An Expository Paper on Optimal Design

Article *in* Quality Engineering · July 2011

DOI: 10.1080/08982112.2011.576203

3 authors, including:

Bradley Jones
SAS Institute
94 PUBLICATIONS **1,603** CITATIONS

Quality Engineering, 23:287–301, 2011

ISSN: 0898-2112 print/1532-4222 online
DOI: 10.1080/08982112.2011.576203

Taylor & Francis
Taylor & Francis Group

An Expository Paper on Optimal Design

Rachel T. Johnson[1], Douglas C. Montgomery[2], Bradley A. Jones[3]

[1]Naval Postgraduate School, Monterey, California
[2]Arizona State University, Tempe, Arizona
[3]SAS Institute, Cary, North Carolina

ABSTRACT There are many situations in which the requirements of a standard experimental design do not fit the research requirements of the problem. Three such situations occur when the problem requires unusual resource restrictions, when there are constraints on the design region, and when a nonstandard model is expected to be required to adequately explain the response. This article provides an introduction to optimal design for these types of situations. Optimal designs are computer-generated experiments that are aimed at satisfying specific research problem requirements. We show that the optimal design approach is applicable to any design problem and necessary when there are situations involving resource constraints or nonstandard design regions or models. The mathematical formulations of several design optimality criteria are presented along with examples of optimal design applications.

KEYWORDS *A*-optimality, constrained region of experimentation, *D*-optimality, design of experiments, factorial experiments, *G*-optimality, *I*-optimality, response surfaces

INTRODUCTION

The use of designed experiments has grown dramatically over the last 20 years. In addition to the traditional applications of manufacturing process development and product/process quality improvement, applications in engineering design, marketing, service industries, finance, and e-commerce have grown rapidly. There are several reasons for this, including greater awareness on the part of engineers, scientists, and other quantitative analysts of the power of designed experiments, better education for these individuals at the university level, the adoption of widespread business improvement strategies such as Six Sigma, and the increasing availability of computer software that supports application of experimental design.

Most practitioners are exposed to designed experiments through a course, either at a university as part of their academic training or increasingly through a short course or company-sponsored program such as Six Sigma green belt or black belt training. These courses usually emphasize factorial designs and fractional factorial designs (often the 2^{k-p} regular fractions), discuss blocking, and introduce response surface designs such as the central composite and Box-Behnken designs. These standard designs are

Address correspondence to Douglas C. Montgomery, Arizona State University, School of Computing, Informatics and Decision Systems Engineering, Tempe, AZ 85287. E-mail: doug.montgomery@asu.edu

the foundation of many experimentally based research and development programs.

Standard designs work well when the research problem and the design are a good match. But there are many situations where the requirements of a standard design and the research problem are not a good fit. Some of these include the following:

1. The experimenter has unusual resource restrictions, so either the number of runs that can be made in the experiment or the size of the blocks required is different from the sample size and/or block sizes required by a standard design. For example, suppose that the experimenter needs to fit a second-order model in four factors. A central composite design requires between 25 and 30 runs, depending on the number of center points. But the experimenter can only afford to do 20 runs. Examples of optimal designs for cases where unusual resource restrictions are required are found in the section on unusual resource restrictions.
2. There are restrictions or constraints on the design region. That is, the standard cuboidal regions for factorial and fractional factorial designs and spherical or cuboidal regions for response surface designs are not appropriate either because it is impossible to experiment in some portions of the factor space (such as temperatures and pressures that are simultaneously beyond certain boundaries lead to unsafe operating conditions) or there are infeasible combinations of some factors. An example of an optimal design for a problem with a constrained design region is presented in the section on constrained design regions.
3. The experimenter needs to fit a nonstandard model. Models contain a mix of factors of different types. For example, suppose that the experimenter is interested in fitting a full quadratic model in two variables x_1 and x_2, but there is a third two-level categorical factor z that is also of interest. The model that the experimenter wants to entertain is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \gamma z + \delta_1 z x_1 + \delta_2 z x_2 + \delta_{12} z x_1 x_2 + \delta_{11} z x_1^2 + \delta_{22} z x_2^2 + \varepsilon$$

This is a full quadratic model in the two continuous factors and it also contains the main effect of the categorical factor plus all interactions between the categorical factor and the linear, interaction, and pure quadratic effects of the continuous factors. If this full 12-parameter model is the final model for the experiment, then the model describes two completely different response surfaces at the two different levels of the categorical factor. Assuming that the experimenter can only conduct 15 runs, there is no standard response surface design for this problem. The closest standard design that would work for this problem would be the $3 \times 3 \times 2$ factorial, which requires 18 runs. Examples of optimal designs for nonstandard models are contained in the section on nonstandard models.

Designing experiments for these types of problems requires a different approach. We can't look in the textbook or course notes and try to match the designs we find there to the problem. Instead, we need to create a custom design that fits our specific problem. Creating this custom design requires the following:

1. Information about the problem; specifically the model that the experimenter wants to entertain, the region of interest, the number of runs that can be performed, and any requirements about blocking, covariates, etc.
2. Choosing an optimality criterion; that is, a criterion for selecting the design points to be run. In the next section we will give a brief overview of optimality criterion for designed experiments.
3. A software package to construct the design. Sometimes optimal designs are called *computer-generated designs*. Several standard software packages do a good job of finding optimal designs.

It is always better to create a custom design for the actual problem that you want to solve than to force your problem to fit a standard design. Fortunately, it has been relatively easy to construct optimal design for about the last 15 years. The early research work on the theory of design optimality began with Kiefer (1959, 1961) and Kiefer and Wolfowitz (1959). The first practical algorithm for construction of optimal designs was developed by Mitchell (1974). This was a point exchange method, in which runs from

a candidate set of all possible runs that the experimenter would consider running were systematically exchanged with the runs in a current design until no further improvement in the optimality criterion could be achieved. Several variations of the point exchange approach were developed and implemented over the next 20 years. Meyer and Nachtsheim (1995) developed a coordinate exchange algorithm in which individual design coordinates were systematically searched to find the optimal settings. No candidate set of runs was required. This approach quickly became the standard one and today almost all efficient optimal design software makes use of the coordinate exchange approach.

