

# Response surface methodology

André I. Khuri<sup>1\*</sup> and Siuli Mukhopadhyay<sup>2</sup>

The purpose of this article is to provide a survey of the various stages in the development of response surface methodology (RSM). The coverage of these stages is organized in three parts that describe the evolution of RSM since its introduction in the early 1950s. Part I covers the period, 1951–1975, during which the so-called classical RSM was developed. This includes a review of basic experimental designs for fitting linear response surface models, in addition to a description of methods for the determination of optimum operating conditions. Part II, which covers the period, 1976–1999, discusses more recent modeling techniques in RSM, in addition to a coverage of Taguchi's robust parameter design and its response surface alternative approach. Part III provides a coverage of further extensions and research directions in modern RSM. This includes discussions concerning response surface models with random effects, generalized linear models, and graphical techniques for comparing response surface designs. © 2010 John Wiley & Sons, Inc. *WIREs Comp Stat* 2010 2 128–149

## PART I. THE FOUNDATIONAL YEARS: 1951–1975

### An Introduction and Some Preliminaries

Response surface methodology (RSM) consists of a group of mathematical and statistical techniques used in the development of an adequate functional relationship between a response of interest,  $y$ , and a number of associated control (or input) variables denoted by  $x_1, x_2, \dots, x_k$ . In general, such a relationship is unknown but can be approximated by a low-degree polynomial model of the form

$$y = f'(x)\beta + \epsilon \quad (1)$$

where  $x = (x_1, x_2, \dots, x_k)'$ ,  $f(x)$  is a vector function of  $p$  elements that consists of powers and cross-products of powers of  $x_1, x_2, \dots, x_k$  up to a certain degree denoted by  $d (\geq 1)$ ,  $\beta$  is a vector of  $p$  unknown constant coefficients referred to as parameters, and  $\epsilon$  is a random experimental error assumed to have a zero mean. This is conditioned on the belief that model (1) provides an adequate representation of the response. In this case, the quantity  $f'(x)\beta$  represents

the mean response, that is, the expected value of  $y$ , and is denoted by  $\mu(x)$ .

Two important models are commonly used in RSM. These are special cases of model (1) and include the first-degree model ( $d = 1$ ),

$$y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \epsilon \quad (2)$$

and the second-degree model ( $d = 2$ )

$$y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i < j} \beta_{ij} x_i x_j + \sum_{i=1}^k \beta_{ii} x_i^2 + \epsilon. \quad (3)$$

The purpose of considering a model such as (1) is threefold:

1. To establish a relationship, albeit approximate, between  $y$  and  $x_1, x_2, \dots, x_k$  that can be used to predict response values for given settings of the control variables.
2. To determine, through hypothesis testing, significance of the factors whose levels are represented by  $x_1, x_2, \dots, x_k$ .
3. To determine the optimum settings of  $x_1, x_2, \dots, x_k$  that result in the maximum (or minimum) response over a certain region of interest.

\*Correspondence to: ufakhuri@stat.ufl.edu

<sup>1</sup>Department of Statistics, University of Florida, Gainesville, FL 32611, USA

<sup>2</sup>Department of Mathematics, Indian Institute of Technology Bombay, Powai, Mumbai 400076, India

DOI: 10.1002/wics.73

In order to achieve the above three objectives, a series of  $n$  experiments should first be carried out, in each of which the response  $y$  is measured (or observed) for specified settings of the control variables. The totality of these settings constitutes the so-called response surface design, or just design, which can be represented by a matrix, denoted by  $\mathcal{D}$ , of order  $n \times k$  called the design matrix,

$$\mathcal{D} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{22} & x_{22} & \dots & x_{2k} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix} \quad (4)$$

where  $x_{ui}$  denotes the  $u$ th design setting of  $x_i$  ( $i = 1, 2, \dots, k; u = 1, 2, \dots, n$ ). Each row of  $\mathcal{D}$  represents a point, referred to as a design point, in a  $k$ -dimensional Euclidean space. Let  $y_u$  denote the response value obtained as a result of applying the  $u$ th setting of  $\mathbf{x}$ , namely  $\mathbf{x}_u = (x_{u1}, x_{u2}, \dots, x_{uk})'$  ( $u = 1, 2, \dots, n$ ). From Eq. (1), we then have

$$y_u = f'(\mathbf{x}_u)\boldsymbol{\beta} + \epsilon_u, \quad u = 1, 2, \dots, n \quad (5)$$

where  $\epsilon_u$  denotes the error term at the  $u$ th experimental run. Model (5) can be expressed in matrix form as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (6)$$

where  $\mathbf{y} = (y_1, y_2, \dots, y_n)'$ ,  $\mathbf{X}$  is a matrix of order  $n \times p$  whose  $u$ th row is  $f'(\mathbf{x}_u)$ , and  $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)'$ . Note that the first column of  $\mathbf{X}$  is the column of ones  $\mathbf{1}_n$ .

Assuming that  $\boldsymbol{\epsilon}$  has a zero mean and a variance-covariance matrix given by  $\sigma^2\mathbf{I}_n$ , the so-called ordinary least-squares estimator of  $\boldsymbol{\beta}$  is (see e.g., Ref 1, Section 2.3)

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}. \quad (7)$$

The variance-covariance matrix of  $\hat{\boldsymbol{\beta}}$  is then of the form

$$\begin{aligned} \text{Var}(\hat{\boldsymbol{\beta}}) &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\sigma^2\mathbf{I}_n)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}. \end{aligned} \quad (8)$$

Using  $\hat{\boldsymbol{\beta}}$ , an estimate,  $\hat{\mu}(\mathbf{x}_u)$ , of the mean response at  $\mathbf{x}_u$  is obtained by replacing  $\boldsymbol{\beta}$  by  $\hat{\boldsymbol{\beta}}$ , that is,

$$\hat{\mu}(\mathbf{x}_u) = f'(\mathbf{x}_u)\hat{\boldsymbol{\beta}}, \quad u = 1, 2, \dots, n. \quad (9)$$

The quantity  $f'(\mathbf{x}_u)\hat{\boldsymbol{\beta}}$  also gives the so-called predicted response,  $\hat{y}(\mathbf{x}_u)$ , at the  $u$ th design point ( $u = 1, 2, \dots, n$ ). In general, at any point,  $\mathbf{x}$ , in an experimental region, denoted by  $\mathcal{R}$ , the predicted response  $\hat{y}(\mathbf{x})$  is

$$\hat{y}(\mathbf{x}) = f'(\mathbf{x})\hat{\boldsymbol{\beta}}, \quad \mathbf{x} \in \mathcal{R}. \quad (10)$$

Since  $\hat{\boldsymbol{\beta}}$  is an unbiased estimator of  $\boldsymbol{\beta}$ ,  $\hat{y}(\mathbf{x})$  is an unbiased estimator of  $f'(\mathbf{x})\boldsymbol{\beta}$ , which is the mean response at  $\mathbf{x} \in \mathcal{R}$ . Using Eq. (8), the variance of  $\hat{y}(\mathbf{x})$  is of the form

$$\text{Var}[\hat{y}(\mathbf{x})] = \sigma^2 f'(\mathbf{x})(\mathbf{X}'\mathbf{X})^{-1}f(\mathbf{x}). \quad (11)$$

The proper choice of design is very important in any response surface investigation. This is true because the quality of prediction, as measured by the size of the prediction variance, depends on the design matrix  $\mathcal{D}$  as can be seen from formula (11). Furthermore, the determination of the optimum response amounts to finding the optimal value of  $\hat{y}(\mathbf{x})$  over the region  $\mathcal{R}$ . It is therefore imperative that the prediction variance in Eq. (11) be as small as possible provided that the postulated model in Eq. (1) does not suffer from lack of fit (for a study of lack of fit of a fitted response model, see e.g., Ref 1, Section 2.6).

