be adapted to provide a more general approach.

3. THEORETICAL APPROACH

Working under the idea that simple criteria are the best, several criteria were developed for use in this research to search globally for new points to add, and locally for points near features of interest. Different approaches were considered to solve the resulting multicriteria optimization problem. These approaches included conducting a direct optimization on different functional combinations of these criteria, with weights determined by a cooling schedule, and employing an alternating

criteria cooling schedule. A method was developed and applied to a set of test functions for evaluation.

3.1. CRITERIA

Three criteria were developed for this research. One criterion is for a global search, based on Equation 8, while the remaining two criteria focused on different aspects of a local search. All three criteria are defined so that a minimization optimization problem is formed.

The first criterion, referred to as a variance criterion, v , performs a global search. Derived from Equation 7, this criterion represents the difference between the variance for the entire data set and the variance at the point. The resulting criterion is large near data points, and small far away from data points. The form of the criterion is shown in Equation 11.

$$v = \sigma_z^2 \left(\{r_x\}^T [R]^{-1} \{r_x\} \right) \quad (11)$$

The local criteria were based upon the current iteration of the metamodel. One criterion, referred to as the optimality criterion, b , located the current predicted optima, while the second criterion, referred to as the slope criterion, s , searched for regions of rapid change in the model (defined by large slopes). These criteria are shown in Equations 12 and 13 respectively.

$$b = m(x) \quad (12)$$

$$s = 1 - (m'(x))^2 \quad (13)$$

where $m(x)$ represents the norm of the current model of the unknown function and $m'(x)$ is its derivative.

All three criteria were normalized to run between 0 and 1, to eliminate scaling issues in their use. Furthermore, all criteria were defined consistent with a minimization objective function (i.e. small values are preferable).

The slope and optimality criteria will dramatically change as points are added, and are unlikely to be particularly accurate during early sampling iterations. This suggests that the variance criterion should be the initial focus of any search.

3.2. OPTIMIZATION

Several optimization strategies were considered. Fundamentally, these strategies included direct optimization with a cooling schedule, and alternating criteria strategies. Each type of formulation will be discussed independently.

3.2.1. DIRECT OPTIMIZATION APPROACHES

Direct optimization attempted to use optimization techniques to determine the optimum location for each sample. This formulation is shown in Equation 14.

$$\begin{aligned} \min f(x) &= \alpha v(x) + \beta s(x) + \gamma b(x) & (14) \\ \text{s.t. } g1: & x_l - x \leq 0 \quad (\text{Provides a lower bound on } x) \\ g2: & x - x_h \leq 0 \quad (\text{Provides an upper bound of } x) \end{aligned}$$

where $v(x)$ is the variance criterion, $s(x)$ is the slope criterion, and $b(x)$ is the optimality criterion. α, β, γ are the weights corresponding to these criteria.

This form requires a technique to determine the weights of the individual criteria for each optimization run. A cooling schedule approach was adopted in this research.

A multiplicative form of the objective function defined in Equation 14 was also considered, but ultimately discarded due to the fact that a multiplicative formulation eliminated the advantages of the multi-criteria optimization approach.

3.2.2. COOLING SCHEDULES

By using a cooling schedule, as opposed to optimizing the acquisition of data points through a single criterion, we were able to obtain results where we sampled a point that was the best compromise of our goals. Thus, we generally did not sample the “optimal” point for any single criterion. This is entirely acceptable from our point of view due to our imperfect knowledge of the criteria to begin with. The resulting cooling schedule was developed from the basis functions used to define a quadratic Bezier curve. The resulting weights are shown in Figure 2. Equations 15, 16, and 17 define the criteria used.

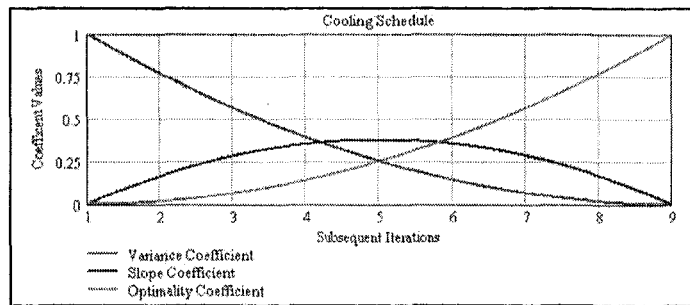


Figure 2. Criteria Cooling Schedule, for $C=1.5$ and $t_{max} = 9$.

$$\alpha(t) = \left(1 - \frac{t-1}{t_{max}-1}\right)^2 \quad (15)$$

$$\beta(t) = C \left(\frac{t-1}{t_{max}-1}\right) \left(1 - \frac{t-1}{t_{max}-1}\right) \quad (16)$$

$$\gamma(t) = \left(\frac{t-1}{t_{max}-1}\right)^2 \quad (17)$$

where C is a positive weighting factor on the influence of β , relative to α and γ , t is the current sample number (starting at $t=1$), and t_{max} is the limit on the samples of iterations we can afford. For $C \ll 1$, the associated slope criterion never becomes dominant. Experimentation with various values of C suggests that values between 1.25 and 1.5 yielded reasonable results. C should have a value greater than 2.

