Proceedings of DETC'02
ASME 2002 Design Engineering Technical Conferences
and Computer and Information in Engineering Conference
Montreal, Canada, September 29-October 2, 2002

DETC2002/DAC-34141

BAYESIAN EVALUATION OF ENGINEERING MODELS

Natesh Chandrashekar
Graduate Research Assistant

Sundar Krishnamurty
Associate Professor

Department of Mechanical and Industrial Engineering
University of Massachusetts
Amherst, MA 01003

ABSTRACT

This paper deals with the development of simulation-based design models under uncertainty, and presents an approach for building surrogate models and validating them for their efficacy and relevance from a design decision perspective. Specifically, this work addresses the fundamental research issue of how to build such surrogate models that are computationally efficient and sufficiently accurate, and meaningful from the viewpoint of its subsequent use in design. Towards this goal, this work presents a Bayesian analysis based iterative model building and model validation process leading to reliable and accurate surrogate models, which can then be invoked in the final design optimization phase. The resulting surrogate models can be expected to act as abstractions or idealizations of the engineering analysis models and can mimic system performance in a computationally efficient manner to facilitate design decisions under uncertainty. This is accomplished by first building initial models, and then refining and validating them over many stages, in line with the iterative nature of the engineering design process. Salient features of this work include the introduction of a novel preference-based design screening strategy nested in an optimally-selected prior information set for validation purposes; and the use of a Bayesian evaluation based model-updating technique to capture new information and enhance model's value and effectiveness. A case study of the design of a windshield wiper arm is used to demonstrate the overall methodology and the results are discussed.

Keywords: Simulation-based Design under Uncertainty, Surrogate Models, Model Validation, Preferential Screening, and Bayesian Updating.

INTRODUCTION

This work identifies a programmatic structured approach that can act as the basis for establishing necessary and sufficiency conditions to choose optimal size and type of predictive models that would be of use in simulation-based design under uncertainty. A main motivation for this work is the recognition that any system or process is hard to understand completely in reality.

In the context of engineering design, analysis models play the role of providing structure to the design problem. Typically, models are simplified representations of reality, primarily because reality is very hard to model precisely due to a lack of sufficient knowledge about the system. Secondly, it may be practically impossible to build an elaborate model and use it in the design optimization cycle simply because of the high computation and the expense involved. Furthermore, even data collection to build such elaborate models may be prohibitively expensive. It is then apparent that some type models, be it iconic, symbolic, or surrogate meta-models become necessary to achieve design decisions [1]. The fundamental question then is how to build such engineering models that are computationally efficient, sufficiently accurate, and meaningful, especially from the viewpoint of its utilization in the subsequent engineering design phase. To start with, physics based engineering analysis models can often be difficult to build and validate, and the problem becomes more complicated and computationally intensive in the case of predictive models that can enable reliable and accurate design mapping of the system. This is particularly true in numerically generated analysis models such as the engineering mechanics models using finite element analysis, the empirically constructed

process models in manufacturing, and the models of fluid flow used in computational fluid mechanics. Therefore, a preferred approach is to build and use surrogate models, that are interpolation models like regression or kriging functions, constructed with results from only a few such expensive runs of the simulation models. These interpolation models, which are models of models or metamodels, can then act as predictive models in design situations in lieu of the actual simulation models.

As the use of computer simulations in engineering design decision-making is growing in importance, surrogate models have become more popular. Though easiest to handle among all models, predictive surrogate models can be granular and approximate. Recognizing that models are built and used to predict the system performance and then make design decisions accordingly, use of such models has to take into account the many causes for the uncertainty in the prediction of system performance. Therefore, a major source of error in the building and the subsequent use of surrogate models is that they can only be expected to be as accurate as the set of data available to build them. A fundamental challenge to building such models then lies in gathering information, its encoding, and its subsequent use in building that can yield most efficient and effective predictive models with least effort. However, though it has been very common to acknowledge the parameter uncertainty through noise arrays and Monte Carlo Simulations (MCS), it is not so common to account for the uncertainty caused due by such structural assumptions done while building the model [2,3].

Often, model validation is the only way of ensuring accuracy and reliability and avoiding possible errors due to simplified, inaccurate, or inappropriate models. Literature on verification and validation of simulation models can be found in [3-7]. A model validation process will need to address the basic questions of: 1) how to initiate the information gathering phase and initial model building; 2) how to incorporate preference information that will ensure that resulting design decisions using such models will be robust, reliable, and accurate; 3) how to select optimally informative sample points to test the model, recognizing that the model cannot be tested at every location in the design set; and 4) how to capture new information and use it to update model fidelity.

ENGINEERING MODEL BUILDING

Towards model building, designers routinely resort to regression analysis or kriging models to estimate the system performance by interpolation [8,9], where interpolation models using some form of optimization are constructed to capture the *basic trend* in the performance [10]. However, though statistical experimentation-based design methods like central composite design explore the input space effectively, research has shown such surrogate models can tend to be coarse owing to sparse data used in creating the response surfaces. These interpolation models are constructed to capture the *basic trend* in the performance space. Any 'noise' throughout the performance space in the surrogate models can lead to a very inefficient the optimization process. Moreover, constructions of such models are typically independent of the preference information, which can further affect the performance of the resulting decision making.

Alternatively, one can adopt an iterative model building strategy, where existing information in the form of input-output relationship can be used to choose optimal samples for conducting future simulations [8,9]. Here, proven techniques such as entropy criteria can be used for choosing such optimal samples or experiments. Related literature in this area can be found in the works by Sacks and associates [11]. While these criteria are now being used in building surrogates extracting spatially optimal information, they still do not take into account the preference information. This may result in spending time and effort in building and refining a surrogate in regions of design space that are not relevant to any decision-making process. Addressing these critical issues, this paper introduces a preference-information based sampling strategy that can be expected to be optimally informative for model validation purposes.

ENGINEERING MODEL VALIDATION

Models are built and used to predict the system performance and then make design decisions accordingly. However, many causes can lead to errors in the prediction of system performance using models. While it is common to acknowledge parametric impreciseness and the incomplete information on design parameters such as material properties, geometric parameters, the modeling errors are rarely acknowledged and dealt with in a comprehensive manner [2]. Errors in the performance estimation caused due to simulation-based predictive modeling approximations is referred to as model uncertainty in this work. As stated before, a model is realistic with least number of assumptions and at the same time it is more complicated with a large number of model parameters that monitor its performance. A fundamental dilemma is then if and when to use more complicated models, or to get more information about a system, or if it is sufficient to take the risk of designing the system using a simplified model that may or may not give accurate results. Naturally, the logical step of model validation becomes essential in the engineering decision making process.

