AIAA-2000-4801

COMPARATIVE STUDIES OF METAMODELING TECHNIQUES UNDER MULTIPLE MODELING CRITERIA

Ruichen Jin* and Wei Chen†
Department of Mechanical Engineering
University of Illinois at Chicago
Chicago, Illinois 60607-7022

Timothy W. Simpson‡
Department of Mechanical & Nuclear Engineering
The Pennsylvania State University
University Park, PA 16802

Abstract

Despite the advances in computer capacity, the enormous computational cost of complex engineering simulations makes it impractical to rely exclusively on simulation for the purpose of design optimization. To cut down the cost, surrogate models, also known as metamodels, are constructed from and then used in lieu of the actual simulation models. In the paper, we systematically compare four popular metamodeling techniques—Polynomial Regression, Multivariate Adaptive Regression Splines, Radial Basis Functions, and Kriging—based on multiple performance criteria using fourteen test problems representing different classes of problems. Our objective in this study is to investigate the advantages and disadvantages these four metamodeling techniques using multiple modeling criteria and multiple test problems rather than a single measure of merit and a single test problem.

1 Introduction

Simulation-based analysis tools are finding increased use during preliminary design to explore design alternatives at the system level. In spite of advances in computer capacity and speed, the enormous computational cost of complex, high fidelity scientific and engineering simulations makes it impractical to rely exclusively on simulation codes for the purpose of design optimization. A preferable strategy is to utilize approximation models which are often referred to as metamodels as they provide a "model of the model" (Kleijnen, 1987), replacing the expensive simulation model during the design and optimization process. Metamodeling techniques have been widely used for design evaluation and optimization in many engineering applications; a comprehensive review of metamodeling applications in mechanical and aerospace systems can be found in (Simpson, et al., 1997) and will therefore not be repeated here. For the interested reader, a review of metamodeling applications in structural optimization can be found in (Barthelemy and Haftka, 1993); for metamodeling applications in multidisciplinary design optimization, see (Sobieszczanski-Sobieski and Haftka, 1997).

A variety of metamodeling techniques exist; the Response Surface Methodology (Box, et al.; 1978; Myers and Montgomery, 1995) and Artificial Neural Network (ANN) methods (Smith, 1993; Cheng and Titterington, 1994) are two well known approaches for constructing simple and fast approximations of complex computer codes. An interpolation method known as Kriging is becoming widely used for the design and analysis of computer experiments (Sacks, et al., 1989; Booker, et al., 1999). Finally, other statistical techniques that hold a lot of promise, such as Multivariate Adaptive Regression Splines (Friedman, 1991) and radial basis function approximations (Hardy, 1971; Dyn, et al., 1986) are beginning to draw the attention of many researchers.

An immediate question that a designer may have is on what basis the various techniques should be used? Moreover, is one technique superior to the others? Numerous examples that demonstrate the application of one metamodeling technique or the other, typically for a specific application exist; however, our survey reveals a lack of comprehensive comparative studies of the various techniques, let alone standard procedures for testing the relative merits of different methods. In Simpson, et al. (1998), kriging methods are compared against polynomial regression models for the multidisciplinary design optimization of an aerospike nozzle. Giunta, et al. (1998) also compare kriging models and polynomial regression models for two 5 and 10 variable test problems. In Varadarajan, et al. (2000), Artificial Neural Network (ANN) methods are compared with polynomial regression models for the engine design problem in modeling the nonlinear thermodynamic behavior. In Yang, et al., (2000), four approximation methods—enhanced Multivariate Adaptive Regression Splines (MARS), Stepwise

* Graduate research assistant.
† Assistant Professor, Member AIAA, corresponding author, weichen1@uic.edu, 312-996-6072.
‡ Assistant Professor, Member AIAA.

Regression, ANN, and the Moving Least Square—are compared for the construction of safety related functions in automotive crash analysis, for a relative small sampling size. Simpson (1999) presents the results of ongoing work investigating different metamodeling techniques—response surfaces, kriging models, radial basis functions, and MARS models—on a variety of engineering test problems. Although the existing studies provide useful insights into the various approaches considered, a common limitation is that the tests are restricted to a very small group of methods and test problems, and in many cases only one problem due to the expenses associated with testing. Moreover, when using multiple test problems, it is often difficult to make comparisons between the test problems when they belong to different classes of problems (e.g., linear, quadratic, nonlinear, etc.).

It is our belief that various factors contribute to the success of a given metamodeling technique, ranging from the nonlinearity of the model behavior, to the dimensionality and data sampling technique, to the internal parameter settings of the various techniques. *We contend that instead of using accuracy as the only measure, multiple metrics for comparison should be considered based on multiple modeling criteria*, such as accuracy, efficiency, robustness, model transparency, and simplicity. Overall, the knowledge of the performance of different metamodeling techniques with respect to different modeling criteria is of utmost importance to designers when trying to choose an appropriate technique for a particular application.

In this work, we present preliminary results from a systematic comparative study which provides insightful observations into the performance of various metamodeling techniques under different modeling criteria, and the impact of the contributing factors to their success. A set of mathematical and engineering problems has been selected to represent different classes of problems with different degrees of nonlinearity, different dimensions, and noisy versus smooth behaviors. Relative large, small, and scarce sample sets are also used for each test problem. Four promising metamodeling techniques, namely, Polynomial Regression (PR), Kriging (KG), Multivariate Adaptive Regression Splines (MARS), and Radial Basis Functions (RBF), are compared in this study. Although ANN is a well-known technique, it is not included in our study due to the large amount of trail-and-error associated with the use of this technique.

2 Metamodeling Techniques

The principle features of the four metamodeling techniques compared in our study are described in the following sections.

2.1 Polynomial Regression (PR)

PR models have been applied by a number of researchers (Engelund, et al, 1993; Unal, et al., 1996; Vitali, et al., 1997; Venkataraman, et al., 1997; Venter, et al., 1997; Chen, et al., 1996; Simpson, et al., 1997) in designing complex engineering systems. A second-order polynomial model can be expressed as:

$$\hat{y} = b_o + \sum_{i=1}^{k} b_i x_i + \sum_{i=1}^{k} b_{ii} x_i^{\,2} + \sum_i \sum_j b_{ij} x_i x_j \qquad (1)$$

When creating PR models, it is possible to identify the significance of different design factors directly from the coefficients in the normalized regression model. For problems with a large dimension, it is important to use linear or second-order polynomial models to narrow the design variables to the most critical ones. In optimization, the smoothing capability of polynomial regression allows quick convergence of noisy functions (see, e.g., Guinta, et al., 1994). In spite of the advantages, there is always a drawback when applying PR to model highly nonlinear behaviors. Higher-order polynomials can be used; however, instabilities may arise (Barton, 1992), or it may be too difficult to take sufficient sample data to estimate all of the coefficients in the polynomial equation, particularly in large dimensions. In this work, linear and second-order PR models are considered.