OPTIMAL DESIGN CRITERIA

All designed experiments can be written in a linear regression model form. See Montgomery (2009) and Myers et al. (2009) for details. Therefore, it is easy to describe design optimality criteria in terms of a regression model. The linear regression model in matrix notation is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad [1]$$

where $\mathbf{y}$ is an $n \times 1$ vector of the observation (the response vector), $\mathbf{X}$ is an $n \times p$ model matrix of the levels of the independent variables, $\boldsymbol{\beta}$ is a $p \times 1$ vector of the regression coefficients, $\boldsymbol{\epsilon}$ is an $n \times 1$ vector of random errors, n is the number of runs in the experimental design, and p is the number of regression model parameters. The model matrix $\mathbf{X}$ is the design matrix expanded to the form of the model that the experimenter wants to fit. The number of regression coefficients, p, depends on the number of input factors and the assumed form of the model. For example, if there are four input factors and the assumed form for the model is a second-order linear regression model, there will be a total of 15 regression coefficients associated with the input factors: an intercept, four main effects, six two-way interactions, and four squared or pure quadratic terms.

The theory of linear models can be used to derive the least squares estimator of $\boldsymbol{\beta}$, the fitted regression model, and the variance of both $\hat{\boldsymbol{\beta}}$ and the predicted response. The least squares estimator of $\hat{\boldsymbol{\beta}}$ is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad [2]$$

The fitted regression model is

$$\hat{\mathbf{y}} = X\hat{\boldsymbol{\beta}} \quad [3]$$

where $\hat{\mathbf{y}}$ is the predicted response. The covariance matrix of the regression coefficients is

$$\mathrm{Var}[\hat{\boldsymbol{\beta}}] = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \quad [4]$$

The variance of the predicted response at a specific point in the design space relative to the error variance σ^2, also know as the *scaled prediction variance*, is

$$\mathrm{Var}[\hat{\mathbf{y}}]/\sigma^2 = \mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x} = v(\mathbf{x}) \quad [5]$$

where $\mathbf{x}$ is a vector containing the coordinates of the point of interest in the design space expanded to model form.

The optimal designs discussed in this article fall into two categories: designs that are optimized with respect to the regression coefficients and designs that are optimized with respect to the prediction variance of the response.

Optimization with Respect to the Regression Coefficients

Precision in terms of estimating the regression coefficients is important for screening experiments, where the goal is to the find the effects that significantly influence the response variable. More precise estimates can translate into fewer Type I and Type II errors. Here we discuss two optimality criteria: the *D*-optimality the *A*-optimality criteria.

The *D*-optimal design employs a criterion on the selection of design points that results in the minimization of the volume of the joint confidence region of the regression coefficients. This is achieved by maximizing the determinant (hence the *D* in *D*-optimal) of the $\mathbf{X}'\mathbf{X}$ matrix. That is, the quantity $|\mathbf{X}'\mathbf{X}|$ is maximized over all possible designs with N runs. The dispersion matrix, $(\mathbf{X}'\mathbf{X})^{-1}$, contains the variances and covariances of the regression coefficients and it can be shown that the square of the volume of the confidence region is inversely proportional to $|\mathbf{X}'\mathbf{X}|$. Controlling the volume of the confidence region is related to the precision of the regression coefficients; a smaller confidence region, for the same level of confidence, means more precise estimates.

Similar to the *D*-optimal design, the *A*-optimal design has an optimization objective function that deals with the variance of the regression coefficients. Whereas the *D*-optimal design considers both variance and covariance of the regression coefficients, the *A*-optimal criterion deals only with the individual variances of the regression coefficients. This is achieved by minimizing the trace of the $(\mathbf{X}'\mathbf{X})^{-1}$ matrix. The trace is the sum of the diagonal elements of the matrix, which in $(\mathbf{X}'\mathbf{X})^{-1}$ are the variances of the regression coefficients.

Optimization with Respect to the Predicted Response

Though estimating the regression coefficients with accuracy and precision is highly important for screening experiments, in response surface modeling, where the overall prediction model is most important, criteria on the experimental design aimed at precision of the model are desirable. Precision of the model can be measured in terms of the prediction variance metric. There are two main optimality criteria that are used to control the experimental design in such a way that aspects of prediction variance of the model are minimized. These are the *I*-optimal design and the *G*-optimal design.

The *I* in the *I*-optimal design represents the integrated prediction variance of the regression model. The integrated variance optimality criterion minimizes the integrated prediction variance of the regression model over a particular region of interest *R*. The objective function of the *I*-optimal design is

$$\frac{1}{K}\int_R v(\mathbf{x})dx \qquad [6]$$

where $v(\mathbf{x})$ is the scaled prediction variance given in Eq. [5] and K is the volume of the design region R. The region of interest R is generally the region specified by the boundaries of the ranges of the input variables. The integration of the prediction variances creates a single measure of prediction performance, which is interpreted as the average scaled prediction variance.

The *G*-optimal design also makes use of an objective function relating to the prediction variance. Instead of minimizing the average prediction variance over the design region, *R*, the *G*-optimal criterion minimizes the maximum value of prediction variance in the design region, *R*. Constructing *G*-optimal designs has historically proved difficult because two optimization problems must be solved: finding the best coordinate value to change in the current design and determining the maximum value of the scaled prediction variance for each new design that is evaluated. Rodriguez et al. (2010) described a commercially viable algorithm for constructing *G*-optimal designs and compare the performance of several *G*-optimal designs to their *I*-optimal and *D*-optimal counterparts.

Of the optimal designs in this section, *D*-, *A*-, *I*-, and *G*-optimal, the *D*-optimal and *I*-optimal are the two most widely used. Experimental designs that are created with respect to both *D*- and *I*-criteria are available in many commercially available software packages. Creating these designs requires an optimization algorithm. Techniques such as the coordinate exchange method of Meyer and Nachtsheim (1995) have been developed that minimize the computational burden and reduce the time required to find the optimal design. These techniques do not always guarantee a global optimal but the efficiency, a metric that quantifies the quality of an optimal design in terms of the best possible design is reported by the software programs.

STANDARD DESIGNS ARE OFTEN OPTIMAL DESIGNS

Many of the standard designs studied in basic experimental design courses are optimal designs. To illustrate, consider the 2^k factorial design. The 2^k factorial and its many variants are probably the most widely used family of designs in industrial research and development. Consider a special case; the 2^2 factorial with a single replicate. This is a four-run experiment and it can be used to fit the first-order model with interaction

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon$$

For the 2^2 design, the determinant of the matrix is $|\mathbf{X}'\mathbf{X}| = 256$. This is the maximum possible value of the determinant for a four-run design on the design space bounded by ± 1. Therefore, this is a *D*-optimal design. We can also show (see Montgomery 2009) that the maximum value of the prediction variance

occurs when $x_1 = \pm 1$ and $x_2 = \pm 1$ and is equal to σ^2. To determine how good this is, we need to know the best possible value of prediction variance that we can attain. It turns out that the smallest possible value of the maximum prediction variance over the design space is $p\sigma^2/N$, where p is the number of model parameters and N is the number of runs in the design. The 2^2 design has $N = 4$ runs and the model has $p = 4$ parameters, so the model that we fit to the data from this experiment minimizes the maximum prediction variance over the design region. So the 2^2 design is a *G*-optimal design. Finally, consider the average prediction variance. The average prediction variance over the design space for the 2^2 design is $4\sigma^2/9$, and this is the smallest possible value of the average prediction variance that can be obtained from a four-run design used to fit a first-order model with interaction on this design space. Therefore, the 2^2 design is also *I*-optimal. Similar results can be found for all members of the 2^k family and for the 2^{k-p} family of fractional factorials so long as the design resolution is sufficient to support the assumed model.