Model validation is usually defined to mean 'substantiation that a computerized model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model [3]. Published works related to model validation from a risk and reliability perspective can be found in [12-14]. A model is usually developed for a specific purpose or application, and its validity should therefore be determined with respect to that goal. It then becomes apparent that if a simulation-based predictive model is used to obtain results for evaluating the value of a particular design, then that model should be validated with regards to its ability to evaluate the values of all potential design solutions with sufficient reliability and accuracy. The accuracy and reliability here will refer to the ability of the model/surrogate to mimic the expected reality closely. More data used in the construction of the surrogate model implies larger validity and higher accuracy for the model. However, this also implies higher cost for the designer and thus there is a clear 'trade-off' between cost and accuracy in a model-building and validation process [Figure 1].

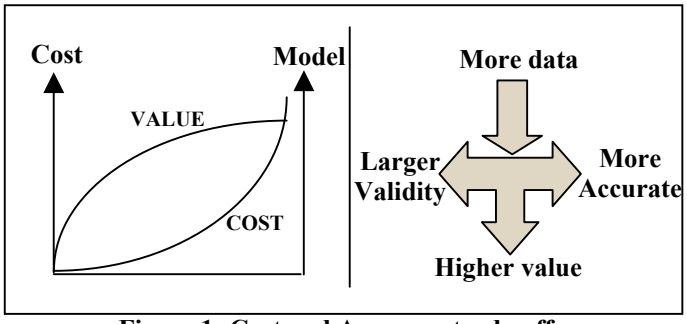

Figure 1: Cost and Accuracy trade-offs

Validation of simulation models is usually through the application of statistical techniques and the type of technique usually depends on the availability of data from the real system. In the case of surrogates, computer simulations need to be performed to validate the accuracy of the interpolating functions. Smooth algebraic approximations of response surfaces may become necessary to help eliminate the *wrinkles* on the performance space. Recognizing the complexities involved in the data acquisition process and that models cannot be checked at all locations in the entire design space, this work process for constructing surrogate functions with only a few scattered data points, yet one that can lead to models with sufficient reliability and accuracy from a design perspective. The validated models will thus focus on the regions of design interest, and using them in optimization or sensitivity analysis will lead to a better understanding of the resulting solutions.

BAYESIAN EVALUATION OF ENGINEERING MODELS

Development of surrogate models has been proven useful to deal with problems with high computational complexities. In a general sense, such models can be considered as a mapping of the set of analysis codes and an associated vector of model parameters or variables on the performance measures, such that the cost and estimation uncertainty are minimized. The fundamental research issue is then how to build such simulation based surrogate models that are computationally efficient, sufficiently accurate, and meaningful from a design perspective.

Previous sections have highlighted the limitations associated with the straightforward application of statistical experiments in surrogate model building process, such as granularity, lack of a preferential sampling strategy and the absence of a validation scheme from a design perspective. From a practical implementation point of view, validated models should be computationally efficient, sufficiently accurate, and meaningful from a design perspective. For example, selecting high fidelity analysis models early in the design process or for variables that have high variance associated with them is clearly not optimal. Whereas, the use of coarse grained models in the final design induces significant design error in performance estimation. Addressing these issues, this paper attempts to improvise on proven methods such as Bayesian analysis by introducing a combined model building and validation framework, which can then be invoked in the final decision making phase.

Research on the applicability of Bayesian analysis to engineering design focusing on the reliability problems in the context of selecting an optimal number of experiments can be found in the works by Howard [15,16]. Recently, Amon and associates [8,9] have discussed the use of Bayesian analysis in the development of surrogate response surfaces that can replace expensive computer simulations in an iterative design optimization process. Application of Bayesian predictive models in deterministic functions with applications to computer-aided experimental designs can be found in the works by Currin [17] and Welch [18]. Recognizing the parallel to decision making in engineering model selection, Doraiswamy and Krishnamurty [19] have studied the Bayesian network in the context of finite element analysis as a means of evaluating the value of information. Here, models were designed by attempting the best trade-off between the quest for more accurate results and reduction of analysis costs under conditions of uncertainty, while considering the expected payoffs from the resulting design using such a model. Their results indicate that an engineering model assessment framework based on Bayesian evaluation can offer a unique approach to handle stochastic problems from a decision based design perspective enabling identification of most optimal model selection. Building on these results, this paper presents the development of a Bayesian analysis based model validation strategy to understand and deal with the uncertainty inherent in simulation and model based design process. Furthermore, it will be used to better understand the value and influence of engineering models in design decision making. The basic tasks are to develop a rigorous framework for designing simulation experiments, and to integrate the new information gathered with existing information to build and enhance the computer-simulated surrogate predictive models. In this scenario, such data sampling has to be coupled with judicial screening of design space using preference information to ensure surrogate accuracy and fineness in the regions of useful design space. In this context, an engineering model building process can be viewed as an information-gathering strategy, using which predictive information regarding the performance of a design can be acquired by means of mathematical simulation studies. An overview of the proposed procedure for building and validating models is shown in Figure 2.

The three major tasks of this methodology include information gathering, optimal sampling with preferential screening, and evaluation and model updating strategies. In this work, intelligent data gathering is identified as a prime need to ensure that useful information can occur with minimal computational effort. An experimental design setup is used to act as an information gathering system to extract initial information and to map it mathematically in the form of response surface models. This approach includes a preferential screening process to generate Pareto-optimal validation grid locations, and a screening strategy using only the informatively optimal points from such validation grids for model evaluation. Model assessment at any stage of data collection and its updating with new information are achieved through a Bayesian evaluation-based systematic procedure.

Information Gathering

Prior information: Current state of knowledge about the performance of the system is known as prior information. Prior information refers to the information obtained before data

collection begins and this can be obtained from first principles, application of physical laws to simplified systems, or empirical correlation obtained from previous experiments on the system. Designer's decisions are based on his current state of knowledge of the system thus it is imperative that the prior information be captured mathematically in the form of models. In an iterative model building process, models are built by designing experiments that make best use of the current state information, and then seeking new information and using it in an optimal manner to further improve model fidelity. As such, system performance can be treated as a nuisance parameter over the entire design space, and the current state of the system can be modeled as a prior distribution, $p(\mu, \sigma^2)$. This is supposed to express the state of information about the mean and variance of the system performance. In this work, normality is hypothesized for convenience and the response is assumed to be the realization of a Gaussian stochastic process at all times.

