SAND2000-0333J

A Methodology for Selecting an Optimal Experimental Design for the Computer Analysis of a Complex System

Brian Rutherford
Sandia National Laboratories

Abstract

Investigation and evaluation of a complex system is often accomplished through the use of performance measures based on system response models. The response models are constructed using computer-generated responses supported where possible by physical test results. The general problem considered is one where resources and system complexity together restrict the number of simulations that can be performed. The levels of input variables used in defining environmental scenarios, initial and boundary conditions and for setting system parameters must be selected in an efficient way. This report describes an algorithmic approach for performing this selection.

Key Words: Computer-aided design; Experimental design; Large-scale simulation; Uncertainty analysis; Supercomputing.

I. Introduction

Analyses of complex systems performed in support of scientific and engineering projects are frequently based to a large extent on results generated through computer (or computational) simulation. Suppose we have specific concerns about a system or process that might operate in one of a number of environmental settings. Initial and boundary conditions specifying the systems interface with the environment vary from application to application and the system description itself requires specification of a number of parameters related to dimensions, material properties etc. Computer 'experiments', performed over a subset of the range of these 'inputs', provide information toward a better understanding of what might be expected of system performance. Alternatively, consider an engineering design problem. Some of the variables or inputs might fit into the description given above; others are to be optimized to determine the most appropriate, efficient or perhaps most robust design. Again, computer experiments can provide insight -- but the computer runs are generally expensive and only a restricted number of these experiments can be performed. As the complexity of the analysis increases it becomes more difficult to choose informative input levels by intuition alone. Consequently, more relevant information can often be obtained when experimentation is performed using a statistically designed set of input levels.

The use of computer algorithms to support experimental design for complex problems began when computing resources became readily available to most statisticians. Wynn (1970) and Mitchell (1974) provide early examples addressing D-Optimal designs for multiple regression problems. More recently, Sacks et al (1989a, b) use computer algorithms in the selection of simulation experiments designed for fitting a response model to limited computer generated data in several dimensions. Designs are selected that reduce functions of the expected mean square error in the fitted response model.

The purpose of the proposed methodology is to extend the development of algorithmic support in experimental design to accommodate still more complex situations where interest is focused on performance measures. Performance measures are assumed here to be the system characteristics

DISCLAIMER

that directly address the objectives of the analysis. They might be complex functions of a set of system responses perhaps measured over time. The performance measures, themselves, have a distribution function reflecting variability in the input variables and modeling uncertainty associated with the computer response. This report describes an algorithmic approach for choosing an experimental design (a set of supplemental computer runs) that strives to maximize the amount of information obtained concerning the performance criteria as measured through increased precision in their distribution.

The remainder of the report is organized into four sections. Section II gives a more detailed description of the problem, a review of relevant literature and an example. Section III gives the details of the system experimental design methodology proposed here. Section IV addresses the example in more detail providing results for the design constructed for a base case analysis and for designs reflecting changes in the assumptions. The report concludes with a Summary and Discussion Section.

II. Problem Description

General Formulation and Notation

A general formulation of the problem consists of the following. A computer experiment is performed through execution of the code after specification of a vector of inputs **x**. Some or all of the components of **x** may be elements of an assumed multivariate distribution $F_x(\mathbf{x})$. Assume that each p-vector of inputs $\mathbf{x}_i$ yields a q-dimensional vector of simulated responses $\mathbf{r}_i$. A set of n experiments will yield the n X q response data matrix. For the analyses considered here, it is assumed that the computer model or code is correct. Modeling uncertainty related to the physics or conceptual model providing the code is an important issue, but one not considered further here. Consequently, reference to the response model is used to address the response surface and not the computational simulation model. It is assumed further that (computer) experimentation is expensive enough that an effort to model the system response is useful -- this as opposed to assessing system performance on the basis of averages or weighted averages of the discrete responses alone (as one would with Monte Carlo simulation for example). This modeling might be accomplished (as in Figure 1) using a set of response surfaces $\mathbf{r}^*(\mathbf{x}) = r^{*j}(\mathbf{x})$, j=1,...,q for estimating the system response **r**. These response surfaces may be supplemented by some measure of uncertainty $\mathbf{h}(\mathbf{x})$ associated with the estimate. Uncertainty here is assumed to be the result of modeling uncertainty in the estimation of **r** at untested 'locations' in the input space but not with the choice of model form. It is assumed that an appropriate response model has been selected. Formally, each response model $r^{*j}(\mathbf{x})$, j=1,...,q is a function defined over some subset of Euclidean p-space.

The analysis is focused on understanding of the system or process through a set of performance measures $g(\mathbf{r})$ that are based, in some generally complex way, on the set of system responses. To simplify this presentation, I will assume a single performance measure $z=g(\mathbf{r})$. Performance measures can usually be constructed so that the objectives of the analysis are best accomplished by reducing their uncertainty. For example, if the objective of the analysis is to evaluate expected performance (as in the example given later), this comparison would be facilitated by reducing the uncertainty with which the performance criteria (expected total degradation in the example) can be predicted. In this paper, precision is assessed through variability in g, a quantity that will reflect the uncertainty resulting from probabilistically defined inputs as well as the response modeling uncertainty mentioned above. Note that it is the efficient reduction of this second component that provides the focus for the experimental design.

(Figure 1 goes here)

Figure 1 illustrates these relationships. In this figure, p = 2, x_1 and x_2, the two inputs, are shown as the axis of the three sets of contour plots. The figure also indicates q = 2 by the two overlapping response contours and the two overlapping contour maps providing estimates of their pointwise uncertainty. The third set of contour maps provides the relative probabilities associated with $F_x(\mathbf{x})$. The arrow providing feedback from the right side of the diagram to the experimental design indicates that these are three of the sources of information contributing to the design. In order for this approach to work, prior information (system knowledge) is required. Typically this prior knowledge can be obtained through more traditional statistical design in the process of screening and scaling experiments for the large array of inputs.

The information contained in these contour plots and their linkage to the performance measures will clearly impact the degree to which further experimentation at particular points (in the input space) will prove informative. More specifically, three considerations are important in determining the information content provided through further computer experimentation:

1. The sensitivity of the performance measure to the responses corresponding to the experimental points in the input space;
2. The 'closeness' in the input space of previous experimentation (and hence response model uncertainty at the point); and
3. The relative likelihood of the inputs at that point (as specified through $F_x(\mathbf{x})$, where applicable).

Each of these factors will influence the choice of an efficient experimental design.

