Sampling Strategies for Computer Experiments: Design and Analysis

Timothy W. Simpson
Departments of Mechanical & Nuclear Engineering and Industrial & Manufacturing Engineering
The Pennsylvania State University, University Park, Pennsylvania 16802, USA

Dennis K. J. Lin [*]
Department of Management Science and Information Systems
The Pennsylvania State University, University Park, Pennsylvania 16802, USA

Wei Chen
Department of Mechanical Engineering
University of Illinois at Chicago, Chicago, IL 60607-7022, USA

Abstract. Computer-based simulation and analysis is used extensively in engineering for a variety of tasks. Despite the steady and continuing growth of computing power and speed, the computational cost of complex high-fidelity engineering analyses and simulations limit their use in important areas like design optimization and reliability analysis. Statistical approximation techniques such as design of experiments and response surface methodology are becoming widely used in engineering to minimize the computational expense of running such computer analyses and circumvent many of these limitations. In this paper, we compare and contrast five experimental design types and four approximation model types in terms of their capability to generate accurate approximations for two engineering applications with typical engineering behaviors and a wide range of nonlinearity. The first example involves the analysis of a two-member frame that has three input variables and three responses of interest. The second example simulates the roll-over potential of a semi-tractor-trailer for different combinations of input variables and braking and steering levels. Detailed error analysis reveals that uniform designs provide good sampling for generating accurate approximations using different sample sizes while kriging models provide accurate approximations that are robust for use with a variety of experimental designs and sample sizes.

Key Words : *design of experiment, kriging, Latin hypercube, multivariate adaptive regression spline, radial basis functions, response surface, uniform design.*

* E-mail address : DKL5@ psu.edu

NOMENCLATURE

APPROX	approximation model type:
krg	kriging approximation
mar	multivariate adaptive regression splines
rbf	radial basis functions
rs2	second-order polynomial response surface
DOE	experimental design type:
hss	Hammersley sequence sampling
lhd	Latin hypercube
oay	orthogonal array
rnd	random design
uni	uniform design
FCN	function number for first example
MAX	maximum absolute error
RMSE	root mean square error
SAMP	number of sample points in an experimental design
x	design (input) variable
y	actual output (response) value
$\hat{y}$	predicted output (response) value from approximation model

1. FRAME OF REFERENCE: COMPUTER EXPERIMENTS

Computer-based simulation and analysis is used extensively in engineering to predict the performance of a system or product. For example, engineers use finite element models to predict the performance of a structure, computational fluid dynamics models to visualize the flow over a body, and Monte Carlo simulation to estimate the reliability of a product due to uncertainty in loading conditions or material parameters. Despite the steady and continuing growth of computing power and speed, single evaluations of aerodynamic or finite element analyses can take minutes to hours, if not longer. The high computational costs of performing these analyses limit their use in design optimization and reliability analysis.

Design of experiments (e.g., Montgomery, 1997) and statistical approximation techniques such as response surface methodology (e.g., Box and Draper, 1987; Box, et al., 1978; Myers and Montgomery, 1995) are becoming widely used in engineering to minimize the computational expense of running such computer analyses (Barthelemy and Haftka, 1993; Barton, 1992; Barton, 1994; Barton, 1998; Simpson, et al., 2001b; Sobieszczanski-Sobieski and Haftka, 1997). The basic approach is to construct approximations of the computationally expensive simulation and analysis codes to provide surrogate models that are sufficiently accurate to replace the original code. These surrogate models are then used in lieu of the original analysis or simulation code, facilitating design space exploration, optimization, and reliability analysis.

Building approximations for these computer simulations involves (a) choosing an experimental design to sample the region of interest and (b) constructing an approximation model to the observed sample data as shown in figure 1. As shown in Figure 1a, the region

of interest is often referred to as the "design space," which is bounded by the upper and lower limits of each of the design (input) variables being studied. Design of experiments strategies are often used to sample the design space to generate sample data to fit an approximate model to each of the output variables (responses) of interest. Experimental designs can also be used for "screening" experiments to identify significant factors and reduce the dimensionality of the problem (Box and Draper, 1987; Gangadharan, et al., 1995; Goldsman and Nelson, 1998; Koch, et al., 1997; Welch, et al., 1992). In Figure 1b, a second-order response surface is used to approximation the relationship between the design (input) variables x1 and x2, and y, the output variable (response).

(a) Sample Region of Interest   (b) Construct Approximation Model
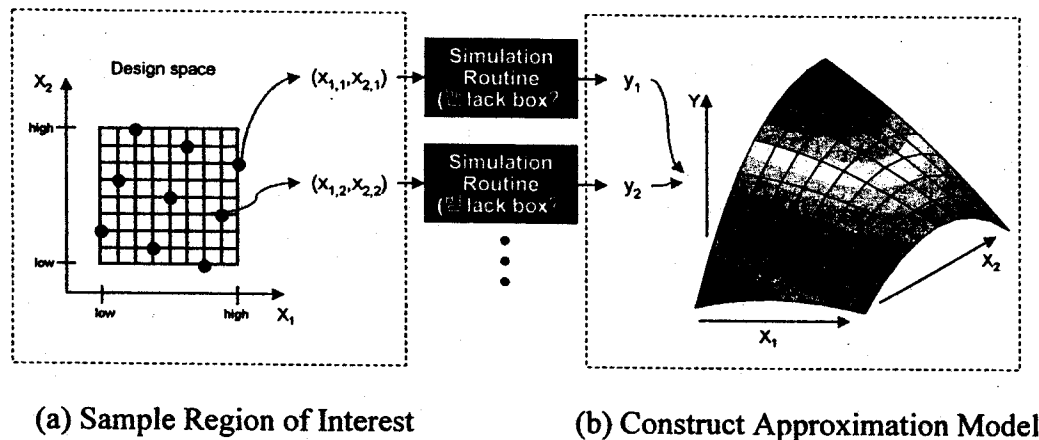
Figure 1. Design and Analysis of Computer Experiments

The growing use of computers in design optimization has given rise to considerable research in the design and analysis of computer experiments. The primary research thrusts are to improve:

1. the *efficiency* with which the design space is sampled either by using fewer sample points or seeking better coverage of the design space, and
2. the *accuracy* of the resulting surrogate model by using more complex approximations that are capable of fitting both linear and non-linear functions.

In this paper, we systematically compare several experimental design types and surrogate model types, which are widely used in the engineering design community, in terms of their capability to generate accurate approximations for computer experiments. In the next section, an overview of experimental design strategies for computer experiments is offered. This is followed in Section 1.2 with a summary of the different types of surrogate models that are being used for approximating computer experiments. Descriptions of the experimental design types and surrogate models employed in this study are given in Sections 2.1 and 2.2, respectively. Sections 3 and 4 contain two example problems wherein the different experimental design types and surrogate models are compared in terms of their capability to generate accurate approximations. The first example involves the design of a two-member frame subject to out-of-plane loading—it is a small three variable problem that is relatively inexpensive to analyze yet is characteristic of many engineering analyses used in structural optimization. The second example simulates roll-

over of a semi-tractor-trailer for a given design and driving conditions—the computational expense and size of this example (14 variables) yields a large complex problem, involving a highly nonlinear response. Based on the results of these examples, recommendations for efficient and accurate approximation model building are given in Section 5.

1.1 Experimental Designs for Computer Experiments

Properly designed experiments are essential for effective computer utilization. Experimental design techniques, which were initially developed for physical experiments, are finding considerable use for the design of computer experiments to increase the efficiency of these analyses. In the "classical" design and analysis of physical experiments (i.e., using central composite and factorial designs), random variation is accounted for by spreading the sample points out in the design space and by taking multiple data points (replicates) as shown in Figure 2a, Sacks, et al. (1989) state that the "classical" notions of experimental blocking, replication, and randomization are irrelevant when it comes to *deterministic* computer experiments; thus, sample points should be chosen to fill the design space for computer experiments. Consequently, many researchers advocate the use of "space filling" designs when sampling deterministic computer analyses to treat all regions of the design space equally. For instance, Sacks, et al. (1989) suggest minimizing the integrated mean squared error (IMSE) over the design region by using an IMSE-optimal design as shown in Figure 2b.
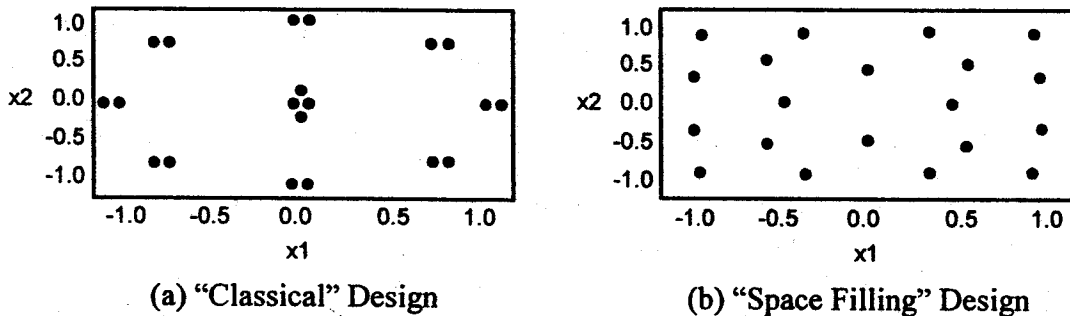

(a) "Classical" Design   (b) "Space Filling" Design

Figure 2. "Classical" and "Space Filling" Designs (Booker, 1998)

Koch, Mavris and Mistree (1998) investigate the use of a modified central composite design (CCD) that combines half-fractions of an inscribed CCD with a face-centered CCD to distribute points more evenly throughout the design space. Koehler and Owen (1996) describe several Bayesian and Frequentist "space filling" designs, including maximum entropy designs, mean squared-error designs, minimax and maximin designs, Latin hypercubes, randomized orthogonal arrays, and scrambled nets. Minimax and maximin designs were originally proposed by Johnson, Moore and Ylvisaker (1990) specifically for use with computer experiments. Sherwy and Wynn (1987; 1988) and Currin, et al. (1991) use the maximum entropy principle to develop designs for computer experiments. Tang (1993; 1994) describes orthogonal array-based Latin hypercubes which he asserts are more suitable for computer experiments than general Latin hypercubes. Park (1994) introduces optimal Latin hypercube designs for computer experiments which either

minimize IMSE or maximize entropy, spreading the points out over the design region. Beattie and Lin (1997) present a method to construct Latin hypercubes via rotated factorial designs. Fang and his co-authors (Fang, et al., 2000; Fang and Wang, 1994) use number-theoretic methods to develop uniform designs for use with computer experiments. Morris and Mitchell (Mitchell and Morris, 1992; Morris and Mitchell, 1992) propose maximin distance designs found within the class of Latin hypercube arrangements since they "offer a compromise between the entropy/maximin criterion, and good projective properties in each dimension." Owen (1992) advocates the use of orthogonal arrays as suitable designs for computer experiments, numerical integration, and visualization.

1.2 Approximation Models for Computer Experiments

As with experimental designs, a variety of approximation models and techniques exist for constructing "surrogates" of computationally expensive computer analysis and simulation codes. Response surface methodology (see, e.g., Box and Draper, 1987; Box, et al., 1978; Draper and Lin, 1990; Myers and Montgomery, 1995) and artificial neural network methods (see, e.g., Cheng and Titterington, 1994; Haykin, 1994; Smith, 1993) are two well-known approaches for constructing simple and fast approximations of complex computer analyses. An interpolative model known as kriging is also becoming widely used for the design and analysis of computer experiments (see, e.g., Barton, 1998; Booker, 1998; Currin, et al., 1991; Sacks, et al., 1989). Multivariate adaptive regression splines (Friedman, 1991) and radial basis function approximations (Dyn, et al., 1986; Powell, 1987) are also beginning to draw the attention of many researchers. Radial basis functions and multivariate adaptive regression splines are discussed in more detail in Section 2.2 along with response surface and kriging models.