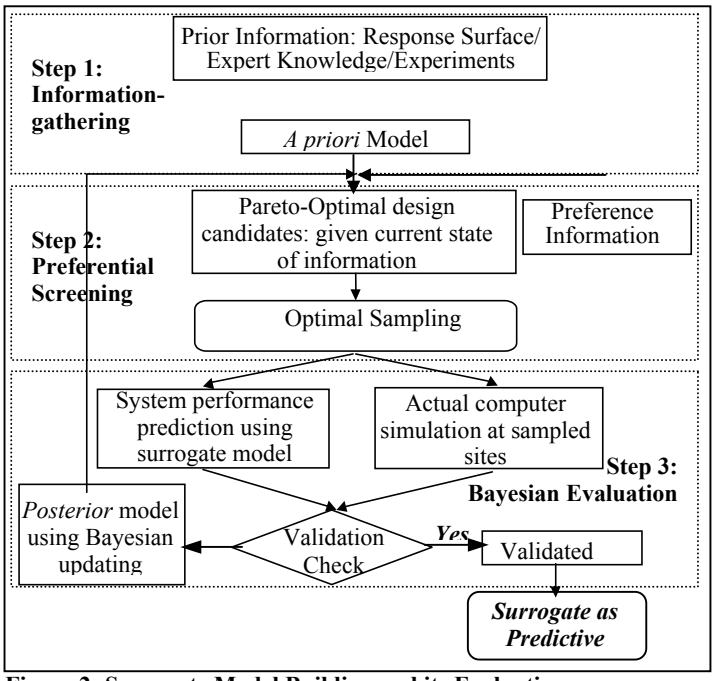

Figure 2: Surrogate Model Building and its Evaluation

DOE Setup: An experimental design setup is used to obtain initial information about the dynamics of the system by simulating the system performance at a few selected combinations of the input design variables. It is assumed that the designer has a good feel for the design variables and has a design space within which he would like to understand the system better as a function of the design variables he has chosen. In this work, a fractional design of experiment setup is suggested for the gathering of information through computer simulations. This forms initial information base for the designer, which is then be used to construct standard response surface models.

Response surface model: Most experimental designs, including those mentioned above, are based on an algebraic regression-model assumption about the way the input factors affect the outputs. For instance, if there are two factors (X_1 and X_2, say) that are thought to affect an output response Y, one may approximate this relationship by the regression model,

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \beta_4 X^2_1 + \beta_5 X_2^{\,2} + \varepsilon$$

where β_y coefficients are unknown and is to be estimated, and ε is a random error term representing whatever inaccuracy such a model might have in approximating the actual simulation-model response Y. The parameters of the model are estimated by making simulation runs at various input values for the X_j's, recording the corresponding responses, and then using standard least-squares regression to estimate the coefficients. Exactly which sets of input values are used is itself an experimental design question, and a more comprehensive reference on this subject can be found in the works by Box and Draper [20]. The prior information about the system now locked in the form of a mathematical response surface model is used in the next step described below.

Preferential screening:

The methodology concentrates on design decision making with multiple attributes and thus the performance space is expected to consist of a set of non-inferior design alternatives. Note that this view is simply to identify a spectrum of non-inferior solutions for the purposes of evaluating the appropriateness of engineering models. Therefore, the ongoing debate on the merits of single-criterion and multi-criteria formulation in reaching the best design decision is not relevant here. However, what is important is that the approach is internally consistent and is within the limits of multi-criteria formulation.

In such a scenario, a preferential screening of design candidates can be performed within the design space and the surrogate model can be validated and enhanced in these targeted regions of the design space. This would ensure that the model would be accurate and would predict with great resolution in the region of the design space, which is of interest to the designer. The *P*areto *E*fficient *F*rontier (PEF) methodology based on Data Envelopment Analysis (DEA) is a very efficient technique to generate Pareto-optimal design candidates within the design space. As part of his graduate research work, Krishnamoorthy [21] has successfully developed and illustrated the PEF algorithm to engineering design problems. Using this algorithm, preference information can be gathered and encoded with standard single attribute utility models for each of the multiple attributes involved. Specifically, attribute preferences can be quantified by means of lottery questions using standard forms such as exponential utility functions to generically scale the values of all the design attributes between zero and one [22,23]. At this point, PEF and the cross-evaluation matrix can be used to rate and rank each design alternative using an efficiency criterion as follows:

$$E_{ks} = \Sigma_y f(x)_{sy} W_{ky}$$

where E_{ks} is the rating for alternative s, using multipliers of the 'test' alternative k; $f(x)_{sy}$ is the value of maximizing criteria y for alternative s; W_{ky} the multiplier assigned to alternative k for the maximizing criteria y. Note that $f(x)_{sy}$ in this work will be the regression model developed in the first stage and the updated models based on posterior prediction equations in the later stages. In this setup, the objective of finding the design space of interest can be achieved by maximizing the rating E_{kk} of a test alternative k, from among a reference set of alternatives 's', and by selecting the optimal multipliers

associated with that maximizing criteria. Mathematically, this can be formulated as:

$$\text{Max } E_{kk} = \sum_{y} f(x)_{ky} W_{ky}$$

subject to $E_{ks} <= 1 \quad \forall$ alternatives s, $W_{ky} >= 0$

The result of such a formation is an optimal efficiency value E_{kk} that is at most equal to 1. In the event E_{kk} =1 (which corresponds to the best rating), it can be concluded that no other alternative is more efficient (dominates) than alternative k for its selected optimal multipliers. That is, if $E_{kk} < 1$, then alternative k does not lie on the optimal frontier and there is at least one other alternative that is more efficient for the optimal set of weights.

The ratings assessed by the PEF methodology for each alternative by itself and the other alternatives in the design space are conveniently arranged in Cross-Evaluation Matrix (Table 1). This average CEM rating can be interpreted as the fitness value of the alternative and can be read into a parallel processing scheme such as genetic algorithm for further selection and crossover operations, resulting in highly 'fit' alternatives with fitness values greater than a predetermined cutoff value.

Optimal Sampling:

The crux of the methodology lies in sampling points within the design space for further data collection (simulation) at points where information is most needed. For this purpose, the Pareto-optimal validation set generated in the last section needs to be subjected to an optimal sampling criterion to determine which of those design candidates if analyzed with a computer simulation would yield maximum information from a design perspective. Most optimal sampling methods seek to minimize metrics of the posterior correlation matrix. For example, Sacks [10] have proposed an integrated mean square error (IMSE) criterion that essentially minimizes the trace of the posterior correlation matrix. Another criterion is the minimization of the maximum mean square error (MMSE), which is equivalent to the entropy based optimality criterion that minimizes the determinant of the posterior correlation matrix and maximizes the determinant of the prior correlation matrix.

