Proceedings of the 2000 Winter Simulation Conference
J. A. Joines, R. R. Barton, K. Kang, and P. A. Fishwick, eds.

LOW COST RESPONSE SURFACE METHODS FOR AND FROM SIMULATION OPTIMIZATION

Theodore Allen

The Ohio State University
210 Baker Systems
1971 Neil Avenue
Columbus, OH 43210, U.S. A.

Liyang Yu

i2 Technologies, Inc.
909 East Las Colinas Boulevard
Sixteenth Floor
Irving, TX 75039, U.S. A.

ABSTRACT

We propose "low cost response surface methods" (LCRSM) that typically require half the experimental runs of standard response surface methods based on central composite and Box Behnken designs but yield comparable or lower modeling errors under realistic assumptions. In addition, the LCRSM methods have substantially lower modeling errors and greater expected savings compared with alternatives with comparable numbers of runs, including small composite designs and computer-generated designs based on popular criteria such as D-optimality. Therefore, when simulation runs are expensive, low cost response surface methods can be used to create regression meta-models for queuing or other system optimization. The LCRSM procedures are also apparently the first experimental design methods derived as the solution to a simulation optimization problem. For these reasons, we say that LCRSM are "for and from" simulation optimization. We compare the proposed LCRSM methods with a large number of alternatives based on six criteria. We conclude that the proposed methods offer attractive alternative to current methods in many relevant situations.

1 INTRODUCTION

Many engineers and scientists use design of experiments techniques to construct empirical regression or "response surface" models. An important application of response surface models is meta-models for optimizing a simulated system. When simulation runs are expensive, e.g., if a system with a large number of queues is being modeled with a high degree of realism, response surface methods (RSM) permit the user to develop an inexpensive surrogate or "meta-model" to facilitate the understanding and optimization of the system being studied. Kelton (1999) provides a recent tutorial on applications of RSM to simulation meta-modeling.

Popular choices for experimental designs include Box Behnken (1960), central composite, and small central composite designs, e.g., see Draper and Lin (1996). These designs have several important justifications but are only available for numbers of experimental runs that may, for many relevant applications, be considered too large. In the common situation in which the experimenter has only a fixed budget, he or she must simply drop factors until the corresponding number of runs meets the budget. Because these procedures clearly can result in models of limited scope and poor engineering results, there has been considerable interest in alternative methods with fewer runs for a given number of factors. For reviews of some of this research, see Box and Draper (1987), and Draper and Lin (1996 and 1990), and Myers and Montgomery (1995).

This paper proposes low cost response surface methods (LCRSM), which provide simple-to-use alternatives to standard response surface methods with approximately half the runs of Box Behnken and central composite designs and substantially fewer runs than small composite designs in most situations. The proposed methods derive from design criteria based on assumptions, model selection techniques, and diagnostic methods that have recently become possible to implement because of improvements in optimization methods and computing power. The main justifications of the methods are that (1) the expected accuracy of the empirical models derived using LCRSM is comparable to the accuracy derived from more expensive methods under realistic assumptions about the experimental conditions, and (2) the methods are simple to use, requiring no special software and limited training. The importance of this second justification is established by the widely cited fact, e.g., Myers and Montgomery (1995), that simple, mechanical methods based on central composite and Box Behnken (1960) designs are used much more than any other response surface methods.

In the next section, we formally propose the LCRSM methods and illustrate their application to an automotive design problem. We present and illustrate the LCRSM methods first in order to highlight their ease of application before we discuss their derivation and justifications. Next, we review recent related progress in the areas of experimental design criteria and optimization methods. Then, we present justifications for the chosen candidate model selection and regression diagnostics, and we compare LCRSM methods with alternative methods based on cost, final empirical model accuracy, and the remaining Draper and Lin (1996) "value for money" criteria. Finally, we summarize the contributions.

2 LOW COST RESPONSE SURFACE METHODS

The application of low cost response surface methods (LCRSM) is very similar to that of ordinary response surface methods except multiple models are fit instead of one and the diagnostic test is different. The four major steps in the application of any response surface methodology are: 1) experimental setup and testing, 2) modeling, 3) diagnosing whether the model is sufficiently accurate, and 4) additional testing, if needed. We use the application of LCRSM to aid in decision-making aimed at increasing profits and reducing customer lead-times of a fictitious facility to illustrate the methods. Example applications of LCRSM in engineering design contexts include Allen, Yu, and Bernshteyn (2000) and Koc, Allen, Jiratheranat, and Altan (2000). In the example we use in this paper, the goal is to allocate engineering resources to reduce processing times at various machine centers in a production facility. Imagine that each simulation run requires greater than 1 hour of computer time and nontrivial preparation time and that there is time pressure on the allocation decision. Therefore, we only have enough time for 14 simulation runs. We have four factors which correspond to possible centers to invest in and the correspondence between processing time distribution and expenditure is built into the simulation. Since four factors are of interest to the engineers, we choose to use LCRSM rather than to drop a factor, as would be required using methods based on Box Behnken (1960) or central composite designs.

2.1 The LCRSM Procedure

The precise procedure is defined as follows.

Step 1: (Setup and Experimentation) Choose the experimental factors. Set up the experiment by taking the experimental design from an appropriate table, either *Table 1(a)* or *Table 2(a)* for three and four factors respectively and perform the specified tests. These experimental arrays are derived in Section 3 by minimizing the expected integrated means squared modeling error, proposed in Allen and Yu (2000), as evaluated through a simplified simulation of the multi-model, potentially sequential analysis process described in *Steps 2-4*. In our example, we use the design in *Table 2(a)*. *Table 3* also shows the inputs in thousands of dollars along with the data from the 14 distinct simulation runs for the two responses, which are profits in thousands of dollars/shift and lead-time in hours.

Step 2: (Model Selection) Create the regression model(s) of each response by fitting the appropriate set of candidate model forms. For the 3 factor design in *Table 1(a)*, this is the set in *Table 1(b)*. Similarly, see *Table 2* for 4 factors. Select the fit model form with the lowest sum of squares error. Scaled (-1,1) units are used until the last stage when the chosen model forms are fit in the engineering units. The primary justification of these choices of fit models, as described below, relates to the pragmatic need to keep the number of candidate models small in order to maintain reasonable computation times for the practitioners during analysis and for us during design generation. Also, while LCRSM procedures have so far only been characterized formally for the specific sets of models described in the tables, we have used linear combinations of fit models for prediction following engineering judgment in specific cases. In our example, we fit the four model forms in *Table 2(b)* to each of the responses and selected the one with the lowest sum of squares error. The selected models are: $y_{1,est}$=72.0 + 9.0A + 14.1B + 13.4C + 11.8D + 8.52A^2 - 6.15B^2 + 0.86C^2 + 3.95AB - 0.462AC - 0.744BC and $y_{2,est}$ = 14.63 + 0.821A + 1.49B - 0.302C - 3.66D - 0.453A^2 - 1.666C^2 + 7.89D^2 - 2.221AC - 0.307AD + 1.36CD.