In other work, Rasmussen (1990) offers an accumulated approximation technique for structural optimization which refines the approximation of objective and constraint functions by accumulating the function values of previously visited points. Similarly, Balling and Clark (1992) describe weighted and gradient-based approximations for use with optimization which utilize weighted sums of exact function values at sample points. Wavelet modeling uses a special form of a basis function which is especially effective in modeling sharp jumps in a response surface (Mallet, 1998). Friedman and Steutzle (1981) introduce projection pursuit regression which works well in high-dimensional (< 50) data and with large data sets (can handle 200,000+ data points). Projection pursuit regression takes the data and generates different projections of it along linear combinations of the variables; an optimizer finds the best projections and builds a predictor by summing them together with arbitrary levels of precision. Multivariate Hermite approximations for multidisciplinary design optimization are introduced in (Wang, et al., 1996). A comprehensive review of applications of approximation models and techniques in mechanical and aerospace systems can be found in (Simpson, et al., 2001b); a review of metamodeling applications in structural optimization can be found in (Barthelemy and Haftka, 1993) while applications in multidisciplinary design optimization can be found in (Sobieszczanski-Sobieski and Haftka, 1997).

Despite the variety of approximations that are available, comparative studies of these approaches are limited. Kriging methods are compared against polynomial regression models for the multidisciplinary design optimization of an aerospike nozzle in (Simpson,

et al., 2001a); kriging models and polynomial regression models are compared using two 5 and 10 variable test problems in (Giunta and Watson, 1998). In (Varadarajan, et al., 2000), artificial neural network methods are compared with polynomial regression models for modeling the nonlinear thermodynamic behavior of an engine design problem. In (Yang, et al., 2000), four approximation methods—enhanced multivariate adaptive regression splines, stepwise regression, neural networks, and the moving least squares— are compared for the construction of safety related functions in automotive crash analysis, for a relative small sampling size. In (Jin, et al., 2000), response surface models, radial basis functions, kriging models, and multivariate adaptive regression splines are systematically compared on a variety of test problems based on multiple measures of merit (e.g., accuracy, robustness, transparency, etc.). While no one approximation model dominated, recommendations based on problem size, degree of nonlinearity, and availability of sample data are given. The study conducted by Jin, et al. (2000) did not account for different types of experimental design strategies—only different sample sizes; therefore, our objective in this paper is to compare both experimental design strategies and approximation model types. The details of our approach are described next.

2. TECHNICAL APPROACH

Our objective in this paper is to systematically compare several experimental design types and surrogate modeling techniques in terms of their capability to generate accurate approximations of complex engineering analyses. In total, five experimental design types and four surrogate model types are utilized to build approximations for two example problems.

2.1 Experimental Designs

Four different types of "space filling" experimental design strategies are considered in this study: (1) Latin hypercubes, (2) Hammersley sequence sampling, (3) orthogonal arrays, and (4) uniform designs. A fifth type of design, namely, a set of randomly generated points, is also considered for each example. To enable direct comparisons, each design will employ comparable sample sizes. An overview of each experimental design type is offered next.

2.1.1 Latin Hypercubes

Latin hypercubes were the first type of design proposed specifically for computer experiments (McKay, et al., 1979). A Latin hypercube is a matrix of n rows and k columns where n is the number of levels being examined and k is the number of design (input) variables. Each of the k columns contains the levels 1, 2, ..., n, randomly permuted, and the k columns are matched at random to form the Latin hypercube. Latin hypercubes offer flexible sample sizes while ensuring stratified sampling, i.e., each of the input variables is sampled at n levels. These designs can have relatively small variance when measuring output variance (Sacks, et al., 1989).

2.1.2 Hammersley Sequence Sampling

Latin hypercubes are designed for uniformity along a single dimension where subsequent columns are randomly paired for placement on a k-dimensional cube. Hammersley sequence sampling provides a low-discrepancy experimental design for placing n points in a k-dimensional hypercube (Kalagnanam and Diwekar, 1997), providing better uniformity properties over the k-dimensional space than Latin hypercubes. A low discrepancy implies a uniform distribution of points in space.

2.1.3 Orthogonal Arrays

An orthogonal array is a matrix of n rows and k columns with every element being one of q symbols: $0, ..., q-1$(Owen, 1992). An orthogonal array has an associated strength t depending on the number of combinations of l levels appearing in any of the r columns of the array. The strength of the array and the number of levels combine to form the number of samples within the array. Orthogonal arrays provide an attractive class of sparse designs because they provide balanced (full factorial) designs for any projection into r factors (Barton, 1994).

2.1.4 Uniform Designs

A uniform design provides uniformly scatter design points in the experimental domain. A uniform design is a type of fractional factorial design with an added uniformity property; they have been popularly used since 1980 (see, Fang, 1980). If the experimental domain is finite, uniform designs are very similar to Latin hypercubes. When the experimental domain is continuous, the fundamental difference between these two designs is that in Latin hypercubes, points are selected at random from cells, whereas in a uniform design, points are selected from the center of cells. Furthermore, a Latin hypercube requires one-dimensional balance of all levels for each factor, while a uniform design requires one-dimensional balance and n-dimensional uniformity. Thus these designs are similar in one-dimension, but they can be very different in higher dimensions. Several uniform designs can be obtained from the website: <http://www.math.hkbu.edu.hk /UniformDesign>; for a recent review of uniform designs and their applications, see (Fang, et al., 2000).

In addition to the four specific types of experimental designs, sets of randomly generated points of equal sample size are considered for each example. The sample sizes for each experimental design are chosen based on the number of design variables and are discussed when each example problem is introduced. The approximation models employed in this study are discussed next.

2.2 Approximation Models

Four types of approximation models are investigated in this study: (1) polynomial response surfaces, (2) kriging models, (3) radial basis functions, and (4) multivariate adaptive regression splines. An overview of each type of approximation model is given in the following sections.

2.2.1 Response Surfaces

Originally developed for the analysis of physical experiments (Box and Wilson, 1951), polynomial response surface models have been used effectively for building approximations in a variety of applications. A second-order polynomial response surface model has the form:

$$\hat{y} = \beta_o + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_i \sum_j \beta_{ij} x_i x_j , \qquad (2.1)$$

where the β parameters are computed using least squares regression. Least squares regression minimizes the sum of the squares of the deviations of predicted values, $\hat{y}$ (x), from the actual values, y(x), using the equation:

$$\beta = [X'X]\text{-}1X'y \qquad (2.2)$$

where **X** is the design matrix of sample data points, **X'** is its transpose, and **y** is a column vector that contains the values of the response at each sample point. Polynomial response surface models can be easily constructed, and the smoothing capability allows quick convergence of noisy functions in optimization; however, there is always a drawback when applying polynomial response surfaces to model highly nonlinear or irregular behaviors.

2.2.2 Kriging

Originally developed for applications in geostatistics (see, e.g., Cressie, 1989; Cressie, 1993), a kriging model postulates a combination of a polynomial model and departures of the form:

$$\hat{y} = \sum_{j=1}^{k} \beta_j f_j(x) + Z(x) , \qquad (2.3)$$

where Z(x) is assumed to be a realization of a stochastic process with mean zero and spatial correlation function given by:

$$\text{Cov}[Z(x_i), Z(x_j)] = \sigma^2 R(x_i, x_j), \qquad (2.4)$$

where σ^2 is the process variance and R is the correlation. A variety of correlation functions can be chosen; however, the Gaussian correlation function proposed in (Sacks, et al., 1989) is the most frequently used. Furthermore, $f_j(x)$ is typically taken as a constant term. In our study, we use a constant term for $f_j(x)$ and a Gaussian correlation function with $p=2$ and $k\,\theta$ parameters, one θ for each of the k dimensions in the design space. Determining the maximum likelihood estimates of the $k\,\theta$ parameters used to fit the model is a k-dimensional optimization problem, which can require significant computational time if the sample data set is large, see (Simpson, et al., 1998; Simpson, et al., 2001a) for more details. The correlation matrix, **R**, can also become singular if multiple sample points are spaced close to one another or if the sample points are generated from particular designs. Fitting problems have been observed with some factorial designs and central composite designs when using kriging models (Meckesheimer, et al., 2001; Wilson, et al., 2001). Kriging methods are extremely flexible, however, due to the wide range of correlation

functions that have small amount of unknown coefficients. They can provide accurate predictions of highly nonlinear or irregular behaviors.

2.2.3 Radial Basis Functions

Radial basis functions were developed by Hardy (1971) and use linear combinations of a radially symmetric function based on Euclidean distance or similar metric to build approximation models. A simple radial basis function form is:

$$\hat{y} = \phi(\mathbf{x}) = \sum_i \beta_i \left\| \mathbf{x} - \mathbf{x}^i \right\| \tag{2.5}$$

where $\| \cdot \|$ represents the Euclidean norm, and the sum is taken over an observed set of system responses, $\{(\mathbf{x}^i, f(\mathbf{x}^i))\}$, $i = 1, ..., n$. Replacing $\phi(\mathbf{x})$ with $f(\mathbf{x}^i)$, and solving the resulting linear system yields the β_i coefficients. As commonly applied, the method is an interpolating approximation. Radial basis function approximations have produced good fits to arbitrary contours of both deterministic and stochastic response functions (Powell, 1987).

2.2.4 Multivariate Adaptive Regression Splines

Multivariate Adaptive Regression Splines (MARS) adaptively selects a set of basis functions for approximating the response function through a forward/backward iterative approach (Friedman, 1991). A MARS model can be written as:

$$\hat{y} = \sum_{m=1}^{M} a_m B_m(\mathbf{x}) \tag{2.6}$$

where a_m is the coefficient of the expansion, and B_m, the basis functions, can be represented as:

$$B_m(\mathbf{x}) = \prod_{k=1}^{K_m} \left[s_{k,m} \left(x_{v(k,m)} - t_{k,m} \right) \right]_+^q \tag{2.7}$$

where K_m is the number of factors (interaction order) in the m-th basis function, $s_{k,m} = +/-1$, $x_{v(k,m)}$ is the v-th variable, $1 \le v(k,m) \le n$, and $t_{k,m}$ is a knot location on each of the corresponding variables. The subscript '+' means the function is a truncated power function:

$$\left[s_{k,m} \left(x_{v(k,m)} - t_{k,m} \right) \right]_+^q = \begin{cases} \left[s_{k,m} \left(x_{v(k,m)} - t_{k,m} \right) \right]^q & s_{k,m} \left(x_{v(k,m)} - t_{k,m} \right) > 0 \\ 0 & otherwise \end{cases} \tag{2.8}$$

The major advantages of using the MARS procedure, however, appear to be accuracy and a major reduction in computational cost associated with constructing the approximation model. Compared to other techniques, the use of MARS for engineering design applications is relatively new. The algorithm described in (Chen, 1999) is utilized to build MARS models in this paper.

2.3 Assessing Model Accuracy

Since many of these approximation models interpolate the sample data, additional validation points are collected for each example to assess the accuracy of each

approximation model over the region of interest. For each set of validation points, the maximum absolute error (MAX) and root mean square error (RMSE) are computed as:

$$MAX = \max \{|y_i - \hat{y}_i|\}_{i=1,...,n_{error}} \qquad (2.9)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n_{error}} (y_i - \hat{y}_i)^2}{n_{error}}} \qquad (2.10)$$

where n_{error} is the number of additional validation points. While RMSE provides good estimates of the "global" error over the region of interest, MAX gives a good estimate of the "local" error by measuring the worst error within the region of interest, where a good approximation will have low RMSE and low MAX values. Finally, the average absolute error and the correlation determination (R^2) were also computed using the additional validation points. Error analysis revealed that average absolute error and R^2 were both highly correlated with RMSE for these two examples; therefore, neither measure is included in this paper, and only RMSE and MAX are used when analyzing the results.