### Some Common Design Properties

The choice of design depends on the properties it is required, or desired, to have. Some of the design properties considered in the early development of RSM include the following:

#### Orthogonality

A design  $\mathcal{D}$  is said to be orthogonal if the matrix  $\mathbf{X}'\mathbf{X}$  is diagonal, where  $\mathbf{X}$  is the model matrix in Eq. (6). The advantage of this approach is that the elements of  $\hat{\boldsymbol{\beta}}$  will be uncorrelated because the off-diagonal elements of  $\text{Var}(\hat{\boldsymbol{\beta}})$  in Eq. (8) will be zero. If the error vector  $\boldsymbol{\epsilon}$  in Eq. (6) is assumed to be normally distributed as  $N(0, \sigma^2\mathbf{I}_n)$ , then these elements will be also stochastically independent. This makes it easier to test the significance of the unknown parameters in the model.

#### Rotatability

A design  $\mathcal{D}$  is said to be rotatable if the prediction variance in Eq. (11) is constant at all points that are equidistant from the design center, which, by a proper coding of the control variables, can be chosen to be the point at the origin of the  $k$ -dimensional coordinates

system. It follows that  $\text{Var}[\hat{y}(x)]$  is constant at all points that fall on the surface of a hypersphere centered at the origin, if the design is rotatable. The advantage of this property is that the prediction variance remains unchanged under any rotation of the coordinate axes. In addition, if optimization of  $\hat{y}(x)$  is desired on concentric hyperspheres, as in the application of ridge analysis, which will be discussed later, then it would be desirable for the design to be rotatable. This makes it easier to compare the values of  $\hat{y}(x)$  on a given hypersphere as all such values have the same variance.

The necessary and sufficient condition for a design to be rotatable was given by Box and Hunter.<sup>2</sup> More recently, Khuri<sup>3</sup> introduced a measure of rotatability as a function of the so-called moments of the design under consideration (see e.g., Appendix 2B in Ref 1). The function is expressible as a percentage taking large values for a high degree of rotatability. The value 100 is attained when the design is rotatable. The advantages of this measure are:

1. The ability to compare designs on the basis of rotatability.
2. The assessment of the extent of departure from rotatability of a design whose rotatability may be 'sacrificed' to satisfy another desirable design property.
3. The ability to improve rotatability by a proper augmentation of a nonrotatable design.

### Uniform Precision

A rotatable design is said to have the additional uniform precision property if  $\text{Var}[\hat{y}(x)]$  at the origin is equal to its value at a distance of one from the origin. This property, which was also introduced by Box and Hunter,<sup>2</sup> provides for an approximate uniform distribution of the prediction variance inside a hypersphere of radius one. This helps in producing some stability in the prediction variance in the vicinity of the design center.

### Design Robustness

Box and Draper<sup>4</sup> listed several additional design properties that pertain to detection of lack of fit, generation of satisfactory distribution of information throughout the experimental region, estimation of the error variance, insensitivity to outliers and to errors made in the actual implementation of the settings of the control variables. These properties provide guidelines for the choice of a design (i.e., a 'wish list'). It is not, however, expected that a single design will satisfy all of these properties. A design is said to

be robust if its properties are not severely impacted by failures to satisfy the assumptions made about the model and the error distribution.

### Design Optimality

Optimal designs are those that are constructed on the basis of a certain optimality criterion that pertains to the 'closeness' of the predicted response,  $\hat{y}(x)$ , to the mean response,  $\mu(x)$ , over a certain region of interest denoted by  $\mathcal{R}$ . The design criteria that address the minimization of the variance associated with the estimation of model (1)'s unknown parameters are called variance-related criteria. The most prominent of such criteria is the D-optimality criterion that maximizes the determinant of the matrix  $X'X$ . This amounts to the minimization of the size of the confidence region on the vector  $\beta$  in model (6). Actually, this criterion results in the so-called discrete D-optimal design as compared with the continuous D-optimal design, which was introduced by Jack Kiefer. The latter design is based on the notion that a design represents a probability measure defined on the region  $\mathcal{R}$ . A discrete design is then treated as a special case consisting of a collection of  $n$  points in  $\mathcal{R}$  that are not necessarily distinct.

Another variance-related criterion that is closely related to D-optimality is the G-optimality criterion which requires the minimization of the maximum over  $\mathcal{R}$  of the prediction variance in Eq. (11). Kiefer developed a continuous counterpart of this criterion and showed that it is equivalent to the continuous D-optimality criterion. This result is based on the equivalence theorem proved by Kiefer and Wolfowitz.<sup>5</sup> Other less-known variance-related criteria include A-optimality and E-optimality. See Ref 6, Chapter 4, for a description of these criteria.

These variance-related criteria are often referred to as alphabetic optimality. They are meaningful when the fitted model in Eq. (1) represents the true relationship connecting  $y$  to its control variables.

### Designs for First- and Second-Degree Models

As was pointed out earlier in the Section on *Introduction and Some Preliminaries*, the first-degree model in Eq. (2) and second-degree model in Eq. (3) are the most-frequently used approximating polynomial models in classical RSM. Designs for fitting first-degree models are called first-order designs and those for fitting second-degree models are referred to as second-order designs.

### First-Order Designs

The most common first-order designs are  $2^k$  factorial ( $k$  is the number of control variables), Plackett–Burman, and simplex designs.

#### The $2^k$ Factorial Design

In a  $2^k$  factorial design, each control variable is measured at two levels, which can be coded to take the values,  $-1, 1$ , that correspond to the so-called low and high levels, respectively, of each variable. This design consists of all possible combinations of such levels of the  $k$  factors. Thus, each row of the design matrix  $\mathcal{D}$  in Eq. (4) consists of all 1s, all  $-1$ s, or a combination of 1s and  $-1$ s and represents a particular treatment combination. In this case, the number,  $n$ , of experimental runs is equal to  $2^k$  provided that no single design point is replicated more than once. For example, in a chemical experiment, the control variables are  $x_1 =$  temperature of a reaction measured at 250, 300 ( $^{\circ}\text{C}$ ),  $x_2 =$  pressure set at 10, 16 (psi), and  $x_3 =$  time of the reaction taken at 4, 8 (minutes). The coded settings,  $\pm 1$ , for  $x_1, x_2, x_3$  are attained through the linear transformation,

$$x_1 = \frac{\text{temperature} - 275}{25} \quad (12)$$

$$x_2 = \frac{\text{pressure} - 13}{3} \quad (13)$$

$$x_3 = \frac{\text{time} - 6}{2}. \quad (14)$$

The corresponding  $2^3$  design matrix is of order  $8 \times 3$  of the form

$$\mathcal{D} = \begin{bmatrix} -1 & -1 & -1 \\ 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ -1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}. \quad (15)$$

If  $k$  is large ( $k \geq 5$ ), the  $2^k$  design requires a large number of design points. Since the number of unknown parameters in Eq. (2) is only  $k + 1$ , fractions of  $2^k$  can be considered to fit such a model. For example, we can consider a one-half fraction design that consists of one-half the number of points of a  $2^k$  design, or a one-fourth fraction design that consists of one-fourth the number of points of a  $2^k$  design. In general, a  $2^{-m}$ th fraction of a  $2^k$  design consists of  $2^{k-m}$  points from a full  $2^k$  design. Here,  $m$  is a positive integer such that  $2^{k-m} \geq k + 1$  so that all the

$k + 1$  parameters in model (2) can be estimated. The construction of fractions of a  $2^k$  design is carried out in a particular manner, a description of which can be found in several experimental design textbooks, such as Refs 7–9. See also Chapter 3 in Ref 1.