Related Research

A substantial literature has been produced addressing issues related to selecting input levels for simulation or test levels for physical testing of large-scale systems. The statistical design literature for system testing has followed more classical lines of analysis focusing primarily on designs providing specific information regarding individual effects, interactions and low order polynomial response surfaces. Classical methods are used in the process discussed in this report to screen the inputs and to address questions of sensitivity that will be used to help reduce the dimension of and scale the input space. Application of these methods in the early stages of analysis generally provides the prior information required of the algorithmic procedures. A review of classical response surface methods is given in Box and Draper (1987).

The statistical design literature for system simulation on the other hand has focused on two, more general areas related to large-scale systems analysis – sampling methods and response modeling methods. These areas and some of the contributing literature are outlined in the following two paragraphs. Sampling approaches include Monte Carlo analyses, stratified approaches and directed approaches to random sampling. Strict Monte Carlo in the present notation would yield an experimental design taken as a random sample of (say) n locations in the input space based on probabilities specified through $F_x(\mathbf{x})$. For most applications, more informative designs can be obtained, on average, by using a stratified sample to assure information is distributed appropriately across the range of the inputs. Latin Hypercube sampling, McKay et al (1979) presents one method for accomplishing this task. In Latin Hypercube sampling, stratification is accomplished by taking one sample from each of n equal-probable intervals (or regions). Iman and Conover (1982), Tang (1993), Owen (1994) and Ye (1998) propose techniques for extending the Latin Hypercube for use with multivariate correlated or uncorrelated input distributions. Directed sampling approaches such as importance sampling can be used when portions of the region are thought to be more informative than others. In importance sampling, specific regions of the input space are given higher probability of selection (when compared to probabilities specified through $F_x(\mathbf{x})$) to

increase, on average, the relevant information obtained for a fixed number of experiments. The results are then weighted to remove the bias introduced. Ripley (1987), for example, includes a discussion of this sampling approach.

Many large-scale analyses, where (computer) experimentation is expensive, resort to constructing a model of the response function. The resulting surface is then used to estimate system performance. This is usually accomplished through one of two approaches that are referred to here as analytical or reliability methods and response modeling methods. Analytical approaches have been developed in the engineering literature primarily for reliability problems. These approaches consist of estimating a region of the response with a low order polynomial expansion based on values and derivatives calculated or estimated at a specific point. Ang and Tang (1984) provide an introduction to this approach; Robinson (1998) provides a review of significant developments. While this approach is popular in special areas of engineering, it is not general enough for the problems addressed here. Response modeling methods have been applied to a wide variety of applications requiring experimental design for computer experiments. The terminology response model is used here to include polynomial response surfaces, common in many earlier applications and also the more flexible surfaces based on local criterion such as kriging, see O'Connell and Wolfinger (1997) for example. A review of developments in response surface methodology is provided in Myers et al (1989) and Barton (1994).

Sacks et al (1989a,b) introduce an algorithmic approach to constructing a computer experimental design when the objective is to reduce some function of the expected response mean square error. Their approach uses universal kriging to estimate the response surface and also to obtain an approximation to the relative pointwise uncertainty. The approach takes advantage of the fact that when the (spatial) covariance function of the responses is assumed known – the reduction in variability provided through a specific experimental design can be predicted independently of the values obtained in the experiment; hence a response-variance based objective function depends only on the input locations of the design itself. This approach was expanded in Morris et al (1993) to utilize the results of derivative calculations at the points of experimentation.

The approach proposed in this report is a further attempt to take advantage of the ability to evaluate potential information provided through an experimental design but here, this is accomplished by evaluating the likely impact of a candidate design on the precision with which the performance criterion can be estimated. The approach is useful, not only for uncertainty analyses and prediction, but for engineering design and other large-scale system analyses, as the following example will demonstrate.

Example

This simplified example has been taken from a reliability problem involving a response variable related to the mechanical failure for a subassembly contained in a space vehicle. The subassembly is expected to encounter severe vibration during reentry into the atmosphere. The purpose of the analysis is to predict the expected performance of a new subassembly and compare it to the well-known performance of the current design. Experiments are to be conducted through a computational simulation model for the new design. The complexity of the simulation model leads to computational costs that limit the number of simulations that can be performed to n runs.

A 'true' (analytical) response surface is available for this example that permits evaluation of the proposed system experimental design methodology. A three-dimensional plot of this (generally unknown) response **r** is shown in Figure 2. Two inputs are involved in the analysis; both input

axes are transformed to [0,1] for convenience. The x_1-axis in Figure 2 indicates the frequency for the horizontal vibration where the magnitude is assumed to be constant. The scenario specified for the comparison with the present subassembly is assumed to include all frequencies in equal proportion during reentry. The recorded responses are adjusted so that integration across this range of frequencies will be comparable to a typical reentry sequence. The problem, then, is one-dimensional except that 'running the code' will only reveal a single response on the two-dimensional surface shown in Figure 2. The x_2-axis represents temperature at an interior point during reentry a quantity assumed to follow a uniform distribution on [0,1] for the base case analysis. These somewhat arbitrary settings are considered realistic enough to provide a fair basis for comparison with the present subassembly.

(Figure 2 goes here)

The response shown in Figure 2 represents a variable relating to degradation that could contribute to failure of the subassembly. Only the response values $\mathbf{r}(x_1,x_2)$ above a threshold D^* are considered damaging and those values greater then T are assumed to contribute to degradation at a rate proportional to $(\mathbf{r}(\mathbf{x}) - D^*)^c$. The performance measure 'expected integrated damage' is computed as

$$z = \int\int I\big(r(x_1,x_2),D^*\big)\big(r(x_1,x_2)-D^*\big)^c dx_1 dF_{x_2}(x_2) \qquad (1)$$

and where I is an indicator function of the form $I(x,y)$=1 if x>y and 0 otherwise.

The objective of the analysis is to decide whether of not to replace the subassembly with the new design on the basis of expected integrated degradation – clearly, the precision with which the performance measure can be estimated will impact the ability to make this decision. This example is pursued further in Section IV where results are presented. The damage threshold parameter D^* , the design size n, the assumed distribution of temperatures and the objectives of the analysis are varied to illustrate their impact on the experimental design.

III. System Experimental Design Methodology

Outline for the Proposed Methodology

For the system experimental design methodology described here, it is assumed that enough previous experimentation has been performed (or information is available from related sources) to provide an initial approximation of system knowledge – the system responses together with an assessment of uncertainty. As more information is obtained through computer experimentation, the approximation is updated. This approximation is accomplished here through 'realizations' modeling possible system responses as a function of the system inputs. An ensemble of these realizations is generated using 'stochastic simulation' to estimate and capture the uncertainty in the system response. Stochastic simulation used in this way is terminology common to the geostatistics literature; see, for example, Deutsch and Journel (1998) and its references. This construct (the ensemble) replaces the simpler representation of system knowledge given through a mean response and corresponding estimated pointwise uncertainty (as in Figure 1). It is used as an approximation to a probability measure over the system response space. Specifically this representation is used to accomplish two tasks in the proposed approach:

1. To provide a method for propagating response surface uncertainty through the function g to the performance criterion permitting a probabilistic estimate of the performance criterion; and
2. To provide a discrete approximation to a probability measure for the response that can be used to generate (hypothetical) data. These data are then used to evaluate the potential information provided by candidate experimental designs.