3. EXAMPLE 1: STRUCTURAL ANALYSIS OF A TWO-MEMBER FRAME

3.1 Overview of Two-Member Frame Example

Our first example for testing the five experimental design and four surrogate modeling types comes from (Arora, 1989) and is a typical engineering analysis conducted during structural optimization. This example involves the design of a two-member frame subject to out-of-plane loads as shown in Figure 3. There are three design variables of interest: frame width (d), height (h) and wall thickness (t). The length, L, of each member is 100 in, and the load at node n_2 is P = -10,000 lbs. The stresses are calculated using the finite element method where U_1 is the vertical displacement at node n_2, U_2 is the rotation about bar n_3-n_2 and U_3 is the rotation about bar n_1-n_2.
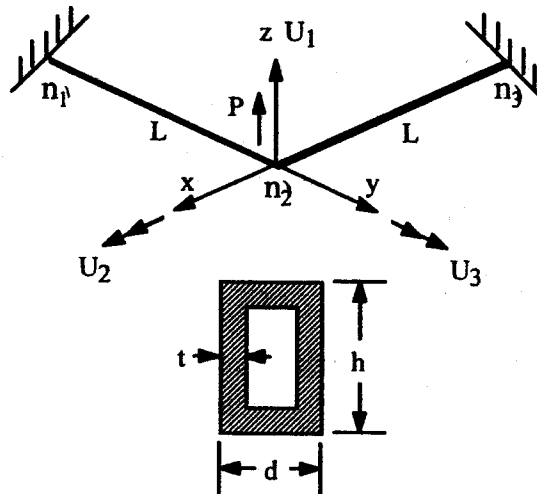

Figure 3. Two-Member Frame

The objective is to minimize the volume of the frame subject to stress constraints and bounds:

$$\text{Min.}_{d,h,t} \quad V(d, h, t) = 2L(2dt + 2ht - 4t^2) \quad (3.1)$$

$$\text{s.t.} \quad g_1(d, h, t) = 1 - \sigma_{e,n_1}/\sigma_{max} \geq 0 \quad (3.2)$$

$$g_2(d, h, t) = 1 - \sigma_{e,n_2}/\sigma_{max} \geq 0 \quad (3.3)$$

$$2.5 \text{ in.} \leq d \leq 10 \text{ in.}$$

$$2.5 \text{ in.} \leq h \leq 10 \text{ in.}$$

$$0.1 \text{ in.} \leq t \leq 1.0 \text{ in.}$$

The maximum allowable stress $\sigma_{max} = 40,000$ psi, and the effective stresses at nodes n_1 and n_2, σ_{e,n_1} and σ_{e,n_2}, are determined using finite element analysis as detailed in (Arora, 1989).

The objective is to build surrogate approximations of the objective function, Eqn. (3.1), and the two stress constraints, Eqns. (3.2) and (3.3), that are sufficiently accurate to be used in place of the original finite element analyses. Three-D grid plots of each equation are shown in Figure 4 to gain insight into their behavior over the region of interest. In all six plots, the horizontal axes are h and d varied over their range of interest (i.e., 2.5 to 10) while t is fixed at its lower (0.1) and upper bound (0.9).
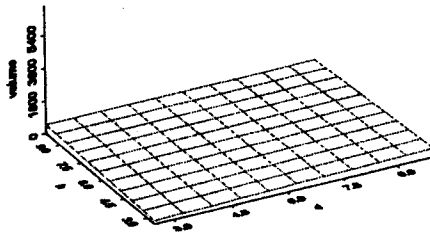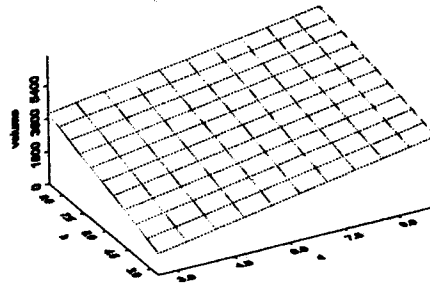

(a) Volume [FCN=1], t=0.1
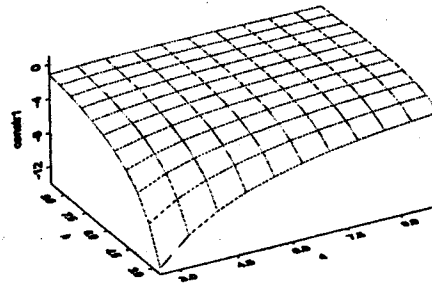
(b) Constraint 1 [FCN = 2], t=0.1
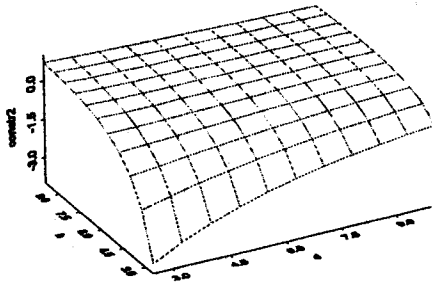
(c) Constraint 2 [FCN = 3], t=0.1

(d) Volume [FCN=1], t=0.9

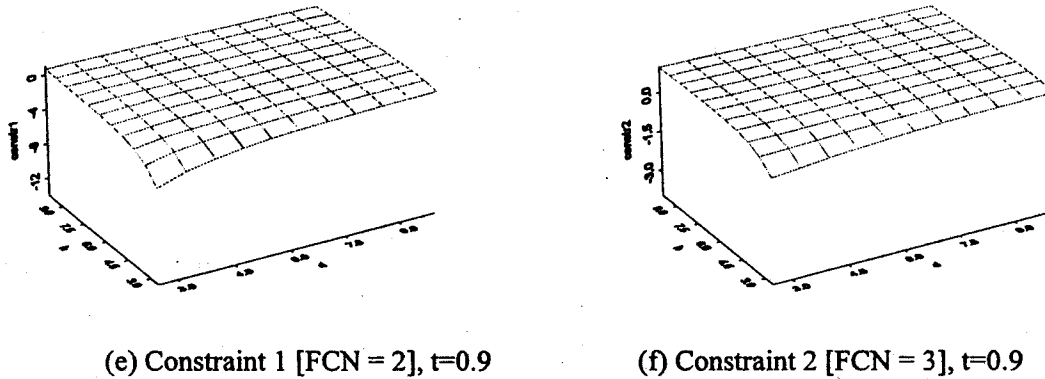(e) Constraint 1 [FCN = 2], t=0.9   (f) Constraint 2 [FCN = 3], t=0.9

Figure 4. 3-D Grid Plots of Objective Function and Constraints

From Figure 4, we see that the volume is a fairly smooth function, increasing gradually as t increases. Meanwhile, the two stress constraints are fairly flat for large values of t (i.e., the constraints are well satisfied since the stresses are low when t is large); however, for small values of t, there is a steep drop-off in one corner of the design space. The magnitude of the drop-off for the first stress constraint is nearly double that of the second (compare Figure 4b and 7c).

3.2 Experimental Set-Up for Two-Member Frame

As stated previously, the objective in this first example is to construct approximation models for the volume and the two constraints. There are three design variables—h, d, and t—whose ranges of interest are listed in the previous section. The experimental designs and approximations used in this first example are summarized as follows.

- *Experimental Design (DOE): 5 types* – Hammersley sequence (hss), Latin hypercube design (lhd), orthogonal array (oay), random set of points (rnd), uniform design (uni).
- *Sample size (SAMP): 6 sizes* – 9, 16, 25, 32, 49, 64.
- *Approximation Model (APPROX): 4 types* – kriging model (krg), radial basis function (rbf), second-order response surface (rs2), multivariate adaptive regression splines (mar).
- *Function (FCN): 3 types* – volume, stress constraint 1, stress constraint 2.

Note that a total of (5)(6)(4)(3) = 360 approximation models are constructed for this example based on the number of experimental design types, sample sizes, approximation model types, and functions being approximated. The number of sample sizes is based on available sample sizes of the orthogonal arrays and the minimum number of points needed to fit a second-order polynomial response surface. For the 9 point designs, the response surface models consist of only first-order effects and two-factor interactions since there is insufficient data to fit a full second-order model. So for each SAMP size and each DOE type, four different APPROX models are constructed for each of the three functions. To validate each approximation a set of 8000 additional validation points is used to compute MAX and RMSE, using Eqns. (2.9) and (2.10). The entire data set is available on the web

at <http://edog.me.psu.edu/IJORA/>. In the next section, each function is examined independently of the other functions in terms of MAX and RMSE due to the different magnitudes of each response.

3.3 Analysis of Results

Bubble plots for RMSE and MAX are given in Figure 5. In these plots, the strip across the top of each quadrant indicates the conditioning factor (i.e., SAMP size), and the size of the circle qualitatively depicts the magnitude of the corresponding value of RMSE or MAX. Since we desire minimum values of both error measures, smaller circles indicate a more accurate fit. We will first make some preliminary observations from the bubble plots and then provide in-depth interpretations through detailed analyses.

Consistent trend of sample size: As an initial consistency check, we note in Figure 5 that metamodel accuracy improves as sample size increases—the size of the circles in each graph get smaller and smaller as sample size increases from 9 to 64. This trend is most noticeable in the plots for RMSE, but also exists in the plots for MAX as the smallest circles occur at the large sample sizes, namely, 49 and 64. Overall, this trend is consistent with intuition as the accuracy of the metamodel should improve as more sample data becomes available.

Comparison of RMSE and MAX results: In Figure 5a and Figure 5b, we note that the combinations of SAMP size, DOE type, and APPROX type that yield low RMSE values for volume (FCN=1) also yield low MAX values. The same does not hold true for the other two functions, however. For instance, while the RMSE values for the 49 and 64 point designs for FCN=2 are low regardless for all DOE and APPROX types, only the orthogonal array (oay) designs yield consistently low MAX values for the 49 and 64 point designs as indicated by the smaller circles in Figure 5d. Meanwhile, the remaining designs yield sporadic results in that no DOE type dominates, nor does any combination of DOE type and APPROX type dominate.

Factors contributing to accuracy: Regarding APPROX type, it appears that all four types yield low RMSE and MAX values for FCN=1 except for the multivariate adaptive regression splines with the lowest sample size. Based on the smoothness of the volume as noted in Figure 4a, it is not surprising to have most APPROX types accurately model this function. However, the performance of the metamodels for the two stress constraints is not nearly as good. As noted earlier, SAMP size has a very strong impact on the accuracy. The type of DOE also appears to have some impact on accuracy; the Hammersley sampling sequence (hss) designs and uniform (uni) design tend to yield the smallest circles, while the random (rnd) sets of points and Latin hypercube designs (lhd) yield some of the largest circles. The response surface (rs2) models tend to do well regardless of SAMP size and DOE type, except for the smallest sample size. The kriging (krg) models approximate the stress constraints well, particularly in terms of RMSE; however, the MAX values appear to be about the same as those obtained from the other model types. The multivariate adaptive regression splines (mar) appear to perform very well for large sample sizes, particularly in terms of RMSE. The radial basis functions (rbf), on the