Ranking Alternative	**Rated alternative**					
	1	**2**	**3**	**……**	**…s^t.**	**n**
1	E_{11}	E_{12}	E_{13}	……	……	E_{1n}
2	E_{21}	E_{22}	E_{23}	……	……	E_{2n}
3	E_{31}	E_{32}	E_{33}	……	……	E_{3n}
:	:	:	:			:
kth	:	:	:		E_{ks}	:
:	:	:	:			:
N	E_{n1}	E_{n2}	E_{n3}	……	……	E_{nn}
Aver. CEM Rating	e_1	e_2	e_3		e_s	e_n

Table 1:Cross Evaluation Matrix

The criterion considered here for optimal sampling uses the product correlation matrix that accounts for multi-dimensional design space. The correlation matrix for any design variable is constructed as a square matrix with spatial correlations between different points in the design space for that variable. The principal diagonal of this matrix refers to the correlation between the same points for the design variable and thus will always have a value of one. The cross correlations are calculated using the spatial correlation function, which exponentially decreases, as the distance between two points in the design space increases. Mathematically, this correlation function can be expressed as follows:

$$R(d,\theta) = \exp\left(-\theta * abs(d)^{\beta}\right)$$

where, β=2 (for Gaussian correlation); θ>0 to be updated every stage after new addition of data, and 'd' is simply represents the distance between two design sites. The product correlation matrix is thus generated by multiplying corresponding elements of each of the individual correlation matrices from different dimensions.

where: $c(\vec{t},\vec{s},\vec{\theta})$ represents the product correlation

$$c(\vec{t},\vec{s},\vec{\theta}) = \prod_{j=1}^{k} R(|t^j - s^j|,\theta^j) = \prod_{j=1}^{k} R(|d^j|,\theta^j)$$

This ensures that a small correlation is assigned between two design points, which are spatially far away at least in one dimension. The correlation parameters (θ) in the function are updated at every stage of data collection so that sampled points always reflect a Gaussian process with redefined parameters for the process. This new process, the posterior process, must now reflect the current updated state of information, and should have a better predictive capability than the one before the data collection. To ensure this, criterion based on Maximum Likelihood Estimate (MLE) is used, which is equivalent to minimizing the predictive deficiency. This is equivalent to maximizing the log likelihood function numerically [8]:

$$MLE(\cdot) = 1/2\{m_p \log(2\Pi) + m_p \log(\sigma^2 p-1) + \log(|C|) + 1/\sigma^2 (y_p - \mu_{p-1}(D_p))^T C^{-1} (y_p - \mu_{p-1}(D_p))\}$$

where:

m_p represents the m sampling sites of each data collection stage p

σ^2_{p-1}: represents the data variance pth stage data collection

C: prior correlation matrix

$(y_p-\mu_{p-1}(D_p))$: error vector, which represents the error of prediction at points D_p using (p-1) stage surrogate from actual computer simulation, y_p

MLE a function of the parameter vector (θ^j) and the data collected, and it is sufficient to update the parameter vector and the correlation matrices just once at every stage. A key distinction in this work is that this optimal sampling criterion is applied only at the preferentially screened validation set. Thus, the top *Informatively Optimal* solutions are identified from the validation set, the actual computer simulations are then run at these design points, and the results are used to further refine and assess model fidelity.

Bayesian Evaluation:

Bayesian analysis provides a rational basis for making inferences and projections using current state of information. Combining prior experience about a system with the likelihood of system behavior, it enables prediction of system behavior at instances without any prior experience. For completion purposes, a brief overview of Bayesian methods is listed below. The heart of Bayesian techniques lies in the celebrated

inversion formula which states that the belief we accord in a hypothesis H upon obtaining evidence e [probability that a hypothesis H is true, given evidence 'e'] can be computed by multiplying our previous belief P(H) by the likelihood P(e|H) that e will materialize if H is true. P(H|e) is sometimes called the posterior probability (or simply posterior), and P(H) is called the prior probability (or prior). Mathematically, this can be shown as follows:

$$P(H \mid e) = \frac{P(e \mid H)P(H)}{P(e)}$$

One of the attractive features of Bayes' updating rule is its amenability to recursive and incremental computation schemes. Let H denote a hypothesis, $e_n=e^1$, $e^2 \ldots e^n$ denote a sequence of data observed in the past, and e denote a new fact/evidence. A brute-force way to calculate the belief in H given past data and the new evidence e, $P(H|e_n,e)$, would be to append the new datum e to the past data e_n and perform a global computation of the impact on H of the entire data set $e_{n+1}=\{e_n, e\}$. Such a computation would be uneconomical for several reasons. First, the entire stream of past data must be available at all times. In addition, as time goes on and the set e_n increases, the computation of $P(H|e_n,e)$ becomes increasingly complex. Alternatively, this computation can be significantly curtailed under certain conditions by incremental updating that will enable discarding of the past data once $P(H|e_n)$ has been computed, will result in the following expression for the computing of the impact of the new datum:

$$P(H \mid e_n, e) = P(H \mid e_n)\frac{P(e \mid e_n, H)}{P(e \mid e_n)}$$

Under this setup, the old belief $P(H|e_n)$ assumes the role of the prior probability in the computation of new impact. This representation completely summarizes the past experience, and for updating it only need to be multiplied by the likelihood function $P(e|e_n,H)$ that measures the probability of the new datum e, given the hypothesis and the past observations. Using this model for information updating, this work combines Gaussian likelihood functions with the design data from computer simulations in an iterative setup to revise and refine model accuracy and reliability by comparing simulated performance and the predicted performance results. Accuracy and resolution, which are two basic aspects that we expect from any model, are explained.

Accuracy: There have been many suggested ways of measuring the accuracy of any model. A simple measure of accuracy would be percentage errors of the prediction (model) from the actual computer simulation. Standard error metrics used extensively include the Mean Squared Error of the prediction (MSE). The squared error in this case is averaged over a set of validation data and its magnitude is minimized as more data is collected from simulations.

Resolution: Resolution is the ability of a model to distinguish between 2 design alternatives. Resolution of the model in the context of our methodology is captured in the uncertainty band for the prediction of the design performance from the model. Different design alternatives are well resolved by the model if it can predict the performance of each with less uncertainty [Figure 3].

