Proceedings of the 2003 Winter Simulation Conference
S. Chick, P. J. Sánchez, D. Ferrin, and D. J. Morrice, eds.

OPTIMAL EXPERIMENTAL DESIGN FOR SYSTEMS INVOLVING BOTH QUANTITATIVE AND QUALITATIVE FACTORS

Navara Chantarat
Ning Zheng
Theodore T. Allen
Deng Huang

The Department of Industrial, Welding, and Systems Engineering
The Ohio State University
Columbus, OH 43210, U.S.A.

ABSTRACT

Often in discrete-event simulation, factors being considered are qualitative such as machine type, production method, job release policy, and factory layout type. It is also often of interest to create a Response Surface (RS) metamodel for visualization of input-output relationships. Several methods have been proposed in the literature for RS metamodeling with qualitative factors but the resulting metamodels may be expected to predict poorly because of sensitivity to misspecification or bias. This paper proposes the use of the Expected Integrated Mean Squared Error (EIMSE) criterion to construct alternative optimal experimental designs. This approach explicitly takes bias into account. We use a discrete-event simulation example from the literature, coded in ARENA™, to illustrate the proposed method and to compare metamodeling accuracy of alternative approaches computationally.

1 INTRODUCTION

Many real world systems of interest to practitioners are too complicated to be modeled analytically. Also gathering real data from the systems can be too expensive to support thorough optimization. Discrete-event simulation is widely considered as useful for studying behavior of those complex systems. Simulations enable practitioners to better understanding of the expected performance of actual systems. For example, in this paper we will focus on a simulation model of a manufacturing system, derived from Seila, Ceric, and Tadikamalla (2001, p. 347), involving optimization of three quantitative inputs and one qualitative input.

Simulation models of those real world systems may themselves be complex and expensive to construct. Therefore, simple mathematical models that approximate the outputs from simulation models are often constructed to clarify the system input-output relationships. These *metamodels* or *surrogate* models, which can be polynomial forms fitted using least squares regression, can help engineers to make system design decisions (see, e.g., Kleijnen, 1987, for a general reference). Figure 1 shows the relationship between the real system, the computer simulation model, and the mathematical metamodel. The metamodel, simulation model, and real physical system all have the same quantitative inputs or "factors", $x_1,\ldots,x_k$ and qualitative inputs or "factors", $z_1,\ldots,z_r$. The outputs associated with each may be different, $y_1,\ldots, y_n$, for the metamodel $y_{1,S},\ldots, y_{n,S}$, for the simulation model, and $y_{1,R},\ldots, y_{n,R}$, for the real system. The metamodel is expected to give the worst prediction accuracy in terms of an ability to predict the real system input-output relationships. Yet, because metamodels are the cheapest in terms of deriving outputs for a given set of inputs, they are widely used. For example, one can quickly obtain the approximately 100 outputs needed to create a contour plot using a metamodel.

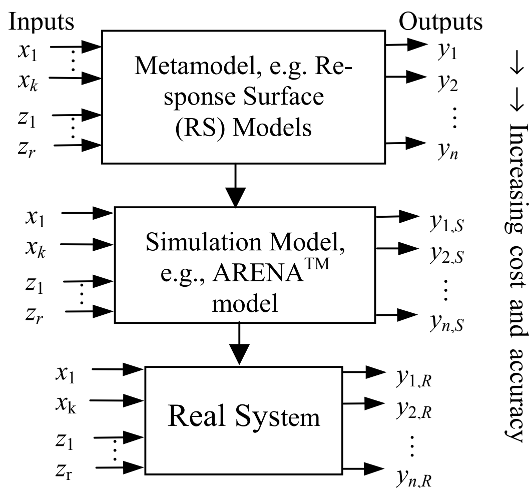

Figure 1: Relationships in Simulation Metamodeling

The present paper offers a new "design of experiments" (DOE) approach for selecting which input combinations of the simulation models to test. The goal is to minimize the prediction errors of the final metamodel generated by a providing a specific experimentation and curve fitting process. In our case, we minimize only the errors of the metamodel related to outputs from the simulated system, i.e., the $y_{i,S}$ and not necessarily errors associated with predicting real system outputs (the $y_{i,R}$).

Barton (1994) divides issues related to constructing metamodels into those that relate to the choice of functional form (e.g., the terms in the fitted model), design of experiments, and assessment of adequacy of metamodels. This paper emphasizes the issues pertaining the choice of functional form and design of experiments. So-called "bias errors" result from a difference between the fitted metamodel forms and the true, usually unknown model form. The main motivation of the new methods is to improve metamodel prediction errors in the context of common situations in which bias errors are a concern because the true model form is unknown. Further, in the context of a real problem from the literature, we will show that the proposed approaches derive more accurate metamodels than approaches in Draper and John (1988) and Wu and Ding (1998).