other hand, give spotted performance and yield some of the largest MAX values even for FCN=1, see Figure 5b. More detailed analysis of the impact of each factor and their interactions on metamodel accuracy is presented in Figures 6-9.
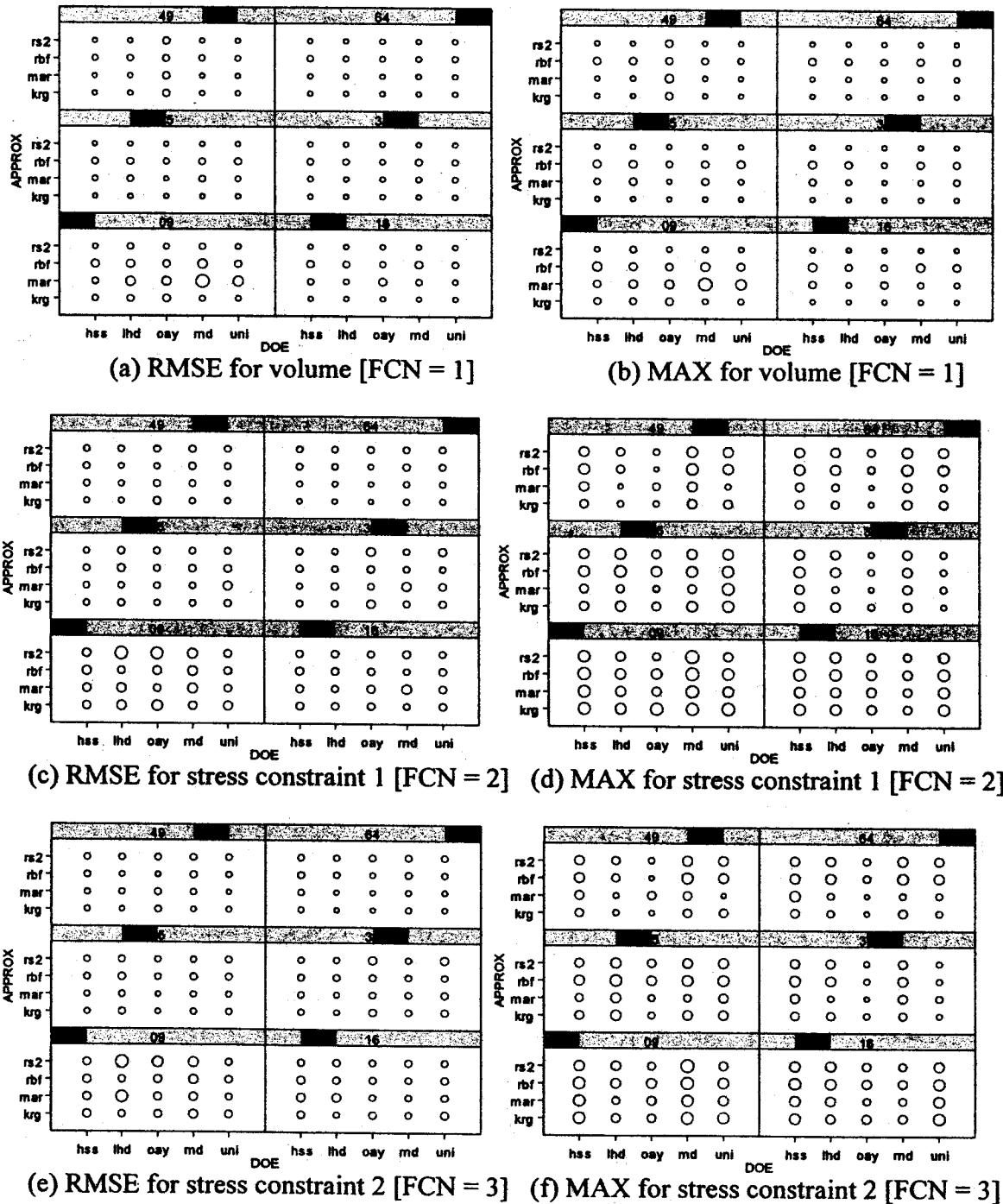

(a) RMSE for volume [FCN = 1]   (b) MAX for volume [FCN = 1]

(c) RMSE for stress constraint 1 [FCN = 2]   (d) MAX for stress constraint 1 [FCN = 2]

(e) RMSE for stress constraint 2 [FCN = 3]   (f) MAX for stress constraint 2 [FCN = 3]

Figure 5. Effects of DOE, APPROX, and SAMP on RMSE and MAX

Contributions of individual factors: The individual factor contributions for each of the three functions are plotted in figure 6. On the left-hand side of Figure 6, we plot the average effect of each factor on RMSE; on the right-hand side, we plot the average effect of each factor on MAX. For both measures, a lower value indicates a better fit. Our observations on the trend of sample size, the consistency between RMSE and MAX results, and the impact of various factors on accuracy from Figure 6 are consistent with the trends observed in Figure 5.

(a) RMSE for volume [FCN = 1]   (b) MAX for volume [FCN = 1]

(c) RMSE for stress constraint 1 [FCN = 2]   (d) MAX for stress constraint 1 [FCN = 2]

(e) RMSE for stress constraint 2 [FCN = 3]   (f) MAX for stress constraint 2 [FCN = 3]
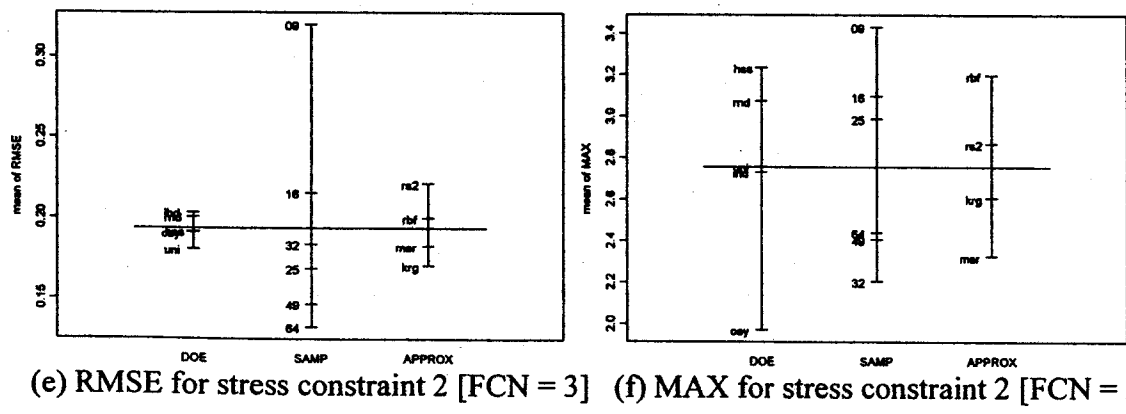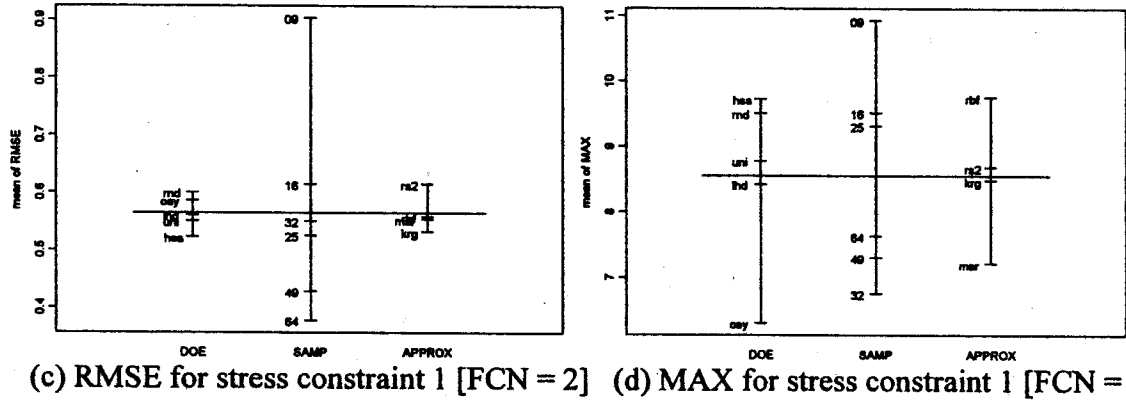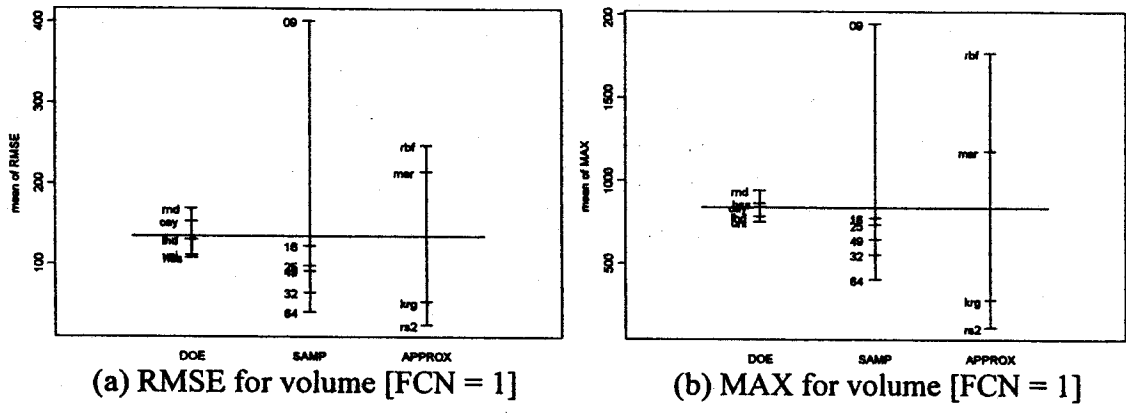
Figure 6. Individual Factor Contributions on RMSE and MAX

Impact of DOE type: We notice that DOE types are tightly spaced in terms of RMSE for all functions and MAX for FCN=1; however, a much wider spread exists in the impact of DOE type on MAX for two stress constraints. The uniform (uni) designs and Hammersley sampling sequences (hss) are consistently among the best performers at providing accurate models with low RMSE. However, as noted in the plots for MAX in Figure 6, the orthogonal arrays (oay) tend to yield the most accurate approximations with the Hammersley sampling sequences giving the worst or next to worst performance in all cases. The uniform (uni) designs tend to be average performers for the stress constraints when measured by MAX value. As expected, the random sets of points (rnd) give some of the worst results, particularly for RMSE. This is primarily because the random points do not guarantee good coverage of the design space when compared to a uniform design, Hammersley sampling sequence, or orthogonal array. Meanwhile, while a uniform design and Hammersley sampling sequence provide good coverage of the design space, and hence low RMSE, they do not position points at the corners as one finds in an orthogonal array. This is the main reason why the resulting approximations from these two designs do not capture the drop-offs of the two stress functions which lead to higher MAX errors, even though the global accuracy indicated by the low RMSE values is still good. This also accounts for the fairly tight grouping for RMSE—all of the DOE types provide reasonably accurate approximations from a global perspective (i.e., low RMSE value) with specialized designs (e.g., uniform designs) offering slight improvements over random sets of points and Latin hypercubes.

Impact of sample size: The overall effect of SAMP size is consistent with intuition (i.e., larger sample sizes yield more accurate models). A few discrepancies among the 32, 49, and 64 sample sizes are more difficult to explain and are investigated in more detail when discussing the interactions in Figure 8 and Figure 9. In Figure 6 we observe that the 9 and 16 point designs tend to yield the worst RMSE values for all three functions while the 64 point designs yield the best. The improvement gains in RMSE appear to lessen as sample size increases above 25 points, with the biggest gains occurring when moving from 9 to 16 to 25 points. The 25, 32, and 49 point designs alternate their rank ordering in terms of their impact on RMSE and MAX. It is also interesting to note that the 32 point designs yield some of the lowest MAX values. This is primarily due to the superior performance of the orthogonal arrays for this sample size as indicated by the small circles in the bubble plot Figure 5. We believe that the poor performance of the 49 and 64 point designs is due primarily to over-fitting the functions because they are so smooth—had the volume and stress constraints been more nonlinear, taking more sample points would most likely continue to improve the accuracy of the approximation.

Impact of approximation type: From Figure 6 we note that the kriging (krg) models tend to yield the most accurate approximations as measured by both RMSE and MAX. The second-order response surface (rs2) models yield very accurate predictions for FCN=1 since Eqn. 25 is quadratic; however, as expected, their performance in modeling the stress constraints is not as good, particularly in terms of MAX. The radial basis functions (rbf) yield some of the worst approximations, compared to the other APPROX types, especially when MAX is considered. Finally, it is interesting to note that the multivariate adaptive regression splines (mar) provide much better approximations for the stress constraints

(FCN=2 and 3) than they do for volume, and they yield the lowest MAX values for both stress constraints. This is consistent with previous observations regarding the bubble plots in Figure 5.

Contributions of Interactions between Factors: Having looked at the individual factor contributions, the next step is to examine the interactions between different pairs of factors. Three interactions are studied: (1) DOE and APPROX, (2) DOE and SAMP, and (3) APPROX and SAMP, and the effect of each interaction is plotted in Figure 7, Figure 8, and Figure 9, respectively. The first set of interactions—APPROX type and DOE type—plotted in Figure 7, shows the impact of each APPROX type for a given DOE type, averaged over all sample sizes. The results are segmented by RMSE value and MAX value and plotted independently for each function to be consistent with previous graphs.