3.2.3. ALTERNATING CRITERIA

Sasena’s approach with alternating criteria is effectively an “either-or” cooling schedule. The cooling schedule conducts a single (global) criteria search for a predetermined number of iterations, followed by a second single (local) criterion search until the local search has defined the local optimum. The search then reverts to a global search for a predetermined number of iterations. An appropriate alternating strategy with three or more criteria is not readily apparent (as it is with two criteria), which led us to the idea of periodically resetting the cooling schedule. This is not an unusual idea for problems in the domain of simulated annealing, where the cooling schedule may be occasionally reset to assist the algorithm in overcoming local optima. Trial #5 uses this approach.

3.3. APPLIED METHODOLOGY

The methodology adopted for this project is based on the discussions in the previous two sections. MathCAD™ is used to define a multi-criteria optimization problem formulated to locate the optimum of an additive objective function. The coefficients on the criteria were based upon a cooling schedule defined at the beginning of the problem, with the capability to reset after a fixed number of iterations. It was assumed that the function would be initially sampled at two locations, serving as upper and lower bounds for the search. Subsequent samples would be taken within these bounds. The fundamental problem is defined in Figure 3.

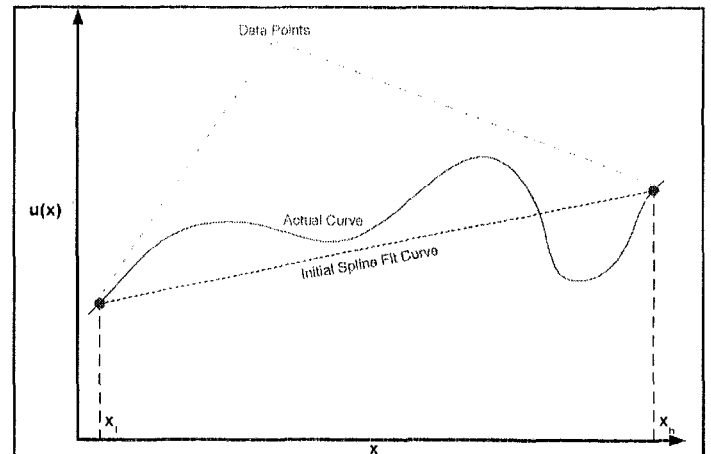


Figure 3. Initial problem formulation.

The unknown function in the problem was supplied by a function call to a subroutine so the algorithm did not have any

direct knowledge of the function. As far as the algorithm was aware, new data points were added as new points of interest were identified.

As a result of the spline fitting algorithm, the initial BSM is a line between the endpoints of the model, which becomes a parabola through the three points known after a single additional sample was taken at $t=1$. In successive iterations the BSM remains locally quadratic, while continuing to fit additional data points. With each iteration, a calculation of the deviation of the new data from previously predicted data could be made and used to determine if the point should be used to further refine the model based on a user-defined threshold. Due to limitations in MathCAD™, this threshold approach was not used and so all new data points were incorporated into subsequent BSM iterations.

At the end of each iteration, a calculation of the correlation of the models with the real function was made using Pearson's Correlation Coefficient defined in Crow [1960].

3.4. TRIAL PROBLEMS

Five trial problems were used in this research, based on four functions. Function 1 was obtained from Sasena [2002a, 2002b] and served as the basis for his initial trials. In this trial, Sasena's current technique will be compared with the results of the proposed technique. This function is defined in Equation 19 and shown in Figure 4.

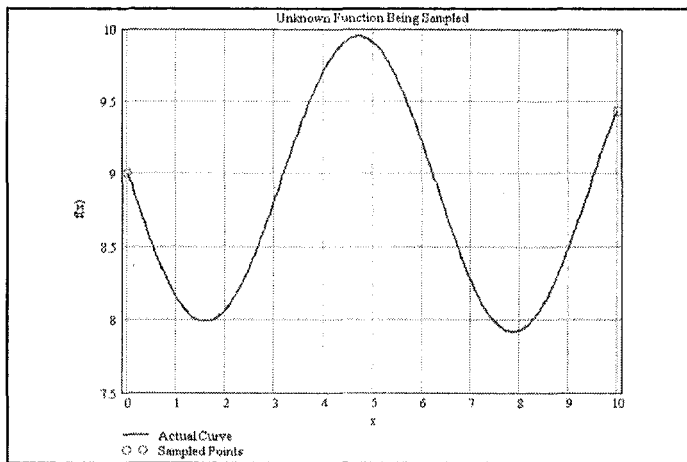


Figure 4. Trial Function #1 [Sasena, 2002a and 2002b].

$$f(x) = 10 - \sin(x) - e^{x/100} \quad (18)$$

This function is an excellent test of the metamodels, with two distinct, but very similar local optima. Accurately modeling these optima is very important if the metamodel is to be later used for optimization.