Table 1: LCRSM with 3 Factors: (a) the Start-up Design in Scaled (-1,1) Units, Referred to as ξ_1, (b) the Model Forms, and (c) the Optional Follow-up Runs, ξ_2

(a)

Run	A	B	C
1	1	-1	0
2	0	-1	1
3	1	1	1
4	-1	-1	-1
5	-1	0	0.5
6	0	0	0
7	-0.5	1	-0.5
8	0.5	0.5	-1
9	0.5	0.5	-1

(b)

Form #1: $\beta_0+\beta_{\mathrm{A}}\mathrm{A}+\beta_{\mathrm{B}}\mathrm{B}+\beta_{\mathrm{C}}\mathrm{C}+$
$\beta_{\mathrm{A}}2\mathrm{A}^2+\beta_{\mathrm{B}}2\mathrm{B}^2+\beta_{\mathrm{AB}}\mathrm{AB}$

Form #2: $\beta_0+\beta_{\mathrm{A}}\mathrm{A}+\beta_{\mathrm{B}}\mathrm{B}+\beta_{\mathrm{C}}\mathrm{C}+$
$\beta_{\mathrm{A}}2\mathrm{A}^2+\beta_{\mathrm{C}}2\mathrm{C}^2+\beta_{\mathrm{AC}}\mathrm{AC}$

Form #3: $\beta_0+\beta_{\mathrm{A}}\mathrm{A}+\beta_{\mathrm{B}}\mathrm{B}+\beta_{\mathrm{C}}\mathrm{C}$
$\beta_{\mathrm{B}}2\mathrm{B}^2+\beta_{\mathrm{C}}2\mathrm{C}^2+\beta_{\mathrm{BC}}\mathrm{BC}$

(c)

Run	A	B	C
A1	1	-0.5	1
A2	1	1	-0.5
A3	-0.5	1	1

Step 3: (The Least Squares Coefficient Based Diagnostic) Calculate

$$\beta_{q,est} = \left(\sum_i^q \beta_{i,est}^2\right)^{1/2}(q-1)^{-1/2} \tag{1}$$

where $\beta_{i,est}$ are the least squares estimates of the q second order coefficients in the model chosen in *Step 2*. Include coefficients of terms like A^2 and BC, but not first order terms such as A and D. Estimate the maximum acceptable standard error of prediction or "plus or minus" accuracy goal, $\sigma_{prediction}$. If $\beta_{q,est} \leq \ 1.0\sigma_{prediction}$, refit the model form in the engineering units. Stop. Otherwise, or if there is any special concern with the accuracy, continue to *Step 4*. Special concerns might include mid-experiment changes to the experimental design. The primary justification for this rule is that, by permitting the same degrees of freedom to be used for both diagnosis and model fitting, the LSCB diagnostic is able to achieve substantially lower expected modeling errors than the standard regression diagnostics which perform poorly in this context. Also, the fact that $\beta_{q,est}$ provides an approximate, conservative estimate of the prediction errors is established below through examination of the quartiles of the distribution of the integrated mean squared error conditioned on $\beta_{q,est}$. The default assumption for $\sigma_{prediction}$ is that it equals 2.0 times the estimated standard error, because then the achieved expected "plus or minus" accuracy approximately equals the error that would be expected if the experimenter applied substantially more expensive methods based on composite designs. The standard error can be estimated in practice using s/c_4, where s is the standard deviation of data from the repeated runs and c_4 is the bias correction (=0.80 when 2 repeated points are used, e.g., *Table 1*, and 0.89 when 3 points are used, e.g., *Table 2*).

Table 2: LCRSM with 4 Factors: (a) the Start-up Design in Scaled (-1,1) Units, Referred to as ξ_1, (b) the Model Forms, and (c) the Optional Follow-up Funs, ξ_2

(a)

Run	A	B	C	D
1	-0.5	-1	-0.5	1
2	1	1	-1	1
3	-1	1	1	1
4	1	-1	-0.5	-0.5
5	0	0	-1	0
6	0	1	0	0
7	-0.5	-1	1	-0.5
8	-1	0	0	0
9	1	1	1	-1
10	-1	1	-1	-1
11	0	0	0	-1
12	0.5	-0.5	0.5	0.5
13	0.5	-0.5	0.5	0.5
14	0.5	-0.5	0.5	0.5

(b)

Form #1: $\beta_0+\beta_A A+\beta_B B+\beta_C C+\beta_D D+$
$\beta_A 2A^2+\beta_B 2B^2+\beta_C 2C^2+$
$\beta_{AB} AB+\beta_{AC} AC+\beta_{BC} BC$

Form #2: $\beta_0+\beta_A A+\beta_B B+\beta_C C+\beta_D D+$
$\beta_A 2A^2+\beta_B 2B^2+\beta_D 2D^2+$
$\beta_{AB} AB+\beta_{AD} AD+\beta_{BD} BD$

Form #3: $\beta_0+\beta_A A+\beta_B B+\beta_C C+\beta_D D+$
$\beta_A 2A^2+\beta_C 2C^2+\beta_D 2D^2+$
$\beta_{AC} AC+\beta_{AD} AD+\beta_{CD} CD$

Form #4: $\beta_0+\beta_A A+\beta_B B+\beta_C C+\beta_D D+$
$\beta_B 2B^2+\beta_C 2C^2+\beta_D 2D^2+$
$\beta_{BC} BC+\beta_{BD} BD+\beta_{CD} CD$

(c)

Run	A	B	C	D
A1	-1	1	-1	1
A2	-1	-1	-1	-1
A3	-1	1	1	-1
A4	1	1	-1	-1

Table 3: Example – The Left shows the Inputs which are Investments in $K to Reduce Times at the Four Machine Centers Labeled A-D, and the Right Shows the Estimated Profits in Dollars and Lead Times in Hours