2.2 Kriging Method (KG)

A kriging model postulates a combination of a polynomial model and departures of the form:

$$\hat{y} = \sum_{j=1}^{k} b_j f_j(x) + Z(x), \qquad (2)$$

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

$$\mathrm{Cov}[Z(x_i), Z(x_j)] = \sigma^2 R(x_i, x_j), \qquad (3)$$

where σ^2 is the process variance and R is the correlation. A variety of correlation functions can be chosen (cf., Simpson, et al., 1998); however, the Gaussian correlation function proposed in (Sacks, et al., 1989) is the most frequently used. Furthermore, $f_j(x)$ in Eqn. 2 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 θ parameters, one θ for each of the k dimensions in the design space.

In addition to being extremely flexible due to the wide range of the correlation functions, the kriging method has advantages in that it provides a basis for a stepwise algorithm to determine the important factors, and the same data can be used for screening and building the predictor model (Welch, et al., 1992). The major disadvantage of the kriging process is that model construction can be very time-consuming. Determining the maximum likelihood estimates of the θ 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. Moreover, the correlation matrix can 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 full factorial designs and central composite designs when using kriging models (Meckesheimer, et al., 2000; Wilson, et al., 2000). Finally, the complexity of the method and the lack of commercial software may hinder this technique from being popular in the near term (Simpson, et al., 1997).

2.3 Multivariate Adaptive Regression Splines (MARS)

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

$$\hat{y} = \sum_{m=1}^{M} a_m B_m(\mathbf{x}), \tag{4}$$

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

$$B_m(\mathbf{x}) = \prod_{k=1}^{K_m} \left[s_{k,m} \left(x_{v(k,m)} - t_{k,m} \right) \right]_+^q \tag{5}$$

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:

$$\left[s_{k,m} \left(x_{v(k,m)} - t_{k,m} \right) \right]_+^q = \begin{cases} \left[s_{k,m} \left(x_{v(k,m)} - t_{k,m} \right) \right]^q & s_{k,m} \left(x_{v(k,m)} - t_{k,m} \right) > 0 \\ 0 & \textit{otherwise} \end{cases} \tag{6}$$

Compared to other techniques, the use of MARS for engineering design applications is relatively new. Buja, et al. (1990) use MARS for extensive analysis of data concerning memory usage in electronic switches. Wang, et al. (1999), compare MARS to linear, second-order, and higher-order regression models for a five variable automobile structural analysis. Friedman (1991) uses the MARS procedure to approximate behavior of performance variables in a simple alternating current series circuit. The major advantages of using the MARS procedure appears to be accuracy and major reduction in computational cost associated with constructing the metamodel.

2.4 Radial Basis Functions (RBF)

Radial basis functions (RBF) have been developed for scattered multivariate data interpolation (Hardy, 1971; Dyn, et al., 1986). The method uses linear combinations of a radially symmetric function based on Euclidean distance or other such metric to approximate response functions. A radial basis function model can be expressed as:

$$\hat{y} = \sum_i a_i \left\| \mathbf{x} - \mathbf{x_{0i}} \right\| \tag{7}$$

where a_i is the coefficient of the expression and $\mathbf{x_{0i}}$ is the observed input.

Radial basis function approximations have been shown to produce good fits to arbitrary contours of both deterministic and stochastic response functions (Powell, 1987). Tu and Barton (1997) found that RBF approximations provide effective metamodels for electronic circuit simulation models. Meckesheimer, et al. (2000) use the method for constructing metamodels for a desk lamp design example, which has both continuous and discrete response functions.

3 Test Problems and Test Scheme

3.1 Features of Test Problems

To test the effectiveness of various approaches to different classes of problems, 14 test problems are selected and classified based on the following representative features of engineering design problems.

- *Problem Scale*. Two relative scales are considered: *large* (the number of variables ≥ 10 and *small* (the number of variables = 2,3).
- *Nonlinearity of the performance behavior*. For convenience, we classify the problems into two categories: *low-order nonlinearity* (if the square regression ≥ 0.99 when using first or second-order polynomial model) and *high-order nonlinearity* (otherwise).
- *"Noisy" versus "smooth" behavior* . In some cases, numerical simulation error or other noisy causes cannot be eliminated. In our study, the noisy behavior is artificially created using local variations of a smooth function.

A summary of the features of the 14 test problems is given in Table 1; the test problems are described in more detail in the next section.

Table 1. Features of Test Problems

Type	Problem No.	Nonlinearity Order	Scale (# of Inputs)	Noisy Behavior
Mathematical	1	High	Large (n=10)	NO
	2	Low	Large (n=10)	NO
	3	High	Large (n=10)	NO
	4	Low	Large (n=10)	NO
	5	Low	Large (n=16)	NO
	6	High	Small (n=2)	NO
	7	High	Small (n=2)	NO
	8	Low	Small (n=2)	NO
	9	High	Small (n=3)	NO
	10	High	Small (n=3)	NO
	11	Low	Small (n=3)	NO
	12	Low	Small (n=2)	NO
	13	Low	Small (n=2)	YES
Vehicle Handling	14	High	Large (n=14)	NO

3.2 Description of Test Problems

Thirteen mathematical problems are utilized in our study (see Appendix). These test problems are chosen from (Hock and Schittkowski, 1981) which offers 180 problems for testing nonlinear optimization algorithms. Provided in the Appendix are the mathematical functions for each of these problems. While some of the functions exhibit smooth low-order nonlinear behavior, the others are highly nonlinear functions that pose challenges for many metamodeling techniques. Figure A.1 are the grid plots of highly nonlinear problems except problem 7, of which the plots from different approaches are compared in Figure 12.

Meanwhile, Problem 14 is a real engineering problem that calls for better vehicle design to improve a vehicle's handling characteristics, particularly the prevention of rollover. The problem statement is first given in (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. The use of ArcSim for the purpose of optimization demands heavy computational costs. In this study, 14 input variables are considered which include nine suspension and vehicle parameters as design variables and five uncontrollable factors for steering and braking. The response of interest is the vehicle handling performance, which is measured by the rollover metric. The previous studies (Chen, et al., 1999) indicate that the rollover metric has a highly nonlinear dependence on the control and noise variables, especially for brake and steering levels.