The standard deviation of the prediction in this case is high, thus rendering the resolution low. Therefore, the model is not able to clearly predict whether performance at 1 is better than the performance at 2 or vice-versa. 1^* and 2^* may represent the actual/true performance and thus this may lead to optimization problems. The inverse of the uncertainty bandwidth is the resolution. Higher this quantity, higher is the resolution. For example, for a 95% confidence interval, this can be expressed in terms of standard deviation (σ) as follows:

$$R = \frac{1}{\left[2 * 1.96 * \sigma\right]}$$

Thus, accuracy and resolution criteria can be generated for the evaluation of the surrogate models at any stage of information collection and updating.

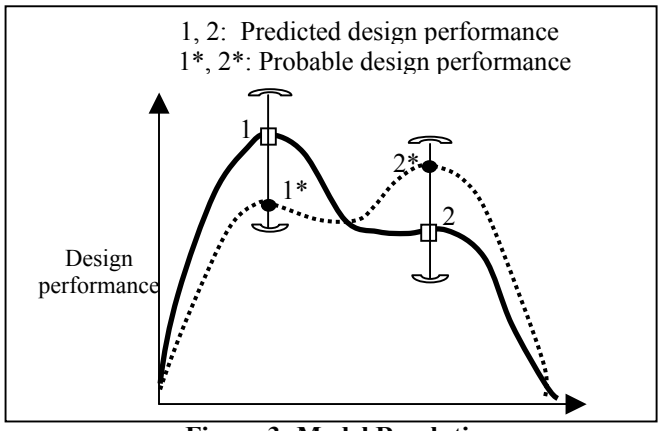

Figure 3: Model Resolution

Surrogate Model Updating:

In this methodology, we are modeling the system performance (as simulated by computer models) as a Gaussian stochastic process with the system performance as a nuisance parameter over the entire design space. The parameters of this process (which manifest as the surrogate parameters) can be predicted better as we have more data characterizing the process. The sequential information updating is done by making use of statistical techniques of *Bayesian Inference*, where the prior information is taken as the initial response surface in the first iteration. Accordingly, the prior information is described as:

$$E\left[Y\left(\vec{t}\right)\right] = \mu(\vec{t}) \quad | \quad \text{cov}\left[Y(\vec{t}), Y(\vec{s})\right] = K(\vec{t}, \vec{s})$$

where Y (t) is a vector with finite collection of computer simulations; and $\text{cov}\left[Y(\vec{t}), Y(\vec{s})\right]$ represents the covariance for any two points t and s. The resulting Gaussian posterior process, obtained once the new data (p^{th} stage data) is collected, will have the mean function:

$$\mu_p(\vec{t}) = E\left[Y_p(\vec{t})\right] = \mu_{p-1}(\vec{t}) + k(\vec{t}) * \Sigma_{p-1}^{-1} * (y_p - \vec{\mu}_{p-1}(D_p)$$

where:

$E\left[Y_p(\vec{t})\right]$ is the expected value of the design process given information Y_p

Σ_{p-1}: Covariance matrix for p-1 stage (prior covariance matrix)

$(y_p-\mu_{p-1}(D_p))$: error vector representing the error of prediction at points D_p using (p-1) stage surrogate from actual computer simulation, y_p

Corresponding posterior covariance can be expressed as:

$$K_p(\vec{t},\vec{s}) = K_{p-1}(\vec{t},\vec{s}) - k_{p-1}(\vec{t}) * \Sigma_{p-1}^{-1} * k_{p-1}(\vec{s})$$

where: K_p (t, s): covariance for 2-design sites t and s after p^{th} stage data collection

K_{p-1} (t, s): Prior covariance for same two sites t and s

k_{p-1}: prior correlation vector

Σ_{p-1}: covariance matrix for p-1 stage (prior covariance matrix)

Thus at any stage of data collection, there will be a posterior mean and a posterior standard deviation (See Equations above) describing the surrogate model's fidelity as a measure of its error in prediction. That can be used to build new response surfaces (new model parameters) at each stage. These model parameters will approach their final values as more and more data is collected and the Gaussian process to model becomes more refined.

CASE STUDY: WINDSHIELD WIPER BLADE MECHANISM

The application of the proposed methodology is illustrated with the Finite Element Analysis of the primary arm of a windshield wiper blade mechanism [Figure 4]. For simplicity, the arm is subject to only a horizontal force about the hole in the center, and the two holes at the ends are treated as held fixed and the maximum deflection and maximum stress are considered the design performance parameters of interest.

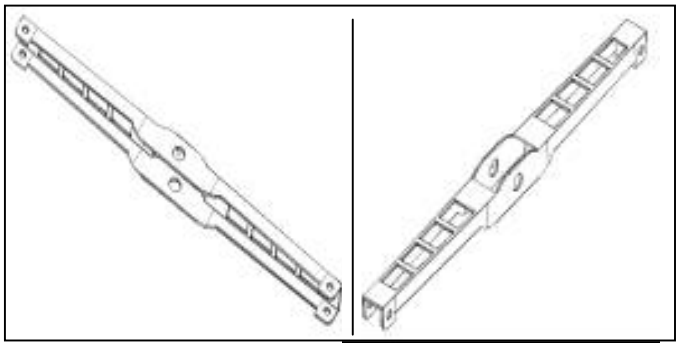

Figure 4: Finite element model of windshield wiper arm

The design constraints on the parameters are:
$Stress_{max} < 90$ MPa
$Deflection_{max} < 0.2$ mm

In this process, the first step is to identify the design space. Three design variables identified as important for the windshield wiper redesign are:

1. Radius of the hole in the center (3-4mm)-'r'
2. The separation between the trapezoidal cutouts on the top (4-6mm)-'s'
3. Thickness of the entire part (0.8-1.2mm)-'t'

In the absence of real-world data, the most realistic model (a finite element model with solid elements) is used for populating the output columns of the DOE array.

RESULTS AND DISCUSSION

Initially, 18 points were selected throughout the design space [fractional DOE] to collect the performance data. In the presented work this is the prior information, and this information is used to scan the design space for collecting further information at points in the design space which promise to increase the knowledge about the system the most. Considering the size of the design space from the constraints on the design variables, two levels for thickness is considered sufficient [Table 2].