Run	*A*	*B*	*C*	*D*	y_1	y_2
1	1.25	1.7	12.5	10.00	55.95	15.39
2	2.00	2.1	10.0	10.00	101.76	19.92
3	1.00	2.1	20.0	10.00	101.23	21.02
4	2.00	1.7	12.5	6.25	52.93	18.55
5	1.50	1.9	10.0	7.50	59.93	13.42
6	1.50	2.1	15.0	7.50	80.54	15.90
7	1.25	1.7	20.0	6.25	60.87	14.70
8	1.00	1.9	15.0	7.50	72.02	13.51
9	2.00	2.1	20.0	5.00	102.70	22.81
10	1.00	2.1	10.0	5.00	51.36	23.79
11	1.50	1.9	15.0	5.00	59.42	26.33
12	1.75	1.8	17.5	8.75	81.94	13.50
13	1.75	1.8	17.5	8.75	81.94	13.50
14	1.75	1.8	17.5	8.75	81.94	13.50

Another relevant assumption is that the standard deviation of the random error is much smaller than the accuracy needed, which holds for many types of computer experiments. Then, the user needs to specify the desired accuracy to avoid unnecessary experimental expense. In our example, we select $\sigma_{prediction}$ based on financial needs to equal $5.0K, or "±5K" accuracy, for the profit and $\sigma_{prediction}$=5.0 hrs. for the lead time. The square roots of the {sum of squares of the 6 quadratic coefficients divided by 5} for the two responses are $\beta_{q,est}$=$5.0K and $3.8 hours respectively. Since these are less than or equal to their respective cutoffs, we stop. No more experiments are believed necessary to achieve adequate meta-model prediction errors.

Step 4: (Additional Runs, If Necessary) If needed, perform additional experimental runs specified in part "*(c)*" in the table appropriate for the number of factors used. After the experiment, fit a full quadratic polynomial regression model as in ordinary response surface methods (RSM). Then, the fit model is expected to have comparable errors as if an expensive composite design had been applied and a quadratic model fit.

3 THE DERIVATION OF LOW COST METHODS

In this section, we review and extend the results from the studies of experimental design criteria and optimization used in the method development from Allen and Yu (1999*a* and *b*). In the next section, we describe new methods for model discrimination and regression diagnostics that make LCRSM possible.

3.1 Experimental Design Criterion

LCRSM use relatively few experimental test runs as compared with alternatives. Therefore, it is imperative to minimize the risk that the LCRSM procedure will derive inaccurate empirical models by capitalizing on the benefits from optional sequential experimentation and model selection. In this section, we begin by developing an experimental design criterion or optimization objective that can estimate the expected loss from model inaccuracy or "risk" in the Bayesian sense, e.g., see Pilz (1991), taking into account errors from important interactions and other terms not included in the final fit model, the possibility of a follow-up experimentation, and model selection.

Box and Draper (1959 and 1987) and others have pointed out the limitations of many popular experimental design criteria, such as D-optimality. These criteria ignore the often-dominant bias errors from fitting a model form that does not allow accurate approximation of the true response. Therefore, we base our criterion on the Box and Draper (1959) integrated mean squared error (IMSE) criterion, which includes bias errors in the estimation of the model errors. Allen and Yu (2000*a*) proposed the semi-Bayesian extension of the IMSE called the expected integrated mean squared error (EIMSE) to permit its application in realistic situations when the true model is not known. Allen and Yu (2000*a*) showed that, in the single fit model non-sequential case, EIMSE optimal designs performed nearly optimally for a variety of assumptions and criteria, which was not true for other criteria such as D-optimality and minimum bias. Further, they showed that minimum bias and integrated variance optimal designs could be generated using the EIMSE criteria.

The advantages of the EIMSE criterion include that the square root of the EIMSE provides a direct and comprehensive estimate of the plus or minus prediction errors that experimenters can expect to achieve. Also, it is one of few criteria that has been extended to apply to sequential experimentation, see Allen and Yu (2000*b*). Moreover, because we will only have enough runs to fit first and selected second order terms and we believe (see below) that the true model can only be accurately approximated by a third order polynomial, it is imperative that we include both variance and bias errors which come from model misspecification in our criterion. The EIMSE criterion is the only usable criterion that we are aware of which does this. Finally, Allen and Yu (2000*a*) showed that, EIMSE optimal designs yielded good performance for a variety of assumptions and other criteria such as D-efficiency and minimum bias, which was not true for alternative designs.

Next, we propose a straightforward extension to the EIMSE as defined in Allen and Yu (2000*b*) to estimate and compare the plus or minus errors in cases in which multiple models are fit and a mechanistic analysis procedure examines the data and selects the best model. We write the expression below for the EIMSE independently of our assumptions about the prior distributions for the coefficients, $\boldsymbol{\beta}$, and the random errors for the first and second experiment (if needed), $\boldsymbol{\varepsilon}_1$ and, $\boldsymbol{\varepsilon}_2$. In the next subsection, we will describe the distributions used to generate LCRSM. Fortunately, as we will describe in later sections, the LCRSM designs give good performance for a variety of assumptions. We define the startup experimental design matrix, $\boldsymbol{\xi}_1$, and the followup design matrix, $\boldsymbol{\xi}_2$, which is optional. The function $y_{i,est}$ is the fit model based on the model form i =1,2,..,s, and $y_{full,est}$ is the fit model after augmentation which is assumed to be a full quadratic functional form. The indicator function $\pi_{stop}(\boldsymbol{\beta},\varepsilon_1,\boldsymbol{\xi}_1)$ equals 1 if augmentation is not needed as decided after the first experiment by the chosen model diagnostic, i.e., if $\beta_{q,est} < 2\sigma_{est}$, and 0 otherwise. Similarly, $\pi_i(\boldsymbol{\beta},\varepsilon_1,\boldsymbol{\xi}_1)$ equals 1 if model i is selected by the model selection method, i.e., in our case if $SSE_i < SSE_j$ for all model pairs $i\neq j$, and 0 otherwise. Defining these "π" functions of random variables is necessary in order to preserve the interpretation of the EIMSE as the expected prediction errors. With these definitions, the EIMSE objective is:

$$\min_{\boldsymbol{\xi}_1,\boldsymbol{\xi}_2} \text{EIMSE}(\boldsymbol{\xi}_1,\boldsymbol{\xi}_2) = \underset{\boldsymbol{\beta},\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2}{\mathrm{E}} \left\{ \begin{array}{l} \pi_{\text{stop}}(\boldsymbol{\beta},\boldsymbol{\varepsilon}_1,\boldsymbol{\xi}_1)\sum_{\text{i}=1}^{\text{s}} \pi_{\text{i}}(\boldsymbol{\beta},\boldsymbol{\varepsilon}_1,\boldsymbol{\xi}_1)\int_{\text{R}} \rho(\mathbf{x})[\text{y}(\mathbf{x},\boldsymbol{\beta}) - \text{y}_{\text{i,est}}(\mathbf{x},\boldsymbol{\beta},\boldsymbol{\varepsilon}_1,\boldsymbol{\xi}_1)]^2 \,\text{d}\mathbf{x} \\ + [1 - \pi_{\text{stop}}(\boldsymbol{\beta},\boldsymbol{\varepsilon}_1,\boldsymbol{\xi}_1)]\int_{\text{R}} \rho(\mathbf{x})[\text{y}(\mathbf{x},\boldsymbol{\beta}) - \text{y}_{\text{full,est}}(\mathbf{x},\boldsymbol{\beta},\boldsymbol{\varepsilon}_1,\boldsymbol{\xi}_1,\boldsymbol{\varepsilon}_2,\boldsymbol{\xi}_2)]^2 \,\text{d}\mathbf{x} \end{array} \right\} \quad (2)$$

where by default $\rho(\mathbf{x})=1/V$ where V is the volume of the region of interest and the expression generally requires numerical simulation for its evaluation.

Defined in this way, the EIMSE has the intuitive interpretation of being the expected IMSE of the final fit model derived from the experimental process. Also, the square root of the EIMSE may be defined as the "standard error of prediction" or simply the "plus or minus prediction error" which is of immediate relevance to the practitioner. The trade-off for this interpretability and other advantages is that the integration in (2) cannot be evaluated analytically, complicating the minimization. Fortunately, recently it has become feasible in relevant cases to perform this minimization using the heuristics described in Section 3.3 and modern computer capabilities.

3.2 Assumptions about the True Model

We begin by making what we believe is an often realistic assumption that a *third* order polynomial with $N(0,\sigma^2)$ experimental random errors, with unknown σ, well approximates the true model. If one feels that the true model is highly nonlinear, i.e., fourth or higher order is needed, then it is not clear whether any empirical modeling methods with comparable numbers of runs to standard response surface methods will provide an adequate fit model. Also, while several authors have investigated the effects of outliers on empirical modeling techniques, for a review see Beckman and Cook (1983), at present we urge the user to re-perform runs believed to be outliers. Note that the assumption of a full cubic polynomial is substantially more realistic than the assumption of a quadratic polynomial made in common implementations of optimal experimental design, e.g., applications of D-optimality, which are based on a single criterion and do not include bias errors.

Next, we specify the necessary assumptions about the coefficients of the cubic polynomial. We make no assumptions about the first order terms and only limited assumptions about the second order terms. This follows because these terms may or may not be very large compared with the standard deviation of the random error, σ. It is not difficult to show analytically that if we make the specific choices about candidate model terms described in the next section then model selection probabilities and virtually all properties of the fit model are independent of the values of the first order coefficients. The main assumption about the quadratic coefficients that we make is conservative. This is that all the quadratic coefficients of our terms, with the standard [1-,1] scaling of all factor, are roughly of the same order of magnitude, i.e., N(0,β_q^2). Clearly, if some of the terms are exactly zero, this would tend to favor methods like ours that omit some quadratic terms. Then, we assume β_q is a U(0,L) hyper parameter to allow for our uncertainty about the degree of curvature. We admit that this scheme may seem somewhat arbitrary, however, in our simulation investigations, we have found that the choice of distributions have a small effect on the relative performance of alternative methods. An important achievement of LCRSM is, we believe, that the EIMSE is highly insensitive to L (and β_q). Thus, the methods are robust to uncertainty about the quadratic curvature. Also, the EIMSE values used in the comparison are based on almost worst-case values of L and so, we believe, are conservative.

In Allen and Yu (2000 *a* and *b*), we investigated assumptions about the third order terms and concluded that assuming that the third order coefficients are N(0,β_c^2) with β_c equal to 0.5 is in accordance with the implied assumptions of the users of response surface methods. As noted by Box and Draper (1987), "An investigator might typically employ a fitted approximating function such as a straight line, if he believed that the average departure from the truth induced by the approximating function were no worse than that induced by the process of fitting." Therefore, it is reasonable to believe that one will tend to choose the degree of this approximating function in such a way that the integrated variance error and the expected integrated systematic error are about equal. It is easy to check that 0.5 makes the expected bias errors approximately equal to the integrated variance errors when a full central composite design is applied. Fortunately, it is easy to show that the relative effectiveness of alternative methods is largely independent of the distribution of β_c.

3.3 The Method for Experimental Design Generation

In this section, we review briefly how search techniques that have traditionally been used for generating experimental designs cannot be used to generate LCRSM and how a new class of simulation optimization or "stochastic" search techniques can; see Allen, Bernshteyn, and Yu (1999) for more details. Each evaluation of the EIMSE in (1) requires time-consuming numerical simulation to calculate the expected values for all but trivial choices of diagnostics and model selection procedures, i.e., $\pi_{stop}(\boldsymbol{\beta},\varepsilon_1,\boldsymbol{\xi}_1)$ and $\pi_i(\boldsymbol{\beta},\varepsilon_1,\boldsymbol{\xi}_1)$ functions. Exchange algorithms, which are used in the majority of commercial optimal experimental design algorithms, obtain excellent efficiencies for linear optimality criteria that do not require simulation evaluation, such as D-optimality, by minimizing the number of time-consuming function calls. To do this, they use recursive formulas and large numbers of computationally inexpensive evaluations, see, e.g., Meyer and Nachtsheim (1995). At present, no recursive formulas exist to aid in the evaluation of the EIMSE. Therefore, every evaluation is expensive and exchange algorithms are extremely inefficient.

Optimization where the objective value must be estimated using numerical integration is called "simulation optimization" for surveys see, e.g., Plug (1996) and Andradottir (1996). We used the optimization heuristic in Allen, Bernshteyn, and Yu (2000) to produce the start up and follow-up designs shown in *Table 1*, *Table 2*, and *Table 3* by minimizing the EIMSE in (3.1), using the proposed hierachical prior, and the model selection and diagnostics described in the next section. This optimization method combines population based and multiple comparison based search methods and uses variation reduction techniques. In brief, a three stage eliminating procedure is used to approximately sort the population in each stage (generation) of the genetic algorithm and to guarantee with an assignable probability that the best solution is not lost.