Section 2 reviews selected response surface metamodel forms from the literature for situations involving both quantitative and qualitative factors. Section 3 reviews the Expected Integrated Mean Squared Errors (EIMSE) criterion and discusses its extension to situations involving qualitative factors. Section 4 describes examples of available experimental designs from the existing literature. A new class of EIMSE-optimal response surface designs for simulation metamodeling of systems involving qualitative factors to minimize prediction errors is then proposed in Sections 5. Section 6 describes a hypothetical manufacturing system being modeled using ARENA^{TM} simulation package and implements the proposed design for metamodeling of the ARENA^{TM} simulation model. Section 7 provides comparison results of expected prediction errors from the proposed methods and alternative methods from the literature. Section 8 summarizes the contributions.

2 RESPONSE SURFACE METAMODELING

In this section, we review selected response surface models from the literature. In later sections, we will refer to these models in the context of the proposed methods and alternatives. So-called "response surfaces" are based on either second-order or third-order polynomials fitted using least squares regression. A second-order model is as shown in equation (1). A third-order model contains additional third-order terms.

$$E(y) = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i=1}^{k} \sum_{j=1}^{k} \beta_{ij} x_i x_j \quad (1)$$

The β's are parameters, x_i's are inputs, and y is output or "response" of interest. The above model form of equation (1) only includes quantitative inputs. When there are some input variables z_i's that are qualitative, more general model forms are relevant.

Draper and John (1988) proposed the model forms shown in equation (2) and (3) that include both quantitative and qualitative input variables. The function, f, is a polynomial of second-order and with z as the subscript that refers to a combination of the qualitative factors, $z_1,\ldots,z_r$. Therefore, $z = 1,\ldots,m$ where m is the number of qualitative factor combinations. Then, W_z for $z = 1, \ldots, m$ represent the selected levels of m dummy variables, chosen so that all possible combinations of qualitative variables are potentially distinguished, see, e.g., Draper and John (1988).

$$E(y) = \sum_{z=1}^{m} W_z f_z(x, \beta_z) \quad (2)$$

$$f_z(x, \beta_z) = \beta_{0z} + \sum_{i=1}^{k} \beta_{iz} x_i + \sum_{\substack{i=1 \\ j \geq i}}^{k} \sum_{j=1}^{k} \beta_{ijz} x_i x_j \quad (3)$$

Wu and Ding (1998) proposed a more concise and restrictive model form shown in equation (4).

$$E(y) = \sum_{z=1}^{m} W_z \left(\beta_{0z} + \sum_{i=1}^{k} \beta_{iz} x_i \right) + \sum_{i,j=1}^{k} \beta_{ij} x_i x_j \quad (4)$$

Where m is, again, the number of combinations of possible qualitative factor settings. Also, W_j is 1, when y is taken at level j of the variable z and 0 otherwise. The coefficients, β_{0z}, is the constant term and β_{1z} is the slope of x_i, both depending on the choice of z. Note that, unlike the model form proposed in Draper and John (1988), the model form proposed in Wu and Ding (1998) does not allow tailored interaction term $\sum_{z=1}^{m} W_z \left(\sum_{i,j=1}^{k} \beta_{ij} x_i x_j \right)$ for all combinations of qualitative factors. Wu and Ding (1998) also showed many examples of models of the type in equation (4) argued to be relevant when the run size is small.

The above notation is useful for displaying concisely assumptions. In the practice of fitting models to data, however, it is common to use the notation scheme in Myers and Montgomery (1995) which separates out the different effect of qualitative factors. In this approach, for the l levels of each qualitative factor, one uses $l - 1$ indicator variables or "contrasts". The i^{th} factor has its j^{th} contrast value equal to zero if the corresponding run is at the j^{th} level of the qualitative factor. Otherwise, the contrast value is zero. For the example in Table 1 below, there are two qualitative factors, z_1 and z_2. Each has two levels. Therefore, both have $2 - 1 = 1$ indicator variables associated. These are represented by 1 when the run is at level 1 and 0 otherwise. The total of four level combinations are then represented by the coding scheme shown in Table 1.

Table 1: Example Combinations (z) and Contrasts

Levels of Qualitative Factors	z	Contrast #1 (z_1)	Contrast #2 (z_2)
1st level of z_1, 1st level of z_2	1	1	1
1st level of z_1, 2nd level of z_2	2	1	0
2nd level of z_1, 1st level of z_2	3	0	1
2nd level of z_1, 2nd level of z_2	4	0	0

In the new notation,

$$E(y) = \beta_0 + \sum_i \beta_i x_i + \sum_{i \le j} \beta_{ij} x_i x_j + \sum_i \gamma_i z_i + \sum_i \sum_j \delta_{ij} x_i z_j + \sum_l \sum_{i \le j} \sum \rho_{ijl} x_i x_j z_l \quad (5)$$

is equivalent to (2) and (3).