3.3 Data Sampling

We are interested in examining the performance of various metamodeling techniques when different sizes of data samples are used for model formation, as shown in Table 2. For different problem scales, different sets of sample data such as scarce, small, and large sets are used as "training" points for model formation. For large-scale problems, Latin Hypercubes (McKay, et al., 1979) are used to generate the "training" points in all cases because this method provides good uniformity and flexibility on the size of the sample. The second order polynomial models have k = (n+1)(n+2)/2 coefficients for n design variables. Giunta, et al. (1994) and Kaufman, et al. (1996) found that 1.5k sample points for 5-10 variable problems to 4.5k sample points for 20-30 variable problems are necessary to obtain reasonably accurate second-order polynomial models. Therefore, for large-scale problems, 3k sample points are selected and are referred to as a large sample set. For complex and time-consuming problems, it is preferable to use fewer samples. For this reason, scarce sample sets with 3n points are tested. In addition to large and scarce sample sets, small sample sets with 10n are also used. For small-scale problems, only small and large sample sets are considered. Also shown in Table 2 is the number of confirmation points used for checking the accuracy of each model. The Monte Carlo method is used to generate confirmation points.

Table 2. Experimental Designs for Test Problems

		Large Scale Problem	Small Scale Problem
Training Points	**Scarce Set**	Latin Hypercube (3n)	N/A
	Small Set	Latin Hypercube (10n)	Latin Hypercube (9 if n=2, 27 if n=3)
	Large Set	Latin Hypercube $(\frac{3(n+1)(n+2)}{2})$	Latin Hypercube (100 if n=2, 125 if n=3)
Confirmation Points		Monte-Carlo Method (500 for vehicle problem, 1000-1200 for others)	

3.4 Metrics for Performance Measures

In accordance with having multiple metamodeling criteria, the performance of each metamodeling technique is measured from the following aspects.

- *Accuracy* – the capability of predicting the system response over the design space of interest.
- *Robustness* – the capability of achieving good accuracy for different the different problem types and sample sizes.
- *Efficiency* – the computational effort required for constructing the metamodel and for predicting the response for a set of new points by metamodels.
- *Transparency* – the capability of illustrating explicit relationships between input variables and responses.
- *Conceptual Simplicity* – ease of implementation. Simple methods should require minimum user input and be easily adapted to each problem.

For accuracy, the goodness of fit obtained from "training" data is not sufficient to assess the accuracy of newly predicted points. For this reason, additional confirmation samples (see Table 2) are used to verify the accuracy of the metamodels. To provide a more complete picture of metamodel accuracy, three different metrics are used: R Square, Relative Average Absolute Error, and Relative Maximum Absolute Error. The equations for these three measures are given in Eqns. (8) to (10), respectively.

a) R Square

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} = 1 - \frac{MSE}{Variance} \quad (8)$$

where $\hat{y}_i$ is the corresponding predicted value for the observed value y_i; $\bar{y}$ is the mean of the observed values. While *MSE* (Mean Square Error) represents the departure of the metamodel from the real simulation model, the variance captures how irregular the problem is. *The larger the value of R Square, the more accurate the metamodel.*

b) Relative Average Absolute Error (RAAE)

$$RAAE = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{n * STD}, \quad (9)$$

where *STD* stands for standard deviation. *The smaller the value of RAAE, the more accurate the metamodel.*

c) Relative Maximum Absolute Error (RMAE)

$$RMAE = \frac{\max(|y_1 - \hat{y}_1|, |y_2 - \hat{y}_2|, ..., |y_n - \hat{y}_n|)}{STD} \quad (10)$$

While the *RAAE* is usually highly correlated with *MSE* and thus R Square, RMAE is not necessarily. Large RMAE indicates large error in one region of the design space even though the overall accuracy indicated by R Square and RAAE can be very good. Therefore, *a small RMAE is preferred*; however, since this metric cannot show the overall performance in the design space, it is not as important as R Square and RAAE.

4 Results and Comparison

Based on the proposed scheme for comparative study, 136 metamodels are created for the 14 test problems (see Table 1), using different sets of sample data (see Table 2), and based on four different metamodeling techniques (see Sections 2.1-2.4). Different techniques are compared based on the results from confirmation points.

4.1 Accuracy and Robustness

To illustrate the performance of the metamodeling techniques under different circumstances (e.g., nonlinearity, problem size, and sample size), multiple bar-charts are provided. While the mean indicates the average accuracy of a technique, the variance illustrates the robustness of the accuracy. Finally, while the height of a bar indicates the magnitude of accuracy, the differences between heights of multiple bars illustrate the impact of a particular contributing factor.

4.1.1 Overall Performance Illustrated in Figures 1 and 2 are the mean and variance of the three accuracy metrics for all metamodels which consider different orders of nonlinearity, different problem sizes, and different sample sizes. For the mean values, the larger the R Square, the better the accuracy is; however, for both RAAE and RMAE, the smaller value indicates better accuracy. For variance, a smaller value always indicates higher robustness.

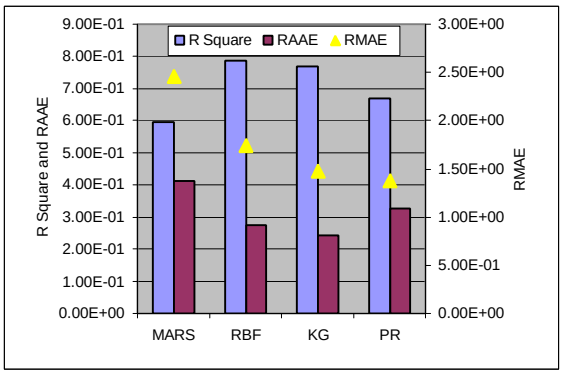


Figure 1. The Mean of Accuracy Metrics

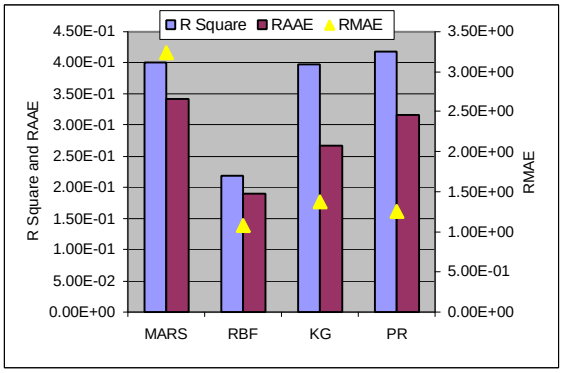


Figure 2. The Variance of Accuracy Metrics

Figure 1 shows that the average accuracies of RBF and KG for all test cases are among the best in the group; their values are very close to each other. RBF is slightly better than KG in R Square (all close to 0.8), but KG is better than RBF in both RAAE and RMAE. The average accuracy of PR is third in the group. As revealed in Section 4.1.3, the poor average performance of MARS is due to the deficiency of using MARS when scare set of samples is available. In terms of the robustness of the accuracy for all test cases, RBF is distinctly the best for all three accuracy measures. Overall, RBF is shown to be the best approach in terms of its average accuracy and robustness when handling all types of problems for any amount of samples.