These uses are described further in the next two subsections: Characterizing the State of System Knowledge and The Design Evaluation Process.

(Figure 3 goes here)

Figure 3 provides a flow diagram illustrating the proposed methodology. Candidate experimental designs are selected using a random search procedure. Each candidate design consists of a set of n input locations. For each selection, the potential for providing relevant information is evaluated using hypothetical results obtained using the generated realizations. An algorithmic outline of this portion (the 'design evaluation iteration') of the approach follows.

Step 1. Use the initial set of generated realizations to get hypothetical response data at the (candidate) input locations.
Step 2. Evaluate this additional information together with the actual data and construct a set of 'augmented realizations' using the same methods as for the initial ensemble.
Step 3. Determine the impact of this new (hypothetical) information on the performance measures by evaluating the change in precision for g.
Step 4. Repeat Steps 1 through 3 for each or for some subset of the realizations (depending on the resources available).

This sequence is repeated for each set of n candidate input locations. No actual (computational) simulation is performed. The resulting distributions of the performance measure are computed and compared to determine the estimated change in precision and hence, estimate the information provided by the candidate. Finally the simulations that indicate the most potential relevant information through this iterative process are selected to be performed in support of the analysis. Sets of experiments that don't affect the system response estimates or those that effect them in ways that don't impact the distribution of the performance measure are eliminated from consideration.

This outline provides a brief overview of the components of the system experimental design methodology. More detail is provided in the following subsections. The first subsection consists of further discussion of methods for constructing an approximate probability measure over the response space that will be used to characterize system knowledge. The second, related, subsection addresses the assessment of potential information at unsampled input locations. Here, the approach to evaluating candidate designs is described in more detail. The third subsection is focused on selecting the candidate designs.

Characterizing the State of System Knowledge

The construction of an accurate representation of the state of system knowledge involves modeling based, in general, on results obtained through previous computer experimentation or related data. This step of going from limited sample information to a probabilistic specification of the system response requires statistical techniques. Most approaches rely on providing a smoothed estimate of the system response together with a pointwise estimate of the related uncertainty. Figure 1 illustrates this approach. The *'s on the upper two contour maps in Figure 1 indicate locations of previous physical testing and simulation, providing results to which the surface was fit. The second contour map indicates uncertainty in the surface as a function of these same experiments. One problem with this structure is that the response surface is generally a smoothed version of the response – perhaps not representing a likely response at all. A second problem relates to the characterization of response uncertainty. Even when an accurate pointwise estimate of uncertainty in the response surface can be computed, it may be difficult to propagate this uncertainty through a complex functional relationship g to determine its impact on the precision of the performance criteria. These problems lead to a different representation of system knowledge used for the proposed approach.

Simulation of response surfaces to construct an approximate probability measure over the response is an alternative approach that avoids the problems mentioned above. The objective of the simulations is to generate an ensemble of multi-dimensional surfaces that represent possible surfaces for the responses of interest. These generated realizations are usually conditioned to the prior data from previous simulation and/or physical testing. If generated correctly, the fraction of realizations in the ensemble (on average) that are within a small neighborhood of the response space will be consistent with the actual probability of occurrence based on the data. The ensemble of generated realizations will then capture the uncertainty in the true system response. An appropriate interpretation of this ensemble, analyzed pointwise, is that the probability of a response value at a specific input location smaller than any response level T_1, can be estimated by the fraction of realizations that have a value less than T_1 at this location. This interpretation can be extended to multiple locations or to the entire region. The probability that no response value in the entire region exceeds the level T_2, for example, can be estimated by the fraction of realizations where T_2 is not exceeded at any point.

This interpretation of the set of realizations is used to provide justification for the probabilistic interpretation of the state of system knowledge. Figure 4 illustrates this proposed framework for quantifying system knowledge. It shows an example of three generated realizations based on two input variables. Note that the realizations can provide an estimate of a mean response and of the pointwise uncertainty, but in addition, they retain likely relationships between neighboring input locations.

(Figure 4 goes here)

Formally, let $\mathbf{R} = R_i,\ i = 1,...,m$ represent the initial ensemble of realizations. This set is used to produce the initial estimated distribution for the performance measure. The ensemble will yield m values of the performance measure if it is a single value such as the probability of failure or optimal response value over the input space. It will yield m distributions of the performance measure if it is a distributed function of the responses. In the later case, these distributions are used to construct an equally weighted mixture distribution for g. It is the precision of the mixture that is of interest. In the next subsection, an ensemble of realizations will be used again to evaluate the distribution of the performance measure but after augmenting the data set with new (hypothetical) data at the candidate design locations.

The Design Evaluation Process

In the subsection above, the ensemble of realizations was used to provide an estimate of performance measure precision based on the current state of system knowledge. In this section, it will be used to provide hypothetical data that are in turn used to evaluate a candidate experimental design. First assume a candidate design of n input locations has been selected. Hypothetical data are generated by selecting a realization from the initial ensemble (see Figures 3 and 4) and assuming the appropriate response values at the desired candidate design input locations. Next, these hypothetical data are combined with the original data to examine their impact on the performance measure precision. Assume, for example, that realization j was selected. The augmented data set consisting of the original data together with the values at the n locations of R_j are used to generate a new ensemble of realizations $\mathbf{R}'_j = R'_{ji},\ i = 1,...,m'$. This data-augmented ensemble $\mathbf{R}'_j$ can be used in the manner described earlier to represent the new state of system knowledge and to estimate the resulting distribution of the performance measure.

Designs that provide information relevant to the performance measure will increase precision in its distribution.