Trials 2 through 5, compare the performance of Kriging models and BSMs using the proposed sequential sampling technique introduced in Section 3. Trial Function 2 is based on Equation 19 with a discontinuous step input to the exponential term added at $x = 6$. The resulting function is therefore more challenging to fit and is shown in Figure 5.

$$f(x) = 10 - \sin(x) - if\left(x < 6, e^{x/100}, 0\right) \quad (19)$$

Trial Function 3 is a simple step input, which represents a considerable challenge in accurately fitting the function due to the discontinuity at $x = 6$ as defined in Equation 20.

$$f(x) = 4 + if(x < 6, 0, 1) \quad (20)$$

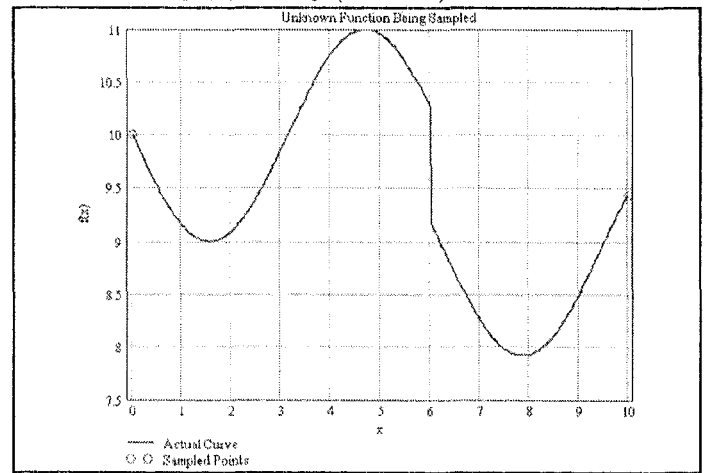


Figure 5. Trial Function #2.

Trial Functions 4 and 5 are defined with an identical function representing the Jacobian Condition Number of a 3 Degree of Freedom planar robotic manipulator derived in Turner [2002] as a test problem for comparing metamodel performance. Notably, this is a data set from a real problem, and is a data set representing a function that does not have a known closed form algebraic solution. The data is obtained from an exhaustive search of x as shown in Figure 6.

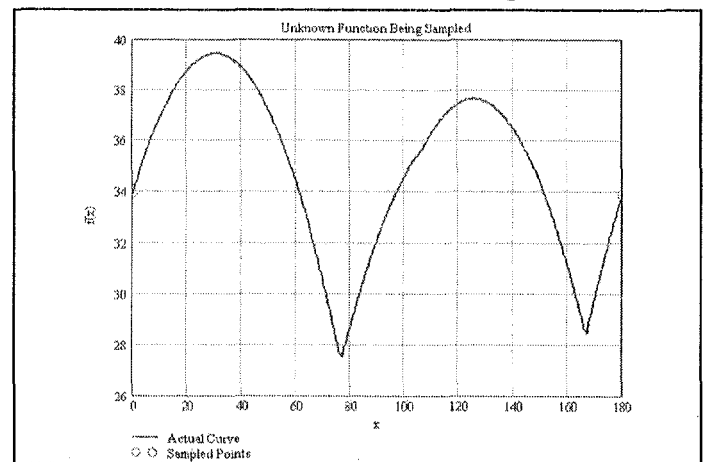


Figure 6. Trial Functions #4 and #5 [Turner, 2002].

The difference between the last two trials is the periodic reset feature added to the cooling schedule in Trial #5. Otherwise, they are based upon the same data set.

4. MODEL COMPARISONS

The resulting metamodels were evaluated during the sequential sampling process by examining the correlation coefficient between the model and the actual function. The goal was to obtain correlation coefficients in excess of 98% with a minimum number of data points. In addition, the accuracy of the metamodel with respect to the optimum was compared to the actual optimum found from the original function.

4.1. TRIAL 1: SMOOTH FUNCTIONS

Trial #1 used the smooth function from Sasena [2002a and 2002b] to compare the switching criterion for a Kriging

metamodel to the multi-criteria cooling schedule approach with BSM metamodels. The resulting metamodels for 6 data points are shown in Figure 7.