In this paper, we focus on a problem involving 3 quantitative and 1 qualitative factor. Therefore, there is only 1 contrast z for the factor z1. Then, equation (4) in the notation of Myers and Montgomery (1995) can be written:

$$y = \beta_0 + \sum_{i=1}^{3} \beta_i x_i + \sum_{i=1}^{3} \beta_{ii} x_i^2 + \sum_{i \le j} \beta_{ij} x_i x_j + \gamma z + \sum_{i=1}^{3} \delta_i x_i z + \varepsilon. \quad (6)$$

To accommodate error due to potential model inadequacy, we assume equation (5) describes the true model form of the system of interest. We use these assumed true models and the fitted models to generate new classes of designs based on the "expected integrated mean squared error" (EIMSE) criterion as defined in the next section.

3 EXPECTED INTEGRATED MEAN SQUARED ERRORS (EIMSE) CRITERION

In this section, we review the expected integrated mean square error (EIMSE) criterion proposed in Allen, Yu, and Schmitz (2003) and adapt it to the context of the system involving both qualitative and quantitative input variables. Let n be the number of runs of the experimental plan, **D**, being evaluated. Assume that the responses from experiments, **y**, derive from the following model:

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}, \quad (7)$$

where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are k_1 and k_2 dimensional vectors of coefficients, $\mathbf{X}_1$ and $\mathbf{X}_2$ are $n \times k_1$ and $n \times k_2$ design matrices respectively, and $\boldsymbol{\varepsilon}$ is a n vector of experimental random errors with standard deviation σ. Define $\mathbf{f}_1(\mathbf{x})$ and $\mathbf{f}_2(\mathbf{x})$ as possible rows of the design matrix $\mathbf{X}_1$ and $\mathbf{X}_2$ corresponding to the point $\mathbf{x} = \{x_1, \ldots, x_q, z_1, \ldots, z_{m-q}\}$ Then, the model to be used to make predictions at the point **x** after the experiment is:

$$\hat{y}(\mathbf{x}) = \mathbf{f}_1(\mathbf{x})\hat{\boldsymbol{\beta}} \quad \text{where} \quad \hat{\boldsymbol{\beta}} = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y}. \quad (8)$$

This assumes that, after the model fitting, the experimenter will use an unedited model to make predictions. With these assumptions, the EIMSE criterion is:

$$\text{EIMSE}(\mathbf{D}) = \underset{\mathbf{x},\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\varepsilon}}{E}\left\{\left[\hat{y}(\mathbf{x},\mathbf{D},\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\varepsilon}) - y(\mathbf{x},\boldsymbol{\beta}_1,\boldsymbol{\beta}_2)\right]^2\right\} = \sigma^2 \text{Tr}[\boldsymbol{\mu}_{11}(\mathbf{X}_1'\mathbf{X}_1)^{-1}] + \text{Tr}[\mathbf{K}_2\Delta] \quad (9)$$

where

$$\Delta = \mathbf{A}'\boldsymbol{\mu}_{11}\mathbf{A} - \boldsymbol{\mu}_{12}'\mathbf{A} - \mathbf{A}'\boldsymbol{\mu}_{12} + \boldsymbol{\mu}_{22} \text{ and } \mathbf{A} = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2 \quad (10)$$

and where $\mathbf{K}_2$ is the assumed prior covariance matrix, $E[\boldsymbol{\beta}_2'\boldsymbol{\beta}_2]$. In the example problem that follows, we use the DuMouchel and Jones (1994) assumptions scheme with the parameter γ adjusted so that the two terms in (9) are roughly equal for the recommended design. The motivation for this approach is described in Allen, Yu, and Schmitz (2003). Also, the so-called moment matrices are defined using:

$$\boldsymbol{\mu}_{ij} = \int_R \mathbf{f}_i(\mathbf{x})\mathbf{f}_j(\mathbf{x})'\rho(\mathbf{x})\,d\mathbf{x} \quad \text{for } i,j = 1,2 \text{ and } i \le j \quad (11)$$

where R is the region of interest and ρ is a weighting function describing the distribution of points where predictions will be expected after experimentation and model fitting. A standard assumption is that $\rho(\mathbf{x}) = 1/V$ with V being the volume of the region of interest.

Note that the EIMSE value does not depend upon the distribution of the true values, $\boldsymbol{\beta}_1$, of any fitted parameter. The EIMSE defined by equations (9), (10), and (11) is an extension of the Box and Draper (1959) integrated mean squared error criterion because there is an additional integration over the distribution of true model coefficients. Allen, Bernshteyn, and Kabiri (in press) show that EIMSE optimal designs produce relatively low prediction errors in the context of computer experiment case studies from the literature. Also, the EIMSE has already been used to achieve useful engineering results as described in Allen, Yu, and Schmitz (2003), Allen, Yu, and Bernshteyn (2000), and Koc, Allen, Jirathearanat, and Altan (2000).