The fact that many widely used standard designs are optimal designs suggests that the optimal design approach is applicable in any design situation. If the problem turns out to be a standard one, an optimal design algorithm will generate the required standard design. But if not, then the optimal design approach will be necessary to construct the appropriate design for this specific research problem.

As another example, consider a situation in which there are two categorical factors with three levels each. An example scenario where this situation might occur is in missile testing. For example, assume that the White Sands missile range testing center wants to compare distance from target for three different types of missiles ($x_1 =$ L1, L2, L3), each containing a slightly different metal alloy encasing, and three different launching mechanisms ($x_2 =$ L1, L2, L3). Using this example, both *D*- and *I*-optimal designs can be created for the main effects–only model and main effects plus two-factor interaction model. Let us assume that the experimenter is interested in a nine-run design. The *D*- and *I*-optimal designs for the main effects–only model were found using JMP (SAS Institute, Cary, NC) and are presented in Tables 1 and 2, respectively. Notice that both designs are identical; in fact, they are both 3^2 factorial designs. If we augment the model to include the two-factor interaction term and construct nine-run *D*- and *I*-optimal designs we get the same results.

TABLE 1 The Nine-Run *D*-Optimal Design for Two Three-Level Categorical Factors

Run	X1	X2
1	L1	L1
2	L2	L1
3	L3	L1
4	L1	L2
5	L2	L2
6	L3	L2
7	L1	L3
8	L2	L3
9	L3	L3

Both of the designs in Tables 1 and 2 are unreplicated factorials. Usually the experimenter would like to replicate the design in order to obtain an estimate of experimental error that would support statistical testing. If the design is replicated twice, this would require a total of 18 runs. We would still have a standard design. However, because each run requires firing a missile, and these runs are likely very expensive, the experimenter would probably be interested in a design with fewer runs. Suppose that the experimenter wants to fit the main effects plus interaction model and can afford a total of 12 runs. Where should these replicate runs be made? An optimal design approach can be used to determine the best place to allocate replicate runs to a design.

Tables 3 and 4 present the *D*- and *I*-optimal designs from JMP, respectively. From inspection of the tables we see that both designs are full 3^2 factorials with three replicated runs.

TABLE 2 The Nine-Run *I*-Optimal Design for Two Three-Level Categorical Factors

Run	X1	X2
1	L1	L1
2	L2	L1
3	L3	L1
4	L1	L2
5	L2	L2
6	L3	L2
7	L1	L3
8	L2	L3
9	L3	L3

TABLE 3 The 12-Run *D*-Optimal Design for Two Three-Level Categorical Factors

Run	X1	X2
1	L1	L1
2	L2	L1
3	L3	L1
4	L1	L2
5	L1	L2
6	L2	L2
7	L2	L2
8	L3	L2
9	L3	L2
10	L1	L3
11	L2	L3
12	L3	L3

The *D*-optimal design replicates the treatment combinations (L1, L2), (L2, L2), and (L3, L2). Thus, L2 appears six times in column X2 and L1 and L3 appear only three times. By contrast, the *I*-optimal design replicates combinations (L1, L1), (L2, L2), and (L3, L2). In column X2 of this design, L1 appears three times, L2 appears five times, and L3 appears four times. The only difference between the two designs is that one of the replicated runs is different in each design. The average scaled prediction variance is 0.833 for both designs.

CREATING OPTIMAL DESIGNS

As noted previously, *D*- and *I*-optimality are the two optimality criteria most widely available for creating designs in commercial software packages. This section gives several examples of both *D*- and *I*-optimal designs for situations where the requirements of a standard design and the research problem are not met. The three main categories these situations are found in were described in the Introduction. Each of the three categories will be presented in this section along with examples that illustrate the choice of optimal design and some properties about those designs.

TABLE 4 The 12-Run *I*-Optimal Design for Two Three-Level Categorical Factors

Run	X1	X2
1	L3	L3
2	L2	L2
3	L2	L1
4	L1	L1
5	L3	L1
6	L3	L2
7	L2	L3
8	L3	L2
9	L1	L3
10	L2	L2
11	L1	L1
12	L1	L2

Unusual Resource Restrictions

In this section two examples of situations that require restrictions on resources are presented.

A Second-Order Model

Suppose that an experimenter wants to fit a second-order model in four factors. This model contains an intercept, four linear terms, four pure quadratic terms, and six cross-product (two-factor interaction) terms. The usual design choices for fitting this model include the central composite design and the Box-Behnken design. These designs would require between 25 and 30 runs, depending on the number of center points. However, suppose that because of equipment availability, the experimenter is restricted to a design with $n = 16$ runs. Because second-order models are typically used for optimization and prediction, the *I*-criterion is a good choice for this situation.

TABLE 5 A 16-Run *I*-Optimal Design for a Second-Order Model in Four Factors

Run	X1	X2	X3	X4
1	1	0	0	−1
2	1	1	0	1
3	−1	1	−1	1
4	0	1	1	−1
5	−1	0	1	1
6	−1	−1	1	−1
7	0	0	0	0
8	−1	0	−1	−1
9	1	−1	−1	1
10	1	1	−1	−1
11	0	−1	−1	−1
12	0	0	−1	0
13	1	−1	1	0
14	−1	−1	−1	0
15	−1	1	0	0
16	0	−1	0	1

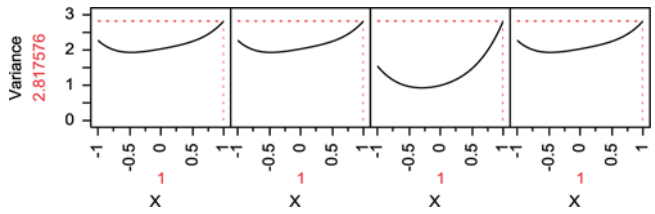

FIGURE 1 **Prediction variance for the 16-run *I*-optimal design shown in Table 5.**

Table 5 shows the 16-run *I*-optimal design created by the JMP v9.0 software. Figure 1 contains output from the JMP prediction variance profiler. The point of maximum scaled prediction variance is shown on the graphs. Notice that the maximum value of the scaled prediction variance is 2.818. Figure 2 contains the fraction of design space (FDS) plot for this design. An FDS plot (Zaharan et al. 2003) graphically presents the ordered scaled prediction variance over the volume of the design region. An ideal design has a very flat FDS plot with a small value of the scaled prediction variance. The average scaled prediction variance over the design region is 0.596255. When using the linear model to make predictions, the lower the prediction variance, the better.