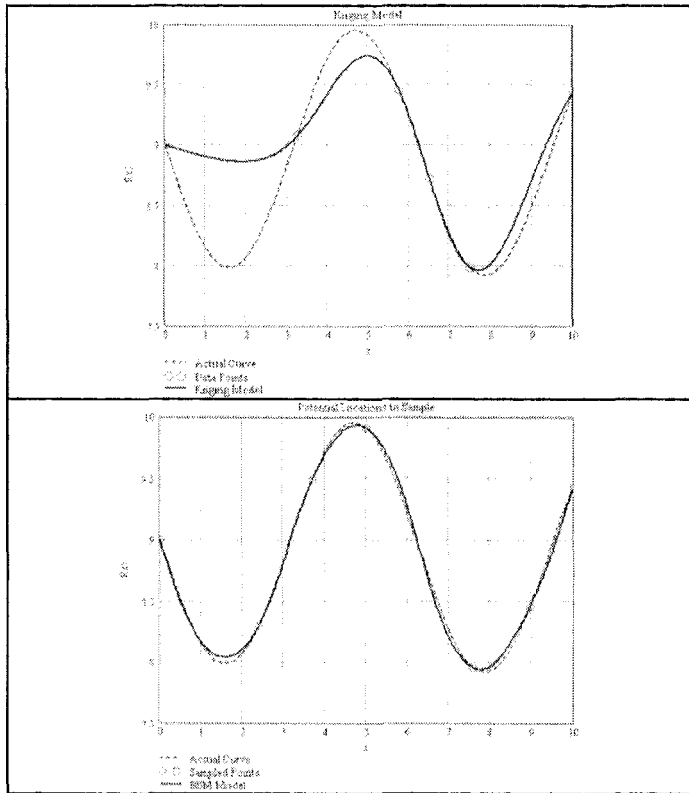


Figure 7. Resulting metamodels for Trial #1, each with 6 data points. The Kriging model is on the top, and the BSM model is on the bottom. The actual curve is shown with a dashed line.

With 6 data points, optimized as per the above discussion the BSM metamodel achieves more than a 99% correlation to the actual function while the Kriging metamodel achieves a correlation of 88%. With 10 optimized points, both models achieve correlations greater than 99%. Both models also predict optima within 2% of the actual optimal value of $x^* = 7.85$.

4.2. TRIAL 2: DISCONTINUOUS FUNCTIONS

Trial #2 introduces a discontinuity into the function used in Trial #1 at $x = 6$. Accurately modeling a discontinuity is an important property for a metamodel, and a substantial challenge. In this trial, both the Kriging and BSM metamodels utilized the same sequential sampling process. The results are shown in Figure 8 for 6 data points, and Figure 9 for 7 data points.

This trial dramatically demonstrates the affect of the slope criterion upon the system. The point added in these iterations is immediately above and to the left of the discontinuity, which attracts the slope criterion. The mixing of the variance criterion with the optimality criterion reduces the chance that the function will be sampled directly on the step, but also allows the function to “zero-in” on the discontinuity from both sides.

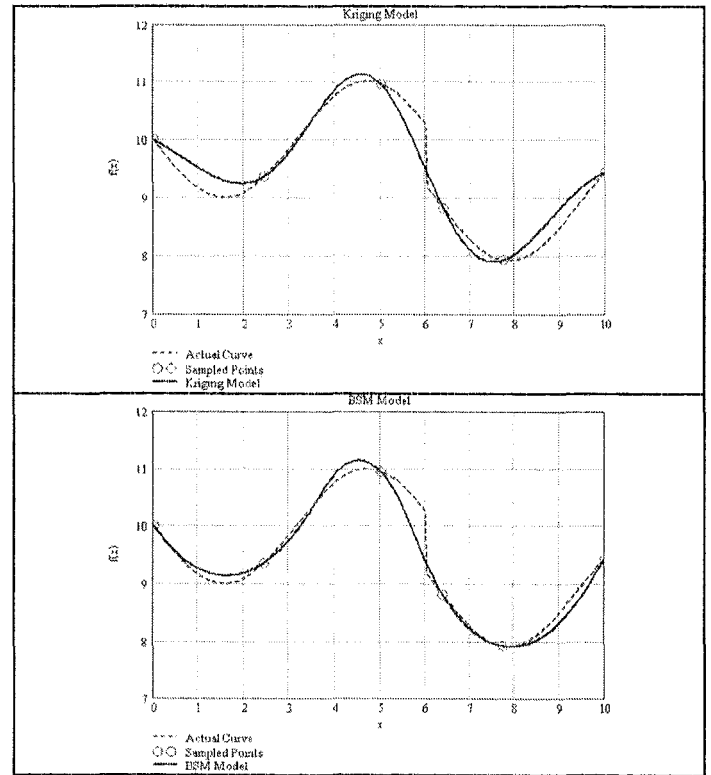


Figure 8. Resulting metamodels for Trial #2, each with 6 data points. The Kriging model is on the top, and the BSM model is on the bottom. The actual curve is shown with a dashed line.