4 MODEL SELECTION AND DIAGNOSTICS

The EIMSE objective, prior distributions, and stochastic optimization methods can be used to generate and evaluate experimental designs based on many types of model selection and diagnostic procedures. In this section, we motivate the choice of the procedures used in LCRSM which, in the case of model selection, are hybrids and extensions of approaches discussed in Srivastava (1996) and Meyer, Steinberg, and Box (1995) adapted to the response surface context. It would be ideal, perhaps, to carry out the optimization of the EIMSE simultaneously over the space of experimental designs, $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$, and over the space of possible model selection and diagnostic decision rules. However, other considerations besides model accuracy such as simplicity and analytical results relating to robustness of the methods constrain the choices and provide the motivation for the proposed hybrid procedures. These procedures, which define the functions $\pi_{\text{stop}}(\boldsymbol{\beta},\varepsilon_1,\boldsymbol{\xi}_1)$ and $\pi_i(\boldsymbol{\beta},\varepsilon_1,\boldsymbol{\xi}_1)$ used to evaluate the EIMSE in (3.1), are largely responsible for the success of LCRSM by permitting designs with few runs to achieve small values of the EIMSE, ease of use, and other desirable properties.

4.1 The Model Selection Procedure

In this section, we motivate the proposed model selection approach. We begin by briefly reviewing the proposed method. Then, we discuss the criteria used to evaluate methods, use these criteria to characterize the relevant alternatives from the literature, and describe how the proposed methods address the criteria. In the proposed method models, all factors are present at first order in all candidates, e.g., for m=3 all candidate models contain $\beta_0+\beta_A A+\beta_B B+\beta_C C$. The m models differ by the combinations of m-1 of the m factors present at second order, e.g., for m=3, one of the three models would also contain $\beta_{A2}A^2 +\beta_{C2}C^2 +\beta_{AC}AC$, missing factor B at second order. See *Tables 1-3b* for other examples.

There has been considerable interest on the topic of experimental design for model discrimination including Atkinson and Fedorov (1975), Pukelsheim and Rosenberger (1993), and Meyer, Steinberg, and Box (1995) work discussed in Srivastava (1996). Two of the most influential types of candidate models include main effect plus P (MEP+P) plans where P is an integer Srivastava describes and what we term the "all combinations of active factors" model sets proposed in Box and Meyer (1993) and used in Myers, Steinberg, and Box (1995). Both types of approaches have been proposed in the context of two level experiments and require generalization to be applied for response surface exploration.

MEP+P plans involve candidate models in which all candidates include all first order terms and the candidates differ by the combination of P terms from a chosen list of R higher order terms, with all R choose P combinations included. The central problem with MEP+P plans is that the large number of candidate models makes modeling difficult. For example, with m=4 factors, the number of possible combinations of P=6 of the possible R=10 second order terms is 210. While selecting the true model from among 210 choices might, depending upon assumptions, drive down the *EIMSE* compared with fitting the 4 candidate models in *Table 2 (b)*, it is not, at present, reasonable to request that the user routinely perform so many regressions. Also, times for generating the arrays in *Tables 1-3* are roughly linear in the number of candidates. Using a Pentium 450 MHz machine, computation time to generate the array in *Table 3* involving 10 candidate models required roughly 1 day. Because this optimization must be run several times to ensure a thorough solution, we feel that 10 candidates is, at least temporarily, a practical limit. The "all combinations of active factors" method uses all possible choices of active factors and includes in each model all possible interactions between the "*a*" active factors including interactions of order *a*. Benefits of this type of approach include that it conforms to the plausible assumptions of "factor sparsity" and "effect heredity". Box and Meyer (1986) defined "factor sparsity" as the assumption that not all factors are "active." This means that some or all of the terms in the Taylor series expansion associated with at least some of the factors are negligible. Hamada and Wu (1992) defined "effect heredity" as the assumption that certain 2nd order terms are only present only if the related 1st order terms being non-zero. If these assumptions hold, then the all combination of active factors approach has a non-negligible probability of determining the "true" model exactly. However, this approach results in 2^m candidate models of different orders, which is usually too many for easy calculation. Also, depending upon the number of active factors, some of the candidates may not be estimable using ordinary least squares requiring relatively difficult Bayesian modeling. Finally, in the "all combinations" method all candidate models do not contain all factors at first order. It is easy to confirm empirically that this leads to high values of the EIMSE because of the high probability that first order coefficients are large.

The proposed strategy has the following justifications. First, all candidate sets contain ten or fewer models facilitating calculation both for the practitioner and for ourselves during design generation. Second, all candidates contain all first order terms to make expected model errors independent of first order term distribution. Third, the strategy capitalizes on the benefits of both types of plans and addresses both of our criteria. It is similar to the main effect plus P (MEP+P) plans because all the candidate models contain all first order terms. Another benefit that is shared with the MEP+P plans is that all candidates include greater than three terms so that the diagnostic described in the next section can be applied. However, we only use m candidate models, so our method is simpler to use. Finally, like the all combinations of active factors plan, our proposed strategy capitalizes on the possibility of factor sparsity and effect heredity. It should be noted that our candidate models do not even approximately minimize the EIMSE under the assumption that the true model is a full cubic with normally distributed coefficients in Section 3.2. Since this assumption does not take into account factor sparsity and heredity, which would tend to favor our choices of models, we feel that the EIMSE based on our earlier assumptions provides a conservative estimate of the prediction errors.

4.2 The Least Squares Coefficient Based (LSCB) Diagnostic

In this section, we justify the proposed LSCB diagnostic by comparing it with the standard regression F-test diagnostic described in, e.g., Myers and Montgomery (1995). We begin by describing the role of diagnostics in the low cost response surface context. Then, we present empirical evidence of the benefits of the LSCB diagnostic. Both LCRSM and small composite designs involve estimating a test statistic based on small numbers of start-up experimental runs with the purpose of deciding whether additional, follow-up experimental runs are needed after the first set of experiments have been performed so that the final model will have acceptable accuracy. For example, when small composite designs are used, the experimenter performs the start-up experiments on the cube and center points and calculates the test statistic that is the mean squared error divided by the variance of the repeated points, σ_{est}. Then, the practitioner performs an F-test using this statistic usually with α=0.05 or 0.25 to decide whether the so-called "star point" follow-up runs and a second order model form are needed, e.g., see Myers and Montgomery (1995). Similarly, in LCRSM, the experimenter performs the first set of experiments, e.g., *Table 1 (a)*, selects the appropriate model containing one subset of the possible quadratic terms, and uses the LSCB test statistic to decide whether or not to perform additional pre-tabulated, follow-up runs, e.g., *Table 1 (c)*. As defined in equation (1), the LSCB test statistic is the square root of (the sum of the squared quadratic coefficients divided by the number of quadratic terms minus one), $\beta_{q,est}$. If $\beta_{q,est}$ is less than two times the desired plus or minus error of prediction, $\sigma_{prediction}$, stop. Otherwise, perform the additional, follow-up runs and fit a full quadratic. By default, we assume that $\sigma_{prediction}$, = $2\sigma_{hat}$, where σ_{hat} is the estimated standard error derived from the repeated points. We show later that with this choice, LCRSM procedures provide expected final model errors comparable to alternatives based on small composite and Box Behknen designs under the realistic assumptions described in Section 3.2.