#### The Plackett–Burman Design

The Plackett–Burman design allows two levels for each of the  $k$  control variables, just like a  $2^k$  design, but requires a much smaller number of experimental runs, especially if  $k$  is large. It is therefore more economical than the  $2^k$  design. Its number,  $n$ , of design points is equal to  $k + 1$ , which is the same as the number of parameters in model (2). In this respect, the design is said to be saturated because its number of design points is equal to the number of parameters to be estimated in the model. Furthermore, this design is available only when  $n$  is a multiple of 4. Therefore, it can be used when the number,  $k$ , of control variables is equal to 3, 7, 11, 15,....

To construct a Plackett–Burman design in  $k$  variables, a row is first selected whose elements are equal to  $-1$  or  $1$  such that the number of 1s is  $\frac{k+1}{2}$  and the number of  $-1$ s is  $\frac{k-1}{2}$ . The next  $k - 1$  rows are generated from the first row by shifting it cyclically one place to the right  $k - 1$  times. Then, a row of negative ones is added at the bottom of the design. For example, for  $k = 7$ , the design matrix,  $\mathcal{D}$ , has eight points whose coordinates are  $x_1, x_2, \dots, x_7$  and is of the form

$$\mathcal{D} = \begin{bmatrix} 1 & 1 & 1 & -1 & 1 & -1 & -1 \\ -1 & 1 & 1 & 1 & -1 & 1 & -1 \\ -1 & -1 & 1 & 1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 & 1 & 1 & -1 \\ -1 & 1 & -1 & -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 & -1 & -1 & 1 \\ -1 & -1 & -1 & -1 & -1 & -1 & -1 \end{bmatrix}. \quad (16)$$

Design arrangements for  $k = 3, 7, 11, \dots, 99$  factors can be found in Ref 10.

#### The Simplex Design

The simplex design is also a saturated design with  $n = k + 1$  points. Its design points are located at the vertices of a  $k$ -dimensional regular-sided figure, or a simplex, characterized by the fact that the angle,  $\theta$ , which any two points make with the design center (located at the origin of the coordinates system) is such that  $\cos \theta = -\frac{1}{k}$ . For example, for  $k = 2$ , the simplex design consists of the vertices of an equilateral triangle whose center is  $(0, 0)$ , and for  $k = 3$ , the design points are the vertices of a tetrahedron centered at  $(0, 0, 0)$ .

Box<sup>11</sup> presented a procedure for constructing a simplex design using a particular pattern of a one-factor-at-a-time design. This procedure is also explained in Ref 1, Section 3.3.5. The simplex design is a less frequently used design than  $2^k$  or Plackett–Burman designs. This is because the actual settings in a simplex design are, in general, difficult to attain exactly in a real experimental situation.

All the above designs ( $2^k$  or fractions of, Plackett–Burman, simplex) share the same property of being orthogonal. For a first-order design, orthogonality can be achieved if the design matrix  $\mathcal{D}$  is such that  $\mathcal{D}'\mathcal{D}$  is diagonal. We may recall that an orthogonal design causes the variance–covariance matrix of  $\hat{\beta}$ , the least-squares estimator of the vector  $\beta$  of unknown parameters in the model, to be diagonal (if the error vector  $\epsilon$  is assumed to have a zero mean and a variance–covariance matrix  $\sigma^2 I_n$ ). This means that the elements of  $\hat{\beta}$  are uncorrelated, hence independent under the normality assumption on  $\epsilon$ . Furthermore, it can be shown that under an orthogonal design, the variances of the elements of  $\hat{\beta}$  have minimum values (see Section 3.3 in Ref 1). This means that an orthogonal first-order design provides maximum precision for estimating the unknown parameters in model (2).

### Second-Order Designs

The number of parameters in the second-degree model in Eq. (3) is  $p = 1 + 2k + \frac{1}{2}k(k - 1)$ . Hence, the number of distinct design points of a second-order design must be at least equal to  $p$ . The design settings are usually coded so that  $\frac{1}{n} \sum_{u=1}^n x_{ui} = 0$  and  $\frac{1}{n} \sum_{u=1}^n x_{ui}^2 = 1$ ,  $i = 1, 2, \dots, k$ , where  $n$  is the number of experimental runs and  $x_{ui}$  is the  $u$ th setting of the  $i$ th control variable ( $u = 1, 2, \dots, n$ ).

The most frequently used second-order designs are the  $3^k$  factorial, central composite, and the Box–Behnken designs.

### The $3^k$ Factorial Design

The  $3^k$  factorial design consists of all the combinations of the levels of the  $k$  control variables which have three levels each. If the levels are equally spaced, then they can be coded so that they correspond to  $-1, 0, 1$ . The number of experimental runs for this design is  $3^k$ , which can be very large for a large  $k$ . Fractions of a  $3^k$  design can be considered to reduce the cost of running such an experiment. A general procedure for constructing fractions of  $3^k$  is described in Montgomery<sup>9</sup> (Chapter 9). See also Ref 12, Appendix 2.

### The Central Composite Design (CCD)

This is perhaps the most popular of all second-order designs. It was first introduced in Ref 13. This design consists of the following three portions:

1. A complete (or a fraction of)  $2^k$  factorial design whose factors' levels are coded as  $-1, 1$ . This is called the factorial portion.
2. An axial portion consisting of  $2k$  points arranged so that two points are chosen on the axis of each control variable at a distance of  $\alpha$  from the design center (chosen as the point at the origin of the coordinates system).
3.  $n_0$  center points.

Thus, the total number of design points in a CCD is  $n = 2^k + 2k + n_0$ . For example, a CCD for  $k = 2$ ,  $\alpha = \sqrt{2}$ ,  $n_0 = 2$  has the form

$$\mathcal{D} = \begin{bmatrix} -1 & -1 \\ 1 & -1 \\ -1 & 1 \\ 1 & 1 \\ -\sqrt{2} & 0 \\ \sqrt{2} & 0 \\ 0 & -\sqrt{2} \\ 0 & \sqrt{2} \\ 0 & 0 \\ 0 & 0 \end{bmatrix}. \quad (17)$$

We note that the CCD is obtained by augmenting a first-order design, namely, the  $2^k$  factorial with additional experimental runs, namely, the  $2k$  axial points and the  $n_0$  center-point replications. Thus, this design is developed in a manner consistent with the sequential nature of a response surface investigation in starting with a first-order design, to fit a first-degree model, followed by the addition of design points to fit the larger second-degree model. The first-order design serves in a preliminary phase to get initial information about the response system and to assess the importance of the factors in a given experiment. The additional experimental runs are chosen for the purpose of getting more information that can lead to the determination of optimum operating conditions on the control variables using the second-degree model.

The values of  $\alpha$  (or the axial parameter) and  $n_0$ , the number of center-point replications, are chosen so that the CCD can acquire certain desirable properties. For example, choosing  $\alpha = F^{1/4}$ , where  $F$  denotes the number of points in the factorial portion, causes the CCD to be rotatable. The value of  $n_0$  can

then be chosen so that the CCD can achieve either the orthogonality property or the uniform precision property. Note that orthogonality of a second-order design is attainable only after expressing model (3) in terms of orthogonal polynomials as explained in Box and Hunter<sup>2</sup> (pp. 200–201). See also Khuri and Cornell<sup>1</sup> (Section 4.3). In particular, Table 4.3 in Khuri and Cornell's book can be used to determine the value of  $n_0$  for a rotatable CCD to have either the additional orthogonality property or the uniform precision property.

### The Box–Behnken Design

This design was developed by Box and Behnken.<sup>14</sup> It provides three levels for each factor and consists of a particular subset of the factorial combinations from the  $3^k$  factorial design. The actual construction of such a design is described in the three RSM books Box and Draper,<sup>15</sup> Section 15.4, Khuri and Cornell,<sup>1</sup> Section 4.5.2, and Myers and Montgomery,<sup>16</sup> Section 7.4.7.