This process is repeated for a subset of $k \leq m$ iterations yielding k distributions of the performance measure. In practice letting $k = m$ and using each realization one time appears to be a reasonable strategy. Each iteration of the design evaluation loop will yield a different estimate for the performance criterion. In general, large differences in these estimates will indicate discriminatory potential in the design. Formally, by conditioning on the results at the candidate locations, one can partition variability in the performance measure z into the components $\mathrm{Var}_r(\mathrm{E}_{r'}(z\backslash D))$ and $\mathrm{E}_r(\mathrm{Var}_{r'}(z\backslash D))$ where the subscripts r and r' indicate the operation is performed with respect to the approximate probability measures for the original responses and augmented responses respectively and where D represents the selected candidate design. See Parzen (1963) and McKay (1995) for further description of this partitioning of the variability in z. In the example discussed earlier and further in the next section one can refer to the within sample variance of the expectations:

$$\hat{\sigma}_w^2 = \frac{1}{k'} \sum_{i=1}^{k'} \left(z_i - \bar{z}\right)^2$$

where there are k' realizations in the ensemble used for each inner loop evaluation. We can also estimate the between sample variance of the expectations:

$$\hat{\sigma}_b^2 = \frac{1}{k} \sum_{j=1}^{k} \left(\bar{z}_j - \bar{\bar{z}}\right)^2$$

where there are k realizations in the original ensemble or the k realizations are some subset of that ensemble. Note that an uninformative design would yield roughly $\hat{\sigma}_b^2 = \hat{\sigma}_w^2/k$. The ratio $\hat{\sigma}_b^2/\hat{\sigma}_w^2$ (=$1/k$ for the uninformative design) provides an indication of the information provided by that design. A good design would yield a substantially higher value for $\hat{\sigma}_b^2$ and a lower value for $\hat{\sigma}_w^2$. It is this ratio that is used to select the best design for the example problem. Different project objectives, the design problem, for example (the minimization problem in the example) will require slightly different metrics for comparing designs.

Search Methods for Selecting Candidate Designs

Candidate experimental designs are compared using their estimated potential for increasing precision in the performance measures as an indication of their relative value. These evaluations are cheap compared to running the computational simulation models or physical testing. For large problems, however, the input space can be of high dimension and the number of candidate designs prohibitive. It is worthwhile to develop an efficient approach to selecting these candidate designs.

The selection of candidate sets of experiments presents an np dimensional optimization problem where n sets of p input levels are to be determined. Typically, the problem can be simplified by applying a grid in p dimensional input space and treating each set of n grid locations as a discrete alternative. In this formulation the problem is a combinatorial optimization problem. The objective function or fitness value for each set of candidate input locations is a measure of potential change in performance measure precision as estimated through the design-evaluation portion of the algorithm. Combinatorial optimization problems of the size likely to be encountered in application can be approached in a number of ways.

Two basic approaches mentioned here are: 1) finding a global optimal solution using combinatorial optimization methods; and 2) finding a near optimal solution using random search procedures. Branch and bound algorithms provide one example of combinatorial optimization methods that can be used to find an exact optimal solution over the input grid. These algorithms derive their benefit (compared to an exhaustive search of every possible candidate), by dismissing large numbers of solutions through a single comparison of objective function values. Szidarovszky (1983) applies branch and bound search methodology to problems of this type in the geosciences. Simulated annealing and evolutionary algorithms are random search methods that compare solutions iteratively, providing a near optimal solution where "near" depends on the efficiency of the algorithm and the number iterations involved (see Back (1996) for a description of evolutionary algorithms). Only a hybrid version of the later approach is used for the example discussed in this report.

IV. Example (revisited) – Further Description, Results and Discussion

The base case analysis considered here uses the performance measure specified in the integrand of (1) with parameter values, T=10 and c=2. The input and response dimensions are p=2, q=1 and a uniform distribution was used for x_2 (temperature). Ten initial data were obtained at input locations selected using a Latin Hypercube design with zero (rank) correlation (see Iman and Conover (1986) for the technique used). The response values r were taken from the "true" response shown in Figure 2 as if computer analyses had been performed at these input locations. These data were used to construct the initial response estimate illustrated in Figure 5. Figures 5b and 5c show the average estimated response. Figure 5b can be compared to Figure 2 (repeated as Figure 5a), an estimate of the response. Figure 5c and Figure 5d correspond to the contour maps of Figure 1. These figures include stars indicating the ten input locations of the original computer runs. The standard deviations in Figure 5d, like the average values in Figures 5b and 5c, are computed using the generated realizations in a pointwise fashion. Note that the standard deviation is 0 at the data locations because (for this example) the surfaces were constructed to interpolate the data exactly.

(Figure 5 goes here)

Figure 5e illustrates six of the twenty-five generated realizations that are used to approximate a probability measure over the response surface. The functional form of these surfaces is a model of the form:

$$r^*(\mathbf{x}) = \beta_o + \sum_i \beta_i x_i + \sum_i \sum_j \beta_{ij} x_i x_j + \varepsilon(\mathbf{x}), \qquad (2)$$

where $i = 1, 2$ and $j = 1, 2$. Here, $\varepsilon(\mathbf{x})$ is a stochastic process over the inputs $\mathbf{x}$ defined through a spatial covariance function $C(\mathbf{x}_k, \mathbf{x}_l) = C(|\mathbf{x}_k - \mathbf{x}_l|)$ for any $\mathbf{x}_k$ and $\mathbf{x}_l$ and where the (exponential) covariance function parameters were estimated using weighted least squares, see Cressie (1985).

This model has been used successfully for applications in geology and hydrology for several years. It was introduced as a model for engineering analyses in Sacks, et al (1989), and is discussed in detail in O'Connell and Wolfinger (1997). The approach used to generate the realizations is Sequential Gaussian Simulation. Deutch and Journal (1998) describe this

stochastic simulation approach and provide software that can be used to construct the realizations in two or three dimensions.

It is an open issue how designs might be formally evaluated for these algorithmic approaches. One common "metric" is to compare the information obtained through the experimentation to that obtained through Monte Carlo sampling to determine an equivalent amount of random samples required to obtain the same information. This quantity, however, is a very problem-dependent and provides no basis for comparing approaches applied to different problems. For this reason, I will rely here on a subjective comparison of results to the three criterion of a good design stated in Section II to assess the selected design.

The best design for the base case analysis is shown in Figure 6. The design points are the squares; previous data are the stars. The contours are the original contours based on the initial data, included here to help understand why the design was selected. From this figure, one can conclude that the design is spread through the region of high response variability at locations where the response would be contributing significantly to accumulated degradation. Hence it appears to satisfy two of the three criterion for a good design (the third was not applicable here because of the uniform distribution selected for x_2).

(Figure 6 goes here)

Deviations from the base case were planned to illustrate the flexibility and generality of the proposed approach. The number of design points, the functional form and parameter values associated with the function g and the distribution of the inputs were changed. Each parameter influences the value of information provided through subsequent computer experimentation and hence leads to a different optimal design. Two additional analyses are examined in this section. First, changes to the objective of the analysis to a minimization problem demonstrates the generality of the approach. This analysis might be performed in an engineering design analysis where packaging materials in the subassembly could be selected by the design engineer to control reentry temperatures. Next, changes to the form of the response surface are examined. Responses are modeled using a linear approximation in place of the quadratic model and using a constant as the structural component to be combined with the stochastic element. Because these modeling choices are somewhat arbitrary in actual application, one would hope for results that are somewhat robust to this type of change.