To address the computational challenge associated with the set of qualitative input variables, the concept of optimization over a candidate set will be used. The candidate set is expressible by the matrix, **C**, of feasible points such that the experimental design optimization problem becomes the selection of n choices from this set. Let $\mathbf{C}_j$ refer to the j^{th} row of **C**. In the context of the EIMSE formulation, this approach results in the integer program:

$$\underset{I_1,\ldots,I_n}{\textit{Minimize}}\ \text{EIMSE}(\mathbf{D})$$

$$\text{Subject to: } \mathbf{D} = \begin{bmatrix} \mathbf{C}_{I_1} \\ \vdots \\ \mathbf{C}_{I_n} \end{bmatrix} \text{ where } I_1, \ldots, I_n \subset \{1,\ldots,N\}. \quad (12)$$

The candidate set suggested by Draper and John (1988) and Wu and Ding (1998) included only factorial points, center points, and star points of the region defined by the cuboidal region of quantitative factors and discrete region of qualitative factors. Motivated by the capabilities of modern computers, we propose to use a relatively exhaustive set of points. One easy-to-implement way to do this is to generate N uniformly distributed random samples in the region of interest, where N is typically substantially larger than the number of points used by Draper and John (1988) and Wu and Ding (1998). By implementing our proposed method, we can also create the designs in spherical region. To generate these samples in spherical region and store them in **C**, we use the algorithm described in Fishman (1996).

4 ALTERNATIVE DESIGN METHODS

Standard response surface designs such as Box Behnken and central composite designs (see, e.g., Box and Draper, 1987 and Khuri and Cornell, 1996) are generally considered noncompetitive in the context of experimentation with both qualitative and quantitative factors. Reasons for this include the fact that these designs have the same number of levels for all factors which is often not desired when qualitative factors are involved. For example, the Box Behnken design has 3 levels for all factors, but the experimenter might be interested in exploring a qualitative factor with two or four levels. Further, designs such as Box Behnken would lead to infinite prediction errors if models of the form in equation (4) were fitted with one or more of the factors being qualitative. This follows because the associated information matrices are singular. Repeating the entire design for all combinations of qualitative factors generally leads to prohibitive experimental cost.

Draper and John (1988) were apparently the first to investigate experimental design with quantitative and qualitative factors. They began by focusing on two specific problems and building on the ordinary central composite design structure. They examined central composite designs in a given number of quantitative factors and focused on assignments of the runs to given combinations of qualitative factor levels that preserved the following property. First order models in all quantitative factors could be fitted (are "estimable") using only data from any given combination of qualitative factors. An example involving one qualitative factor at two levels is shown in Table 2 and Figure 1 (from Draper and John, 1988, Figure 8b). The model fitted by those authors is given by equation (6). Note that the last columns in Table 2, Table 3, and Table 4 give response values achieved which are the average of results for 40 replications using the ARENA™ software for the case study described in Section 6.

Table 2: Response Surface Design Proposed in Draper and John (1988) for r = 1 Qualitative Input Variable and k = 3 Quantitative Input Variables

Run	x_1	x_2	x_3	z	*Cost* (**y**)
1	-1	-1	-1	0	7854.4
2	1	-1	-1	0	14170.0
3	-1	1	-1	0	9528.7
4	-1	-1	1	0	8953.0
5	1	1	-1	1	14863.3
6	1	-1	1	1	14846.9
7	-1	1	1	1	9409.9
8	1	1	1	1	14088.2
9	-1.682	0	0	0	6995.0
10	1.682	0	0	0	14706.6
11	0	-1.682	0	0	207086.2
12	0	1.682	0	1	11898.8
13	0	0	-1.682	1	50818.9
14	0	0	1.682	1	12832.4
15	0	0	0	0	11418.0
16	0	0	0	1	11486.3

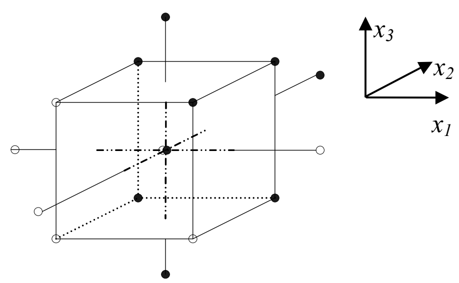

Figure 1: Response Surface Design Proposed in Draper and John (1988) for r = 1 Qualitative Input Variable and k = 3 Quantitative Input Variables with ◦ for z = 0 and • for z = 1

Wu and Ding (1998) proposed an alternative set of designs derived from optimization of the so-called D-optimality objective. More specifically, their proposed designs were constructed to address four prioritized objectives regarding types of models to be fitted and estimability of model terms.

Their most important objective is that the overall design must be efficient for a model that is second order in quantitative factors and has main effects of qualitative factors and interactions between quantitative and qualitative factors. The second most important objective is that at each combination or each level of qualitative factor, the design is an efficient first-order design in quantitative factors. This objective ensures that the first-order effects of quantitative factors that may vary with the levels of qualitative variables can be estimated.

To meet these first two objectives the design points are partitioned into several groups. Each group is associates with a level combination of qualitative factors. Since there can be several partitioning choices, an "optimal" design is the one that best satisfies the objectives mentioned previously.

To be standard in relation to common practice, they built their designs in the quantitative factors on central composite designs. In general, their proposed design consists of $2^{k-p} + 2 + 2k$ runs. The first 2^{k-p} runs were constructed from a high resolution 2-level fractional factorial. Then two center points were added. Finally, $2k$ star points, whose distance from the origin is α, were added. The value of α was chosen so that the design is rotatable.