The use of the Box–Behnken design is popular in industrial research because it is an economical design and requires only three levels for each factor where the settings are  $-1, 0, 1$ . Some Box–Behnken designs are rotatable, but, in general, this design is not always rotatable. Box and Behnken<sup>14</sup> list a number of design arrangements for  $k = 3, 4, 5, 6, 7, 9, 10, 11, 12$ , and 16 factors.

Other second-order designs are available but are not as frequently used as the ones we have already mentioned. Some of these designs include Hoke<sup>17</sup> designs, Box–Draper saturated designs (see Ref 18), uniform shell designs by Doehlert,<sup>19</sup> and hybrid designs by Roquemore.<sup>20</sup>

### Determination of Optimum Conditions

One of the main objectives of RSM is the determination of the optimum settings of the control variables that result in a maximum (or a minimum) response over a certain region of interest,  $\mathcal{R}$ . This requires having a 'good' fitting model that provides an adequate representation of the mean response because such a model is to be utilized to determine the value of the optimum. Optimization techniques used in RSM depend on the nature of the fitted model. For first-degree models, the method of steepest ascent (or descent) is a viable technique for sequentially moving toward the optimum response. This method is explained in detail in Myers and Montgomery,<sup>16</sup> Khuri and Cornell<sup>1</sup> (Chapter 5), and Box and Draper<sup>15</sup> (Chapter 6). Myers and Khuri<sup>21</sup> developed certain improvements regarding the stopping rule used in the execution of this method.

Since the first-degree model is usually used at the preliminary stage of a response surface investigation, we shall only mention here optimization techniques that are applicable to second-degree models. Such models are used after a series of experiments have been sequentially carried out leading up to a region that is believed to contain the location of the optimum response.

### Optimization of a Second-Degree Model

Let us consider the second-degree model in Eq. (3), which can be written as

$$y = \beta_0 + \mathbf{x}'\boldsymbol{\beta}_* + \mathbf{x}'\mathbf{B}\mathbf{x} + \epsilon \quad (18)$$

where  $\mathbf{x} = (x_1, x_2, \dots, x_k)'$ ,  $\boldsymbol{\beta}_* = (\beta_1, \beta_2, \dots, \beta_k)'$ , and  $\mathbf{B}$  is a symmetric matrix of order  $k \times k$  whose  $i$ th diagonal element is  $\beta_{ii}$  ( $i = 1, 2, \dots, k$ ), and its  $(i, j)$ th off-diagonal element is  $\frac{1}{2}\beta_{ij}$  ( $i, j = 1, 2, \dots, k; i \neq j$ ). If  $n$  observations are obtained on  $y$  using a design matrix  $\mathcal{D}$  as in Eq. (4), then Eq. (18) can be written in vector form as in Eq. (6), where the parameter vector  $\boldsymbol{\beta}$  consists of  $\beta_0$  and the elements of  $\boldsymbol{\beta}_*$  and  $\mathbf{B}$ . Assuming that  $E(\epsilon) = 0$  and  $\text{Var}(\epsilon) = \sigma^2\mathbf{I}_n$ , the least-squares estimate of  $\boldsymbol{\beta}$  is  $\hat{\boldsymbol{\beta}}$  as given in Eq. (7). The predicted response at a point  $\mathbf{x}$  in the region  $\mathcal{R}$  is then of the form

$$\hat{y}(\mathbf{x}) = \hat{\beta}_0 + \mathbf{x}'\hat{\boldsymbol{\beta}}_* + \mathbf{x}'\hat{\mathbf{B}}\mathbf{x} \quad (19)$$

where  $\hat{\beta}_0$  and the elements of  $\hat{\boldsymbol{\beta}}_*$  and  $\hat{\mathbf{B}}$  are the least-squares estimates of  $\beta_0$  and the corresponding elements of  $\boldsymbol{\beta}_*$  and  $\mathbf{B}$ , respectively.

### The Method of Ridge Analysis

This is a useful procedure for optimizing the predicted response based on the fitted second-degree model in Eq. (19). It was introduced by Hoerl<sup>22</sup> and formalized by Draper.<sup>23</sup> This method optimizes  $\hat{y}(\mathbf{x})$  in Eq. (19) subject to  $\mathbf{x}$  being on the surface of a hypersphere of radius  $r$  and centered at the origin, namely,

$$\sum_{i=1}^k x_i^2 = r^2. \quad (20)$$

This constrained optimization is conducted using several values of  $r$  corresponding to hyperspheres contained within the region  $\mathcal{R}$ . The rationale for doing this is to get information about the optimum at various distances from the origin within  $\mathcal{R}$ .

Since this optimization is subject to the equality constraint given by Eq. (20), the method of Lagrange

multipliers can be used to search for the optimum. Let us therefore consider the function

$$H = \hat{\beta}_0 + \mathbf{x}'\hat{\beta}_* + \mathbf{x}'\hat{\mathbf{B}}\mathbf{x} - \lambda(\mathbf{x}'\mathbf{x} - r^2) \quad (21)$$

where  $\lambda$  is a Lagrange multiplier. Differentiating  $H$  with respect to  $\mathbf{x}$  and equating the derivative to zero, we get

$$\hat{\beta}_* + 2(\hat{\mathbf{B}}\mathbf{x} - \lambda\mathbf{x}) = 0. \quad (22)$$

Solving for  $\mathbf{x}$ , we obtain

$$\mathbf{x} = -\frac{1}{2}(\hat{\mathbf{B}} - \lambda\mathbf{I}_n)^{-1}\hat{\beta}_*. \quad (23)$$

The solution in Eq. (23) represents just a stationary point of  $\hat{y}(\mathbf{x})$ . A maximum (minimum) is achieved at this point if the Hessian matrix, that is, the matrix  $\frac{\partial}{\partial \mathbf{x}} \left[ \frac{\partial H}{\partial \mathbf{x}'} \right]$  of second-order partial derivatives of  $H$  with respect to  $\mathbf{x}$  is negative definite (positive definite). From Eq. (22), this matrix is given by

$$\frac{\partial}{\partial \mathbf{x}} \left[ \frac{\partial H}{\partial \mathbf{x}'} \right] = 2(\hat{\mathbf{B}} - \lambda\mathbf{I}_n). \quad (24)$$

Therefore, to achieve a maximum, Draper<sup>23</sup> suggested that  $\lambda$  be chosen larger than the largest eigenvalue of  $\hat{\mathbf{B}}$ . Such a choice causes  $\hat{\mathbf{B}} - \lambda\mathbf{I}_n$  to be negative definite. Choosing  $\lambda$  smaller than the smallest eigenvalue of  $\hat{\mathbf{B}}$  causes  $\hat{\mathbf{B}} - \lambda\mathbf{I}_n$  to be positive definite, which results in a minimum. Thus, by choosing several values of  $\lambda$  in this fashion, we can, for each  $\lambda$ , find the location of the optimum (maximum or minimum) by using formula (23) and hence obtain the value of  $\mathbf{x}'\mathbf{x} = r^2$ . The solution from Eq. (23) is feasible provided that  $r$  corresponds to a hypersphere that falls entirely within the region  $\mathcal{R}$ . The optimal value of  $\hat{y}(\mathbf{x})$  is computed by substituting  $\mathbf{x}$  from Eq. (23) into the right-hand side of Eq. (19). This process generates plots of  $\hat{y}$  and  $x_i$  against  $r$  ( $i = 1, 2, \dots, k$ ). These plots are useful in determining, at any given distance  $r$  from the origin, the value of the optimum as well as its location. More details concerning this method can be found in Myers and Montgomery,<sup>16</sup> Khuri and Cornell<sup>1</sup> (Section 5.7), and Box and Draper<sup>15</sup> (Chapter 19).