Modification to the optimal designs based on these changes are illustrated in Figure 7 and discussed in the following subsections. Figure 7 shows the optimal design, the base case design and the origional data for each case.

(Figure 7 goes here)

Increasing n, the Number of Design Points

Changing the problem to include 5 design points yields a design similar to the 3-point design, but where additional points are spread across the high average degradation region (see Figure 7a). Again, the optimal design points are located at locations of relatively high variability, Perhaps the best approach to using this algorithm is to obtain an optimal design for the number of computer experiments that are anticipated (if this is a reasonably low number) and then to choose from these the best $m < n$ experiments to run. At this time, the algorithm would be executed again. Note that the optimal design for a single run may not even be included in a multiple point design.

Changing Parameters of the Performance Measure

The threshold parameter was changed to $T = 12$ from the base case value of $T = 10$. This change resulted in the design focus being relocated in regions of higher estimated degradation (see Figure 7b). The design points were again located at points of relatively high estimated variability.

Changing the Input Distribution

Changing the temperature distribution to a truncated normal with mean .5 and standard deviation .25 resulted in a change in design that was focused on the more highly probably region in the center of the temperature axis (see Figure 7c). This is an example of the influence of the third criterion for a good design algorithm listed in Section II.

Changing Objectives of the Analysis

Changing the problem to a minimization problem resulted in a completely different experimental design as expected (See Figure 7d). The optimal design for this problem is spread across the lower temperature region) where the original data indicated the minimum degradation (when averaged across frequencies) might occur. Horizontal regions, where there were high values of degradation and little uncertainty were not included in the design.

Changing the Response Surface Form

Two changes were considered in the functional form of the response model. The first excluded the third term in Equation (2); the second excluded both the second and third term of that equation. The resulting designs were both similar to the base case design, but neither included the low temperature, low frequency point of that design (See Figures 7e, 7f). Apparently, these reduced models were not influenced to the same degree by the response value obtained in the lower left corner of the input space (note the differences in contours in that region). These changes in design indicate some sensitivity to the choice of response model. It is important for a specific application to choose a flexible class of response models that is reasonable for the problem to which it is applied.

V. Summary and Discussion

This report describes a system experimental design methodology -- an algorithmic approach to experimental design for large-scale computer analyses. This approach can be viewed, as an extension of previous algorithmic methodology developed to accommodate more general problems whose formulation includes a set of system performance measures. The approach consists of an algorithm where simulated results are used to capture possible responses at input locations specified for candidate experimental designs. Candidate designs are selected using random search methods and compared in terms of their potential increases in performance criteria precision. The design selected for the computational simulation runs is the design that is optimal according to this potential information criterion.

The flexibility of the approach is illustrated through the example in Sections II and IV. A design is determined for a base case analysis and the problem is changed in a number of ways to

illustrate how these changes are accommodated by changes in the design. The two dimensional example shows encouraging results for this approach which is now being tested on larger problems.

The methods used in this approach to algorithmic design are quite general. The approach can be applied to physical testing where inputs can be controlled for experimentation, to combined designs with both physical and computer experimentation or to computer analyses involving multiple levels of analysis, differing in their degree of physical detail (and hence cost). In these later types of analysis, the restriction to n tests would be expressed instead as a limitation of resources -- generally, making the selection of candidate designs more difficult.

References

Ang, A., and Tang, W. H. (1984). "Probability Concepts in Engineering Planning and Design - Volume II: Decision, Risk and Reliability", Wiley and Sons, New York.

Back, T. (1996), "Evolutionary Algorithms in Theory and Practice", Oxford University Press, New York.

Barton, R. R. (1994), "Metamodeling: A State of the Art Review", *Proceedings of the 1994 Winter Simulation Conference, Tew, J. D., Manivannan, S., Sadowski, D. A. and Seila, A. F. eds.*

Box, E. P. and Draper, N. R. (1987), "Empirical Model-Building and Response Surfaces", Wiley and Sons, New York.

Deutsch, C. V. and Journel, A. G. (1998), "GSLIB Geostatistical Software Library and Users Guide Second Edition", Oxford University Press, New York.

Iman, R. L. and Conover, W. J. (1982), "A Distribution-Free Approach to Introducing Rank Correlation Among Input Variables", *Communications in Statistics, Part B - Simulation & Computation*, 11, pages 311-334.

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*, Vol. 21, No. 2, pp. 239-245.

McKay, M. D. (1995), "Evaluating Prediction Uncertainty", NUREG/CR-6311, U. S. Nuclear Regulatory Commission, Washington D. C.

Meyers, R. H., Khuri, A. I. and Carter, W. H. Jr. (1989), "Response Surface Methodology: 1966-1988", *Technometrics*, Vol. 31, No. 2, pp. 137-157.

Mitchell T. J. (1974), "Computer Construction of 'D-Optimal' First-Order Designs", *Technometrics*, Vol. 16, No. 2, pp. 211-220.

Morris, M. D., Mitchell, T. J. and Ylvisaker, D. (1993), "Bayesian Design and Analysis of Computer Experiments: Use of Derivatives in Surface Prediction", *Technometrics*, Vol. 35, No. 3, pp. 243-255.

Morris, M. D. (1991), "Factorial Sampling Plans for Preliminary Computational Experiments", *Technometrics*, Vol. 33, No. 2, pp. 161-174.

O'Connell, M. A. and Wolfinger, R. D. (1997), "Spatial Regression Models, Response Surfaces, and Process Optimization", *Journal of Computational and Graphical Statistics*, Vol. 6, No. 2, pp. 224-241.

Owen A. B. (1994), "Controlling Correlations in Latin Hypercube Samples", *Journal of the American Statistical Association*, Vol. 89, No. 428, pp.1517-1522.

Parzen, E. (1962), "Stochastic Processes", Holden Day, San Francisco.

Ripley, B. D. (1987), "Stochastic Simulation", Wiley and Sons, New York.

Robinson, D. G. (1998), "A Survey of Probabilistic Methods used in Reliability Risk and Uncertainty Analysis: Analytical Techniques I", SAND98-1189, Sandia National Laboratories, Albuquerque, New Mexico.

Sacks, J., Schiller, S. B., and Welch, W. J. (1989a), "Designs for Computer Experiments" *Technometrics*, Vol. 31, No. 1, pp. 41-47.

Sacks, J., Welsh, W. J., Mitchell, T. J. and Wynn, H. P. (1989b), "Design and Analysis of Computer Experiments" *Statistical Science*, Vol. 4, No. 4, pp. 409-423.

Szidarovszky, F. (1983), "Multiobjective Observation Network Design for Regionalized Variables", *International Journal of Mining Engineering*, 1.