To address the abovementioned criteria, Wu and Ding (1998) proposed careful assignment of the runs in the central composite design to the combinations of qualitative factor levels. This step required a formal optimization search over possible set of different combinations of levels so that the design objectives are met. Table 3 and Figure 2 shows an example of the design proposed by Wu and Ding (1998) as shown as design no. 1 in their paper, for the case where $r = 1$ qualitative input variable and $k = 3$ quantitative input variables.

Table 3: Response Surface Design Proposed in Wu and Ding (1998) for $r = 1$ Qualitative Input Variable and $k = 3$ Quantitative Input Variables

Run	x_1	x_2	x_3	z	*Cost* (**y**)
1	1	1	1	1	14088.24
2	1	1	-1	1	14863.26
3	1	-1	1	0	12851.63
4	1	-1	-1	1	13720.71
5	-1	1	1	1	9409.86
6	-1	1	-1	0	9528.66
7	-1	-1	1	0	8952.99
8	-1	-1	-1	0	7854.41
9	0	0	0	1	11486.26
10	0	0	0	0	11417.99
11	1.682	0	0	0	14959.76
12	-1.682	0	0	1	8415.81
13	0	1.682	0	0	10508.72
14	0	-1.682	0	1	219424.01
15	0	0	1.682	0	10831.18
16	0	0	-1.682	0	52866.17

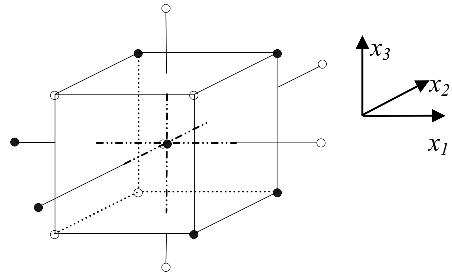

Figure 2: Response Surface Design Proposed in Wu and Ding (1998) for $r = 1$ Qualitative Input Variable and $k = 3$ Quantitative Input Variables with ◦ For $z = 0$ and • For $z = 1$

5 A NEW CLASS OF EIMSE-OPTIMAL DESIGNS

In this section, we propose response surface designs which come directly from the optimization formulation in equation (12) involving both qualitative and quantitative factor. We define the $N \times m$ matrix using $\widetilde{\mathbf{C}} = (\mathbf{C}|\mathbf{C}_2)$ where the matrix $\mathbf{C}$ is the candidate set for the quantitative factors and $\mathbf{C}_2$ is a set of uniform points from the region of interest of the r qualitative factors. The default assumption that we used to generate the EIMSE-optimal design in Table 4 is that the region of interest is a sphere with radius 1.628 to facilitate comparison with Wu and Ding (1998). Therefore, the generated designs minimize the expected squared errors of prediction through combining of candidate points that are themselves uniform samples from the region of interest.

In general, any combination of assumed true and fitted models could be used together with the assumptions about the coefficient distributions. In this paper, the assumed true model is of the general form in equations (2) and (3) and the assumed fitted model is of limited second-order form in equations (4) and (5).

To generate all designs, we used the algorithm described in Hadj-Alouane and Bean (1997). Prior to each design generation, we generated N = 10,000 candidate points. The proposed EIMSE-optimal design for the case where $r = 1$ qualitative input variable and $k = 3$ quantitative input variables is shown in Table 4 and Figure 3.

The EIMSE-optimal design in Table 4 minimizes the expected prediction error for the fitted model form assuming the true model form given that only 16 simulation runs are to be conducted for metamodeling. The EIMSE value of the design is 2.35, which practically implies that the predictions for the mean response values, $\eta(x)$, can be expected to have standard errors of approximately sqrt(2.35) times the standard deviation of the noise. The EIMSE val-

Table 4: EIMSE-Optimal Response Surface Design

Run	x_1	x_2	x_3	z	*Cost* (**y**)
1	-1.41	0.455	-0.022	0	8023.9
2	-0.949	-1.281	-0.012	0	9855.3
3	-0.176	1.355	-0.558	0	10072.3
4	1.089	0.379	-0.631	1	13608.0
5	0.592	0.293	1.424	1	14108.3
6	-0.141	-0.271	-0.927	0	10765.2
7	-0.129	0.324	0.644	0	11181.0
8	0.644	-1.265	0.713	0	13815.7
9	0.224	1.32	0.987	1	13064.6
10	1.363	0.748	0.168	0	15713.9
11	0.795	-1.235	-0.275	1	13102.4
12	-0.882	0.371	0.974	1	8537.7
13	-0.168	-0.041	-1.601	1	20631.6
14	-0.473	0.524	-0.739	1	9462.6
15	-0.084	-1.127	1.014	1	10699.7
16	-0.884	-1.156	-0.697	1	10696.5

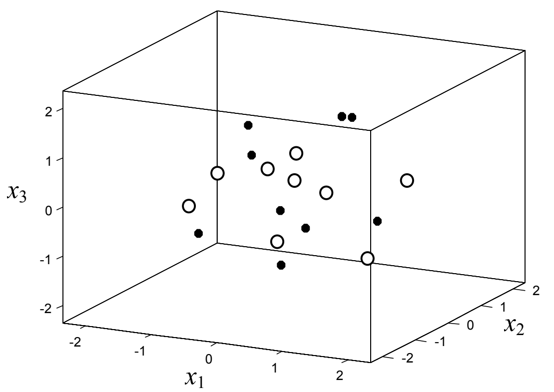

Figure 3: EIMSE-Optimal Design

ues of other alternative designs including the designs proposed in Draper and John (1988) and Wu and Ding (1998) are presented in Section 7.