For comparison, we also used JMP to create a 16-run *D*-optimal design for the problem. This design is shown in Table 6. The prediction variance profile in Figure 3 shows that the maximum scaled prediction variance is 2.078, which is about 26% smaller than the maximum prediction variance from the *I*-optimal design. However, the fraction of design space plot in Figure 4 shows the prediction variance distribution over the design space for both the *I*-optimal and the *D*-optimal design. The *I*-optimal has the

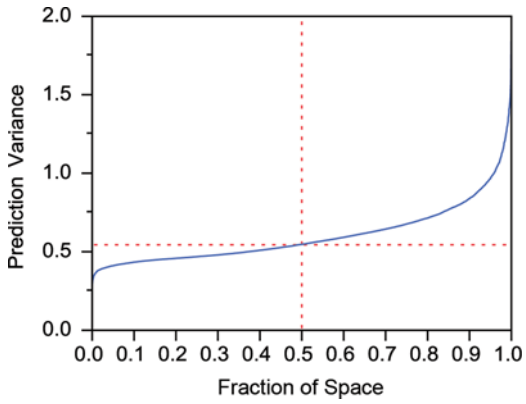

FIGURE 2 **FDS plot for the 16-run *I*-optimal design shown in Table 5.**

TABLE 6 **A 16-Run *D*-Optimal Design**

Run	X1	X2	X3	X4
1	0	1	1	−1
2	0	1	−1	1
3	−1	1	−1	−1
4	−1	1	1	1
5	−1	−1	−1	1
6	1	−1	1	−1
7	1	−1	−1	1
8	−1	−1	1	−1
9	1	0	1	1
10	1	1	−1	−1
11	−1	−1	−1	−1
12	1	1	0	0
13	0	0	−1	0
14	−1	0	0	1
15	0	−1	1	1
16	0	−1	0	−1

lower prediction variance over the region of the design.

The solid line in this figure is the *I*-optimal design. Notice that over almost all of the design space the *I*-optimal design has lower prediction variance. The average scaled prediction variance over the design space for the *D*-optimal design is 0.789988, which is 32% larger than the average scaled prediction variance of the *I*-optimal design. So unless the experimenter is interested in protection against prediction at the "worst" possible place in the design space, the *I*-optimal design is the preferred choice in this example.

This example illustrates that it is useful for the investigator to compare diagnostics for designs being considered. Prediction variance is one good diagnostic tool for comparison. Modern computer software makes generation of alternate designs for a problem easy and the software products can also facilitate design comparisons. For more information on comparing designs, see Anderson-Cook et al. (2009).

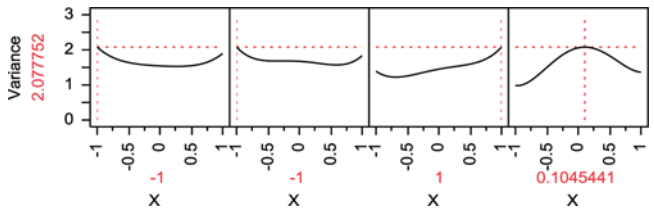

FIGURE 3 **Prediction variance for the 16-run *D*-optimal design shown in Table 6.**

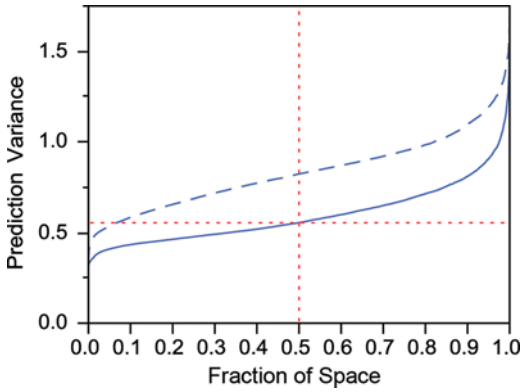

FIGURE 4 FDS plot for the 16-run *D*-optimal design (dashed line) in Table 6 overlaid with the 16-run *I*-optimal design (solid line) in Table 5.

Nonstandard Block Sizes

Suppose an investigator can perform three runs in one day and wishes to run a screening experiment with six continuous factors. The budget for the experiment allows for 12 runs. So, the experiment will involve four days of experimentation with three runs per day.

The 12-run Plackett-Burman design is a natural choice for a main effects model, but there is no blocking scheme for these designs that accommodates blocks of three runs. So, no textbook design quite matches this problem description.

Table 7 shows the factor settings and block assignments for the 12-run *I*-optimal design tailored to fit the problem. The *I*-optimal design is for a main effects model.

This design has a *D*-efficiency of 95.4%. The most notable feature of this design is that each factor has one setting at the middle of its range. This means that if any factor has a strong quadratic effect, there is a good chance of both detecting the curvature and identifying the causal factor. By contrast, a two-level design has no way to detect strong curvature if it exists.

TABLE 7 An *I*-Optimal Design for Six Factors in Four Blocks of Size 3

A	B	C	D	E	F	Day
−1	−1	−1	−1	1	1	1
1	1	1	−1	1	−1	1
1	−1	1	1	−1	1	1
0	−1	1	−1	1	1	2
−1	1	−1	1	−1	−1	2
1	−1	1	0	1	0	2
1	−1	−1	1	1	−1	3
−1	−1	1	−1	−1	−1	3
−1	1	1	1	1	1	3
1	1	0	−1	−1	1	4
−1	−1	1	1	1	−1	4
1	0	−1	−1	0	1	4

Constrained Design Region

A constrained design region implies the application of one or more constraints placed on the values of the design factors. These constraints restrict the experimental design region. Experiments that contain only continuous numeric factors are generally described in coded units on a scale of [−1, 1]. In the example presented, we look at a comparison between eight-run *D*- and *I*-optimal designs for the response surface model in two factors (the RSM linear regression model for two factors includes two main effects, a two-factor interaction, and two squared terms). Tables 8 and 9 present *D*- and *I*-optimal designs for the response surface models for the design region constrained by the equation $x_1 + x_2 \leq 1$. Figure 5 is the FDS plot comparing these two designs. In Figure 5 the solid line represents the *I*-optimal design and the dashed line represents the *D*-optimal design. The *I*-optimal design has smaller prediction variance compared to the *D*-optimal design over about 85% of the design space. If the objective of the experimenter is to build a prediction model, the *I*-optimal design is the preferred choice.