The surrogate models for the individual attributes are constructed using standard regression models. The regression equations that represent the prior information of the designer are the following:

$$\begin{aligned} Str &= 141.394 + 22.533*s - 146.027*t + 4.416s*t - 2.85*s^2 \\ Def &= 0.5086 + 0.006*r - 0.039*s - 0.347*t - 0.004*r*t + 0.021*s*t + 0.001*s^2 \\ Wt &= -1.428 + 0.366*r - 0.103*s + 49.922*t - 0.3666*r*t + 0.816*s*t - 0.053*r^2 \end{aligned}$$

The response surfaces now represent the meta-model of the computer simulated finite element model while offering computationally easy evaluation of the design at various points.

	r	s	t	Stress	Deflection	Weight
1	3.0	4.0	0.8	82.700	0.164	40.45
2	3.0	4.0	1.2	31.600	0.053	61.28
3	3.0	5.0	0.8	83.300	0.151	41.00
4	3.0	5.0	1.2	35.000	0.048	62.16
5	3.0	6.0	0.8	78.300	0.140	41.55
6	3.0	6.0	1.2	31.600	0.045	63.04
7	3.5	4.0	0.8	84.600	0.166	40.31
8	3.5	4.0	1.2	30.100	0.054	61.08
9	3.5	5.0	0.8	83.700	0.152	40.86
10	3.5	5.0	1.2	34.600	0.049	61.95
11	3.5	6.0	0.8	78.400	0.141	41.41
12	3.5	6.0	1.2	32.100	0.045	62.83
13	4.0	4.0	0.8	82.600	0.167	40.15
14	4.0	4.0	1.2	33.800	0.054	60.84
15	4.0	5.0	0.8	83.600	0.153	40.70
16	4.0	5.0	1.2	33.000	0.049	61.71
17	4.0	6.0	0.8	78.600	0.142	41.25
18	4.0	6.0	1.2	27.800	0.046	62.59

Table 2: Initial knowledge; DOE

Preferential Validation: For illustrative purposes, an utility independent model for the stress, deflection and weight attributes are assumed and a lottery elicitation process for the individual attributes are assumed to result in the following single attribute models [23]:

$$U_{wt} = 1.033 + 0.033 * (-e^{0.04621*wt})$$

$$U_{def} = 1.0666 + 0.0666 * (-e^{13.863*def})$$

$$U(str) = k * U(str1) + (1-k) * U(str2)$$

$$U(str1) = -0.09695 + 0.09695 * (e^{34.657/1000*str1})$$

$$U(str2) = e^{-0.13862*str2}$$

It is expected that such a mapping of design performance space on to the utility space will help the designer to pick the model

that is of most interest that can then be used to subsequently find the most optimal design. The surrogate model, which represents the current state of knowledge about the system performance, is used in the PEF algorithm to scan the design space for non-dominated designs [Pareto optimal designs]. The Pareto-optimal candidate designs now form the design space grid within which optimal sampling is performed [Table 3].

	r	s	t		r	s	t
1	3.857	5.905	0.851	6	3.397	5.238	0.851
2	3.619	5.746	0.902	7	3.048	5.968	1.181
3	3.492	5.048	0.927	8	3.889	5.556	1.060
4	3.143	5.556	0.800	9	3.429	5.587	1.016
5	3.397	5.651	0.838	10	3.379	5.124	1.126

Table 3: validation grid: subset

Optimal Sampling: The sampling is done purely based on the spatial correlation among the sampling sites. In accordance with the optimality criterion, a criterion that minimizes the trace of the inverse correlation matrix is identified. The optimal sample set of the designs among the Pareto-optimal designs preferentially screened is listed in Table 4.

	r	s	t
1	3.492	5.048	0.927
2	3.429	5.587	1.016
3	3.397	5.238	0.851
4	3.619	5.746	0.902
5	3.889	5.556	1.060

Table 4: Optimal Validation points

Model Evaluation: These design points form the validation points for the current surrogate. The current surrogate is used to predict the performance of the system at the optimal sample of design sites, and the actual computer simulation of the system is performed at these same design sites. The prediction of the first stage surrogate and the actual system performance (values are measured as performance) at these three validation design sites are shown in Tables 5-7.

					Prediction-Stress Utility			Str U	
	r	s	t	stdev	UB	Mean	LB	actual	sq error
1	3.492	5.048	0.927	0.014	0.947	0.920	0.894	0.655	0.0700
2	3.429	5.587	1.016	0.030	0.615	0.556	0.498	0.373	0.0337
3	3.397	5.238	0.851	0.024	0.444	0.397	0.349	0.834	0.1918
4	3.619	5.746	0.902	0.026	0.978	0.926	0.875	0.708	0.0478
5	3.889	5.556	1.060	0.031	0.508	0.448	0.388	0.301	0.0214
									0.3646
Average resolution for stress prediction model: 10.25 [should not be interpreted since stresses fall out of prediction bounds]									

Table 5: Stress surrogate evaluation

It is apparent from the results that not only does the actual performance of the system falls outside the bounds of prediction by the stress surrogate model, but also the bounds are large. This is a clear indication of a bad surrogate (a highly coarse model) due to the lack of knowledge about the system performance in the whole of the design space. However, this new information about the system performance at the validation points can now be integrated with the existing model to improve our overall fidelity of the resulting model. This construction of the posterior from the prior is 'updating' the information in a Bayesian sense. This will ensure that the information now obtained with computer simulations at the validation points is locked, and that the error from the prediction equation is correctly interpolated irrespective of any error from the regression analysis.

					Prediction-Deflection Utility			Def U	
	r	s	t	stdev	UB	Mean	LB	actual	sq error
1	3.492	5.048	0.927	0.058	0.843	0.729	0.616	0.795	0.0044
2	3.429	5.587	1.016	0.128	1.086	0.836	0.585	0.886	0.0025
3	3.397	5.238	0.851	0.104	0.841	0.638	0.435	0.679	0.0017
4	3.619	5.746	0.902	0.112	0.950	0.730	0.510	0.785	0.0031
5	3.889	5.556	1.060	0.131	1.122	0.866	0.609	0.906	0.0016
									0.0132
Average resolution for deflection prediction model: **2.5**									

Table 6: Deflection surrogate evaluation

					Prediction-weight utility			Weight U	
	r	s	t	stdev	UB	Mean	LB	actual	sq error
1	3.492	5.048	0.927	0.048	0.828	0.733	0.638	0.734	0.000000
2	3.429	5.587	1.016	0.107	0.862	0.653	0.444	0.654	0.000001
3	3.397	5.238	0.851	0.086	0.952	0.782	0.613	0.782	0.000000
4	3.619	5.746	0.902	0.094	0.929	0.746	0.562	0.746	0.000000
5	3.889	5.556	1.060	0.109	0.828	0.613	0.399	0.614	0.000001
									0.000002
Average resolution for weight prediction model: **2.9**									

Table 7: weight surrogate evaluation

Successive surrogates: The new surrogate is then used in the PEF methodology to obtain Pareto-Optimal candidates and the process is continued by comparing surrogate's resolution and accuracy with the computer-simulated performance of the system at the optimally sampled sites [Tables 8-10].