The proposed method offers relatively great flexibility in number of runs depending on the required prediction accuracy. Obviously, more runs result in improved expected prediction accuracy as well as higher costs of the metamodeling process. Using the proposed design method, the experimenter can thoroughly investigate and trade off the prediction accuracy goal and cost budget of experimentation in advance before actual metamodeling experimentations begin.

Also the user of our proposed experimental design procedure has relatively higher flexibility in the number of levels for each factor involved in metamodeling. Since the candidate set is used for construction of the EIMSE-optimal design, each factor can include any number of distinct levels in the set of candidate points from which the optimization procedure selects the recommended design points.

6 CASE STUDY: COMPUTER SIMULATION OF MANUFACTURING SYSTEM

Figure 4 shows a manufacturing system, which is based on a problem presented in Seila, et al. (2001, page 347). An ARENA^{TM} simulation model was constructed and used to generate the response data for each of the relevant metamodeling approaches.

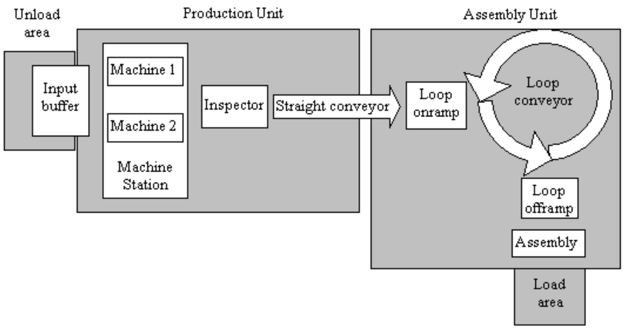

Figure 4: Manufacturing System Example

In the simulated system, there are two machines. When a part arrives in the system, it is kept in a queue until the corresponding machine is available. Two manual carts carry the parts to an inspection station after they are processed at the machine station. After being inspected, some parts that do not pass the inspection go back into a queue and are reworked. The proportion of parts that are reworked is 20%.

If a part passes the inspection, it is put into a buffer to wait for accessing an accumulating straight-line conveyor (with accumulating length 0.8 inches) and then access a non-accumulating loop conveyor at the loop onramp. At the loop offramp, there is another buffer, which can contain four parts. If the buffer is full, the part continues to be conveyed on the loop conveyor until some time the buffer is not full when it come back to the loop offramp again. After going through the assembly station, the part will leave the manufacturing system. The assembly station is maintained using preventive maintenance scheduled at between 2 to 6 hours to try to prevent it from failing. However, if the assembly station failed, it will be repaired and time for preventive maintenance schedule is restarted at time 0.

The cell sizes of both conveyors are 1 foot by 1 foot. In the original system, the velocity of the straight conveyor can be set between 4 to 12 ft/min, and the velocity of the loop conveyor can be set between 3 to 10 ft/min. The length of the straight conveyor is 15 ft. For the loop conveyor, the length from loop onramp to loop offramp is 20 ft and that from loop offramp to loop onramp is 35 ft. The distance from the machine station to the inspection station is 20 ft and the speed of the carts is 25ft/min. Parts enter the system in batches. Batch size contains 4, 5, or 6 parts with probability 0.4, 0.2, 0.4, respectively. The inter-arrival times of the batches are *i.i.d.* exponentially distributed with mean 7.3 minutes. The processing time in the machine is Uniform(30, 72) minutes, the inspection time is Triangular(0.5, 1, 1) minutes, and there are two workers working at the assembly station with assembly time that is *i.i.d.* distributed with distribution Triangular(1.5,1.8,2.5) minutes.

Seila, et al. (2001, page 364) did not include maintenance and queuing policies in their formulation. Therefore, in addition to cost information from their specifications, we assumed that each maintenance will cost $20 (excluding the idle cost of the machine). Also, each failure will cost $100. The system will run 24 hours per day. For everyday, we assume the cost of straight conveyor per day is $(rate * 10) and that of loop conveyor is $(rate * 12). The main output response is the cost of running the system. There are four input variables that are thought to have effects on those responses of interest. The input variables as well as their ranges are shown in Table 5.