Nonstandard Models

Nonstandard models include models with different types of variables—both categorical and continuous—and models that have a specific assumed model form that may not conform to any standard linear model case such as an RSM model.

TABLE 8 A *D*-Optimal Design for a Constrained Region

X1	X2
−1	1
−1	−1
1	0
−1	0.08
1	−1
0	1
0.08	−1
−0.11	−0.11

TABLE 9 An *I*-Optimal Design for a Constrained Region

X1	X2
1	0
−1	0.76
0.07	−0.95
0	1
1	−1
−1	−1
−0.22	−0.10
−0.22	−0.10

Optimal Design for Mixed Categorical and Continuous Factors

Suppose that a quality engineer wants to determine whether five machines on a production line are generating comparable product. Each machine has four two-level controllable variables. The budget for the experiment is 15 runs.

Clearly, there is no orthogonal or textbook design for this situation. Because the budgeted number of runs is odd, there is no way to have the same number of runs at each level of the two-level factors. The number of runs is a multiple of the number of machines, so each machine can produce three parts for the study.

Table 10 shows the factor settings and block assignments for the 15-run *I*-optimal design custom built to match the requirements as specified. The *D*-efficiency of this design is 98.4%.

Intuition might suggest that each of the two-level factors should be as balanced as possible. That is, each should have seven runs at one setting and six runs at the other. The surprise here is that each of the two-level factors has six runs at one setting and nine runs at the other. However, there is a type of hidden balance in this structure. For each machine two of the factors have two runs at +1 and one run at −1. The other two factors have two runs at −1 and one run at +1. The result is that the two-level factors' main effect estimates are equivariant. Each has a variance inflation factor (VIF) of only 1.155. This would correspond to confidence intervals that are less than 7.5% longer than an orthogonal design

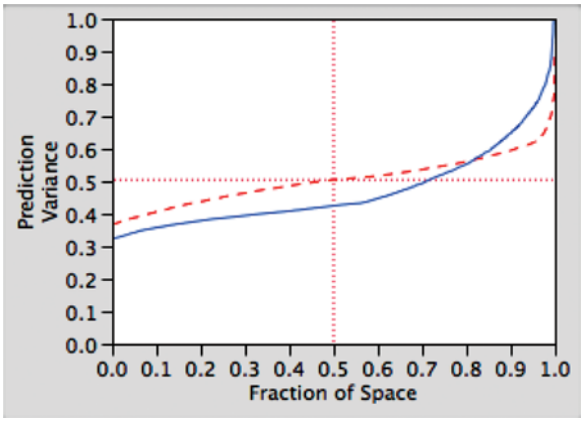

FIGURE 5 FDS plot comparing the *D*-optimal (dashed line) and *I*-optimal (solid line) designs for a constrained region.

TABLE 10 An *I*-Optimal Design for Mixed Categorical and Continuous Factor Design

A	B	C	D	E
1	−1	−1	−1	−1
1	−1	1	−1	1
1	1	1	1	1
2	−1	−1	1	1
2	−1	1	−1	−1
2	1	1	−1	1
3	−1	1	−1	1
3	−1	1	1	−1
3	1	−1	−1	1
4	−1	−1	1	1
4	1	−1	−1	−1
4	1	1	1	−1
5	−1	1	−1	1
5	−1	1	−1	1
5	1	−1	1	−1

TABLE 11 The *D*-Optimal Design for a Nonstandard Linear Model

X1	X2	X3
1	−1	1
−1	1	−1
−1	−1	−1
−1	1	−0.5
1	1	−1
−1	−1	1
1	−1	−1
−1	1	1
1	1	1
1	−1	0.5
−1	0	−1
−1	−1	−0.5
1	0	−0.5
−1	0	0.5
1	0	1
1	1	0.5

TABLE 12 The *I*-Optimal Design for a Nonstandard Linear Model

X1	X2	X3
1	1	−0.5
−1	−1	−1
−1	1	0.5
1	1	1
−1	0	−0.5
−1	0	−0.5
−1	1	−1
1	0	0.5
1	0	0
−1	0	0
1	−1	1
−1	−1	0.5
−1	0	1
1	−1	−0.5
1	0	−1
1	0	0.5

for this problem (if one existed). Here, the lack of balance and orthogonality has a price but it is small.

Optimal Designs for a Nonstandard Linear Model

Consider an experiment with three continuous factors. Suppose that x_1 is expected to be the factor that has the simplest relationship with the response, factor x_2 is expected to have a more complex relationship, and factor x_3 is expected to have the most complex relationship with the response. The experimenter assumes that the model form is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \beta_{22} x_2^2 + \beta x_{33} x_3^2 + \beta_{123} x_1 x_2 x_3 + \beta_{333} x_3^3$$

This is a reduced cubic model. Tables 11 and 12 contain the *D*-optimal and *I*-optimal designs. Figure 6 shows the designs graphically. Notice that factor x_1 has two levels, factor x_2 has three levels, and factor x_3 has four levels in the *D*-optimal design and five levels in the *I*-optimal design.

Figure 7 contains the FDS comparison of the two designs. The dashed line is the *D*-optimal design and the solid line is the *I*-optimal design. Notice that the *I*-optimal design dominates the *D*-optimal design over more than 90% of the volume of the design space. This is an illustration of the importance of evaluating a design before the experiment is actually conducted.