The primary justification for the LSCB diagnostic is that it is able to achieve the objectives of the diagnostic procedures while requiring two or more fewer start-up runs than the standard F-test diagnostic. The diagnostics have two objectives. First, the additional runs generally carry a high cost, which is presumably the motivation for using the low cost methods. Therefore, it is desirable to minimize the probability that the follow-up runs will be used. Second, in some cases the follow-up runs may be necessary to minimize the expected modeling errors to achieve acceptable accuracy. Empirical comparison of the two diagnostics is not possible based on composite designs, because the LSCB cannot be applied. The LSCB diagnostic requires a model containing at least three quadratic terms for the robustness reasons mentioned above which is not possible based on a cube and center point design. However, it is possible to compare the two diagnostics based on LCRSM start-up and follow-up designs and modeling strategies. *Table 5* compares the performance of the two methods. The extremely high EIMSE values show that standard regression diagnostics cannot be used in the context of LCRSM. The results also illustrate the dangers of having fewer than three degrees of freedom for the diagnostic. With the same experimental designs and model selection strategy, the LSCB diagnostic bases decisions on several more degrees of freedom than the standard regression diagnostic. In all cases, the *MSE* has only one degree of freedom while the $\beta_{q,est}$, used in LSCB testing, has three or more.

Table 4: Compares F-Test Based and LCRSM Diagnostics for L=8 and β_c=0.5 and the Designs in Tables 1-2(a)

No.	F-Test with α = *0.05*		F-Test with α = *0.25*		LCRSM diagnostic	
Factors	*EIMSE*	P*(*stop*)*	*EIMSE*	P*(*stop*)*	*EIMSE*	P*(*stop*)*
3	*10.0*	*0.89*	*5.1*	*0.57*	*1.2*	*0.15*
4	*13.1*	*0.84*	*8.0*	*0.59*	*1.4*	*0.17*
5	*25.6*	*0.84*	*15.7*	*0.59*	*2.3*	*0.15*

5 COMPARISONS WITH ALTERNATIVES

In the next two sections, we compare LCRSM with alternatives in order to establish that it provides an attractive alternative to standard methods. Comparison of alternative methods is an important topic in its own right. Much in the next two sections is repeated from Allen and Yu (2000*b*), which contains a more thorough comparison of the alternatives. In this section, we compare a relatively small number of alternatives based on what we consider to be the two most important criteria: 1) the experimental cost which primarily depends upon the number of runs, and 2) the expected accuracy of the derived, final empirical model. We restrict the scope of the comparison so that we can examine a range of assumptions about the conditions.

Comparison of the modeling errors using popular experimental design criteria such as G-efficiency and D-efficiency, see, e.g., Myers and Montgomery (1995), which rely on a pre-known functional form and experimental design, is complicated by the fact that the LCRSM and the methods based on composite designs are sequential. Therefore, the final model form and associated experimental design cannot be known before the experiments are performed and these criteria would need to be generalized to be applied. For this reason and because it has the simple, intuitive interpretation of being the "plus or minus" prediction errors, we base the comparison on the square root of expected integrated mean squared error, *sqrt*(EIMSE), where the formula for the EIMSE is given in (3.1). This formula depends on the parameters *L* and β_c. Therefore, we discuss results for a variety of combinations of these parameters. In Section 3.2, we argued that L=8 and β_c =0.5 represent the implied assumptions of standard response surface methods which are conservative in the sense that they include large quadratic curvature.

Figure 1 shows the plus or minus errors, i.e., *sqrt*(EIMSE) of four methods for a range of values of L and for four factors, which we believe is representative of other numbers of factors. The four methods compared include procedures using the Hartley (1959) small central composite design (SCCD) applied with the standard diagnostic with α=0.25 and without the possibility of stopping, i.e., α=1, LCRSM, and standard response surface methods based on the 27 run Box Behnken design which is not sequential and is based on fitting a full quadratic polynomial. The results show that the model errors of the LCRSM are appreciably below those of small central composite designs whether or not the composite designs are applied sequentially. Also, the probability of stopping with fewer runs, while dependent on L, is substantially larger for LCRSM than for small composite designs applied sequentially. Therefore, the probability of savings is considerably higher and the accuracy is significantly improved. Admittedly, if one stops using the small central composite one has performed only ten tests compared with fourteen using the response surface methods. However, given that the experimenter terminates having performed only the startup runs, the plus or minus errors, (EIMSE)$^{-1/2}$, using the small composite design are 2.7σ compared with 1.5σ for the LCRSM approach. Also, if the degree of nonlinearity as set by the coefficient β_c increases, then the expected bias errors increase roughly proportional to β_c^2. The curves look similar to *Figure 2*, see Allen and Yu (1999), the differences in the accuracy also increase, and the performance compared with small central composite designs improves. Finally, while the Box Behnken design and full quadratic model do reduce the plus or minus errors of prediction, the errors are comparable although the Box Behnken requires 9 or 13 additional runs. In the next section, we discuss other criteria in addition to cost and accuracy.

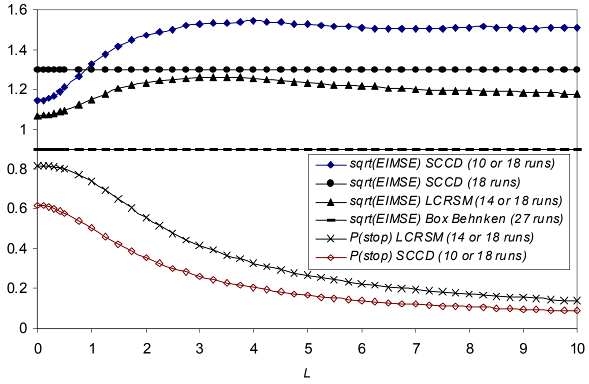

Figure 1: The *sqrt*(EIMSE) or Plus or Minus Prediction Errors for Alternative Methods with Four Factors.