Interaction between DOE and APPROX type (Figure 7): We first notice the tight grouping of DOE types for the response surface (rs2) models in Figures 7a and 7b. This indicates that the response surface models accurately approximate the volume (FCN=1) independent of DOE type, which is not surprising given the quadratic nature of Eqn. 25. No other DOE-APPROX combination is as tightly grouped as found for volume. Looking at DOE types, the uniform designs (uni) appear to provide the least variation among APPROX types, yielding a nearly horizontal line for stress constraint 1 (Figure 7c). All of the DOE types appear to fluctuate when looking at MAX values; however, the orthogonal arrays are consistently among the best performers, independent of APPROX type, as seen previously in Figure 6. It also appears that the Hammersley sampling sequences (hss) work particularly well with the multivariate adaptive regression splines (mar) to yield fairly low RMSE and MAX values for all three functions. The parallel lines in Figure 7d and Figure 7f indicate that the interaction effect of DOE and APPROX type on MAX values is very small for both constraint functions. We also notice that the multivariate adaptive regression splines (mar) method is very DOE type dependent, indicating that it is the least robust method with respect to different DOE types.

Interactions between DOE and SAMP (Figure 8): Despite some jumpiness, the general trend for each DOE type in Figure 8 is to improve accuracy as the SAMP size increases; this is particularly noticeable in the RMSE plots for all three functions. The sharp spike for the 49 point OA seen in Figure 8a is primarily due to poor randomization within the orthogonal array and subsequently poor approximations as indicated by the slightly larger circles in Figure 6a. The Latin hypercube designs (lhd) and random sets of points (rnd) yield some of the worst RMSE and MAX values, regardless of sample size. This can be attributed to the random positioning of points in both types of designs. For instance, we see that the random sets of points (rnd) perform very poorly for low sample sizes when MAX is considered; they also yield high MAX values even with large SAMP sizes for the two stress constraints. Meanwhile, the orthogonal arrays (oay) yield the lowest MAX values for the two stress constraints. In Figure 8, the uniform (uni) designs offer some of the lowest RMSE values for many SAMP sizes; however, their performance in terms of MAX is not as good due mainly to the positioning of points within the design space as discussed previously. The Hammersley sampling sequences (hss) are average performers for low sample sizes, but they start to perform quite well for large SAMP sizes (25 and
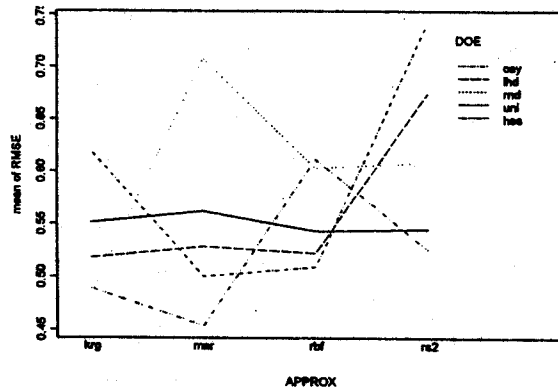
above). Finally, we note that capability of most designs to yield further improvements in
RMSE tends to start to level off at 25 and 32 points, i.e., taking more than 25 or 32 points
does not yield significant gains in the overall accuracy of the resulting approximation.
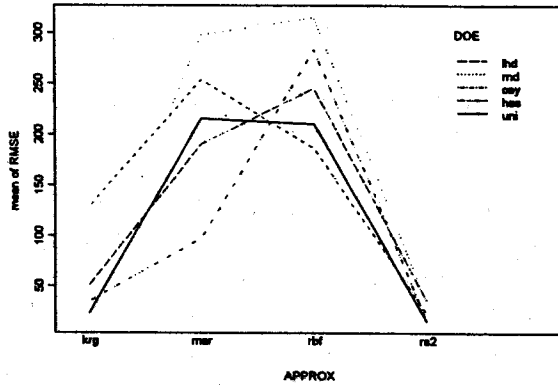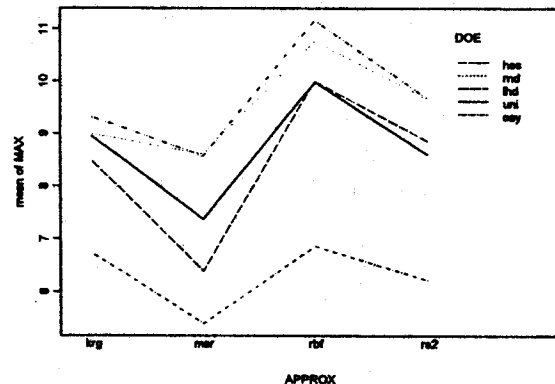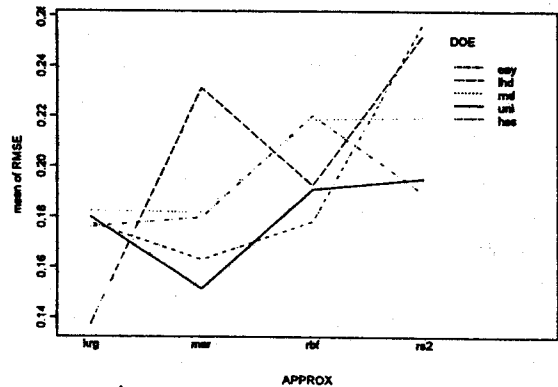

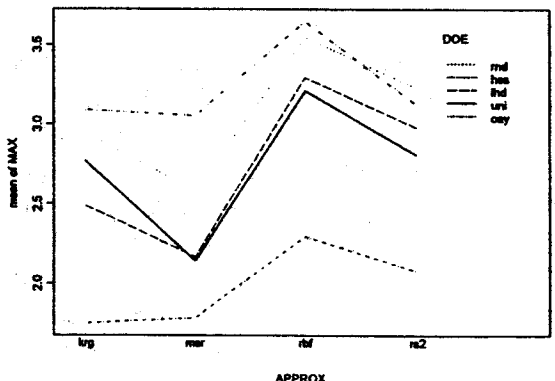(a) RMSE for volume [FCN = 1]   (b) MAX for volume [FCN = 1]
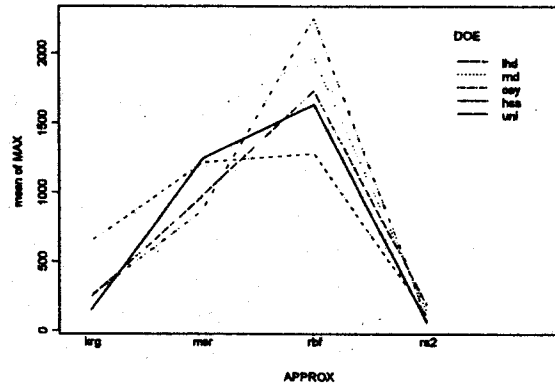
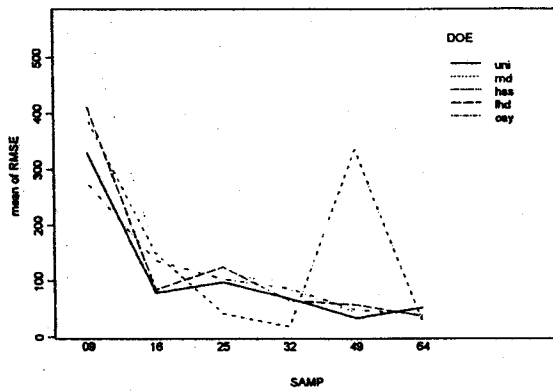(c) RMSE for stress constraint 1 [FCN = 2]   (d) MAX for stress constraint 1 [FCN = 2]
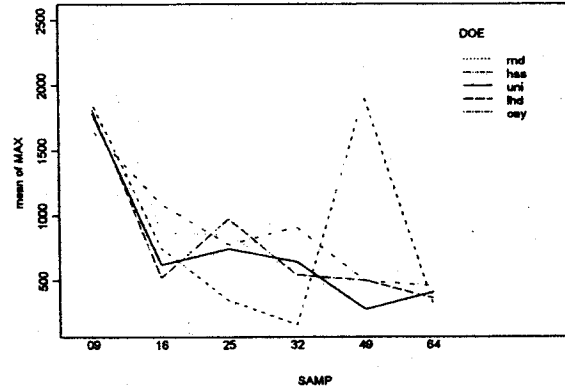
(e) RMSE for stress constraint 2 [FCN = 3]   (f) MAX for stress constraint 2 [FCN = 3]

Figure 7. Interaction of DOE and APPROX on RMSE and MAX

(a) RMSE for volume [FCN = 1]

(b) MAX for volume [FCN = 1]

(c) RMSE for stress constraint 1 [FCN = 2]

(d) MAX for stress constraint 1 [FCN = 2]

(e) RMSE for stress constraint 2 [FCN = 3]

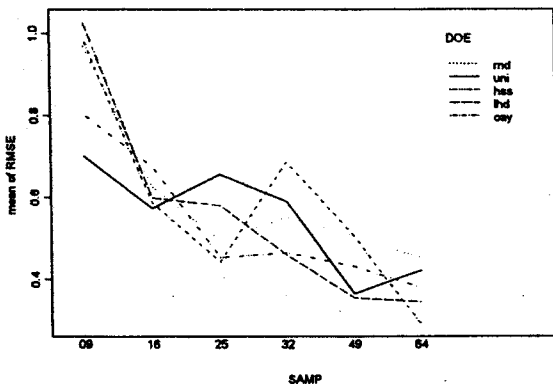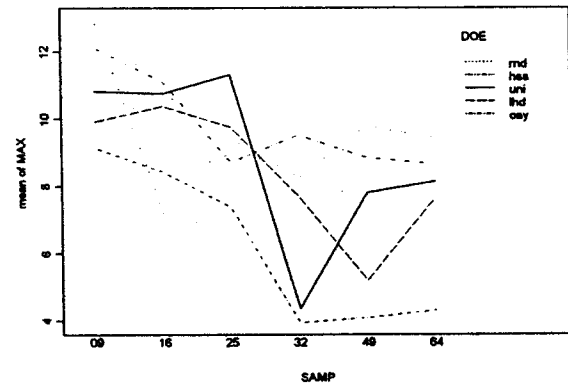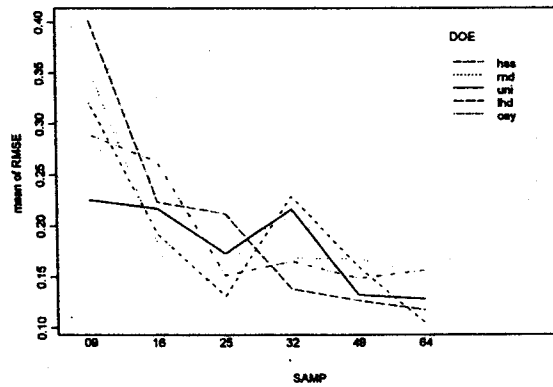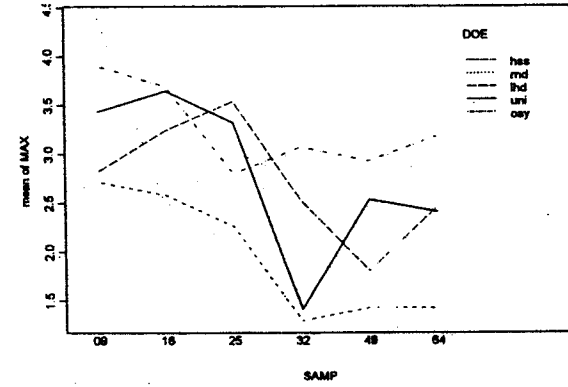(f) MAX for stress constraint 2 [FCN = 3]

Figure 8. Interaction of DOE and SAMP on RMSE and MAX

Interaction between APPROX type and SAMP size (Figure 9): As a first check, all APPROX types seem to improve as SAMP increases in Figure 9, providing consistency with previous findings and intuition. Of all the APPROX type, however, the multivariate adaptive regression splines (mar) are the most affected by SAMP size—the multivariate adaptive regression splines are among the worst performers at small SAMP sizes but are among the best performers at large SAMP sizes based on RMSE and MAX. We also

notice that the kriging models perform well at low SAMP sizes as do the radial basis functions (rbf), which yield the most accurate RMSE values for the two stress constraints at the lowest SAMP size. Meanwhile, the performance of the response surface (rs2) models tend to level off for SAMP sizes greater than 9 when looking at RMSE values, and this is due to the inability to fit a full second-order model when only 9 points are available. Only ten sample points are required to fit a full second-order model for 3 variables, but it appears that once we are able to fit a full second-order response surface (rs2) model that little improvement in RMSE is obtained by increasing SAMP size further.