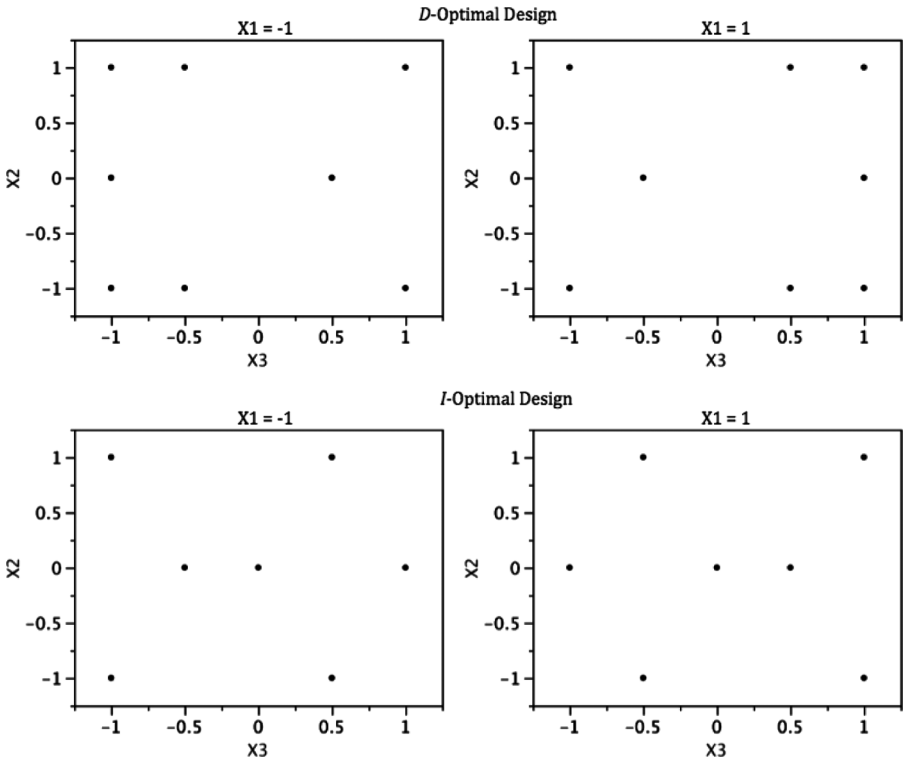

FIGURE 6 Plots of the *D*-optimal and *I*-optimal designs.

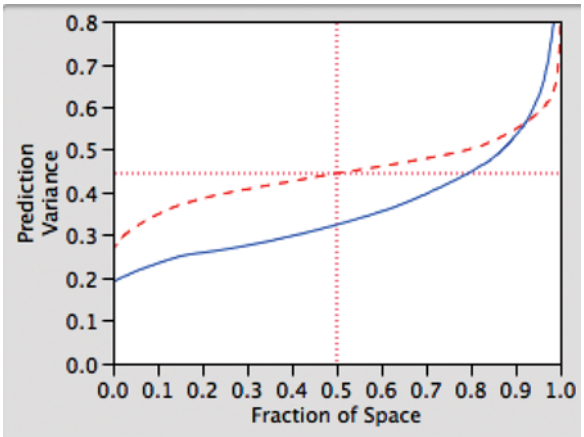

FIGURE 7 FDS plot comparison of the *D*-optimal and *I*-optimal design.

USE OF AN OPTIMAL DESIGN TO CONDUCT A CONFIRMATION EXPERIMENT

Khoramnia et al. (2010) published an article in the *Electronic Journal of Biotechnology* describing a process characterization and optimization case study. The investigators were interested in maximizing the lipase activity of a bacterial culture. Lipase is a water-soluble enzyme that catalyzes the hydrolysis of certain types of chemical bonds and they play a vital role in digestion and processing of triglycerides, fats, and oils. Bioactivity is a way to describe the interaction of the lipase with other materials. There were five factors in the study with ranges shown in Table 13.

To model the lipase production process, the investigators employed a central composite design inscribed within the ranges specified in the table. Note that the central composite design is a spherically symmetric design. There are no points that are simultaneously at or near the extremes of all the factors. Table 14 shows the design and resulting lipase activity. The maximum observed activity was 0.164.

Using a stepwise regression approach including all first-, second-, and third-order terms yields the parameter estimates shown in Table 15. Note the large three-factor interaction involving temperature, size, and agitation. The original investigators missed this effect possibly due to failing to look for it. As a result, they opted to use a neural network model instead.

The root mean squared error of the model shown in Table 15 is 0.088 with an R^2 value of 98%. The average lipase activity at the center of the design region is 0.14. Using the model the expected lipase activity is 0.81 when temperature = 45, pH = 6, size = 1, time = 65, and agitation = 0. This represents a factor of nearly fivefold increase over the best observed activity in the experiment. However, the length of a 95% confidence interval on this expected value is 0.21. This very wide confidence interval is due to the fact that all but one of the factors is at an extreme value. Essentially, this prediction represents an extrapolation outside of the design region.

Suppose that we wish to run a confirmation experiment using the regression model in Table 15

TABLE 13 Factors and Levels for the Lipase Study

Setting	Temperature	pH	Size	Time	Agitation
Low	27	6	1	24	0
High	45	9	5	96	200

TABLE 14 Central Composite Design and Observed Response Data

Run	Temperature	pH	Size	Time	Agitation	Activity
1	40.94	8.32	1.9	79.77	45.09	0.032
2	40.94	6.68	4.1	79.77	45.09	0.08
3	31.06	8.32	4.1	40.23	154.91	0.095
4	40.94	8.32	4.1	40.23	45.09	0.072
5	40.94	8.32	1.9	40.23	154.91	0.057
6	40.94	6.68	1.9	79.77	154.91	0.065
7	31.06	6.68	4.1	79.77	154.91	0.106
8	31.06	8.32	1.9	79.77	154.91	0.119
9	40.94	6.68	4.1	40.23	154.91	0.065
10	31.06	8.32	4.1	79.77	45.09	0.059
11	31.06	6.68	1.9	40.23	45.09	0.044
12	27	7.5	3	60	100	0.125
13	45	7.5	3	60	100	0.164
14	36	6	3	60	100	0.126
15	36	9	3	60	100	0.105
16	36	7.5	1	60	100	0.145
17	36	7.5	5	60	100	0.063
18	36	7.5	3	24	100	0.076
19	36	7.5	3	96	100	0.08
20	36	7.5	3	60	0	0.131
21	36	7.5	3	60	2	0.11
22	36	7.5	3	60	100	0.142
23	36	7.5	3	60	100	0.138
24	36	7.5	3	60	100	0.152
25	36	7.5	3	60	100	0.121
26	36	7.5	3	60	100	0.14

TABLE 15 Parameter Estimates from Stepwise Regression Analysis

Term	Estimate	Std error
Intercept	0.141	0.003
Temperature	0.020	0.006
pH	−0.01	0.005
Size	0.089	0.014
Time	0.001	0.005
Agitation	−0.01	0.005
Temperature × Size	−0.06	0.019
Temperature × Agitation	0.034	0.016
pH × Time	−0.17	0.037
pH × Agitation	0.155	0.019
Size × Agitation	0.044	0.016
Time × Agitation	0.155	0.019
Temperature × Size × Agitation	0.448	0.060
pH × pH	−0.03	0.007
Size × Size	−0.04	0.007
Size × Size × Size	−0.13	0.016
Time × Time	−0.06	0.007
Agitation × Agitation	−0.02	0.007

TABLE 16 *I*-Optimal Response Surface Confirmatory Experiment for a Nonstandard Model

Run	Temperature	pH	Size	Time	Agitation
1	45	6	4	24	0
2	45	8.25	2	96	100
3	27	9	4	96	0
4	45	9	3	24	200
5	27	9	1	24	0
6	27	9	1	78	200
7	27	6	2	96	200
8	27	6	2	60	0
9	45	9	2	60	0
10	45	7.5	4	60	100
11	27	9	5	96	200
12	45	6	5	24	200
13	45	7.5	5	96	0
14	45	6.75	1	96	0
15	27	7.5	2	24	100
16	27	7.5	5	24	0
17	27	7.5	4	42	200
18	45	7.5	4	96	200
19	45	6.75	1	60	200
20	45	9	5	60	100
21	27	7.5	4	60	100
22	45	6	1	24	100
23	27	6	4	96	100
24	27	7.5	2	60	100

as our a priori model. There are 18 terms in this model, including various two-factor interactions, quadratic effects, and two third-order effects—the Temperature × Size × Agitation three-factor interaction and the cubic size effect. If we chose to perform 24 runs, we would have six degrees of freedom for lack of fit. These extra runs allow you to estimate the error variance under the assumption that the model is correct. A second benefit of these extra runs is that they allow for the detection of other active effects that are not in the a priori model. Finally, they also increase the denominator degrees of freedom for significance tests of the model coefficients. This makes for more powerful tests.