Since the method of ridge analysis optimizes  $\hat{y}(\mathbf{x})$  on concentric hyperspheres within the region  $\mathcal{R}$ , its application is meaningful provided that the prediction variance in formula (11) is constant on the surface of any given hypersphere. This calls for the use of a rotatable design to fit model (18). If, however, the design is not rotatable, then the prediction variance can vary appreciably on the surface of a hypersphere, which may lead to poor

estimates of the optimum response. For this reason, Khuri and Myers<sup>24</sup> proposed a modification of the method of ridge analysis whereby the optimization of  $\hat{y}(\mathbf{x})$  is carried out under an added constraint on the size of the prediction variance. This modification can produce better optimization results when the design used to fit model (18) is not rotatable. More recently, Paul and Khuri<sup>25</sup> extended the use of Khuri and Myers' modification to linear models where the error variances are heterogeneous and also to generalized linear models.

## PART II. FURTHER DEVELOPMENTS AND THE TAGUCHI ERA: 1976–1999

### Multiresponse Experiments

In a multiresponse experiment, measurements on several responses are obtained for each setting of a group of control variables. Examples of multiresponse experiments are numerous, for example, a chemical engineer may be interested in maximizing the yield while minimizing the cost of a certain chemical process. Refs 26–28 cited several papers in which multiresponse experiments were studied. While analyzing the data from a multiresponse experiment, special attention should be given to the correlated nature of the data within experimental runs. Usually, it is assumed that the responses are correlated within runs but independent otherwise.

Suppose that  $n$  is the number of experimental runs and  $q$  is the number of responses. Then the  $i$ th response may be modeled as (see Ref 1, pp. 252–254)

$$y_i = \mathbf{X}_i\beta_i + \epsilon_i, \quad i = 1, \dots, q \quad (25)$$

where  $y_i$  is an  $n \times 1$  vector of observations on the  $i$ th response,  $\mathbf{X}_i$  is an  $n \times p_i$  known matrix of rank  $p_i$ ,  $\beta_i$  is a vector of  $p_i$  unknown regression parameters, and  $\epsilon_i$  is a vector of random errors associated with the  $i$ th response. Using matrix notation, the above model can be expressed as

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon \quad (26)$$

where  $\mathbf{Y} = [y'_1, \dots, y'_q]'$ ,  $\mathbf{X}$  is the block-diagonal matrix,  $\text{diag}(\mathbf{X}_1, \dots, \mathbf{X}_q)$ ,  $\beta = [\beta'_1, \dots, \beta'_q]'$ , and  $\epsilon = [\epsilon'_1, \dots, \epsilon'_q]'$ . It is assumed that  $E(\epsilon) = 0$  and the variance-covariance matrix  $\text{Var}(\epsilon) = \Sigma \otimes \mathbf{I}$ , where  $\mathbf{I}$  is the  $n \times n$  identity matrix. The best linear unbiased estimate of  $\beta$  is given by

$$\hat{\beta} = [\mathbf{X}'(\Sigma^{-1} \otimes \mathbf{I})\mathbf{X}]^{-1}\mathbf{X}'(\Sigma^{-1} \otimes \mathbf{I})\mathbf{Y}. \quad (27)$$

An estimate of  $\Sigma$  is used to find  $\hat{\beta}$  when  $\Sigma$  is unknown. A commonly used estimate of  $\Sigma$  is the one proposed by Zellner.<sup>29</sup>

In selecting a design optimality criterion for multiresponse experiments, one needs to consider all the responses simultaneously. Draper and Hunter<sup>30</sup> proposed a criterion for the estimation of the unknown parameters in a multiresponse situation. Their criterion was used for selecting additional experimental runs after a certain number of runs have already been chosen. The approach used was Bayesian, with the unknown parameter vector having a uniform prior distribution. The variance–covariance matrix of the responses was assumed to be known. Box and Draper<sup>31</sup> later extended Draper and Hunter's<sup>30</sup> criterion by considering response surface designs with blocks. The most common criterion for multiresponse designs is the D-optimality criterion. One such criterion was developed by Fedorov<sup>32</sup> for linear multiresponse designs. Fedorov's procedure was sequential in nature and was used for constructing D-optimal designs. However, it required knowledge of the variance–covariance matrix associated with the several response variables. Wijesinha and Khuri<sup>33</sup> later modified Fedorov's procedure by using an estimate of  $\Sigma$  at each step of the sequential process. Some recent works on D-optimal designs for linear multiresponse models include those of Krafft and Schaefer,<sup>34</sup> Bischoff,<sup>35</sup> Chang,<sup>36,37</sup> Imhof,<sup>38</sup> and Atashgah and Seifi.<sup>39</sup> Locally D-optimal designs for describing the behavior of a biological system were constructed by Hatzis and Larntz<sup>40</sup> for nonlinear multiresponse models. Other design criteria for linear multiresponse models include the power criterion of Wijesinha and Khuri<sup>41</sup> and the robustness criterion of Wijesinha and Khuri<sup>42</sup> and Yue.<sup>43</sup>

The model given in Eq. (26) is said to suffer from lack of fit if it does not represent the  $q$  true mean responses adequately. Due to the correlated nature of the responses in a multiresponse situation, lack of fit of one response may affect the fit of the other responses. Khuri<sup>44</sup> proposed a multivariate test of lack of fit that considers all the  $q$  responses simultaneously. His test was based on Roy's union intersection principle<sup>45</sup> (Chapters 4 and 5) and required that replicated observations on all responses be taken at some points of the experimental region. Levy and Neill<sup>46</sup> considered additional multivariate lack of fit tests and used simulations to compare the power functions of these tests.

One important objective of multiresponse experimentation is to determine the optimum operating conditions on the control variables that lead to the simultaneous optimization of the predicted values of

the responses within a region of interest. In multiresponse experiments, the meaning of 'optimum' is sometimes unclear as there is no unique way to order the multiresponse data. Conditions that are optimal for one response may be far from optimal or even physically impractical for the other responses from the experimental point of view. For example, in a dose–response experiment, where both efficacy and toxicity responses are measured at each dose, the experimenter may wish to find the dose level of the drug(s) which simultaneously maximize efficacy while minimizing toxicity. Common knowledge is that as the dose of a drug increases, so do its efficacy and toxic side effects. This implies that the efficacy response is optimized at a higher dose level, whereas the toxicity response is minimized at a lower dose level. Thus, it is difficult to identify dose levels which are optimal for both responses. The problem of simultaneous optimization for linear multiresponse models was addressed in Refs 47–50.

Lind et al.<sup>47</sup> developed a graphical approach in which contours of all the responses were superimposed on each other and the region where operating conditions were 'near' optimal for all the responses was identified. As the number of responses and control factors increases, finding the optimum graphically becomes infeasible. Myers and Carter<sup>51</sup> proposed the dual response system consisting of a primary response and a secondary response. The procedure involved determining the operating conditions for which the primary response was optimized while the secondary response was constrained to be equal to some prespecified value. Refs 52–55 provided various extensions to the dual response approach. Harrington<sup>48</sup> developed the desirability approach to multiresponse optimization. In his algorithm, exponential type transformations were used to transform each of the responses into desirability functions. Derringer and Suich<sup>49</sup> later generalized the transformations and developed more flexible desirability functions. The individual desirability functions were then incorporated into a single function, which gave the desirability for the whole set of responses. Both Refs 48 and 49 used the geometric mean of the individual desirability functions to construct a single overall desirability function, which was maximized to determine the operating conditions. Del Castillo et al.<sup>56</sup> modified the desirability approach of Ref 49 such that both the desirability function and its first derivative were continuous. However, their approach ignored variations and correlations existing among the responses. Wu<sup>57</sup> presented an approach based on the modified double-exponential desirability

function taking into account correlations among the responses.