(a) RMSE for volume [FCN = 1]   (b) MAX for volume [FCN = 1]

(c) RMSE for stress constraint 1 [FCN = 2]   (d) MAX for stress constraint 1 [FCN = 2]

(e) RMSE for stress constraint 2 [FCN = 3]   (f) MAX for stress constraint 2 [FCN = 3]
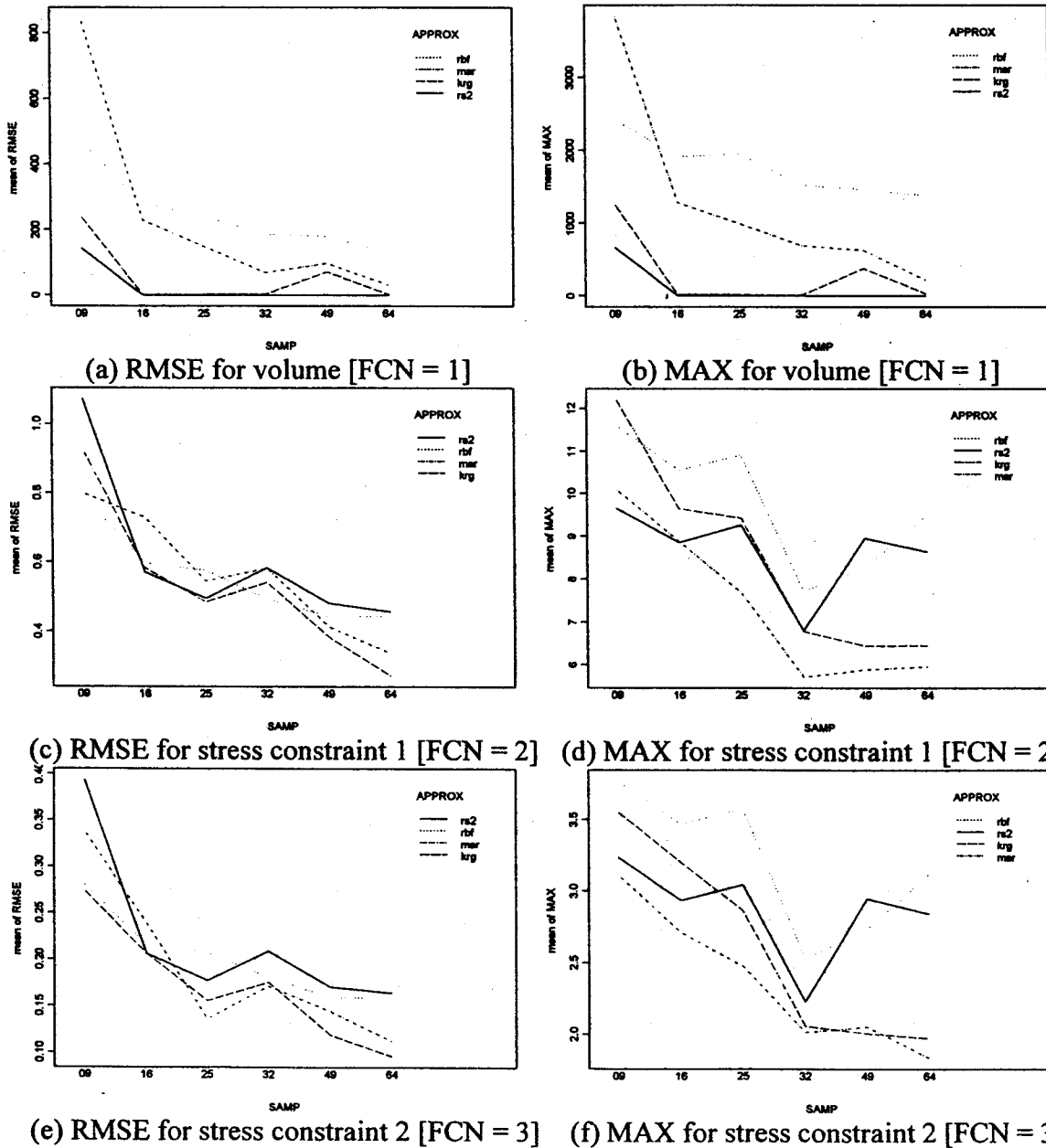
Figure 9. Interaction of SAMP and APPROX on RMSE and MAX

To determine if similar trends exits for larger problems containing many design variables (>10) and a highly nonlinear response, a fourteen variable problem involving the analysis of semi-tractor trailer roll-over is next.

4. EXAMPLE 2: ROLL-OVER ANALYSIS OF A SEMI-TRACTOR TRAILER

Our second example is a real engineering problem that analyzes vehicle design to improve a vehicle's handling characteristics, particularly the prevention of roll-over (Chen, et al., 1999). The simulator used is the integrated computer tool ArcSim (ArcSim, 1997; Sayers and Riley, 1996) developed at the University of Michigan for simulating and analyzing the dynamic behavior of 6-axle tractor-semitrailers. Each simulation takes more than three minutes to run on a Sun UltraSparc 1 workstation; therefore, using ArcSim during optimization imposes heavy computational costs. An overview of ArcSim is given in the next section; details of the example and experimental set-up are discussed in Section 4.2 with result analysis in Section 4.3.

4.1 Overview of ArcSim Example

ArcSim can simulate responses of tractor-semitrailers to user-defined steering and braking inputs on both flat and inclined surfaces (ArcSim, 1997). The program contains a nonlinear 3-D mathematical model with 91 state variables, a nonlinear tire model, and a detailed steering system model with major compliance effects. ArcSim also considers solid-axle suspensions and major suspension effects. In this study, 14 input variables are considered which include nine suspension and vehicle parameters as design variables and five uncontrollable (i.e., noise) factors for steering and braking. The response of interest is the vehicle handling performance, which is measured by the roll-over metric. The previous studies (Chen, et al., 1999) indicate that the roll-over metric has a highly nonlinear dependence on the control and noise variables, especially for different combinations of brake and steering levels. A description and the range of interest for each of the 14 input parameters are summarized in Table 1. All of the variables except brake_end have a range of +/- 20 % from their nominal values (i.e., the values for the baseline design); brake_end varies by +/- 15% to avoid overlap with the steering parameters.

In most cases, roll-over occurs due to extreme steering and braking inputs; therefore, the steering and braking parameters are taken as noise factors. Five noise factors are chosen: three corresponding to the braking inputs, and two corresponding to the steering inputs. Their ranges are also listed in Table 1. The ranges of brake_start and brake are +/- 15% from their nominal values to avoid overlap of the two parameters, whereas the other parameter ranges are +/- 20% from their nominal values. The level of braking is the amount of braking pressure applied. The level of steering is the angle the steering wheel is turned. The starting and ending times define total time of braking and steering, respectively.

Table 1. ArcSim Variables and Ranges of Interest

Design Variable	Description	Lower Bound	Upper Bound
HH1	Height of Hitch above ground	51.2 in	76.8 in
KHX1	Hitch roll torsional stiffness	8e5 in-lb/deg	1.2e6 in-lb/deg
LTS11	Distance between springs on Axle 1	30.4 in	45.6 in
LTS123	Distance between springs on Axles 2 & 3	30.4 in	45.6 in
LTS2123	Distance between springs on Axles 4,5 & 6	30.4 in	45.6 in
M11	Laden load for Axle 1	11540 lbm	17310 lbm
M2123	Laden load for Axles 4, 5 and 6	16274.4 lbm	24411.6 lbm
KT2123	Axles 4, 5 & 6 tire stiffness	4139.20 lb/in	6208.80 lb/in
SCFS11	Axle 1 spring stiffness scale factor	0.8	1.2
Noise Variables	**Description**	**Lower Bound**	**Upper Bound**
brake_start	Time at which braking is applied	1.02 sec	1.38 sec
brake_level	Level of braking that is applied	70 psi	100 psi
brake_end	Time after which braking is no longer applied	1.53 sec	2.07sec
steer_level	Level of steering that is applied	60 deg	100 deg
steer_end	Time after which steering is no longer applied	2.16 sec	3.24 sec

In terms of the vehicle handling response, it is assumed that if the roll-over angle becomes greater than 45°, roll-over will occur. The roll-over metric is one of ArcSim's outputs and is defined as the square root of the integral of the square of the roll-over angle in a 5-second period:

$$R = \sqrt{\int_0^5 roll\_angle^2 dt} \qquad (4.1)$$

Five seconds is chosen as the upper limit of the integration based on previous studies (Chen, et al., 1999). The square of the roll-over angle is used as the metric since the roll-over angle can take negative or positive values depending on whether roll-over is to the left or the right. Based on this definition, we note that the value of the roll-over metric is desired to be as small as possible. The plot of roll-over metric versus brake level and steering level shown in Figure 10 illustrates the high non-linearity, which is further complicated by the high dimensionality of the problem. The experimental design set-up used for this example is described next.
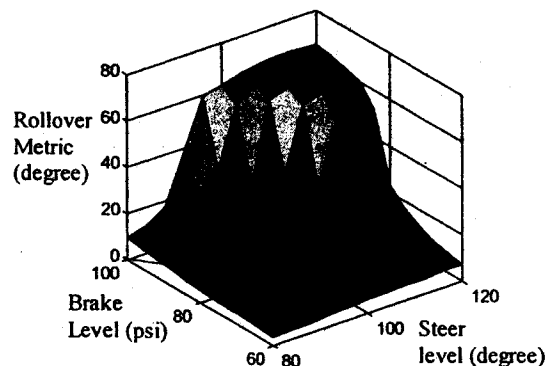

Figure 10. 3-D Plot of Roll-Over Metric Versus Brake Level and Steering Level

4.2 Experimental Set-Up for ArcSim Example

The objective in this second example is to construct an accurate approximation model for the roll-over metric computed by ArcSim. There are a total of fourteen input variables as listed in Table 1, and the ranges of interest for each variable are also listed in Table 1. The factors and levels considered in this example are summarized as follows.

- *Experimental Design (DOE): 5 types* – Hammersley sequence (hss), Latin hypercube design (lhd), orthogonal array (oay), random set of points (rnd), uniform design (uni)
- *Sample size (SAMP): 4 sizes* – 128, 169, 256, 361
- *Approximation Model (APPROX): 4 types* – kriging model (krg), radial basis function (rbf), second-order response surface (rs2), multivariate adaptive regression splines (mar)
- *Function (FCN): 1 type* – roll-over metric

Notice that a total of (4)(5)(4)(1) = 80 approximation models are constructed for this example based on the number of experimental design types, sample sizes, approximation model types, and functions being approximated. The sample sizes are based on available sample sizes of the orthogonal arrays and the minimum number of points needed to fit a second-order polynomial response surface. For each approximation of the roll-over metric, a set of 1000 additional validation points is used to compute MAX and RMSE, using Eqns. 9 and 10. The analysis of the results is next; the complete data set is available on the web at <http://edog.me.psu.edu/IJORA/>.