Table 16 displays the *I*-optimal design with 24 runs. Note that this design has two fewer runs than the central composite design.

Table 17 provides a comparison of the relative standard deviations of the coefficients for the *I*-optimal and central composite designs. These standard deviations are relative to the true value of the standard deviation of the random errors.

The fourth column of Table 17 contains the ratio of the standard deviation of each coefficient for the central composite design to the standard deviation of the *I*-optimal design. A value greater than 1 indicates that the *I*-optimal design more precisely estimates this coefficient. Note that these ratios range from 1.29 to 4.55. So, the *I*-optimal design outperforms the central composite design for every coefficient. A ratio of 4.55 means that a confidence interval for the given coefficient is 4.55 times longer for the central composite design than for the *I*-optimal design. This is despite the fact that the *I*-optimal design has fewer runs.

The primary focus of the *I*-optimal design is prediction variance. Figure 8 compares the prediction variance over the five-dimensional hypercube covered by the *I*-optimal design to the prediction variance for the central composite design over the same region. The average relative prediction variance for the *I*-optimal design is 0.4 compared to 2.0 for the central composite design. Thus, the average prediction variance is five times larger for the central composite design. Toward the vertices of the hypercube, the variance ratio becomes much larger. At the predicted optimum setting, the relative prediction variance of the central composite design is 12 times larger than the *I*-optimal design, meaning that confidence intervals for prediction here are about three and a half times longer.

TABLE 17 **Comparison of Coefficient Variability for Two Designs**

Effect	Std *I*-optimal	Std dev CCD	Ratio
Temperature	0.22	0.61	2.75
pH	0.28	0.51	1.84
Size	0.78	1.32	1.7
Time	0.28	0.51	1.81
Agitation	0.25	0.47	1.86
Temperature × Agitation	0.28	1.17	4.21
pH × Agitation	0.33	1.43	4.37
Size × Agitation	0.33	1.07	3.23
Time × Agitation	0.32	1.46	4.55
pH × pH	0.48	0.75	1.57
Size × Size	0.58	0.75	1.29
Time × Time	0.48	0.75	1.55
Agitation × Agitation	0.47	0.75	1.6
Size × Size × Size	0.88	1.65	1.86
Temperature × Size	0.30	1.10	3.74
Temperature × Size × Agitation	0.33	1.18	3.55
pH × Time	0.36	1.33	3.68

Why is the *I*-optimal design doing so much better for reducing these variances? The main reason is that the central composite design puts its points on a sphere that is inside the hypercube used for the *I*-optimal design. Therefore, the *I*-optimal design is covering a much larger region of the design space. For continuous factors you can always lower the variance of the parameter estimates and predicted responses by increasing the size of the experimental region. Of course, in life you never get something for nothing. The *I*-optimal design is much more vulnerable to bias coming from active terms that are not in the a priori model. Because the proposed experiment is a confirmatory experiment for a model having a substantial amount of data-driven validity, it makes sense to increase the volume of the experimental region at least enough to include the predicted optimum factor settings.

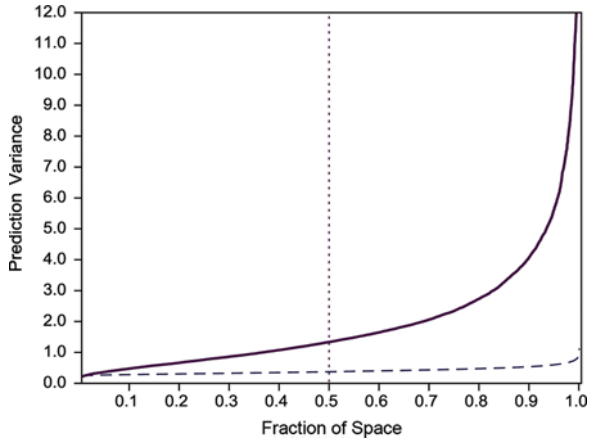

FIGURE 8 **Fraction of design space plot comparing the prediction variances for the central composite design (solid line) and the *I*-optimal design (dashed line).**

EXTENSIONS OF THE OPTIMAL DESIGN APPROACH

We have discussed the use of design optimality for linear models in relatively simple situations. There are several extensions of these ideas that practitioners should be aware of.

A criticism often leveled at the optimal design approach is that the final design depends on the model chosen by the experimenter. DuMouchel and Jones (1994) introduced a Bayesian modification of the *D*-optimal design that affords protection to experimenters against terms that are not in the assumed model. They assumed that the model contains p primary terms but they wanted to obtain protection against q potential model terms. These potential terms are typically of higher order than those in the primary model. Their Bayesian *D*-optimal designs have N runs, where $p < N < p + q$. These designs allow some of the potential model terms to be fit if necessary. Jones et al. (2008) used the Bayesian *D*-optimality approach to construct supersaturated fractional factorial designs. Andere-Rendon et al. (1997) used this approach to design mixture experiments in the presence of model uncertainty.

It is also possible to construct optimal designs for nonlinear models. In linear models the optimal design problem is relatively simple because the model covariance matrix $\mathbf{X}'\mathbf{X}$ does not contain any of the unknown parameters $\boldsymbol{\beta}$. However, if the model is nonlinear this is not the case. To find a *D*-optimal design for a nonlinear model we must find design points that maximize the determinant of $\mathbf{D}'\mathbf{D}$, where $\mathbf{D}$ is a matrix of partial derivatives of the nonlinear model expectation function with respect to each model parameter evaluated at each design point. This matrix is a function of the unknown parameters, so finding a *D*-optimal design would require knowledge of the model parameters. One possible approach to this problem is to assume values for the unknown $\beta's$. This would produce a conditional *D*-optimal design.