Tang, B. (1993), "Orthogonal Array Based Latin Hypercube", *Journal of the American Statistical Association*, Vol. 88, No. 424, pp.1392-1397.

Wynn, H. P. (1970), "The Sequential Generation of D-Optimal Experimental Designs", *Annals of Mathematical Statistics*, 41, 1655-1664.

Ye, K. Q. (1998), "Orthogonal Column Latin Hypercubes and Their Application in Computer Experiments", *Journal of the American Statistical Association*, Vol. 93, No. 444, pp.1430-1439.

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under contract DE-AC04-94AL85000.

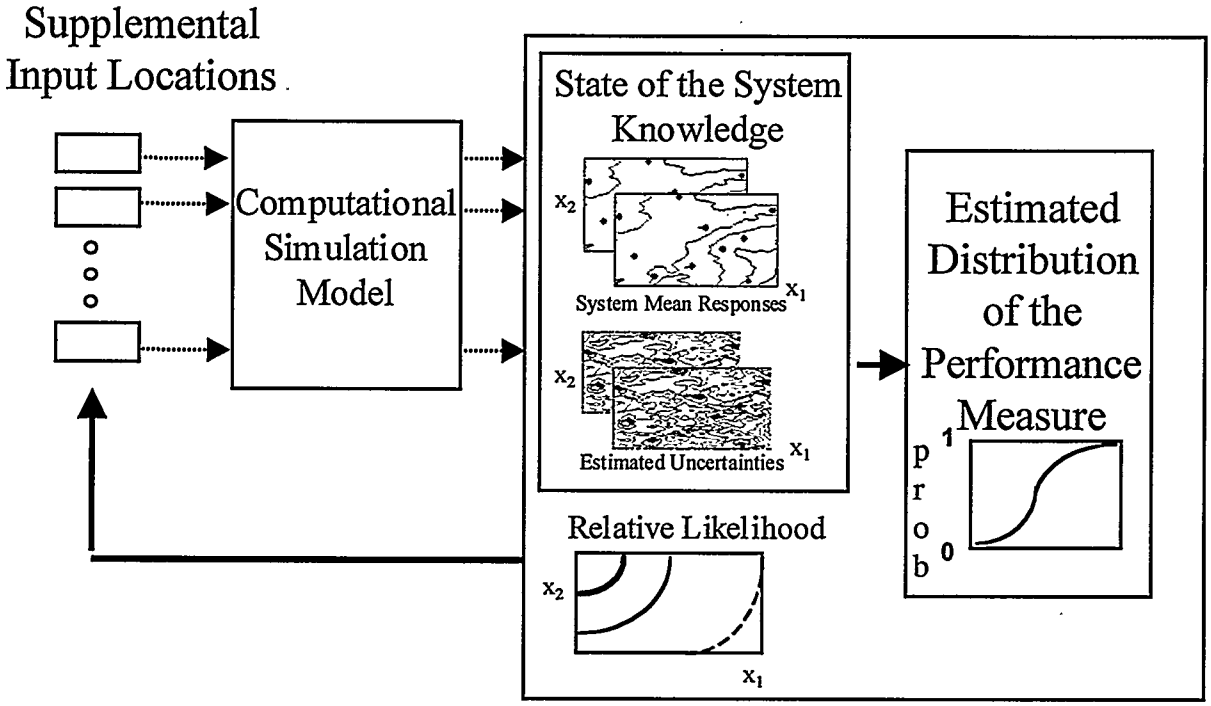

Figure 1.

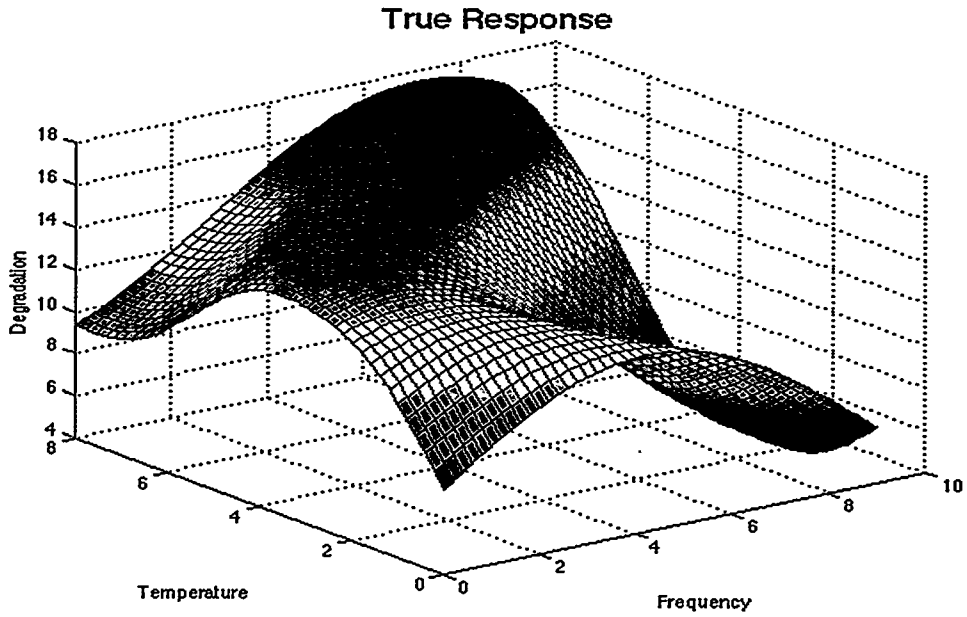

Figure 2.