Multiresponse optimization using the so-called generalized distance approach was introduced by Khuri and Conlon.<sup>50</sup> The main characteristic of this approach was that it took into account the heterogeneity of the variances of the responses and also the correlated nature of the responses. If the individual optima were not attained at the same settings of the control factors, then compromise conditions on the input variables that are 'favorable' to all the mean responses were determined. The deviation from the ideal optimum was measured by a distance function expressed in terms of the estimated mean responses along with their variance-covariance matrix. By minimizing such a distance function, Khuri and Conlon arrived at a set of conditions for a 'compromise optimum'. Vining<sup>58</sup> proposed a mean squared error method to determine the compromise optimum for a multiresponse experiment. Pignatiello<sup>59</sup> and Ames et al.<sup>60</sup> also proposed approaches based on the squared error loss. A comprehensive survey of the various methods of multiresponse optimization was presented in Ref 61.

### Taguchi's Robust Parameter Design

Robust parameter design is a well-established engineering technique to increase the quality of a product by making it robust/insensitive to the uncontrollable variations present in the production process. Since the introduction of parameter design in the United States by Genichi Taguchi during the 1980s, a multitude of papers by Kacker,<sup>62</sup> Taguchi and Wu,<sup>63</sup> Taguchi,<sup>64</sup> Nair and Shoemaker<sup>65</sup> and books authored by Khuri and Cornell (Chapter 11),<sup>1</sup> Taguchi,<sup>66</sup> Phadke,<sup>67</sup> Wu and Hamada,<sup>68</sup> and Myers and Montgomery,<sup>69</sup> and several other authors, have been written on the topic. Review articles by Myers et al.<sup>70</sup> and Robinson et al.<sup>71</sup> cite several papers based on robust parameter designs. Taguchi proposed that the input variables in an experiment were of two types, (1) control factors: easy to control and (2) noise factors: difficult to control. These difficult-to-control noise factors are the cause of variations in a production process. The main aim of parameter design is to determine the settings of the control factors for which the process response is robust to the variability in the system caused by the noise factors. To achieve this goal, Taguchi advocated the use of crossed arrays by crossing an orthogonal array of control variables (inner array) with an orthogonal array of noise variables (outer array). Taguchi identified that there were three specific goals in an experiment:

1. The smaller, the better: minimizing the response.
2. The larger, the better: maximizing the response.
3. Target is best: achieving a given target value.

For each of the different goals, Taguchi defined performance criteria known as signal-to-noise (S/N) ratios that took into account both the process mean and the variance. Each set of settings of the control variables contained  $n$  runs in the noise variables from the outer array. For each of the three different goals, he defined the S/N ratios as follows:

1. The smaller, the better:  $-10 \log[\frac{1}{n} \sum_{i=1}^n y_i^2]$ .
2. The larger, the better:  $-10 \log[1/n \sum_{i=1}^n 1/y_i^2]$ .
3. Target is best:  $-10 \log(\frac{s^2}{\bar{y}^2})$ , where  $\bar{y}$  is the sample mean and  $s^2$  is the sample variance.

All the above S/N ratios are to be maximized. Although the Taguchi method was a significant step toward quality improvement, it received a number of criticisms. It was pointed out (see Ref 69) that in the Taguchi methodology (1), interactions among the control factors were not estimated, (2) large numbers of experimental runs were required, (3) S/N ratios were unable to distinguish between inputs affecting process mean from those affecting the variance. Some of the authors who discussed the Taguchi methodology in detail and offered criticisms were Myers and Montgomery,<sup>69</sup> Box,<sup>72,73</sup> Easterling,<sup>74</sup> Pignatiello and Ramberg,<sup>75</sup> Nair and Pregibon,<sup>76</sup> Welch *et al.*<sup>77</sup>, and Nair.<sup>78</sup>

### Response Surface Approach to Robust Parameter Design

Two response surface approaches to parameter design were introduced during the 1990s. The approaches were (1) the dual response approach and (2) the single model approach. In the dual response approach, separate models were fitted to the process mean and the process variance. While in the single model approach, as the name suggests, a single model containing both the noise and the control variables was fitted to the process response.

Vining and Myers<sup>79</sup> were the first to propose that Taguchi's aim of keeping the mean on target while minimizing the process variance could also be achieved in a response surface framework. They fitted separate second-degree models to the process

mean ( $\mu$ ) and the process variance ( $\sigma^2$ ),

$$\hat{\mu} = b_0 + \mathbf{x}'\mathbf{b} + \mathbf{x}'\mathbf{B}\mathbf{x}, \quad (28)$$

$$\hat{\sigma}^2 = c_0 + \mathbf{x}'\mathbf{c} + \mathbf{x}'\mathbf{C}\mathbf{x} \quad (29)$$

where  $b_0, \mathbf{b}, \mathbf{B}, c_0, \mathbf{c}, \mathbf{C}$  were the estimates of the coefficients. Using the dual response optimization of Ref 51,  $\hat{\sigma}^2$  was minimized while keeping the process mean at target.

Del Castillo and Montgomery<sup>80</sup> suggested the use of nonlinear programming to solve a similar optimization problem as proposed by Vining and Myers,<sup>79</sup> but replacing the equality constraints by inequalities. Criticizing the use of Lagrangian multipliers and equality constraints, Lin and Tu<sup>81</sup> proposed a procedure based on the mean squared error (MSE) criterion. Using the same examples as discussed by Vining and Myers<sup>79</sup> and Del Castillo and Montgomery,<sup>80</sup> they showed that more reduction in the variance was possible by introducing a little bias. One criticism of using the MSE was that no restriction was placed on the distance of the mean from the target. Addressing this problem, Copeland and Nelson<sup>82</sup> minimized the variance while keeping the distance between the mean and the target less than some specified quantity,  $\Delta$ . For processes where it was important to keep the mean near the target,  $\Delta$  was chosen to be small. Various other extensions to the dual response approach were suggested by Fan,<sup>55</sup> Kim and Lin,<sup>83</sup> Del Castillo et al.,<sup>84</sup> Koksoy and Doganaksoy,<sup>85</sup> Kim and Cho,<sup>86</sup> and Tang and Xu.<sup>87</sup>

To overcome the shortcomings (requirement of too many runs and being unable to fit interaction terms) of Taguchi's crossed array, Welch et al.<sup>77</sup> proposed a combined array, which was a single experimental design for both the control and the noise variables. The combined array was shown to be more economical than the crossed arrays of Taguchi (see Refs 77, 88, 89). Myers et al.<sup>90</sup> used the combined array of Welch et al.<sup>77</sup> to fit a single model containing both the control and noise variables to the response variable,

$$y(\mathbf{x}, \mathbf{z}) = \beta_0 + \mathbf{g}'(\mathbf{x})\boldsymbol{\beta} + \mathbf{z}'\boldsymbol{\delta} + \mathbf{g}'(\mathbf{x})\boldsymbol{\Delta}\mathbf{z} + \epsilon \quad (30)$$

where  $\mathbf{x}$  and  $\mathbf{z}$  are the control and noise variables, respectively. In the above model,  $\mathbf{g}'(\mathbf{x})$  is a row of the design matrix containing polynomial and interaction terms in the control variables,  $\boldsymbol{\beta}$  and  $\boldsymbol{\delta}$  are the vectors of regression coefficients for the control and noise variables, respectively, and  $\boldsymbol{\Delta}$  contains the interaction coefficients. They showed that although a single model in the noise and control variables was fitted to the

response, there were still two response surfaces for the process mean and the variance,

$$E[y(\mathbf{x}, \mathbf{z})] = \beta_0 + \mathbf{g}'(\mathbf{x})\boldsymbol{\beta} \quad (31)$$

and

$$\text{Var}[y(\mathbf{x}, \mathbf{z})] = [\boldsymbol{\delta}' + \mathbf{g}'(\mathbf{x})\boldsymbol{\Delta}] \text{Var}(\mathbf{z}) [\boldsymbol{\delta}' + \mathbf{g}'(\mathbf{x})\boldsymbol{\Delta}]' + \sigma_e^2. \quad (32)$$

To solve the parameter design problem, they chose the estimated squared error loss as the performance criterion,

$$E[\hat{y}(\mathbf{x}, \mathbf{z}) - T]^2 \quad (33)$$

where  $T$  was the prespecified target value, and minimized it with respect to  $\mathbf{x}$ . Myers et al.<sup>90</sup> proposed a linear mixed effects approach, in which the elements of  $\boldsymbol{\delta}$  and  $\boldsymbol{\Delta}$  in Eq. (30) were treated as random. Aggarwal and Bansal<sup>91</sup> and Aggarwal et al.<sup>92</sup> considered robust parameter designs involving both quantitative and qualitative factors. Brenneman and Myers<sup>93</sup> considered the single model in the control and noise variables to model the response. In their model, the noise variables were considered to be categorical in nature.