4.1.2 Performance for Different Types of Problems Figures 3-6 show the mean and variance of R Square of the metamodels for different types of problems. In Figures 3 and 4, "High" and "Low" represent the nonlinearity of problems, while "Large"

and "Small" represent problem scale. So for example, "High Large" means a high-order nonlinear and large scale problem.

The values in Figures 3 and 4 are derived based on the data from all sample sizes (large, small, and scare). It is noted that for high-order nonlinear and large scale problems, RBF performs best in terms of both average accuracy and robustness. For low-order nonlinear and large scale problems, KG performs best in terms of both average accuracy and robustness. For high-order nonlinear and small scale problems, RBF performs best in terms of both average accuracy and robustness. For low-order nonlinear and small scale problems, PR performs best in terms of both average accuracy and robustness.

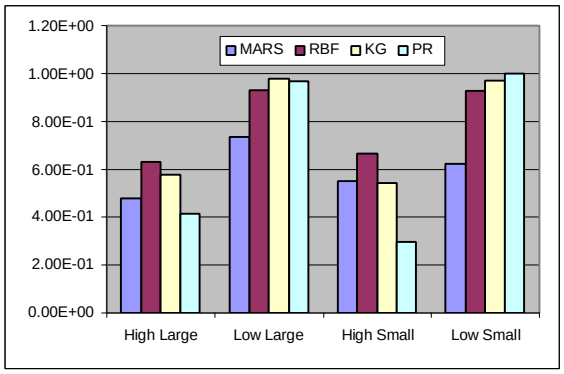


Figure 3. The Mean of R Square for Different Types of Problems (a)

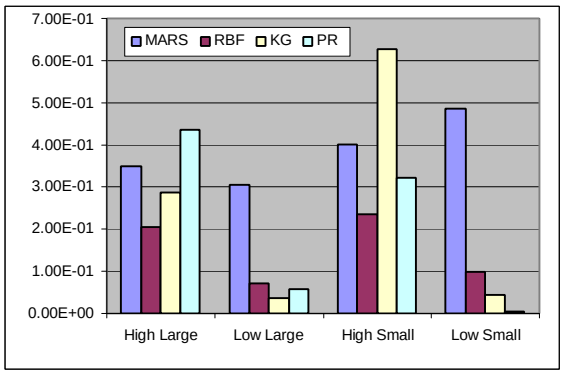


Figure 4. The Variance of R Square for Different Types of Problems (a)

In Figures 5 and 6, the average accuracy and its robustness are derived for single contributing factors (e.g., higher-order nonlinear) based on all the test data belong that that category. It indicates that for high-order nonlinear problems, RBF performs best in terms of both average accuracy and robustness. For low-order nonlinear problems, the average accuracy of KG and PR is very close, while the robustness of KG is slightly better than PR. So, overall, KG is slightly better than PR. We also observe that each method has distinctively better accuracy for low-order nonlinear problems than for high-order nonlinear problems, which matches well with our intuition. The difference is the most significant for PR: while the mean of R Square is close to 1 for low-order nonlinear problems, it is less than 0.35 for high-order nonlinear problems. Except for MARS (due to the deficiency for scare set of samples), the accuracy of the other three methods is acceptable for low-order nonlinear problems with any size sample set. However, the robustness, although small for RBF, KG, and PR when the model is low-order nonlinear, becomes larger when the problems are high-order nonlinear. The impact is the most significant for KG and PR.

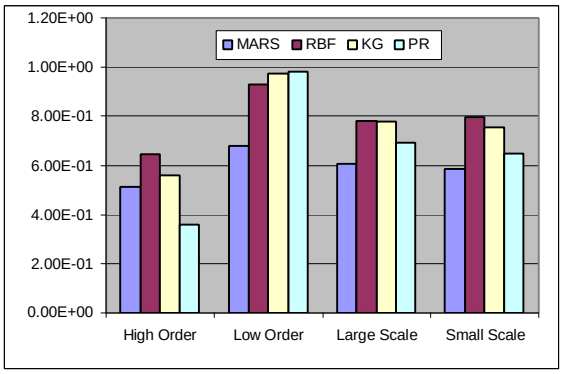


Figure 5. The Mean of R Square for Different Types of Problems (b)

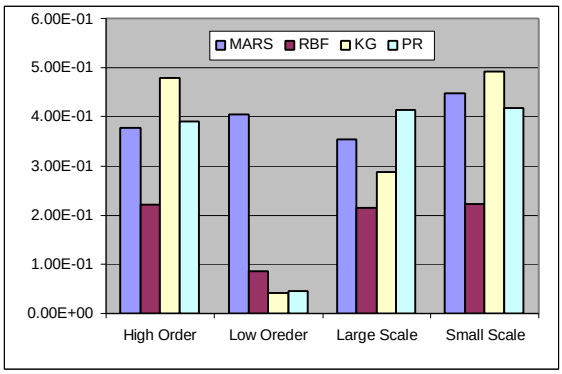


Figure 6. The Variance of R Square for Different Types of Problems (b)

For large scale problems, the average accuracy of RBF and KG is very close, while the robustness of RBF is better than KG. So, overall, RBF is the best. For small scale problems, RBF is the best again in terms of both average accuracy and robustness. It is also found that problem scale has little impact on the performance of RBF. Although the impact of problem scale on the average accuracy of KG is also small, the impact on robustness for KG is rather large.