4.3 Analysis of ArcSim Results

Following a similar order to that in Section 3.3, a bubble plot of the effects of DOE, APPROX, and SAMP on RMSE and MAX is shown in Figure 11. Blank spaces in the bubble plot indicate that an approximation model could not be fit for that particular combination of DOE and SAMP. For instance, all of the Hammersley sequence sampling (hss) designs yielded singular design matrices (X'X), preventing a second-order response surface (rs2) model from being fit. Meanwhile, several of the multivariate adaptive regression splines yielded extremely poor approximations due to numerical round-off error and were consequently removed from the data set. This trend is consistent with previous results wherein large sample sizes are needed in order to fit accurate multivariate adaptive regression splines.

Factors contributing to accuracy: Looking at the data in Figure 11, the radial basis functions (rbf) and kriging (krg) models appear to offer good approximations for a wide variety of DOE and SAMP sizes. These two approximation models provide the majority of the lowest RMSE and MAX values (i.e., smallest circles) for the roll-over metric. Meanwhile, the multivariate adaptive regression splines (mar) and response surface models only yield accurate approximations for large sample sizes. The Latin hypercube designs (lhd) and random sets of points (rnd) yield some of the least accurate models as indicated by the large circles associated with many of the approximations constructed

from these two DOE types. The orthogonal arrays (oay) appear to be the most robust, yielding accurate approximations for a variety of SAMP sizes and APPROX types. The orthogonal arrays are followed closely by the uniform (uni) designs, and the Hammersley sampling sequences based on the sizes of the circles for RMSE and MAX. To further understand the impact of each DOE type, APPROX type, and SAMP size on modeling accuracy, the individual factor contributions for modeling the roll-over metric are shown in Figure 12.

Impact of DOE type: Since we want to minimize both MAX and RMSE, the orthogonal array (oay) and uniform designs (uni) appear to be the experimental designs of choice, followed closely by Hammersely sampling sequences (hss) and Latin hypercubes (lhd). The worst possible DOE types, as expected, is the random sets of points (rnd). It is again reassuring to note that both MAX and RMSE decrease as the sample size increases.
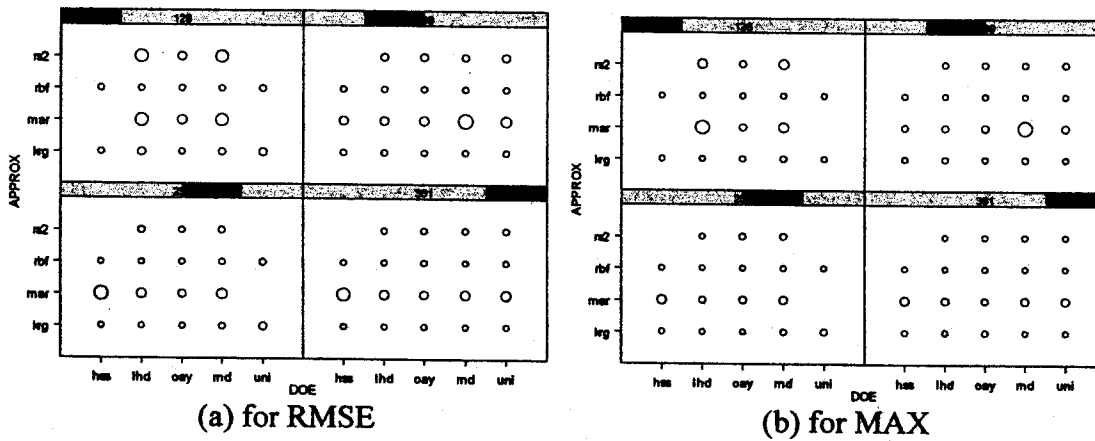

(a) for RMSE   (b) for MAX

Figure 11. Effects of DOE, APPROX, and SAMP on RMSE and MAX

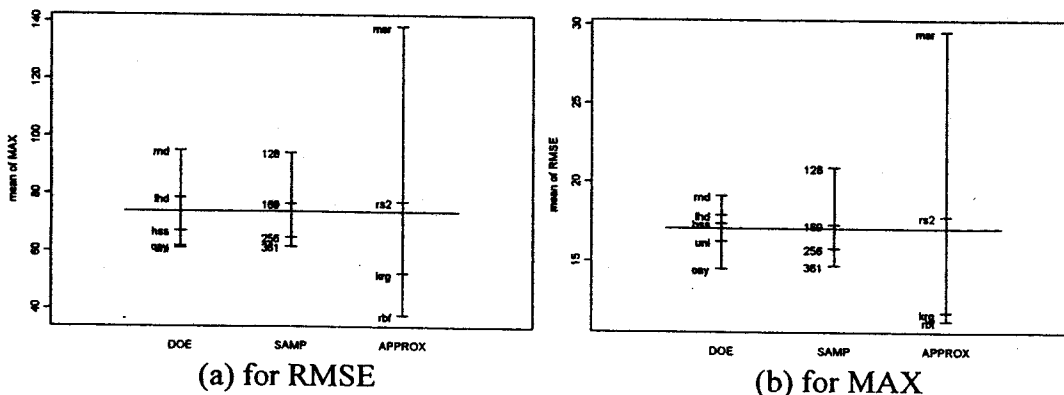(a) for RMSE   (b) for MAX

Figure 12. Factor Contributions on Model Accuracy

Impact of APPROX type: The radial basis function (rbf) and kriging models (krg) both perform well in terms of MAX and RMSE as shown in Figure 12. The radial basis function models offer a slight improvement in MAX error over the kriging (krg) models, while both approximations yield similar RMSE values. The second-order response surface (rs2) models yield average results for both error measures while the multivariate adaptive regression splines (mar) are the most inaccurate, particularly when small sample sizes are used. This is consistent with previous results (Jin, et al., 2000).

Interaction between DOE and SAMP (Figure 13): Consistent with the previous figure, most of the lines in Figure 13 are negatively sloped, indicating increased accuracy as sample size increases; however, there are some exceptions. Most notably, the MAX values for the uniform design (uni) rise slightly as sample size increases, but the corresponding RMSE values tend to decrease. Meanwhile, the Hammersley sampling sequences (hss) show the most pronounced increase in MAX and RMSE when moving to the 361 point designs, while the other DOE types level off at the higher sample sizes.
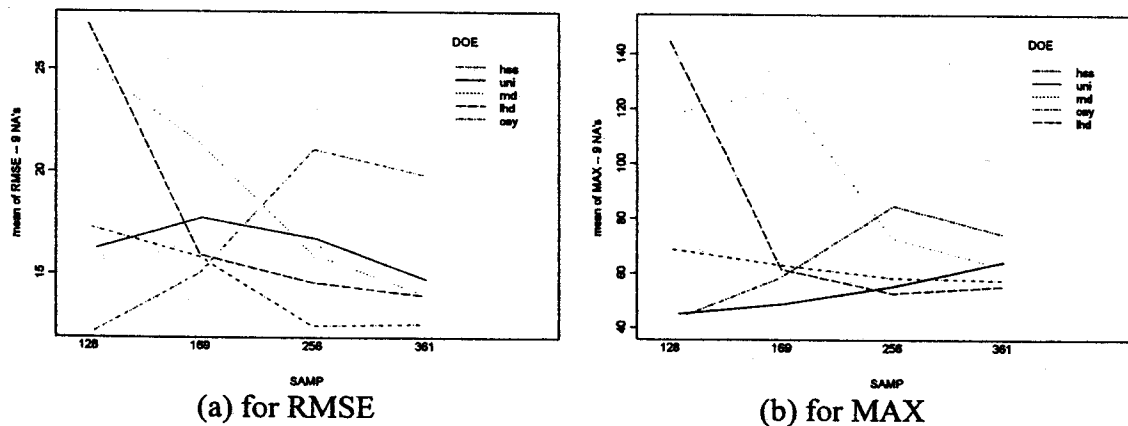

(a) for RMSE   (b) for MAX

Figure 13. Interaction of DOE and SAMP

Interaction between DOE and APPROX (Figure 14): We see from Figure 14 that both the kriging (krg) and radial basis function (rbf) models yield good results, regardless of DOE type; the same does not hold true for the response surface (rs2) models or the multivariate adaptive regression splines (mar). Of the five design types available, the uniform designs (uni) and orthogonal arrays (oay) seem to work well with both multivariate adaptive regression splines (mar) and radial basis functions (rbf) while the random points provide the worst data set for fitting accurate approximations.

Interaction between APPROX type and SAMP size: The interaction between APPROX and SAMP is not plotted for this example because the only interaction that exists has already been captured, namely, as SAMP size increases, each APPROX yields more accurate results. There are no big jumps in accuracy for any particular APPROX type as observed previously.
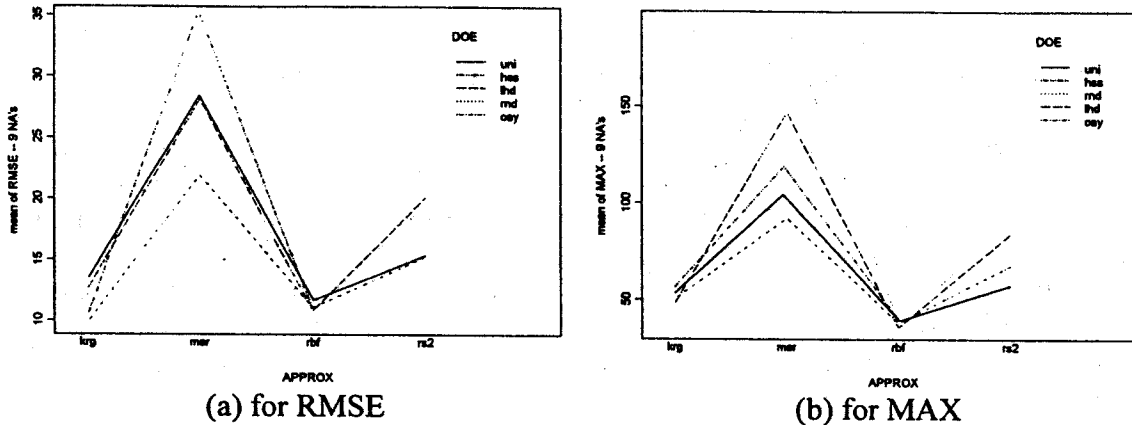
(a) for RMSE   (b) for MAX

Figure 14. Interaction of DOE and APPROX

5. RECOMMENDATIONS AND CLOSING REMARKS

Based on the comparative study results for two engineering example problems, one small dimension ($k = 3$) with both low-order and higher-order nonlinear functions, and the other large dimension ($k=14$) with higher-order nonlinear behavior, some general conclusions can be drawn regarding the selection of DOE type, the approximation type and the sample size.

For DOE type, we find that good design space coverage afforded by the uniform designs and Hammersley sampling sequences tend to yield more accurate approximations globally as indicated by the consistently low RMSE values associated with them. The uniform designs tend to perform well at low sample sizes while the Hammersley sampling sequences tend to fair better when large sample sizes can be afforded, but both offer improvements over standard Latin hypercube designs and random sets of points. The orthogonal arrays do particularly well at giving low MAX values because these designs place points at the corners of the design space which is critical when trying to approximate the two stress constraints in the first example. This type of behavior may not always be present in a system, and we recommend a design that provides good overall coverage (and therefore lower RMSE) be chosen over one that yields low MAX—validation of the approximation during use can always help correct large MAX values.

For APPROX type, the kriging (krg) and radial basis function (rbf) models tend to offer more accurate approximations over a wide range of DOE types and SAMP sizes. The performance of the multivariate adaptive regression splines (mar) is the least stable; its performance varies quite a lot when different sample sizes or DOE types are available. In both examples, large sample sizes are needed to fit accurate multivariate adaptive regression splines. The second-order response surfaces yield average results and perform particularly well when approximating low-order non-linear functions.

For SAMP size, larger sizes generally improve the accuracy, however, for low-order non-linear functions, we also find that taking large samples for many approximation types, with the exception of the multivariate adaptive regression splines, does not improve accuracy that much.