Table 5: Input Variables or Factors for Simulation Model of a Manufacturing System Example

Factors	Name	Range
x_1	Straight-line conveyor speed (ft/min)	4 -12
x_2	Loop conveyor speed (ft/min)	3 - 10
x_3	Maintenance schedule (hours)	2 - 6
z	Queuing Policy	0: Both machines share the same queue; 1: Different machines have independent queues, parts will select the shortest queue

7 COMPARISONS OF ALTERNATIVE DESIGNS

In this section, we compare the proposed designs with alternative designs from the literature including the designs from Draper and John (1988) and Wu and Ding (1998). We describe the application of each of these alternatives and the proposed method to create metamodels for the manufacturing problem in Section 6. We compare both the expected prediction errors before experimentation using the EIMSE from Section 3 and a measure of the actual prediction errors from the case study. We assume that all methods involve 16 runs and involve fitting the model in equations (4) and (5) in keeping with examples in those papers.

There were two alternative designs with similar origins usable in our case study which Draper and John (1988) provided in their Figure 8b and 8c. Also, Wu and Ding (1998) provided three relevant alternative designs in their Table 4. Some of those designs were shown in Section 4 of this paper. Table 6 shows the EIMSE values or expected prediction errors for alternative response surface designs.

Table 6: EIMSE Values for Alternative 16-Run Response Surface Designs

Designs		EIMSE values
Draper and John (1988)	8b.	128.37 (IV = 76.08)
	8c.	4.75 (IV = 3.14)
Wu and Ding (1998)	4.1.1	3.48 (IV = 1.05)
	4.1.2	3.22 (IV = 1.10)
	4.1.3	6.89 (IV = 2.01)
EIMSE-optimal		2.35 (IV = 1.18)

The results in Table 6 confirms that putatively EIMSE-optimal designs can be expected to generate substantially more accurate fitted models than the relevant design alternatives. For example, the standard errors of the response predictions for this problem are expected to be sqrt(128.37/2.35) times greater for Draper and John (1988) 8b design than for our proposed design. The Integrated Squared Error (ISE) calculations were conducted as empirical validation of the claim.

Table 7 shows the results from Integrated Squared Error (ISE) calculations. First, six metamodels in the form of equation (6) were fitted using actual ARENA™ simulation response data from each design in Table 6. Then, fifty random experimental points were generated uniformly in the spherical region of interest. ARENA™ model was run again to gather actual simulation results for those points. Next, all random design points were fitted to each metamodel to gather predictions. The errors were calculated using the different between predictions and actual ARENA™ simulation results.

Table 7: Sqrt(ISE) Values for Alternative 16-Run Response Surface Designs

Designs		Sqrt(ISE) values
Draper and John (1988)	8b.	367,406.9
	8c.	43,737.5
Wu and Ding (1998)	4.1.1	62,294.5
	4.1.2	64,956.8
	4.1.3	56,736.3
EIMSE-optimal		2,811.5

Formally, the Integrated Squared Error (ISE) formula can be written:

$$\text{ISE} = \left(\frac{N}{\sigma^2}\right)\int_o w(x)E\left\{\hat{y}(x)-\eta(x)\right\}^2 d(x) \qquad (13)$$

where $\eta(x)$ includes the random experimental errors, in our case from using only 40 replicates. Since this error includes both contributions from bias, random errors during experimentation, and random errors during testing of the model, it can be expected to be higher in practice than the EIMSE which includes only the former two influences as described in Section 3.

A major finding is that the EIMSE-optimal design is expected to achieve far lower prediction errors before the experiment as calculated using the EIMSE. Also, in our case study, we verified subsequent to experimentation and model fitting that these relatively low prediction errors were achieved.

It is perhaps interesting to note that the estimated standard error from the regression involving the EIMSE design was sqrt(MSE) = 1,136. Therefore, scaling the EIMSE, we would typically expect prediction errors approximately equal to sqrt(2.35)×(1,136) ~ 2,000 which is roughly what was observed. This provides some measure of confirmation that the EIMSE is a relevant measure of prediction accuracy of potential interest before experimentation begins.

A final consideration relates to model editing after data has been collected, i.e., decisions about dropping specific terms from the full second order model. In editing, which may be regarded as optional, orthogonality of columns in the design matrix, $\mathbf{X}_1$, aids in the analysis because estimates of one coefficient do not depend on which other coefficients are included in the model. In this context, designs such as those proposed by Draper and John (1988) and Wu and Ding (1998) have an advantage because more of their associated $\mathbf{X}_1$ matrix columns are orthogonal.

8 CONCLUSIONS

In this paper, we have presented new classes of response surface designs for situations involving both qualitative and quantitative factors, generated using the EIMSE criterion. We also described computational issues related to adapting that criterion to this context. We demonstrated that designs of various sizes that were generated to minimize the EIMSE were able to achieve substantially lower prediction errors than alternatives in the context of an examples and a fitted model from the literature.

We suggest that EIMSE-optimal response surface design methodology constitute a potentially important tool for designing response surface experiments in situation involving both qualitative and quantitative factors. Commercially available software for the planning and analysis based on the proposed methods is available through contacting the authors.