	Second Stage				Third Stage		
	r	s	t		r	s	t
1	3.683	5.746	0.863	1	3.603	5.619	0.825
2	3.476	5.206	1.016	2	3.746	5.651	0.984
3	3.603	5.365	0.952	3	3.238	5.429	0.914
4	3.698	5.841	1.048	4	3.333	5.556	1.048
5	3.333	5.746	1.101	5	3.492	5.302	1.168

Table 8: Optimal sampling: successive stages

The surrogate staging with prior, simulation and posterior is continued until the evaluation criteria are met. As before, the accuracy and resolution of the model are checked after the validation points have been used in the computer simulations. In this case, the accuracy of the third stage surrogate has improved (stress attribute has less than 0.003 average error and deflection, weight attribute have prediction errors less than 0.00001 and the resolution (>5) among the validation points are better and more uniform). Accordingly, the process is terminated and the new surrogate models for the stress, deflection and the weight are identified from the third iteration as follows:

$$Str = 349.09+16.309*s-543.996*t + 4.247*s*t-2.264*s^2 +199.26*t^2$$

$$Def = 0.991 + 0.005*r-0.039*s-1.354*t-0.004*r*t + 0.0203*s*t + 0.001*s^2 + 0.503*t^2$$

$$Wt = -0.263 + 0.385*r-0.102*s + 47.422*t-0.367*r*t + 0.815*s*t-0.055*r^2 + 1.252*t^2$$

				Attribute		Stress		Average
		stdev	UB	Mean	LB	actual	Error	Resolution
Stress	1	0.053	1.036	0.933	0.830	0.910	0.00054	3.08
	2	0.072	0.562	0.422	0.281	0.391	0.00092	
	3	0.099	0.774	0.580	0.385	0.543	0.00135	
	4	0.091	0.500	0.322	0.143	0.308	0.00019	
	5	0.100	0.458	0.262	0.067	0.246	0.00026	
							0.00326	
Deflection	1	0.033	0.787	0.722	0.656	0.726	0.00002	4.86
	2	0.045	0.966	0.877	0.788	0.879	0.00000	
	3	0.063	0.953	0.830	0.707	0.832	0.00000	
	4	0.058	1.016	0.903	0.790	0.904	0.00000	
	5	0.063	1.046	0.922	0.798	0.924	0.00001	
							0.00003	
Weight	1	0.030	0.830	0.7722	0.714	0.7721	0.00000	5.48
	2	0.040	0.738	0.6592	0.580	0.6591	0.00000	
	3	0.056	0.821	0.7121	0.603	0.7120	0.00000	
	4	0.051	0.722	0.6217	0.521	0.6215	0.00000	
	5	0.056	0.672	0.5620	0.452	0.5618	0.00000	
							0.00000	

Table 9: Second stage Surrogate Results

				Attribute		Stress		Average
		stdev	UB	Mean	LB	actual	Error	Resolution
Stress	1	0.050	0.594	0.497	0.399	0.505	0.0001	4.87
	2	0.051	0.562	0.462	0.361	0.438	0.0006	
	3	0.072	0.859	0.719	0.578	0.678	0.0017	
	4	0.043	0.428	0.343	0.258	0.320	0.0005	
	5	0.046	0.316	0.226	0.136	0.207	0.0004	
							0.0032	
Deflection	1	0.045	0.719	0.631	0.543	0.634	0.0000	5.41
	2	0.046	0.955	0.865	0.775	0.863	0.0000	
	3	0.065	0.921	0.794	0.668	0.792	0.0000	
	4	0.039	0.979	0.902	0.826	0.901	0.0000	
	5	0.041	1.016	0.935	0.854	0.932	0.0000	
							0.00002	
Weight	1	0.037	0.868	0.796	0.724	0.796	0.0000	6.61
	2	0.038	0.757	0.684	0.610	0.684	0.0000	
	3	0.053	0.842	0.738	0.635	0.738	0.0000	
	4	0.032	0.686	0.623	0.560	0.623	0.0000	
	5	0.034	0.554	0.487	0.421	0.488	0.0000	
							0.00000	

Table 10: Third stage Surrogate Results

The final values of regression coefficients after the third stage are listed in Table 11.

Some Observations: Since the design performance is viewed as a stochastic process, it is possible to estimate the error in performance prediction at any point in the design space. The uncertainty/variance of prediction is larger in the design space where actual data is absent and prediction is purely based on interpolation. Note that the process is not well known and the entire knowledge about the system is based on the simulation information obtained from the initial 18 points. Naturally, the errors due to model granularity will increase as we move away from region of present information. Since we target regions in the design space preferentially, more data is progressively gathered in these select regions yielding a final model that is fine in those regions of design space.

Results for str			
Variable	Actual Coefficient	p	Interpret
Constant	349.0910	5.9E-34	
s	16.3089	3.3E-08	Significant
t	-543.9964	2.5E-34	Significant
s*t	4.2470	1.5E-02	Significant
s^2	-2.2650	3.9E-04	Significant
t^2	199.2597	7.7E-14	Significant
Statistics			
Std Err	12.273		
R-sq	99.70%		
R-sq(adj)	99.65%		
R-sq(pred)	99.50%		
Press	6583.424		

Results for def			
Var.	Actual Coeff.	p	Interpret
Const.	0.9912	1.6E-53	
r	0.0054	2.3E-13	Significant
s	-0.0388	4.1E-37	Significant
t	-1.3537	1.0E-56	Significant
r*t	-0.0037	7.7E-06	Significant
s*t	0.0203	3.8E-28	Significant
s^2	0.0010	1.5E-09	Significant
t^2	0.5032	3.1E-39	Significant
Statistics			
Std Err	0.002		
R-sq	100.00%		
R-sq(adj)	100.00%		
R-sq(pred)	100.00%		
Press	0		

Results for wt			
Var.	Actual Coeff.	p	Interpret
Const.	-0.2627	9.9E-98	
r	0.3846	4.0E-44	Significant
s	-0.1024	1.0E-58	Significant
t	47.4223	3.6E-87	Significant
r*t	-0.3668	6.2E-26	Significant
s*t	0.8153	1.2E-41	Significant
r^2	-0.0557	1.1E-11	Significant
t^2	1.2527	2.2E-23	Significant
Statistic			
Std Err	0.003		
R-sq	100.00%		
R-sq(adj)	100.00%		
R-sq(pred)	100.00%		
Press	0		

Table 11: Final stage surrogate regression coefficients

One final step that is needed is to make sure that the regression equations fit all the data collected so far with as little error as possible, so that the equations can be used in optimization or sensitivity analysis directly. One important observation from the results is that the methodology forces us to collect more data for design candidates with thickness values in-between the two levels of thickness initially assumed, thus indicating possible nonlinear stress response with change in thickness which is consistent with the inherent understanding of the stress dependency on thickness. In addition to information capture, the error by the surrogate performance estimation can also be identified in this approach.