ACKNOWLEDGEMENT

REFERENCES

ArcSim, (1997), ArcSim User's Guide, *PDF version can be downloaded from http://arc.engin.umich.edu/sw_distri/ARCSIM/arcsim.html,* The University of Michigan, Ann Arbor, MI.

Arora, J. S., (1989), *Introduction to Optimum Design,* McGraw-Hill, New York.

Balling, R. J. and Clark, D. T., (1992), A Flexible Approximation Model for Use with Optimization, *4th AIAA/USAF/NASA/OAI Symposium on Multidisciplinary Analysis and Optimization,* Cleveland, OH, AIAA, 2, 886-894. AIAA-92-4801-CP.

Barthelemy, J.-F. M. and Haftka, R. T., (1993), Approximation Concepts for Optimum Structural Design - A Review, *Structural Optimization,* 5, 129-144.

Barton, R. R., (1992), Metamodels for Simulation Input-Output Relations, *Proceedings of the 1992 Winter Simulation Conference* (Swain, J. J., Goldsman, D., et al., eds.), Arlington, VA, IEEE, 289-299.

Barton, R. R., (1994), Metamodeling: A State of the Art Review, *Proceedings of the 1994 Winter Simulation Conference* (Tew, J. D., Manivannan, S., et al., eds.), Lake Beuna Vista, FL, IEEE, 237-244.

Barton, R. R., (1998), Simulation Metamodels, *Proceedings of the 1998 Winter Simulation Conference (WSC'98)* (Medeiros, D. J., Watson, E. F., et al., eds.), Washington, DC, IEEE, 167-174.

Beattie, S. D. and Lin, D. K. J., (1997), Designing Computer Experiments: Rotated Factorial Designs, *Technical Report No. 97-06,* Department of Statistics, The Pennsylvania State University, University Park, PA.

Booker, A. J., (1998), Design and Analysis of Computer Experiments, *7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis & Optimization*, St. Louis, MO, AIAA, **1**, 118-128. AIAA-98-4757.

Box, G. E. P. and Draper, N. R., (1987), *Empirical Model Building and Response Surfaces*, John Wiley & Sons, New York.

Box, G. E. P., Hunter, W. G. and Hunter, J. S., (1978), *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*, John Wiley & Sons, New York.

Box, G. E. P. and Wilson, K. B., (1951), On the Experimental Attainment of Optimal Conditions, *Journal of the Royal Statistical Society*, Vol. Series B, **13**, 1-38 (with Discussion).

Chen, V. C. P., (1999), Application of MARS and Orthogonal Arrays to Inventory Forecasting Stochastic Dynamic Programs, *Computational Statistics and Data Analysis*, **30**, 317-341.

Chen, W., Garimella, R. and Michelena, N., (1999), Robust Design for Improved Vehicle Handling under a Range of Maneuver Conditions, *Advances in Design Automation*, Las Vegas, NV, ASME, Paper No. 99-DETC/DAC-8580.

Cheng, B. and Titterington, D. M., (1994), Neural Networks: A Review from a Statistical Perspective, *Statistical Science*, **9(1)**, 2-54.

Cressie, N., (1989), Geostatistics, *The American Statistician*, **43(4)**, 197-202.

Cressie, N. A. C., (1993), *Statistics for Spatial Data*, Revised, John Wiley & Sons, New York.

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*, **86(416)**, 953-963.

Draper, N. R. and Lin, D. K. J., (1990), Connections Between Two-Level Designs of Resolutions III and V, *Technometrics*, **32(3)**, 283-288.

Dyn, N., Levin, D. and Rippa, S., (1986), Numerical Procedures for Surface Fitting of Scattered Data by Radial Basis Functions, *SIAM Journal of Scientific and Statistical Computing*, **7(2)**, 639-659.

Fang, K. T., (1980), Experimental Design By Uniform Distribution, *Acta Mathematice Applicatae Sinica*, **3**, 363-372.

Fang, K.-T., Lin, D. K. J., Winker, P. and Zhang, Y., (2000), Uniform Design: Theory and Application, *Technometrics*, **42**, 237-248.

Fang, K.-T. and Wang, Y., (1994), *Number-theoretic Methods in Statistics*, Chapman & Hall, New York.

Friedman, J. H., (1991), Multivariate Adaptive Regression Splines, *The Annals of Statistics*, **19(1)**, 1-67.

Friedman, J. H. and Steutzle, W., (1981), Projection Pursuit Regression, *Journal of the American Statistical Association*, **76(376)**, 817-823.

Gangadharan, S. N., Haftka, R. T. and Fiocca, Y. I., (1995), Variable-Complexity-Modeling Structural Optimization Using Response Surface Methodology, *36th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference and AIAA/ASME Adaptive Structures Forum*, New Orleans, LA, AIAA, AIAA-95-1164.

Giunta, A. and Watson, L. T., (1998), A Comparison of Approximation Modeling Techniques: Polynomial Versus Interpolating Models, *7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis & Optimization*, St. Louis, MO, AIAA, **1**, 392-404. AIAA-98-4758.

Goldsman, D. and Nelson, B. L., (1998), Statistical Screening, Selection, and Multiple Comparison Procedures in Computer Simulation, *Proceedings of the 1998 Winter Simulation Conference (WSC'98)* (Medeiros, D. J., Watson, E. F., et al., eds.), Washington, DC, IEEE, 159-166.

Hardy, R. L., (1971), Multiquadratic Equations of Topography and Other Irregular Surfaces, *Journal of Geophysical Research*, **76**, 905-1915.

Haykin, S., (1994), *Neural Networks: A Comprehensive Foundation*, Macmillan Publishing, New York.

Jin, R., Chen, W. and Simpson, T. W., (2000), Comparative Studies of Metamodeling Techniques under Multiple Modeling Criteria, *8th AIAA/NASA/USAF/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, Long Beach, CA, AIAA, AIAA-2000-4801.

Johnson, M. E., Moore, L. M. and Ylvisaker, D., (1990), Minimax and Maximin Distance Designs, *Journal of Statistical Planning and Inference*, **26(2)**, 131-148.

Kalagnanam, J. R. and Diwekar, U. M., (1997), An Efficient Sampling Technique for Off-Line Quality Control, *Technometrics*, **39(3)**, 308-319.

Koch, P. N., Allen, J. K., Mistree, F. and Mavris, D., (1997), The Problem of Size in Robust Design, *Advances in Design Automation*, Sacramento, CA, ASME, Paper No. DETC97/DAC-3983.

Koch, P. N., Mavris, D. and Mistree, F., (1998), Multi-Level, Partitioned Response Surfaces for Modeling Complex Systems, *7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis & Optimization*, St. Louis, MO, AIAA, 3, 1954-1968.

Koehler, J. R. and Owen, A. B., (1996), Computer Experiments, *Handbook of Statistics* (Ghosh, S. and Rao, C. R., eds.), Elsevier Science, New York, 261-308.

Mallet, C. G., (1998), *A Wavelet Tour of Signal Processing*, Academic Press, Boston, MA.

McKay, M. D., Beckman, R. J. and Conover, W. J., (1979), A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code, *Technometrics*, 21(2), 239-245.

Meckesheimer, M., Barton, R. R., Simpson, T. W., Limayem, F. and Yannou, B., (2001), Metamodeling of Combined Discrete/Continuous Responses, *AIAA Journal*, to appear.

Mitchell, T. J. and Morris, M. D., (1992), Bayesian Design and Analysis of Computer Experiments: Two Examples, *Statistica Sinica*, 2, 359-379.

Montgomery, D. C., (1997), *Design and Analysis of Experiments*, Fourth Edition, John Wiley & Sons, New York.

Morris, M. D. and Mitchell, T. J., (1992), Exploratory Designs for Computer Experiments, *ORNL/TM-12045*, Oak Ridge National Laboratory, Oak Ridge, TN.

Myers, R. H. and Montgomery, D. C., (1995), *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, John Wiley & Sons, New York.

Owen, A. B., (1992), Orthogonal Arrays for Computer Experiments, Integration and Visualization, *Statistica Sinica*, 2, 439-452.

Park, J.-S., (1994), Optimal Latin-Hypercube Designs for Computer Experiments, *Journal of Statistical Planning and Inference*, 39(1), 95-111.

Powell, M. J. D., (1987), Radial Basis Functions for Multivariable Interpolation: A Review, *Algorithms for Approximation* (Mason, J. C. and Cox, M. G., eds.), Oxford University Press, London.

Rasmussen, J., (1990), Accumulated Approximation-A New Method for Structural Optimization by Iterative Improvement, *3rd Air Force/NASA Symposium on Recent*

Advances in Multidisciplinary Analysis and Optimization, San Francisco, CA, 253-258.

Sacks, J., Welch, W. J., Mitchell, T. J. and Wynn, H. P., (1989), Design and Analysis of Computer Experiments, *Statistical Science*, **4(4)**, 409-435.

Sayers, M. W. and Riley, S. M., (1996), Modeling Assumptions for Realistic Multibody Simulations of the Yaw and Roll Behavior of Heavy Trucks, *SAE Paper No. 960173*, Society of Automotive Engineers, Warrendale, PA.

Shewry, M. C. and Wynn, H. P., (1987), Maximum Entropy Sampling, *Journal of Applied Statistics*, **14(2)**, 165-170.

Shewry, M. C. and Wynn, H. P., (1988), Maximum Entropy Sampling with Application to Simulation Codes, *Proceedings of the 12th World Congress on Scientific Computation*, IMAC88, **2**, 517-519.

Simpson, T. W., Allen, J. K. and Mistree, F., (1998), Spatial Correlation Metamodels for Global Approximation in Structural Design Optimization, *Advances in Design Automation*, Atlanta, GA, ASME, Paper No. DETC98/DAC-5613.

Simpson, T. W., Mauery, T. M., Korte, J. J. and Mistree, F., (2001a), Kriging Metamodels for Global Approximation in Simulation-Based Multidisciplinary Design Optimization, *AIAA Journal*, to appear: **40(1)**.

Simpson, T. W., Peplinski, J., Koch, P. N. and Allen, J. K., (2001b), Metamodels for Computer-Based Engineering Design: Survey and Recommendations, *Engineering with Computers*, **17(2)**, 129-150.

Smith, M., (1993), *Neural Networks for Statistical Modeling*, von Nostrand Reinhold, New York.

Sobieszczanski-Sobieski, J. and Haftka, R. T., (1997), Multidisciplinary Aerospace Design Optimization: Survey of Recent Developments, *Structural Optimization*, **14**, 1-23.

Tang, B., (1993), Orthogonal Array-Based Latin Hypercubes, *Journal of the American Statistical Association*, **88(424)**, 1392-1397.

Tang, B., (1994), A Theorem for Selecting OA-Based Latin Hypercubes Using a Distance Criterion, Communications in Statistics, Theory and Methods, **23(7)**, 2047-2058.

Varadarajan, S., Chen, W. and Pelka, C., (2000), The Robust Concept Exploration Method with Enhanced Model Approximation Capabilities, *Engineering Optimization*, **32(3)**, 309-334.

Wang, L., Grandhi, R. V. and Canfield, R. A., (1996), Multivariate Hermite Approximation for Design Optimization, *International Journal for Numerical Methods in Engineering*, **39(5)**, 787-803.

Welch, W. J., Buck, R. J., Sacks, J., Wynn, H. P., Mitchell, T. J. and Morris, M. D., (1992), Screening, Predicting, and Computer Experiments, *Technometrics*, **34(1)**, 15-25.

Wilson, B., Cappelleri, D. J., Frecker, M. I. and Simpson, T. W., (2001), Efficient Pareto Frontier Exploration Using Surrogate Approximations, *Optimization and Engineering*, **2(1)**, 31-50.

Yang, R. J., Gu, L., Liaw, L., Gearhart, C., Tho, C. H., Liu, X. and Wang, B. P., (2000), Approximations for Safety Optimization of Large Systems, *ASME 2000 Design Engineering Technical Conferences - Design Automation Conference (Renaud, J. E., ed.)*, Baltimore, MD, ASME, Paper No. DETC-2000/DAC-14245.