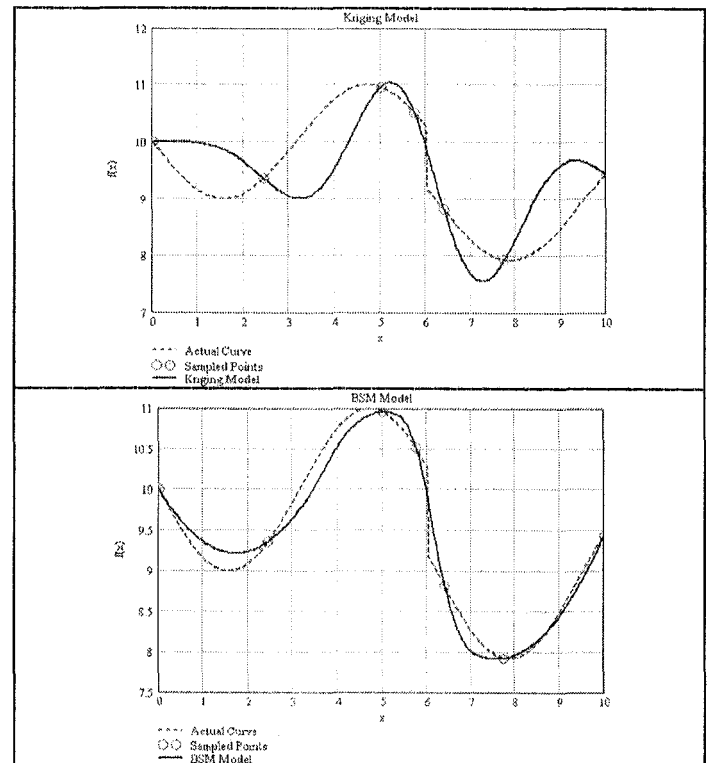


Figure 9. Resulting metamodels for Trial #2, each with 7 data points. The Kriging model is on the top, and the BSM model is on the bottom. The actual curve is shown with a dashed line.

In both cases (with 6 and 7 data points) the BSM fits the actual data to more than a 98% correlation (slightly better with 7 data points than 6), while the Kriging model lags behind. While with 6 data points, the Kriging Model explains the behavior of 97% of the data, the addition of another data point dramatically changes the behavior of the model, as can be seen by comparing Figures 8 and 9. This illustrates a key difference between the local changes that occur in a BSM due to the addition of new data, and the more global changes than may affect a Kriging model in response to new data. Finally, while both models predict the optimum within 10% of the actual value with 6 data points, the BSM is accurate to within 4%.

4.3. TRIAL 3: THE STEP FUNCTION

Even more challenging than the previous function is a true step function. In this case, no one optimal solution exists, since the function is a single value until $x = 6$. Once again, both the BSM and Kriging models were fit with the same sequential sampling criteria. The resulting models for 10 data points are shown in Figure 10.

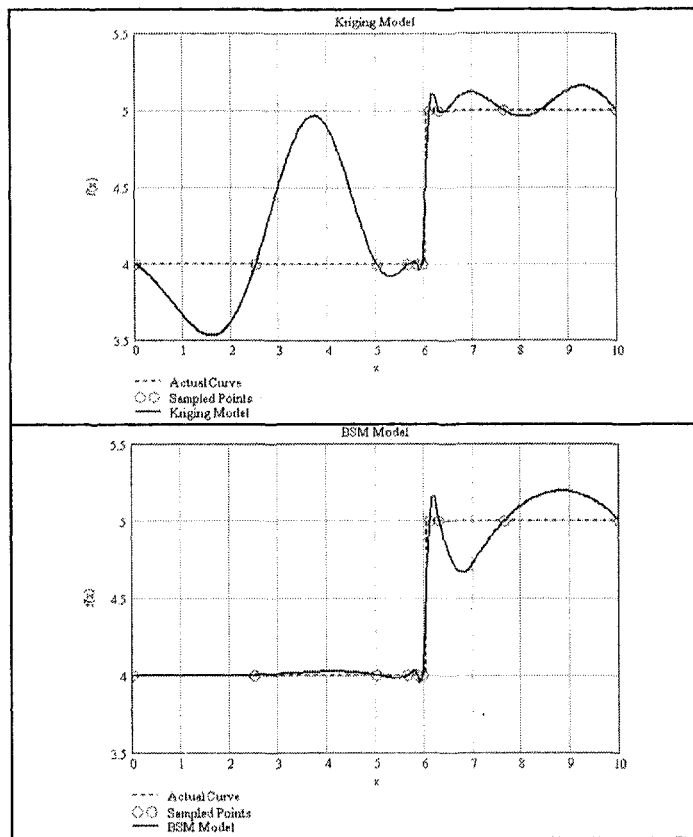


Figure 10. Resulting metamodels for Trial #3, each with 10 data points. The Kriging model is on the top, and the BSM model is on the bottom. The actual curve is shown with a dashed line.

While the final BSM achieves a correlation of almost 98%, the Kriging model never achieves more than 95% correlation to the actual function and achieves only a 79% correlation with 10 data points included.

The BSM model also does a much better job of representing the lower region of the step function, where the optimum is located. Based on the criteria used, it makes a great

deal of sense that the upper region of the step function, which is clearly not near the optimum, is poorly represented by the BSM. This is the result of the optimality criterion in the optimization. In this problem, the optimality criterion draws the sampling method to sample the region of the curve where $x < 6$ more extensively. Because no unique optimum exists, comparisons of the accuracy by which each model located an optimum are not meaningful in this case. It should be noted however, that the BSM represents the value of the optimum much more closely than the Kriging model does.