4.1.3 Performance under Different Sample Size

Figures 7 and 8 show the performance of metamodeling techniques for types of problems under different sample sizes (large, small and scarce). For large set samples, the performances of MARS, RBF and KG are very close not only in average accuracy but also in robustness, while the performance of PR is worse than the others. We cannot tell overall which is the best because KG performs slightly better than RBF and MARS in average accuracy while MARS is more robust than KG and RBF. RBF performs best for small set samples both in average accuracy and robustness. Although KG also performs well in average accuracy, it is not as robust. For scarce sample sets, the average accuracy of RBF, KG and PR are close but the robustness of RBF is the best. Therefore, for scarce set samples, RBF performs best overall.

It is also noted that the sample size has the largest impact on MARS, for both the mean and variance of accuracy. When small or scarce sample sets are used, the accuracy of MARS is low ($R < 0.45$). This is because MARS fails to work when the size of samples is too small. The variances of the accuracy (robustness) are shown to be very small for MARS, RBF, and KG when large sample sets are used.

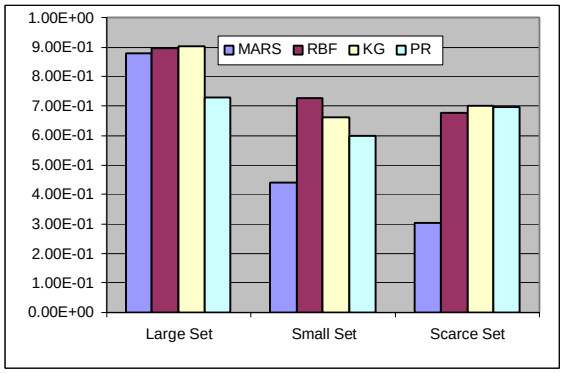


Figure 7. The Mean of R Square under Different Sample Scales

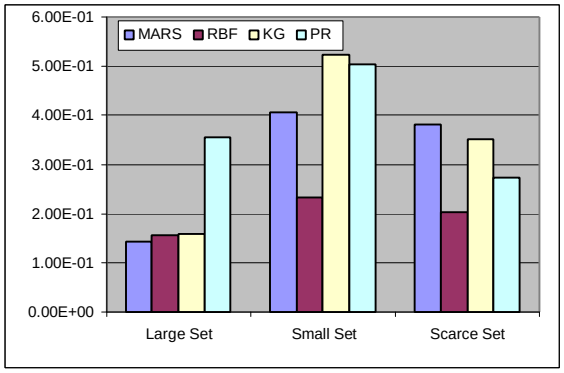


Figure 8. The Variance of R Square under Different Sample Scales

Figures 9 and 10 further illustrate the performance of metamodeling techniques for different sample sizes when handling the most difficult situation, i.e., large scale and high-order nonlinear problems. It shows the average accuracy of MARS is the best when large sample sets are used for this type of problem. For small sample sets, MARS also performs best if average accuracy and robustness are both considered (although RBF performs best in average accuracy). However, its performance deteriorates significantly when the sample size becomes scarce, under which RBF performs best. The impact of sample size on average accuracy and robustness is the smallest for RBF. The accuracy and robustness of PR is not stable, as for high nonlinear problem, its performance is very problem-dependent and sample-dependent (not only the sample size).

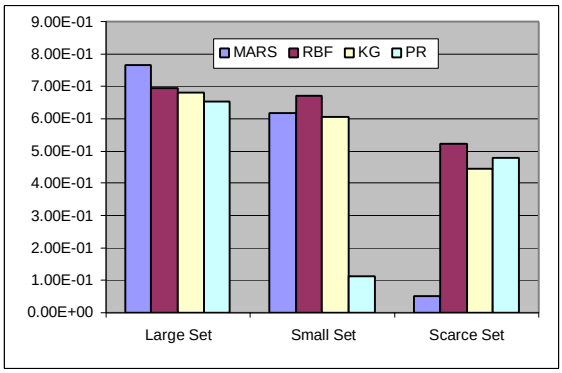


Figure 9. The Mean of R Square under Different Sample Scales for Large Scale and High-Order Nonlinear Problems

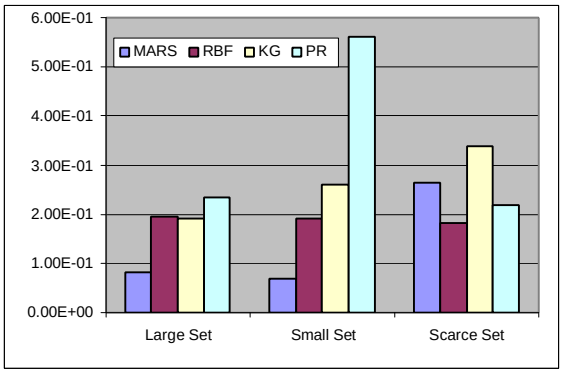


Figure 10. The Variance of R Square under Different Sample Scales for Large Scale an d High-Order Nonlinear Problems

4.1.4 Impact of Noisy Behavior Figure 11 shows the influence of noise on the performance of different metamodeling techniques. Only problems 12 and 13 are compared here since the function in Problem 13 is the result of local variations of the function in problem

12, a low-order nonlinear problem. From Figure 11, it is found that Kriging is very sensitive to the noise since it interpolates the data. Consequently, when estimating the accuracy of the kriging metamodel for Problem 13 using the non-noisy data from Problem 12, the kriging model does not yield good predictions. PR performs the best because its tendency to give a smooth metamodel. MARS and RBF also perform well in this test problem.

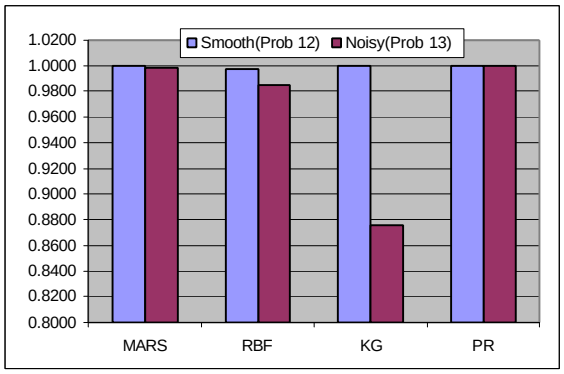
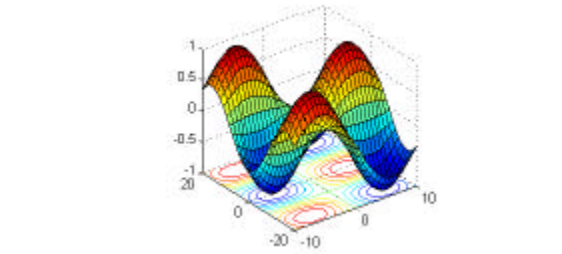
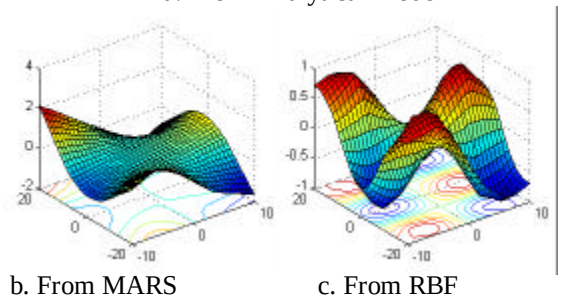