6 COMPREHENSIVE COMPARISONS

In this section, we expand the comparison to include twelve methods in addition to LCRSM and six objectives based on our interpretation of the Draper and Lin (1995) money for value criteria. We admit that this comparison ignores several properties that the practitioner might consider important, e.g., the distribution of the model errors. However, we follow Draper and Lin (1995) in assuming that people interested in LCRSM will be less interested in these properties than the five that we consider. We have attempted to include all of the most popular regression-based methods that have been proposed for response surface investigations in the comparison. These include methods based on standard and small central composite designs reviewed or proposed in Hartley (1959) and Draper and Lin (1990), Box Behnken designs proposed in Box and Behnken (1960), and other computer-generated design methods. Other research on design of experiments for model discrimination, e.g., Atkinson and Fedorov (1975), Pukelsheim and Rosenberger (1993), and Srivastava (1996), has apparently not generated any response surface designs. Therefore, no comparison is possible with other multi-model procedures and all the alternative methods are based on single fit models.

6.1 The Draper and Lin "Value for Money" Criteria

We use the following re-definitions of the 6 Draper and Lin (1996) criteria that we feel are needed to objectify the comparison while conforming to the intent of those authors. First, we compare the model errors over the region of interest using the square root of the expected integrated mean square error, *sqrt*(*EIMSE*), or plus or minus error of prediction, based on default assumptions described in Section 4. Second, we compare the ability to estimate lack of fit by specifying the type of diagnostics that have been proposed for each method and the degrees of freedom available for regression lack of fit tests, v_1. Possible diagnostic procedures include regression lack of fit tests, available whenever v_1>0, and the least squares coefficient based diagnostic proposed in Section 2. Third, we consider whether the implementation that we are evaluating is performed sequentially, with a possibility of stopping with less than the full run number. Fourth, we compare the number of degrees of freedom available for estimation of pure error, v_2, which equals the number of repetitions of one of the points. Fifth, we specify the number of runs taken into account in the evaluation of the *EIMSE*. Sixth, the subjective assessment of the simplicity of the application of the different methods is complicated by uncertainty about practitioner's knowledge of regression. If regression knowledge is limited, all the methods are approximately equivalently difficult since the additional difficulty of fitting more than 1 model is small compared with learning regression. Assuming that fitting one regression model is simple, then the LCRSM is relatively complicated because it requires fitting multiple models. Still, complication is moderate compared with, e.g., Bayesian regression, neural nets, or Kriging modeling which all, at present, require specialized software.

6.2 Comparisons with Alternative Methods

Next, we use these interpretations of the Draper and Lin (1996) "value for money" criteria to compare LCRSM with regression-based alternatives using substantially more runs, i.e., higher cost, and others with comparable numbers of runs. Then, we draw conclusions about when LCRSM should be used. *Table 5* summarizes the comparison.

The Box Behnken (1960) and central composite designs applied non-sequentially (NS) have advantages including approximately 30% lower modeling errors in sqrt(*EIMSE*). These designs also have many more degrees of freedom to estimate the model errors using regression diagnostics and detect lack of fit. However, if large errors are detected, no widely used plan for adding runs to reduce the model errors exists. Also, in order to achieve these benefits, these methods require nine additional runs in the four-factor case and substantial numbers in other cases.

Considering the less expensive alternatives, LCRSM has the smallest modeling error regardless of our assumption about *L*. The only two of these alternatives with approximately the same expected errors are computer-generated designs based on saturated or nearly saturated full quadratic models and Hartley (1959) small composite designs. Even minimum bias designs for fitting forms with missing quadratic terms give rise to extremely high errors compared with LCRSM assuming that model form #1 in Table 2 is used for 4 factors. Standard computer-generated designs, e.g., D-optimal, for nearly saturated models offer limited diagnostic capabilities, no predefined augmentation runs, and no estimate of the pure error. We therefore conclude that these computer-generated designs have few advantages beyond the simplicity of fitting a single model.

As described in Section 5, if the Hartley (1959) small central composite design is applied sequentially (S), i.e., standard regression lack of fit test based on $F_{2,1,0.25}$ determines whether a first order model is fit to the (resolution III) fractional factorial and center points or whether star points are used and the fit model is a full quadratic, the expected errors are high. Also, the probability of stopping with only the start-up runs is considerably smaller than for the LCRSM procedures. The other considerations seem roughly equal for the two methods. For example, while the full small composite design has an extra degree of freedom for diagnostics, the LCRSM has an extra degree of freedom for the pure error. Therefore, LCRSM has important cost and accuracy advantages and no important disadvantages except the complication of fitting multiple models.

In conclusion, LCRSM procedures were designed for cases in which experimental cost is a concern and the experimenter is considering dropping factors in order to meet a budget. Under the assumptions about the realistic conditions, which include substantial third order non-linearity, detailed in Section 3.2, LCRSM yields models with an expected accuracy within 30% of substantially more expensive methods. If the experimenter believes that the surface being modeled is unusually non-linear and accuracy is important, then the more expensive methods such as methods based on Box Behnken designs should be used. In those cases, the modeling error *sqrt*(*EIMSE*) advantages increase approximately proportionally to our parameter β_c, but so does the accuracy domination of LCRSM compared with other alternatives with comparable numbers or runs. We conclude that the results establish that LCRSM is a very attractive alternative with few runs, low model errors, diagnostic capabilities, and moderate simplicity.