4.4. TRIAL 4: JACOBIAN CONDITION NUMBER

The function used in the final two trials was derived from real data for the Jacobian Condition Number, which measures the proximity of a robotic manipulator configuration to a singularity point. This data has no known closed form expression, making this an exceptional test of the capabilities of a metamodel. The results are shown in Figure 11.

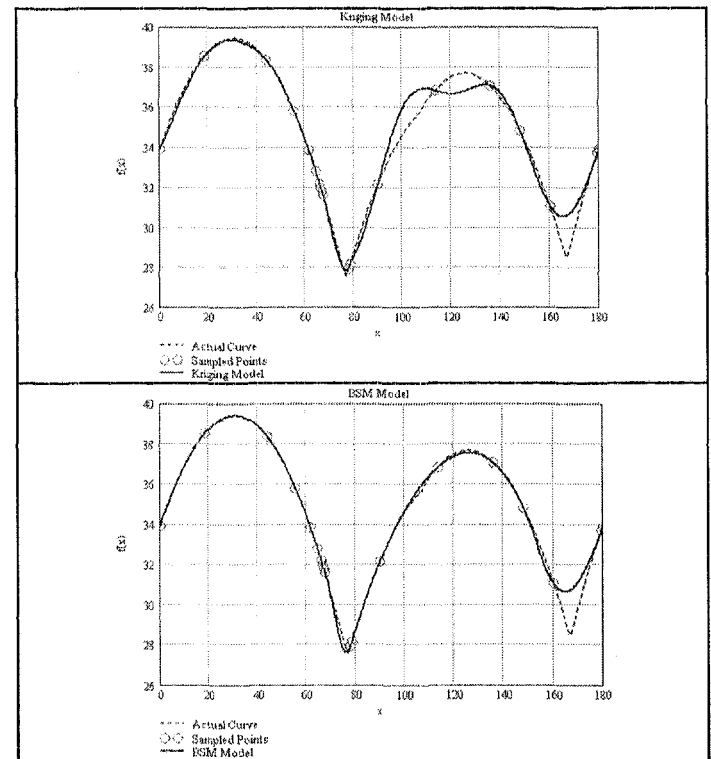


Figure 11. Resulting metamodels for Trial #4, each with 20 data points. The Kriging model is on the top, and the BSM model is on the bottom. The actual curve is shown with a dashed line.

Both metamodels achieve rates of correlation in excess of the goal of 98%. The BSM actually achieves better than 99% correlation. Both metamodels also correctly identify the global minimum located near $x = 75$. Notably, the Kriging model exhibits a little more variation between data points than the BSM model, and neither model did a particularly good job of sampling the second local minimum near $x = 165$. Due to the range of the data, the number of allowed data points was increased to 20 for this trial.

4.5. TRIAL 5: JACOBIAN CONDITION NUMBER

Because of the difficulty of identifying both optima in the previous trial, a cooling schedule restart was considered. In this trial, 13 points were collected during the initial application of the cooling schedule, which was then reset and an additional 13 points were collected for a total of 26 data points. The results prior to resetting the cooling schedule are shown in Figure 12, and for the entire process in Figure 13.

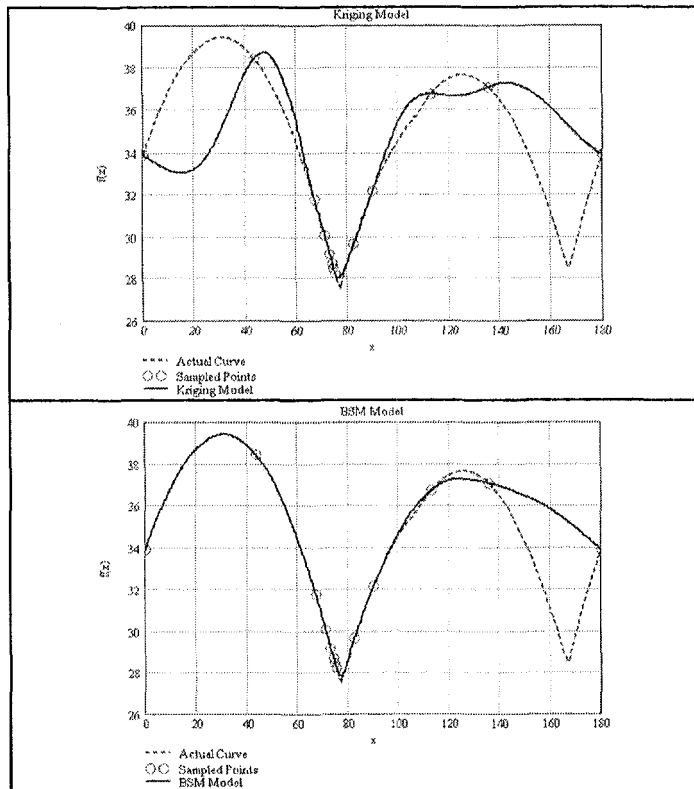


Figure 12. Resulting metamodellers for Trial #5, each with 13 data points. The Kriging model is on the top, and the BSM model is on the bottom. The actual curve is shown with a dashed line.