The design performance is estimated differently during iterations, and the regression itself may have error, as it is merely an approximate mapping of the data. The error captured in the error vector at any stage p given by $(Y_p-\mu_{p-1}(D_p))$, and this error vector updates itself whenever the interpolation equation changes (Y_p). This 'memory-less' property of the error vector corresponds to the relative '*insensitivity'* of the surrogate prediction to small errors from interpolation. Furthermore, many other methods that do not incorporate preference information in the optimal sampling selection process rely upon standard design of experiment grids to start and propagate data collection. This operation could require thousands of design points, as opposed to a few select points in the Pareto-optimal validation grid used in this work. This process also ensures that the grid is dense in regions of design space of interest and where more accuracy from the model is desired. Finally, this approach overcomes a serious limitation with methods that use

no prior information to start their surrogate building process, and thus avoid any potential chance of exhausting the resources kept aside for analysis even before getting the surrogate fine enough in the regions of desirable performance.

SUMMARY

This paper addresses issues related to predictive models in engineering design, and presents a robust model building and preferential screening methodology using Bayesian analysis for the assessment of computer-simulated surrogate models. The three major tasks of the proposed methodology including information gathering, preferential screening, and evaluation and model updating strategies are detailed. Special features of the work includes an experimental design setup to extract initial information using surrogate response surface models; the mapping of system performance as a nuisance parameter over the entire design space represented by prior normal distribution; and the employment of a preferential screening based optimal sampling technique in model validation that enables targeting of model refinement to limited, yet most valuable regions of the performance space. Through an engineering case study, this paper shows how this work can be effectively implemented in a computationally viable environment to recursively and incrementally use Bayesian updating rule to develop probabilistic predictive models that can act as surrogate models in simulation-based design under uncertainty.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of the National Science Foundation under Grant No. DMI-9713382.

REFERENCES

1) Hazelrigg, G.A., 1999, "On the Role and Use of Mathematical Models in Engineering Design," ASME Journal of Mechanical Design, Vol. 121, pp:.336-341.
2) Draper D., 1995, "Assessment and Propagation of Model Uncertainty", Journal of the Royal Statistical Society, Vol. 1, pp. 45-97.
3) Sargent R.G., 1998, "Verification and Validation of Simulation Models", Proceedings of the 1998 Winter Simulation Conference, pp: 121-130.
4) Hills R.G. and Trucano T.G., 1999 " Statistical Validation of Engineering and Scientific Models: Background ", Computational Physics Research and Development, Sandia National Laboratories, 1099-1256.
5) Kelton W.D., 1999, "Designing Simulation Experiments", Proceedings of the 1999 Winter Simulation Conference, pp: 33-38.
6) Kleijnen J.P.C., 1999, "Validation of Models: Statistical Techniques and Data Availability", Proceedings of the 1999 Winter Simulation Conference, pp: 647-654
7) Yesilyurt, S., and Patera, A.T., 1995, "Surrogates for numerical simulations; optimization of eddy-promoter heat exchangers", Computer Methods in Applied Mechanics and Engineering 121, pp. 231-257.
8) Osio I.G., and Amon C.H., 1996 "An engineering design methodology with multistage Bayesian surrogates and optimal sampling" Research in engineering design, Vol. 8, pp: 1893-206.
9) Pacheco J.E., Amon C.H., and Finger S., 2001, "Developing Bayesian Surrogates for Use in Preliminary Design," Proceedings of ASME Design Theory and Methodology, Pittsburgh, PA.
10) Booker A.J., Dennis Jr. J.E., Frank P.D., Serafini D.B., Torczon V., Trosset M.W., 1998, "A Rigorous Framework for Optimization of Expensive Functions by Surrogates," Langley Research Center, http://www.icase.edu/Dienst/Repository/2.0/Body/ncstrl.icase/TR-98-47/pdf.
11) Sacks J. Welch W.J., Mitchell T.J. and Wynn H.P., 1989, "Design and Analysis of Computer Experiments", Statistical Science, Vol. 4, No.4, pp: 409-435
12) French S., 1995, " Uncertainty and Imprecision: Modeling and Analysis", Journal of the Operational Research Society, Vol. 2, pp: 70-79.
13) Haimes Y.Y., 1998, "Risk Modeling, Assessment and Management", A Wiley-Interscience publication, New York, New York.
14) Wahlstrom B., 1994, " Models, Modeling and Modelers: an application to risk analysis ", European Journal of Operational Research, pp: 477-487.
15) Howard R.A., 1965, "Bayesian Decision Models for System Engineering", IEEE Transactions on Systems Science and Cybernetics, pp: 36-40.
16) Howard R.A., 1968, "The Foundations of Decision Analysis", IEEE Transactions on Systems Science and Cybernetics, Vol. 4, pp: 211-219.
17) Currin C., Mitchell T., Morris M., and Ylvisaker D., 1991, "Bayesian Prediction of Deterministic Functions, with Applications to the Design and Analysis of Computer Experiments", Journal of the American Statistical Association, Vol. 86, pp. 953-963.
18) Welch J.W., Buck R.J., Sacks J., Wynn H.P., and Mitchell T.J., & Morris M.D., "Screening, Predicting, and Computer Experiments", Technometrics, 1992, Vol. 34, No.1, pp. 15-25.
19) Doraiswamy S., Krishnamurty S., and Grosse I.R., September 2000, " Bayesian Analysis in Engineering Model Assessment" ASME Design Engineering Technical Conferences, DETC 2000/DTM 14546.
20) Box G.E.P and Tiao G.C., 1973, "Bayesian Inference in Statistical Analysis", Addison Wesley, Reading, MA.
21) Krishnamoorthy, A., 'Pareto Efficient Frontier and Design Alternative Enumeration in Engineering Design', Masters' Project Report, May 2001, University of Massachusetts, Amherst.
22) Keeney R.L. and Raiffa H., 1976, "Decisions with Multiple Objectives: Preferences and Value Trade-offs", Wiley and Sons, New York.
23) Gold S. and Krishnamurty, S., 1997, "Tradeoffs in Robust Engineering Design", *1997 ASME Design Engineering Technical Conferences*, DETC97/DAC-3757.