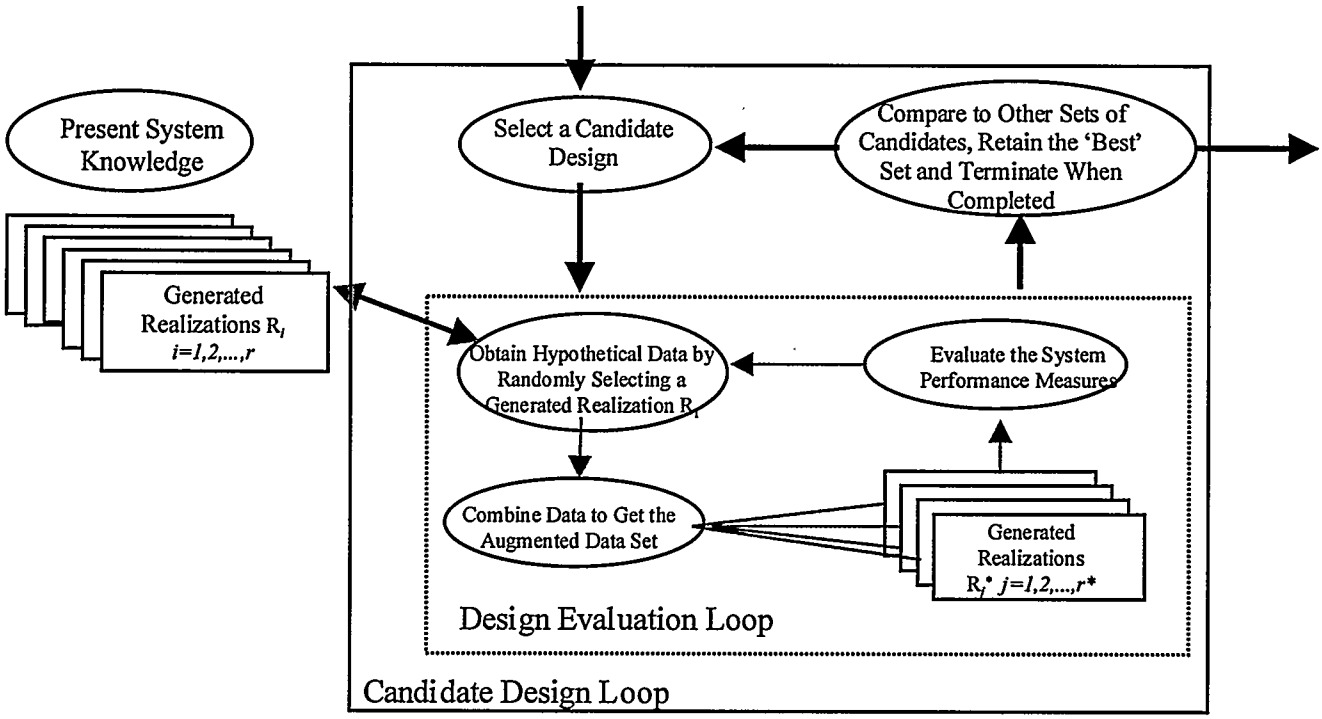

Figure 3.

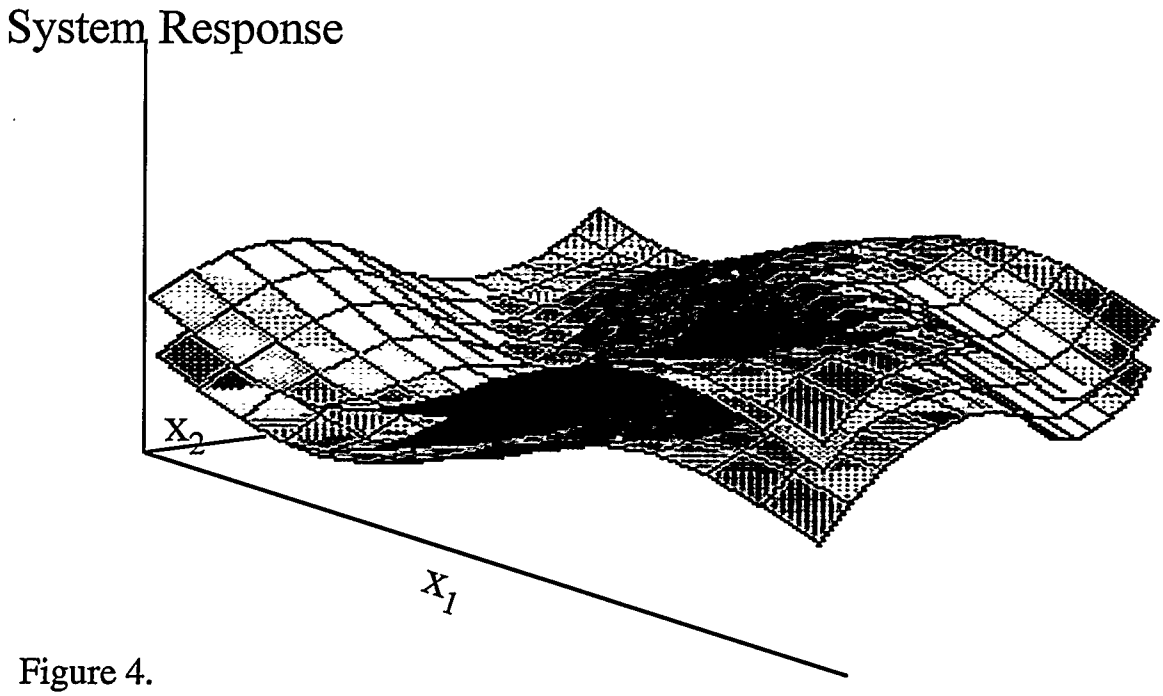

Figure 4.

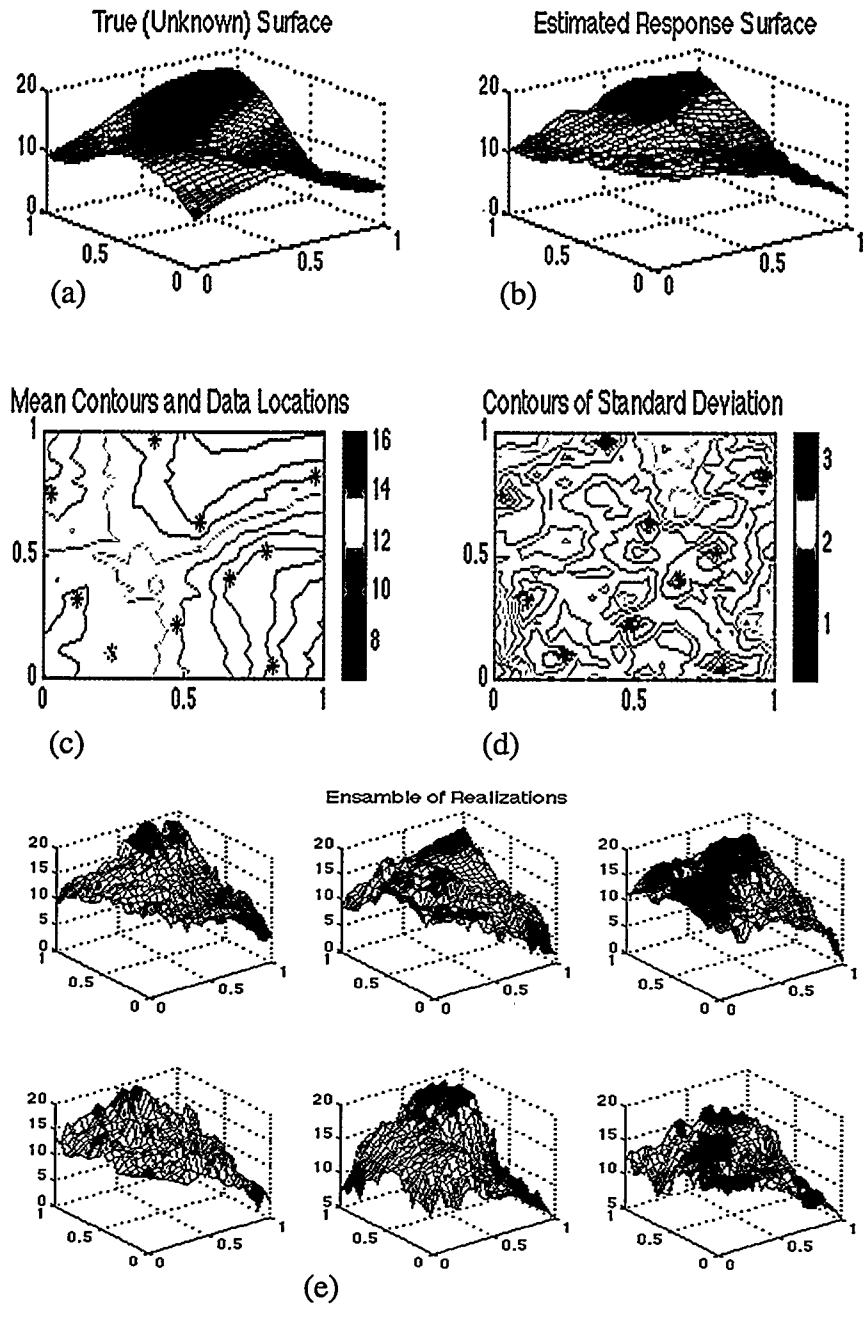
True (Unknown) Surface
Estimated Response Surface
(a)
(b)
Mean Contours and Data Locations
Contours of Standard Deviation
(c)
(d)
Ensamble of Realizations
(e)