## PART III. EXTENSIONS AND NEW DIRECTIONS: 2000 ONWARDS

### Response Surface Models with Random Effects

The response surface models we have considered so far include only fixed polynomial effects. These models are suitable whenever the levels of the factors considered in a given experiment are of particular interest to the experimenter, for example, the temperature and concentrations of various chemicals in a certain chemical reaction. There are, however, other experimental situations where, in addition to the main control factors, the response may be subject to variations due to the presence of some random effects. For example, the raw material used in a production process may be obtained in batches selected at random from the warehouse supply. Because batches may differ in quality, the response model should include a random effect to account for the batch-to-batch variability. In this section, we consider response surface models which, in addition to the fixed polynomial effects, include random effects.

Let  $\mu(\mathbf{x})$  denote the mean of a response variable,  $y$ , at a point  $\mathbf{x} = (x_1, x_2, \dots, x_k)'$ . It is assumed that

$\mu(x)$  is represented by a polynomial model of degree  $d (\geq 1)$  of the form

$$\mu(x) = f'(x)\beta. \tag{34}$$

Suppose that the experimental runs used to estimate the model parameters in Eq. (34) are heterogeneous due to an extraneous source of variation referred to as a block effect. The experimental runs are therefore divided into  $b$  blocks of sizes  $n_1, n_2, \dots, n_b$ . Let  $n = \sum_{i=1}^b n_i$  be the total number of observations. The block effect is considered random, and the actual response value at the  $u$ th run is represented by the model

$$y_u = f'(x_u)\beta + z'_u\gamma + g'(x_u)\Lambda z_u + \epsilon_u, \quad u = 1, 2, \dots, n \tag{35}$$

where  $g'(x)$  is such that  $f'(x) = [1, g'(x_u)]$ ,  $x_u$  is the value of  $x$  at the  $u$ th run,  $z_u = (z_{u1}, z_{u2}, \dots, z_{ub})'$ , where  $z_{ui}$  is an indicator variable taking the value 1 if the  $u$ th trial is in the  $i$ th block and the value 0 otherwise,  $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_b)'$ , where  $\gamma_i$  denotes the effect of the  $i$ th block, and  $\epsilon_u$  is a random experimental error ( $i = 1, 2, \dots, b; u = 1, 2, \dots, n$ ). The matrix  $\Lambda$  contains interaction coefficients between the blocks and the fixed polynomial terms in the model. Because the polynomial portion in Eq. (35) is fixed and the elements of  $\gamma$  and  $\Lambda$  are random, model (35) is considered to be a mixed-effects model. It can be expressed in vector form as

$$y = X\beta + Z\gamma + \sum_{j=2}^p U_j\delta_j + \epsilon \tag{36}$$

where the matrix  $X$  is of order  $n \times p$  and is the same as in model (6),  $Z$  is a block-diagonal matrix of order  $n \times b$  of the form

$$Z = \text{diag}(\mathbf{1}_{n_1}, \mathbf{1}_{n_2}, \dots, \mathbf{1}_{n_b}). \tag{37}$$

$U_j$  is a matrix of order  $n \times b$  whose  $i$ th column is obtained by multiplying the elements of the  $j$ th column of  $X$  with the corresponding elements of the  $i$ th column of  $Z$  ( $i = 1, 2, \dots, b; j = 2, 3, \dots, p$ ),  $\delta_j$  is a vector of interaction coefficients between the blocks and the  $j$ th polynomial term ( $j = 2, 3, \dots, p$ ) in model (35). Note that  $\delta_j$  is the same as the transpose of the  $j$ th row of  $\Lambda$  in Eq. (35).

We assume that  $\gamma, \delta_2, \delta_3, \dots, \delta_p$  are normally and independently distributed with zero means and variance-covariance matrices  $\sigma_\gamma^2 I_b, \sigma_2^2 I_b, \dots, \sigma_p^2 I_b$ , respectively. The random error vector,  $\epsilon$ , is also

assumed to be independent of all the other random effects and is distributed as  $N(0, \sigma_\epsilon^2 I_n)$ . Consequently, the mean of  $y$  and its variance-covariance matrix are given by

$$E(y) = X\beta, \tag{38}$$

$$\text{Var}(y) = \sigma_\gamma^2 ZZ' + \sum_{j=2}^p \sigma_j^2 U_j U_j' + \sigma_\epsilon^2 I_n. \tag{39}$$

On the basis of Eqs. (38) and (39), the best linear unbiased estimator of  $\beta$  is the generalized least-squares estimator (GLSE),  $\hat{\beta}$ ,

$$\hat{\beta} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y \tag{40}$$

where

$$\begin{aligned} \Sigma &= \frac{1}{\sigma_\epsilon^2} \text{Var}(y) \\ &= \frac{\sigma_\gamma^2}{\sigma_\epsilon^2} ZZ' + \sum_{j=2}^p \frac{\sigma_j^2}{\sigma_\epsilon^2} U_j U_j' + I_n. \end{aligned} \tag{41}$$

The variance-covariance matrix of  $\hat{\beta}$  is

$$\text{Var}(\hat{\beta}) = (X'\Sigma^{-1}X)^{-1}\sigma_\epsilon^2. \tag{42}$$

The GLSE of  $\beta$  requires knowledge of the ratios of variance components,  $\frac{\sigma_\gamma^2}{\sigma_\epsilon^2}, \frac{\sigma_2^2}{\sigma_\epsilon^2}, \dots, \frac{\sigma_p^2}{\sigma_\epsilon^2}$ . Because the variance components,  $\sigma_\gamma^2, \sigma_2^2, \dots, \sigma_p^2, \sigma_\epsilon^2$  are unknown, they must first be estimated. Let  $\hat{\sigma}_\gamma^2, \hat{\sigma}_2^2, \dots, \hat{\sigma}_\epsilon^2$  denote the corresponding estimates of the variance components. Substituting these estimates in Eq. (41), we get

$$\hat{\Sigma} = \frac{\hat{\sigma}_\gamma^2}{\hat{\sigma}_\epsilon^2} ZZ' + \sum_{j=2}^p \frac{\hat{\sigma}_j^2}{\hat{\sigma}_\epsilon^2} U_j U_j' + I_n. \tag{43}$$

Using  $\hat{\Sigma}$  in place of  $\Sigma$  in Eq. (40), we get the so-called estimated generalized least-squares estimator (EGLSE) of  $\beta$ , denoted by  $\hat{\beta}^*$ ,

$$\hat{\beta}^* = (X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}y. \tag{44}$$

The corresponding estimated variance-covariance matrix of  $\hat{\beta}^*$  is approximately given by

$$\widehat{\text{Var}}(\hat{\beta}^*) \approx (X'\hat{\Sigma}^{-1}X)^{-1}\hat{\sigma}_\epsilon^2. \tag{45}$$

Furthermore, the predicted response at a point  $\mathbf{x}$  in a region  $\mathcal{R}$  is

$$\hat{y}^*(\mathbf{x}) = \mathbf{f}'(\mathbf{x})\hat{\boldsymbol{\beta}}^* \tag{46}$$