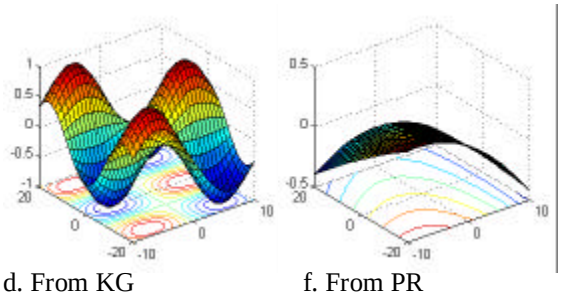
Figure 11. R Square—Smooth Vs Noisy Problems



a. From Analytical Model



b. From MARS c. From RBF



d. From KG f. From PR

Figure 12 Grid Plots for Problem 7

Due to the space limitation, we only provide here the sample grid plots of problem 7, which has a low-order highly nonlinear (waving) behavior, for comparing the accuracy of different approaches. It is noted that KG is extremely accurate for modeling the waving behavior for this particular case, while the RBF is the second best. We also found that PR is not suitable at all for this type of behavior, while MARS captures the general trend but falls short in its accuracy in local regions.

4.2 Efficiency

The efficiency of each metamodeling technique is measured by the time used for model construction and new predictions. The time depends on the problem scale and the sample size, which also depends highly on the computer platform (MARS, PR, RBF are tested on a PC-pentium III 500 MHz machine while KG is run on a Sun Ultra60 workstation). Rough time statistics needed for model construction and new prediction are provided in Tables 3 and 4, respectively.

Table 3. Time Needed to Construct Metamodel

Problem Scale/ Sample size	Large/ Large	Large/ Small	Small/ Large	Small/ Small
MARS	5-10s	2-5s	<1s	<<1s
RBF	5-10s	1-2s	<1s	<<1s
KG	2-3h	10-15m	2-5m	10-30s
PR	1-2s	<1s	<<1s	<<1s

Table 4. Time Needed for 1000 New Predictions

Problem Scale/ Sample size	Large/ Large	Large/ Small	Small/ Large	Small/ Small
MARS	<<1s	<<1s	<<1s	<<1s
RBF	10-20s	2-5s	2-5s	<1s
KG	5-10s	1-3s	1-3s	<1s
PR	<<1s	<<1s	<<1s	<<1s

It is obvious that PR is the most efficient method both in model construction and response prediction. It is found that, relatively speaking, model construction is very time-consuming for Kriging. Kriging requires a k-dimensional optimization to find the maximum likelihood estimates of the parameters used to fit the model, which can become computationally expensive when the problem scale and the sample size are large. For predictions, RBF and kriging are relatively slow because spatial distances are needed for prediction. However, all the reported time scales are considered to be small compared to the time needed for simulations to obtain the "training data" in complex applications.

4.3 Transparency

PR provides the best transparency in terms of the function relationship and the factor contributions (see

Eqn. 1). When using MARS, models can be recast into the form (Friedman, 1991):

$$\hat{y} = a_0 + \sum_{K_m=1} f_i(x_i) + \sum_{K_m=2} f_{ij}(x_i, x_j) + \sum_{K_m=3} f_{ijk}(x_i, x_j, x_k) + ... \quad (11)$$

The first sum is over all basis functions that involve only a single variable. The second sum is over all basis functions that involve exactly two variables, representing (if present) two-variable interactions. Similarly, the third sum represents (if present) the contributions from three-variable interactions and so on. Therefore, MARS also provides some model transparency. For RBF, an explicit function exists (Eqn. 9); however, the factor contributions are not clear. The same holds true for Kriging; however, the theta parameters can be interpreted with some practice—large theta values indicate a highly non-linear function while small thetas indicate a smooth function with little variation. The influence of each input variable on the final output response cannot be readily ascertained from a kriging model, however.

4.4 Simplicity

Applying PR and RBF is relatively straightforward, and no parameters need to be specified by a user. They are both easy to implement. MARS and Kriging are more sophisticated in theory. For MARS, internal parameters can be specified which may improve or deteriorate its performance depending on the problem. For kriging models, the user has the capability to manipulate the optimization parameters used when constructing the model, and there are a variety of choices for the correlation function used in the model and for the underlying global portion of the model. While the Gaussian correlation function and a constant underlying global model are the most frequently used, it is currently unclear the extend to which more complex kriging models improve the accuracy of the model.

5 Summary

The systematic comparative study presented in this paper has provided insightful observations into performance of various metamodeling techniques under different modeling criteria. Based on the discussions in Section 4, we provide here a summary of our observations and conclusions.

In terms of the accuracy and robustness of the various techniques for different types of problems (i.e., different orders of nonlinearity and problem scales), we can summarize our results as shown in Table 5. As shown, RBF excels in most of the categories. When large sample sizes are used (assuming designers have enough computational resource), MARS, RBF, and KG perform equally well in both average accuracy and robustness. For small and scarce sample sets, RBF performs the best when both average accuracy and robustness are considered. It is noted that the performance of MARS deteriorates when small or scarce sample sets are used. For the most difficult problems, i.e., large scale and high-order nonlinear problems, the average accuracy of MARS is the best when the large sample sets are used. MARS also performs best if average accuracy and robustness are both considered. However, its performance deteriorates significantly when the sample size reduces to scare, under which RBF performs best. For RBF, the impact of sample size on average accuracy and robustness is the least. Finally, in our test problem with noise, PR performs the best and MARS also works well; however, KG is very sensitive to the noise because it interpolates the sample data. From the above observations, we can conclude that RBF is the most dependable method in most situations in terms of accuracy and robustness.