Figure 5.

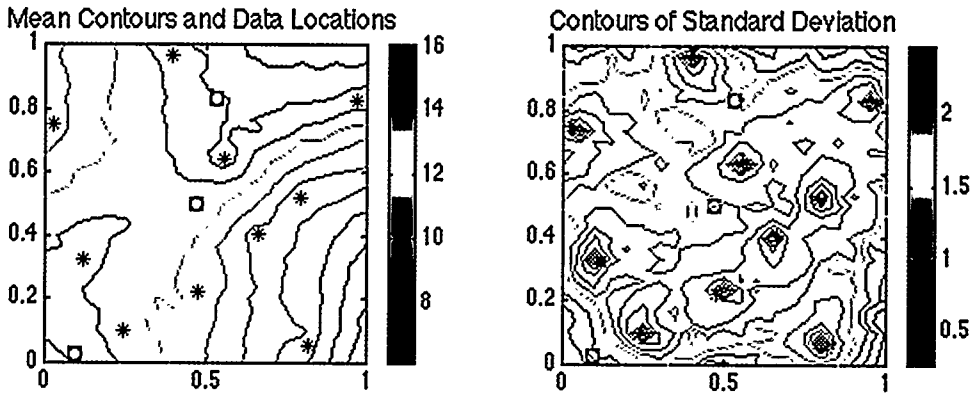
Mean Contours and Data Locations
Contours of Standard Deviation
1
0.8
0.6
0.4
0.2
0
0
0.5
1
16
14
12
10
8
2
1.5
1
0.5
Design Locations
Previous (Actual) Data

Figure 6.

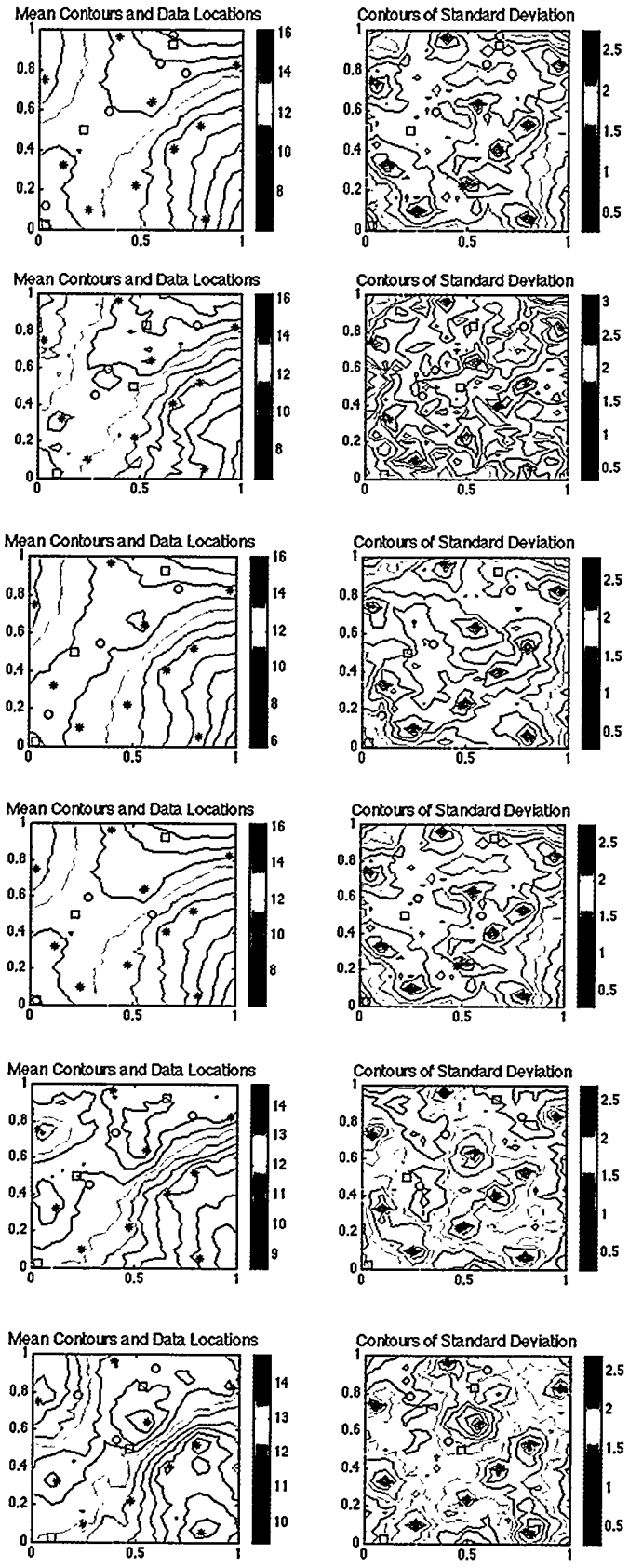
Mean Contours and Data Locations
Contours of Standard Deviation
(a)
(b)
(c)
(d)
(e)
(f)
○ Design location
□ Base Case Design
∗ Previous (Actual) Data

Figure 7.