ACKNOWLEDGMENTS

The authors would like to thank Dr. William Notz, Dr. Angela Dean, and Dr. Steven Gilmour for their useful discussion. Also we would like to thank Deng Huang for helping with the computer programming codes utilized in this paper.

REFERENCES

Allen, T. T., Bernshteyn, M., Yu, L., and Kabiri, K. (in press). A Comparison of Alternative Methods for Constructing Meta-Models for Computer Experiments. *Journal of Quality Technology*.

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

Allen, T. T., Yu, L., and Schmitz, J. 2003. An Experimental Design Criterion for Minimizing Meta-Model Prediction Errors Applied to Die Casting Machine Design. *Journal of the Royal Statistical Society Series C: Applied Statistics*. 52(1). 1-15.

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

Box, G. E. P. and Draper, N. R. 1987. *Expirical Model-Building and Response Surfaces*. John Wiley, NY.

Barton, R. R. 1994. Metamodeling: A State Of The Art Review. In *Proceedings of the 1994 Winter Simulation Conference* ed. J. D. Tew, S. Manivannan, D. A. Sadowski, and A. F. Seila. 237-244. Lake Buena Vista, FL, USA.

Draper, N. R. and John, J. A. 1988. Response-Surface Designs for Quantitative and Qualitative Factors. *Technometrics.* 30(4). 423-428.

DuMouchel, W. and Jones, B. 1994. A Simple Bayesian Modification of D-Optimal Designs to Reduce Dependence on a Assumed Model. *Technometrics*. 36(1). 37-47.

Fishman, G. S. 1996. *Monte Carlo: Concepts, Algorithms, and Applications*. Springer-Verlag.

Hadj-Alouane, A. and Bean, J. 1997. A Genetic Algorithm for the Multiple-Choice Integer Program. *Operations Research*. 45. 92-101.

Khuri, A. I. And Cornell, J. A. 1996. Response Surfaces Designs and Analyses (2nd *edition).* Marcel Dekker, Inc. NY.

Kleijnen, J. P. C. 1987. *Statistical tools for simulation practitioners.* New York: Marcel Dekker.

Kleijnen, J. P. C., and R. G. Sargent. 2000. A methodology for fitting and validating metamodels in simulation. *European Journal of Operational Research* 120: 14-29.

Koc M., Allen T. T., Jirathearanat S., and Altan, T. T. 2000. The use of FEM and experimental design to investigate tube hydroforming of a simple geometry. *The International Journal of Machine Tools and Manufacture*. 40. 2249-2266.

Law, A. M., and Kelton, W. D. 2000. *Simulation modeling and analysis*, 3rd ed. New York: McGraw-Hill.

Montgomery, D. C. 1997. *The design and analysis of experiments,* 4th Ed. New York: John Wiley and Sons.
Myers, R. H. and Montgomery, D. C. 1995. Response Surface Methodology. John Wiley. NY.
Sargent, G. R. 1991. Reseach Issues In Metamodeling. In *Proceedings of the 1991 Winter Simulation Conference* ed. B. L. Nelson, W. D. Kelton, and G. M. Clark. 888-893. Phoenix, AZ, USA.
Seila, A., Ceric, V., and Tadikamalla, P. 2001. *Applied Simulation Modeling*. Thomson Learning. 662 pp.
Wu, C. F. J. and Ding, Y. 1998. Construction of Response Surface Designs for Qualitative and Quantitative Factors. *Journal of Statistical Planning and Inference.* 71. 331-348.

AUTHOR BIOGRAPHIES

NAVARA CHANTARAT is a Ph.D. candidate in the Department of Industrial, Welding, and Systems Engineering at the Ohio State University. He has a B.S. degree in Agro-Industrial Product Development from Kasetsart University, Thailand and M.S. in Operations Research from The Ohio State University. His research interests include applications of statistical and simulation methods to system design and to product design, manufacturing and delivery. His email address is <chantarat.1@osu.edu>.

NING ZHENG is a Master's degree student in the Department of Industrial, Welding, and Systems Engineering at the Ohio State University. He has a B.S. degree in Mechanical Engineering from Tsinghua University, China. His research interests include applications of statistical and simulation methods to system design and to product design, and manufacturing. His email address is <zheng.481@osu.edu>.

THEODORE T. ALLEN is an Assistant Professor in the Department of Industrial, Welding, and Systems Engineering at the Ohio State University. He has B.S. from Princeton University, M.S. from UCLA, and Ph.D. degree in Industrial Engineering from University of Michigan, Ann Arbor. His research interests include applications of statistical and simulation methods to system design and to product design, experimental design, and decision support. His email address is <allen.515@osu.edu>.

DENG HUANG is a Ph.D. candidate in the Department of Industrial, Welding, and Systems Engineering at the Ohio State University. He has a B.E. degree in Material Science from Tsinghua University, China and an M.S. in Material Science from the Ohio State University. His research interests include utilizing variable fidelity data in response surface building and efficient global optimization. He is also a research scientist at Scientific Forming Technologies. His email address is <huang.262@osu.edu>.