Table 5. Summary of Best Methods

	High-Order Nonlinear	**Low-Order Nonlinear**	**Overall**
Large Scale	RBF	Kriging	RBF
Small Scale	RBF	PR	RBF
Overall	RBF	PR	RBF

In terms of efficiency for metamodel construction, KG can be very time-consuming, especially for large scale problems with large sample sizes. Meanwhile, PR takes the least amount of time for model building. For all techniques, the time needed for new predictions is considered to be trivial compared to the time used for simulation and model construction. PR and MARS have good transparency, which means we can obtain the contributions of each input factor and the interaction among them. Both RBF and KG are less transparent in this sense. In terms of simplicity, PR and RBF are the easiest to implement, while the users need to configure the parameters for MARS and KG to achieve better accuracy. Although PR is not accurate for highly nonlinear problems, it is easy to use and is very accurate for low order nonlinearity. *It is therefore proposed that when constructing m etamodels, PR should implemented first to see if a reasonable fit can be obtained.*

It is also observed that experimental design (data sampling) also plays an important role, especially when building KG models. If the samples are not properly selected, KG may not work well and sometimes even fail to obtain the metamodel. Data sampling also has some influence on RBF. It is found for small scale problems, when using full factorial design, the accuracy will improve. Although not tested, data sampling is

expected to impact the performance of other techniques. It is desired to obtain uniform samples not only in one-dimensional projections but also higher dimensional projections, to improve metamodeling performance. Adaptive sampling methods, which sample each variable according its contributions to the response and other variable interactions, should also be investigated.

Finally, to improve our study, more test problems with large dimensions and some medium dimensions (between 3 to 10 variables) need to be considered. For the same size of data sample, the effectiveness of different sample techniques is also of interest.

Acknowledgements

The support from the NSF/DMII 9896300 is gratefully acknowledged. Support from Dr. Kam Ng, ONR 333, through the Naval Sea Systems Command under Contract No. N00014-00-G-0058 is also acknowledged.

References

ArcSim User's Guide, Automotive Research Center, The University of Michigan, Ann Arbor, July 1997.

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

Barton, R. R., 1992, December 13-16, "Metamodels for Simulation Input-Output Relations," *Proceedings of the 1992 Winter Simulation Conference (Swain, J. J., et al., eds.),* Arlington, VA, IEEE, pp. 289-299.

Booker, A. J., Dennis, J. E., Jr., Frank, P. D., Serafini, D. B., Torczon, V. and Trosset, M. W., 1999, "A Rigorous Framework for Optimization of Expensive Functions by Surrogates," *Structural Optimization,* 17(1), pp. 1-13.

Box, G. E. P., Hunter, W. G. and Hunter, J. S., 1978, *Statistics for Experimenter* s, John Wiley & Sons, NY.

Chen, W., Allen, J. K., Mavris, D. and Mistree, F., 1996, "A Concept Exploration Method for Determining Robust Top-Level Specifications," *Engineering Optimization,* Vol. 26, pp. 137-158.

Chen, W., Garimella, R., and Michelena, N., 1999, "Robust Design for Improved Vehicle Handling Under a Range of Maneuver Conditions," 1999 ASME Design Technical Conference, Las Vegas, NV, Paper No. DAC-8580.

Cheng, B. and Titterington, D. M., 1994, "Neural Networks: A Review from a Statistical Perspective," *Statistical Science,* Vol. 9, No. 1, pp. 2-54.

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

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,* Vol. 7, No. 2, pp. 639-659.

Engelund, W. C., Douglas, O. S., Lepsch, R. A., McMillian, M. M. and Unal, R., 1993, "Aerodynamic Configuration Design Using Response Surface Methodology Analysis," *AIAA Aircraft Design, Sys. & Oper. Mngmt.,* Monterey, CA, Paper 93-3967.

Friedman, J. H., 1991, "Multivariate Adaptive Regression Splines," *The Annals of Statistics* , 19(1), pp. 1-141.

Giunta, A.A., Dudley, J.M., Narducci, R., Grossman, B., Haftka, R.T., Mason, W.H, and Watson, L. T., 1994, "Noisy Aerodynamic Response and Smooth Approximation in HSCT Design," *Proceedings Analysis and Optimization* (Panama City, FL), Vol. 2, AIAA, Washington, DC, pp. 1117-1128.

Giunta, A., Watson, L. T. and Koehler, J., 1998, September 2-4, "A Comparison of Approximation Modeling Techniques: Polynomial Versus Interpolating Models," *7th AIAA/USAF/NASA/ISSMO Symposium on Multid isciplinary Analysis & Optimization,* St. Louis, MO, AIAA, Vol. 1, pp. 392-404. AIAA-98-4758.

Hardy, R.L., 1971, "Multiquadratic Equations of Topography and Other Irregular Surfaces," *J. Geophys. Res.*, Vol. 76, pp. 1905-1915.

Hock and Schittkowski, 1981, *Test Examples for Nonlinear Programming Codes* , New York: Springer-Verlag.

Kaufman, M., Balabanov, V., Burgee, S.L., Giunta, A.A., Grossman, B., Mason, W. H., and Watson, L. T., 1996, "Variable-Complexity Response Surface Approximations for Wing Structural Weight in HSCT Design," *34th Aerospace Sciences Meeting and Exhibit,* Reno, NV, AIAA Paper 96-0089.

Kleijnen, J.P.C., 1987, *Statisticall Tools for Simulation Practitioners* , Marcel Dekker, NY.

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), pp. 239-45.

Meckesheimer, M., Barton, R. R., Limayem, F. and Yannou, B., 2000, "Metamodeling of Combined Discrete/Continuous Responses," *Design Theory and Methodology – DTM'00 (Allen, J.K., Ed.),*ASME, Baltimore, MD, Paper No. DETC2000/DTM-14573.

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

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

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

Simpson, T. W., Mauery, T. M., Korte, J. J. and Mistree, F., 1998, September 2-4, "Comparison of Response Surface and Kriging Models for Multidisciplinary Design Optimization," *7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis & Optimization,* St. Louis, MO, AIAA, Vol. 1, pp. 381-391. AIAA-98-4755.

Simpson, T. W., 1999, November 7-11, "A Comparison of Metamodeling Strategies for Computer-Based Engineering Design," *INFORMS Philadelphia Fall 1999 Meeting,* Philadelphia, PA, INFORMS.

Simpson, T. W., Peplinski, J., Koch, P. N. and Allen, J. K., 1997, "On the Use of Statistics in Design and the Implications for Deterministic Computer Experiments," *Design Theory and Methodology - DTM'97,* Sacramento, CA, ASME, Paper No. DETC97/DTM-3881.

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

Smith, M., 1993, "Neural Networks for Statistical Modeling", Von Nostrand Reinhold, New York.