An alternative is to use a Bayesian approach employing a prior distribution $f(\boldsymbol{\beta})$ to specify the uncertainty in the p parameter values. This leads to a design criterion

$$\phi(\mathbf{D}) = \int \log|\mathbf{D}'\mathbf{D}| f(\boldsymbol{\beta}) d\boldsymbol{\beta} \qquad [7]$$

This is the expectation of the logarithm of the information matrix. This criterion was proposed by Chaloner and Larntz (1989) for single-factor logistic regression model. The difficulty in using Eq. [7] as a design criterion is that the p-dimensional integral must be evaluated a very large number of times. Gotwalt et al. (2009) have recently developed a clever quadrature scheme that greatly reduces the computing time to evaluate the integral in Eq. [7] with excellent accuracy. This procedure is implemented in the nonlinear design routine of JMP and allows computationally efficient construction of D-optimal designs for nonlinear models. Gotwalt et al. (2009) presented examples of the use of this technique. Also see Johnson and Montgomery (2010).

A very important type of nonlinear model that occurs frequently in industrial experimentation is the generalized linear model. This is a family of models that unify linear and nonlinear regression models with response distributions that are a member of the exponential family (which includes the binomial, Poisson, normal, exponential, and gamma distributions). Important special cases include logistic regression, Poisson regression, and regression with exponential responses. Often an experimenter will know in advance that the response distribution is binomial (for example). Then a design for a logistic regression model would be appropriate. The method described in Gotwalt et al. (2009) can be used to construct D-optimal designs for this experiment. For examples of designed experiments for generalized linear models, see Johnson and Montgomery (2009) and Myers et al. (2010).

CONCLUSIONS

The optimal design approach is a powerful and useful way for experimenters to approach almost all problems they face. Modern applications of designed experiments often involve nonstandard problems for which the design optimality framework represents the only viable approach. However, as we have shown, many standard designs are also optimal designs, so the use of design optimality for design construction in standard situations leads to the appropriate design choice. It is also important to evaluate designs carefully before conducting the experiment and to answer questions regarding choice of optimality criterion, sample size, and choice of tentative model for the experiment. Modern computer software facilitates this process.

ABOUT THE AUTHORS

Dr. Rachel T. Johnson is an Assistant Professor in the Operations Research Department at the Naval Postgraduate School. She received her B.S. in industrial engineering from Northwestern University and her M.S. and Ph.D. from Arizona State University. Her research and teaching interests are in statistics and operations research with focus in design of experiments. She is a recipient of the Mary G. and Joseph Natrella Scholarship for "leadership, outstanding scholarship and commitment to the pursuit of quality improvement through statistical methods."

Dr. Douglas C. Montgomery is Regents' Professor of industrial engineering and statistics and Foundation Professor of Engineering at Arizona State University. His research and teaching interests are in industrial statistics. Professor Montgomery is a Fellow of the ASA, a Fellow of the ASQ, a Fellow of the RSS, a Fellow of IIE, a member of the ISI, an academician of the IAQ and has received several teaching and research awards.

Dr. Bradley A. Jones is the Principal Research Fellow at SAS Institute and a guest professor at the University of Antwerp. He is the inventor of the prediction profile plot, an interactive graph for exploring multidimensional response surfaces. At SAS Institute he is responsible for the design of experiments capabilities in the JMP software package. He is a Fellow of the American Statistical Association and a winner of both the Nelson Award and the Brumbaugh Award.

REFERENCES

Anderson-Cook, C. M., Borror, C. M., Montgomery, D. C. (2009). Response surface design evaluation and comparison (with discussion). *Journal of Statistical Planning and Inference*, 139:629–674.

Andere-Rendon, J., Montgomery, D. C., Rollier, D. A. (1997). Design of mixture experiments using Bayesian D-optimality. *Journal of Quality Technology*, 29:451–463.

Chaloner, K., Larntz, K. (1989). Optimal Bayesian designs applied to logistic regression experiments. *Journal of Statistical Planning and Inference*, 21:191–208.

DuMouchel, W., Jones, B. (1994). A simple Bayesian modification of *D*-optimal designs to reduce dependence on an assumed model. *Technometrics*, 36:37–47.

Gotwalt, C. M., Jones, B. A., Steinberg, D. M. (2009). Fast computation of designs robust to parameter uncertainty for nonlinear settings. *Technometrics*, 51:86–95.

Johnson, R. T., Montgomery, D. C. (2009). Choice of second-order response surface designs for logistic and Poisson regression models. *International Journal of Experimental Design and Process Optimization*, 1:2–23.

Johnson, R. T., Montgomery, D. C. (2010). Designing experiments for nonlinear models—An introduction. *Quality and Reliability Engineering International*, 26:431–441.

Jones, B., Lin, D. K. J., Nachtsheim, C. J. (2008). Bayesian *D*-optimal supersaturated designs. *Journal of Statistical Planning and Inference*, 138:86–92.

Khoramnia, A., Lai, O. M., Ebrahimpour, A., Tanduba, C. J., Voon, T. S., Mukhlis, S. (2010). Thermostable lipase from a newly isolated *Staphylococcus xylosus* strain; process optimization and characterization using RSM and ANN. *Electronic Journal of Biotechnology*, 13. Available at: http://dx.doi.org/10.2225/vol13-issue5-fulltext-22 (accessed 10 March 2011).

Kiefer, J. (1959). Optimum experimental designs. *Journal of the Royal Statistical Society, Series B*, 21:272–304.

Kiefer, J. (1961). Optimum designs in regression problems II. *Annals of Mathematical Statistics*, 32:298–325.

Kiefer, J., Wolfowitz, J. (1959). Optimum designs in regression problems. *Annals of Mathematical Statistics*, 30:271–294.

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

Mitchell, T. J. (1974). An algorithm for the construction of *D*-optimal experimental designs. *Technometrics*, 16:203–210.

Montgomery, D. C. (2009). *Design and Analysis of Experiments,* 7th ed. Hoboken, NJ: Wiley.

Myers, R. H., Montgomery, D. C., Anderson-Cook, C. M. (2009). *Response Surface Methodology: Process and Product Optimization Using Designed Experiments,* 3rd ed. Hoboken, NJ: Wiley.

Myers, R. H., Montgomery, D. C., Vining, G. G., Robinson, T. J. (2010). *Generalized Linear Models with Applications in Engineering and the Sciences,* 2nd ed. Hoboken, NJ: Wiley.

Rodriguez, M., Jones, B., Borror, C. M., Montgomery, D. C. (2010). Generating and assessing exact *G*-optimal designs. *Journal of Quality Technology*, 42:3–29.

Zahran, A., Anderson-Cook, C. M., Myers, R. H. (2003). Fraction of design space to assess the prediction capability of response surface designs. *Journal of Quality Technology*, 35:377–386.