Table 5. Comparison with 4 Factors Using 6 criteria: 1) sqrt(EIMSE) Assuming L=8, α=0.25, and β_c=0.5, 2) Diagnostic/Degrees of Freedom, 3) Additional Runs Availability, 4) #Degrees of Freedom for Pure Error, 5) Simplicity

Methods/Criteria	*1*	*2*	*3*	*4*	*5*	*6*
D-optimal Fitting Form #1	9.7	Std./1	NA	2	14	Simple
Saturated D-optimal Fitting Form #1	6.3	NA/0	NA	0	11	Simple
IV-optimal Fitting #1	5.6	Std./1	NA	2	14	Simple
Minimum Bias Fitting #1	3.7	Std./1	NA	2	14	Simple
Saturated D-optimal Fitting Full Quadratic	1.6	NA/0	NA	0	15	Simple
Hartley(1959) Small Composite (S)	1.5	Std./2	A	1	10 or 18	Simple
Central Composite (S)	1.4	Std./6	A	2	19 or 27	Simple
IV-optimal Fitting Full Quadratic	1.2	Std./1	NA	2	18	Simple
LCRSM	1.2	Special	A	2	14 or 18	Moderate
Central Composite (NS)	0.9	Std./12	NA	2	27	Simple
Box Behnken (NS)	0.9	Std./12	NS	2	27	Simple

7 SUMMARY

We have proposed low cost response surface methods (LCRSM), which require roughly half the test runs of methods based on central composite and Box Behnken designs. We have reviewed and extended results in experimental design criterion selection and optimization used in the method development. Also, we have introduced approaches for model discrimination and diagnostics and explained how the methods derive from the solution to a simulation optimization problem that involves minimizing the expression in (1) which must be estimated using simulation. To our knowledge, this research constitutes the first application of simulation optimization to create experimental designs. We have compared

LCRSM methods with alternatives that include approaches based on Hartley (1959) small central composites and computer generated designs. Low cost methods allow sequential construction, provide the ability to estimate the pure errors and lack of fit and are simple to use. We conclude that LCRSM methods offer an attractive alternative to regression-based methods appropriate for the common situation in which the experimenter is considering dropping factors in order to meet a budget constraint.

ACKNOWLEDGMENTS

The authors would like to thank Gary Herrin and Max Morris for their many contributions. Also, we would like to thank John Barkley, Mikhail Bernshteyn, Dennis Harwig, Yuan Li, and David Yapp for support. Finally, we would like to thank Barry Nelson for helpful discussions.

REFERENCES

Adradottir, S. 1996. A Global Search Method for Discrete Stochastic Optimization. *SIAM Journal of Optimization* 6: 513-530.

Allen, T., M. Bernshteyn, and L. Yu. 2000. An algorithm for discrete stochastic optimization. OSU-IWSE working paper, Columbus, OH.

Allen, T. and L. Yu. 2000*a*. The Expected Integrated Mean Squared Error Criterion. OSU-IWSE working paper, Columbus, OH.

Allen, T. and L. Yu. 2000*b*. Graphical Comparison of the Sequential Application of Second Order Designs Based on the Expected Integrated Mean Squared Error. OSU-IWSE working paper, Columbus, OH (under review).

Allen, T., L. Yu, and Bernshteyn 2000. Low Cost Response Surface Methods Applied to the Design of Plastic Snap Fits. *Quality Engineering*, 12: 583-591.

Atkinson, A. C. and V. V. Fedorov. 1975. Optimal Design: Experiments for Discriminating Between Several Models. *Biometrika* 62: 289-303.

Beckman, R. J. and R. D. Cook. 1983. Outlier…s. *Technometrics* 25: 119--163.

Box, G. and D. W. Behnken. 1960. Some New Three-Level Designs for the Study of Quantitative Variables. *Technometrics* 30: 1-40.

Box, G. and N. R. Draper. 1959. A Basis for the Selection of a Response Surface Design. *Journal of the American Statistical Association* 54: 622-654.

Box, G. and N. R. Draper. 1987. *Empirical Model Building and Responses Surfaces*, Wiley, NY.

Box, G. and R. D. Meyer. 1993. Finding the Active Factors in Fractionated Screening Experiments. *Journal of Quality Technology* 25: 94-105.

Box, G. and R. D. Meyer. 1986. An Analysis for Unreplicated Fractional Factorials. *Technometrics* 28: 11-18.

Draper, N. R. and D. K. J. Lin. 1996. Response Surface Designs.*Handbook of Statistics* 13: 343-376, eds. S. Ghosh and C. R. Rao.

Draper, N. R. and D. K. J. Lin. 1990. Small Response Surface Designs. *Technometrics* 32: 195-202.

Fang, K. T. and Y. Wang. 1994. *Number Theoretic Methods in Stat*. London: Chapman & Hall.

Gupta, S. S. and S. Panchapakesan. 1991. On Sequential Ranking and Selection Procedures in: B. K. Ghosh and P. K. Sen, eds., *Handbook of Sequential Analysis*. Dekker, NY.

Hamada, M. and C. F. J. Wu. 1992. Analysis of Designed Experiments With Complex Aliasing. *Journal of Quality Technology* 24: 130-137.

Hartley, H. 1959. Small Central Composite Designs for Quadratic Response Surfaces. *Biometrics* 15: 611-624.

Koc, M., T. Allen, S. Jiratheranat, and T. Altan. 2000. The use of FEM and experimental design to investigate tube hydroforming of a simple geometry. To appear in *The International Journal of Machine Tools & Manufacture*.

Kelton, W. D. 1999. Designing Simulation Experiments. *Proceedings of the 1999 Winter Simulation Conference*, eds. P. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. W. Evans.

Meyer, R. K. and C. J. Nachtsheim. 1995. The coordinate-exchange algorithm for constructing exact optimal experimental designs. *Technometrics* 37: 60-69.

Meyer, R. D.; D. M. Steinberg; and G. Box. 1995. Follow-Up Designs to Resolve Confounding in Multifactor Experiments. *Technometrics* 38: 303-313.

Myers, R. H. and D. A. Montgomery. 1995. *Response Surface Methodology*, Wiley, NY.

Pflug, G. C. 1996. *Optimization of Stochastic Models: The Interface Between Simulation and Optimization.* Boston, Mass.: Kluwer Academic.

Pilz, J. 1991. *Bayesian Estimation and Experimental Design in Linear Regression Models*. Wiley, NY.

Pukelsheim, F. and J. L. Rosenberger. 1993. Experimental Designs for Model Discrimination. *Journal of the American Statistical Association* 88: 642-649.

Srivastava, J. N. 1996. A Critique of Some Aspects of Experimental Design. *Handbook of Statistics* 13: 309-342. eds. S. Ghosh and C. R. Rao.

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

AUTHOR BIOGRAPHIES

THEODORE ALLEN is an Assistant Professor of Industrial & Systems Engineering at The Ohio State University. He received his Ph.D. from the University of Michigan (Ann Arbor) in 1997. His contact information is: <allen.515@osu.edu> <www-iwse.eng.ohio-state.edu/~facultyp/allen.htm>.

LIYANG YU received his Ph.D. in Industrial & Systems Engineering at The Ohio State University in 2000. He is currently an integration applications engineer at i2 Technologies. His research interests include experimental design, quality engineering, mathematical programming, and simulation. His email is <Liyang_Yu@i2.com>.