Tu, C. H. and Barton, R.R. (1997), "Production Yield Estimation by the Metamodel Method with a Boundary – focused Experiment Design," *Design Theory and Methodology Conference – DTM'97*, DETC97/DTM3870.

Unal, R., Lepsch, R. A., Engelund, W. and Stanley, D. O., 1996, "Approximation Model Building and Multidisciplinary Design Optimization Using Response Surface Methods," *6th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, Bellevue, WA, AIAA, Vol. 1, pp. 592-597.

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

Venter, G., Haftka, R. T. and Starnes, J. H., Jr., 1996, "Construction of Response Surfaces for Design Optimization Applications," *6th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization* Bellevue, WA, AIAA, Vol. 1, pp. 548-564.

Wang, X., Liu, Y. and Antonsson, E. K., 1999, September 12-15, "Fitting Functions to Data in High Dimensional Design Spaces," *Advances in Design Automation*, Las Vegas, NV, ASME, Paper No. DETC99/DAC-8622.

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), pp. 15-25.

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 Design Automation Conference, September 10-13, Baltimore, MD, Paper No. DETC-00/DAC-14245.

Appendix A – Test Problems

1. $f(x) = \sum_{i=1}^{10} [(\ln(x_i - 2))^2 + (\ln(10 - x_i))^2] - (\prod_{i=1}^{10} x_i)^2$

   $2.1 \le x_i \le 9.9$

2. $f(x) = \sum_{j=1}^{10} x_j (c_j + \ln \frac{x_j}{x_1 + \ldots + x_{10}})$, c_j = -6.089,-17.164,-34.054,-5.914,-24.721,-14.986,-24.100,-10.708,-26.662,-22.179; j = 1,2,…,10.

3. $f(x) = \sum_{j=1}^{10} \exp(x_j)(c_j + x_j - \ln(\sum_{k=1}^{10} \exp(x_k)))$,

   c_j = -6.089,-17.164,-34.054,-5.914,-24.721,-14.986,-24.100,-10.708,-26.662,-22.179; j = 1,2,…,10.

4. $f(x) = x_1^2 + x_2^2 + x_1 x_2 - 14x_1 - 16x_2 + (x_3 - 10)^2 + 4(x_4 - 5)^2 + (x_5 - 3)^2 + 2(x_6 - 1)^2 + 5x_7^2 + 7(x_8 - 11)^2 + 2(x_9 - 10)^2 + (x_{10} - 7)^2 + 45$

5. $f(x) = \sum_{i=1}^{16} \sum_{j=1}^{16} a_{ij} (x_i^2 + x_i + 1)(x_j^2 + x_j + 1)$, i, j = 1,2,…, 16.

   $[a_{ij}]$=

```
[1 0 0 1 0 0 1 1 0 0 0 0 0 0 0 1
 0 1 1 0 0 0 1 0 0 1 0 0 0 0 0 0
 0 0 1 0 0 0 1 0 1 1 0 0 0 1 0 0
 0 0 0 1 0 0 1 0 0 0 1 0 0 0 1 0
 0 0 0 0 1 1 0 0 0 0 1 0 1 0 0 0 1
 0 0 0 0 0 1 0 1 0 0 0 0 0 0 1 0
 0 0 0 0 0 0 1 0 0 0 1 0 1 0 0 0
 0 0 0 0 0 0 0 1 0 1 0 0 0 0 1 0
 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 1
 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0
 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0
 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0
 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1];
```

6. $f(x) = (30 + x_1 * \sin(x_1)) * (4 + \exp(-x_2^2))$

7. $f(x) = \sin(\pi x_1 / 12) \cos(\pi x_2 / 16)$

8. $f(x) = \sin(x_1 + x_2) + (x_1 - x_2)^2 - 1.5x_1 + 2.5x_2 + 1$

9. $f(x) = (x_1 - 1)^2 + (x_1 - x_2)^2 + (x_2 - x_3)^4$

   $f(x) = \sum_{i=1}^{99} f_i(x)^2$

10. $f_{i(}x) = -0.01i + \exp(\frac{-1}{x_1}(u_i - x_2)^{x_3})$

    $u_i = 25 + (-50\ln(.01i))^{2/3}$

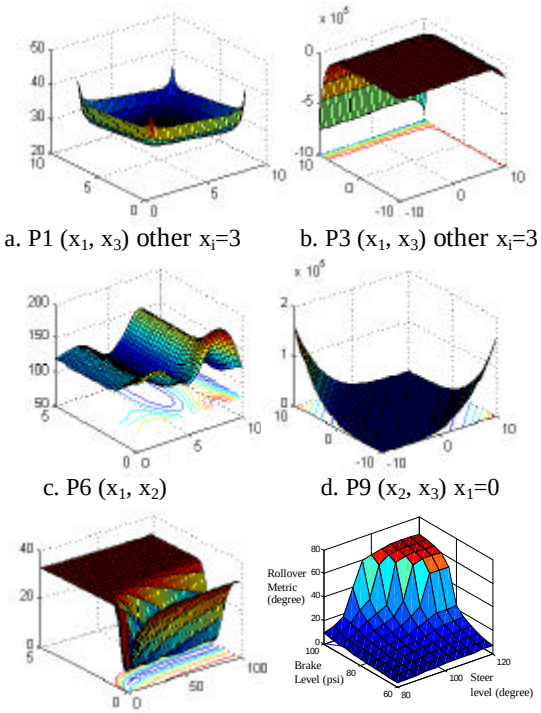
    $i = 1,\ldots,99$

11. $f(x) = 5.3578547x_2^2 + 0.8356891x_1x_3 + 37.293239x_1 - 40792.141$

12. $f(x) = 0.5x_1^2 + x_2^2 - x_1x_2 - 7x_1 - 7x_2$

13. $f(x) = 0.5x_1^2 + x_2^2 - x_1x_2 - 7x_1 - 7x_2 + e(x_1, x_2)$

where $e(x_1,x_2)$ is a normal noise. The mean is equal to 0 and the variance is equal to 1/100 of that of the smooth part of f(x). When -5≤x_1,x_2≤5, the variance is of the noise: 1/100*1254.9=12.55.



a. P1 (x_1, x_3) other x_i=3 b. P3 (x_1, x_3) other x_i=3

c. P6 (x_1, x_2) d. P9 (x_2, x_3) x_1=0

e. P10 (x_1, x_3) x_2=12.8 f. P14 –Vehicle handling

Figure A.1 Gird Plots of Highly Nonlinear Models (P stands for Problem)