Notably, altering the duration of the cooling schedule changes the points selected by the sequential sampling technique. These changes are due to the different criteria weights obtained from the cooling schedule. As a result, the metamodeller appear different between resetting and nonresetting cooling schedules, even for the same number of data points.

After 13 data points have been collected, correlation for the BSM is at 86%, while the Kriging model correlates only 60% of the variation due to the actual function. Resetting the cooling schedule dramatically improves the correlation of both models, both of which eventually surpass the 99% correlation mark. In addition, the second local minimum was better identified and the initial global optimum is better defined than in Trial #4.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we have presented a new, multicriteria optimization technique to sequentially sample an unknown space. This technique was compared to the results of current sequential sampling techniques. This technique was evaluated using BSM metamodelling techniques, and a Kriging

metamodelling technique. Several conclusions can be drawn from these results. First, the multi-criteria solution to this optimization problem improves upon the prior work by Sasena [2002a, 2002b]. Second, in many cases, particularly for nonlinear functions, BSMs matched if not outperformed Kriging models in fitting the unknown functions with this method. However, the level of improvement was often statistically minor.

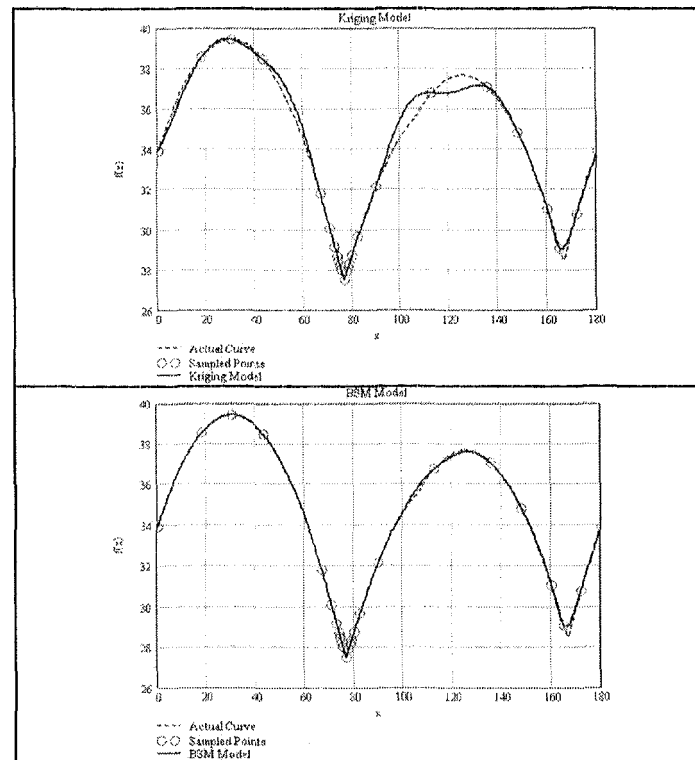


Figure 13. Resulting metamodellers for Trial #5, each with 26 data points. The Kriging model is on the top, and the BSM model is on the bottom. The actual curve is shown with a dashed line.

Additional work on this technique is needed, including the extension of this work to higher dimensional problems, successfully done by Sasena [2002a, 2002b] but not attempted here. For this approach to be successful in large numbers of variables a robust optimization approach for multimodal functions is necessary. As can be seen in Figure 14, the resulting objective function (shown with the dotted line at the top of Figure 14 which is a linear combination of the three lines at the bottom) can be very difficult to optimize. With one variable, it was possible to intelligently select starting points for multistart algorithms to find the true optima. For N variables, this is much more difficult. This challenge is not insurmountable, but simply highlights the need for robust optimization approaches.

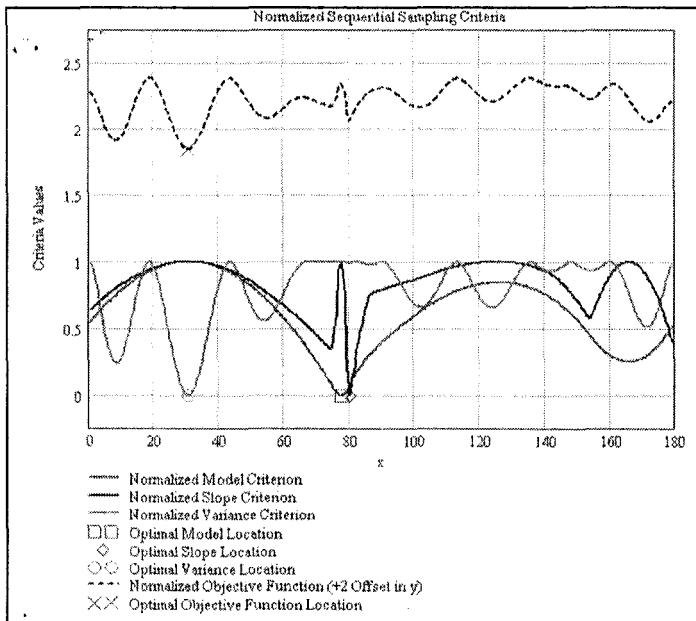


Figure 14. An example of the criteria used in this research, where the variance criterion dominates the objective function (Dotted Line at the Top).

Second, for BSMs to compete with established metamodeling techniques, further work is necessary to improve the performance of the fitting algorithm in order to deal with near singular matrix inversion problems. These issues in our implementation are being resolved with the development of a dedicated algorithm for this application.

Finally, additional criteria (i.e. curvature) should be developed and applied to allow for the exploration of other features of potential interest to the metamodel. The optimality, slope and variance criteria provide a solid basis for defining an accurate metamodel, but may not be adequate for other applications. A wider range of criteria capabilities is very desirable.

ACKNOWLEDGMENTS

This paper is approved for release by Los Alamos National Laboratory under LA-UR#-03-0980. The assistance and support of Robert Hollen, Dr. Peter Pittman and Dr. Christopher James of Los Alamos National Laboratory and the Department of Mechanical Engineering at The University of Texas in completing this work is greatly appreciated.

REFERENCES

Barton, R.R., "Metamodeling: A State-of-the-Art Review," *ACM Proc. of the 1994 Winter Simulation Conference*, Tew, J.D., et al, eds., pp. 237-44.

Barton, R.R., "Simulation Metamodels," *ACM Proc. of the 1998 Winter Simulation Conference*, Medeiros, D.J., et al, eds., pp. 167-74.

Cohen, E., Risenfeld, R., and Elber, G., *Geometric Modeling with Splines: An Introduction*, A.K. Peters, Natick, MA, 2001.

Crow, E., et al., *Statistics Manual*, Dover, Delaware, 1960.

Gopi, M., & Manohar, S., "A Unified Architecture for the Computation of B-Spline Curves and Surfaces," *IEEE Trans. On Parallel & Distributed Systems*, 1997, Vol. 8, pp. 1275-87.

Hussain, M.F., Barton, R.R., & Joshi, S.B., "Metamodeling: Radial Basis Functions, versus Polynomials," *European Journal of Operations Research*, 2002, Vol. 138, pp. 142-54.

Isaaks, E., and Srivastava, R., *An Introduction to Applied Geostatistics*, Oxford University Press, Oxford, UK, 1989.

Jin, R., Chen, W., and Sudjianto, A., "On Sequential Sampling for Global Metamodeling in Engineering Design," DAC-34092, *Proc. of the 2002 ASME IDETC Conferences*, Montreal, Quebec, Canada, September 29-October 2, 2002.

Kirkpatrick, S., C. D. Gelatt Jr., and M. P. Vecchi, 1983, "Optimization by Simulated Annealing," *Science*, Vol. 220, pp. 671-679.

Legault, J., "A Complexity Management Framework for Open Architecture Agile Manufacturing Systems," M.S. Thesis, The University of Texas at Austin, Austin, TX, May 2000.

Montgomery, D., *Design and Analysis of Experiments*, 4th Ed., John Wiley and Sons, NY, NY, 1997.

Piegel, L., and Tiller, W., *The NURBS Book*, 2nd Ed., Springer-Verlag, 1997a.

Rogers, D.F., & Adams, J.A., *Mathematical Elements of Computer Graphics*, 2nd Ed., McGraw Hill, New York, 1990.

Sasena, M., "Optimization of Computer Simulations via Smoothing Splines and Kriging Metamodels," M.S. Thesis, The University of Michigan, Ann Arbor, MI, 1998.

Sasena, M., et al., "Adaptive Experimental Design Applied to an Ergonomics Testing Procedure," DAC-34091, *Proc. of the 2002 ASME IDETC Conferences*, Montreal, Quebec, Canada, September 29-October 2, 2002a.

Sasena, M., "Flexibility and Efficiency Enhancements for Constrained Global Design Optimization with Kriging Approximations," Ph.D. Dissertation, The University of Michigan, Ann Arbor, MI, 2002b.

Simpson, T.W., Peplinski, J.D., Koch, P.N., & Allen, J.K., "Metamodels for Computer-based Engineering Design: Survey and Recommendations," *Engineering with Computers*, 2001, Vol. 17, pp. 129-50.

Turner, C., 2000, "Criteria Development for Actuator Resource Management," M.S. Thesis, The University of Texas at Austin, Austin, TX, August 2000.

Turner, C., "Metamodels for Planar 3R Workspace Optimization," CIE-34500, *Proc. of the 2002 ASME IDETC Conferences*, Montreal, Quebec, Canada, September 29-October 2, 2002.

Wang, N., & Ge, P., "Study of Metamodeling Techniques and Their Applications in Engineering Design," *ASME-MED Manufacturing Science and Engineering*, 1999, Vol. 10, pp. 89-95.

Wu, C., and Hamada, M., *Experiments: Planning Analysis, and Parameter Design Optimization*, John Wiley and Sons, 2000.