This also gives an estimate of the mean response  $\mu(\mathbf{x})$  in Eq. (34). The corresponding prediction variance is approximately given by

$$\widehat{\text{Var}}[\hat{y}^*(\mathbf{x})] \approx \hat{\sigma}_\epsilon^2 \mathbf{f}'(\mathbf{x})(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-1}\mathbf{f}(\mathbf{x}). \tag{47}$$

Estimates of the variance components can be obtained by using either the method of maximum likelihood (ML) or the method of restricted maximum likelihood (REML). These estimates can be easily obtained by using PROC MIXED in SAS.<sup>94</sup>

Tests concerning the fixed effects (i.e., the elements of  $\boldsymbol{\beta}$ ) in the mixed model (35) can be carried out by using the ESGLSE of  $\boldsymbol{\beta}$  and its estimated variance-covariance matrix in Eq. (45). More specifically, to test, for example, the hypothesis,

$$H_0 : \mathbf{a}'\boldsymbol{\beta} = c \tag{48}$$

where  $\mathbf{a}$  and  $c$  are given constants, the corresponding test statistic is

$$t = \frac{\mathbf{a}'\hat{\boldsymbol{\beta}}^* - c}{[\mathbf{a}'(\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X})^{-1}\mathbf{a}\hat{\sigma}_\epsilon^2]^{1/2}} \tag{49}$$

which, under  $H_0$ , has approximately the  $t$ -distribution with  $\nu$  degrees of freedom. Several methods are available in PROC MIXED in SAS for estimating  $\nu$ . The preferred method is the one based on Kenward and Roger's<sup>95</sup> procedure.

Of secondary importance is testing the significance of the random effects in the mixed model (35). The test statistic for testing the hypothesis,

$$H_0 : \sigma_\gamma^2 = 0 \tag{50}$$

is given by the  $F$ -ratio,

$$F = \frac{R(\boldsymbol{\gamma}|\boldsymbol{\beta}, \delta_2, \delta_3, \dots, \delta_p)}{(b-1)MS_E} \tag{51}$$

where  $MS_E$  is the error (residual) mean square for model (35), and  $R(\boldsymbol{\gamma}|\boldsymbol{\beta}, \delta_2, \delta_3, \dots, \delta_p)$  is the type III sum of squares for the  $\boldsymbol{\gamma}$ -effect (block effect). Under  $H_0$ ,  $F$  has the  $F$ -distribution with  $b-1$  and  $m$  degrees of freedom, where  $m = n - b - (p-1)b$ . Similarly, to test the hypothesis

$$H_{0j} : \sigma_j^2 = 0, \quad j = 2, 3, \dots, p \tag{52}$$

we can use the  $F$ -ratio,

$$F_j = \frac{\text{Type III S.S. for } \delta_j}{(b-1)MS_E}, \quad j = 2, 3, \dots, p \tag{53}$$

which, under  $H_{0j}$ , has the  $F$ -distribution with  $b-1$  and  $m$  degrees of freedom. Type III sum of squares for  $\delta_j$  is the sum of squares obtained by adjusting  $\delta_j$  for all the remaining effects in model (35). More details concerning these tests can be found in Ref 96.

It should be noted that in the case of the mixed model in Eq. (35), the process variance (i.e., the variance-covariance matrix of  $\mathbf{y}$ ) in eq. (39) is no longer of the form  $\sigma_\epsilon^2\mathbf{I}_n$ , as is the case with classical response surface models with only fixed effects where the response variance is assumed constant throughout the region  $\mathcal{R}$ . Furthermore, the process variance depends on the settings of the control variables, and hence on the chosen design, as can be seen from formula (39). This lack of homogeneity in the values of the response variance should be taken into account when searching for the optimum response over  $\mathcal{R}$  on the basis of the predicted response expression given in Eq. (46). In particular, the size of the prediction variance in Eq. (47) plays an important role in determining the operating conditions on the control variables that lead to optimum values of  $\hat{y}^*(\mathbf{x})$  over  $\mathcal{R}$ . This calls for an application of modified ridge analysis by Khuri and Myers,<sup>24</sup> which was mentioned earlier. More details concerning such response optimization can be found in Ref 96 and also in Khuri and Cornell<sup>1</sup> (Section 8.3.3).

More recently, the analysis of the mixed model in Eq. (35) under the added assumption that the experimental error variance is different for the different blocks was outlined in Ref 97. This provides an extension of the methodology presented in this section to experimental situations where the error variance can change from one block to another due to some extraneous sources of variation such as machine malfunction in a production process over a period of time.

### Introduction to Generalized Linear Models

Generalized linear models (GLMs) were first introduced by Nelder and Wedderburn<sup>98</sup> as an extension of the class of linear models. They are used to fit discrete as well as continuous data having a variety of parent distributions. The traditional assumptions of normality and homogeneous variances of the response data, usually made in an analysis of variance (or regression) situation, are no longer needed. A classic book on GLMs is the one by McCullagh and Nelder.<sup>99</sup> In addition, the more recent books by Lindsey,<sup>100</sup> Dobson,<sup>101</sup>

McCulloch and Searle,<sup>102</sup> and Myers et al.<sup>103</sup> provide added insight into the application and usefulness of GLMs.

In a GLM situation, the response variable,  $y$ , is assumed to follow a distribution from the exponential family. This includes the normal as well as the binomial, Poisson and gamma distributions. The mean response is modeled as a function of the form,  $\mu(\mathbf{x}) = h[\mathbf{f}(\mathbf{x})\boldsymbol{\beta}]$ , where  $\mathbf{x} = (x_1, \dots, x_k)'$ ,  $\mathbf{f}(\mathbf{x})$  is a known vector function of order  $p \times 1$  and  $\boldsymbol{\beta}$  is a vector of  $p$  unknown parameters. The function  $\mathbf{f}(\mathbf{x})\boldsymbol{\beta}$  is called the linear predictor and is usually denoted by  $\eta(\mathbf{x})$ . It is assumed that  $h(\cdot)$  is a strictly monotone function. Using the inverse of the function  $h(\cdot)$ , one can express  $\eta(\mathbf{x})$  as  $g[\mu(\mathbf{x})]$ . The function  $g(\cdot)$  is called the link function.

Estimation of  $\boldsymbol{\beta}$  is based on the method of maximum likelihood using an iterative weighted least-squares procedure (see Ref 99, pp. 40–43). The estimate of  $\boldsymbol{\beta}$  is then used to estimate the mean response  $\mu(\mathbf{x})$ .

A good design is one which has a low prediction variance or low mean-squared error of prediction (MSEP) (expressions for the prediction variance and MSEP are given in Ref 104). However, the prediction variance and the MSEP for GLMs depend on the unknown parameters of the fitted model. Thus, to minimize either criterion, we require some prior knowledge about  $\boldsymbol{\beta}$ . This leads to the design dependence problem of GLMs. Some common approaches to this problem are listed below. A detailed review of design issues for GLMs can be found in Ref 105.

- Locally optimal designs: designs for GLMs depend on the unknown parameters of the fitted model. Due to this dependence, the construction of a design requires some prior knowledge of the parameters. If initial values of the parameters are assumed, then a design obtained on the basis of an optimality criterion, such as D-optimality or A-optimality, is called locally optimal. The adequacy of such a design depends on how close the initial values are to the true values of the parameters. A key reference in this area is the one by Mathew and Sinha<sup>106</sup> concerning designs for a logistic regression model. Other related work include those of Abdelbasit and Plackett,<sup>107</sup> Minkin,<sup>108</sup> Khan and Yazdi,<sup>109</sup> Wu,<sup>110</sup> and Sitter and Wu.<sup>111</sup>
- Sequential designs: in this approach, experimentation is not stopped at the initial stage. Instead, using the information obtained, initial estimates

of the parameters are updated and used to find additional design points in the subsequent stages. This process is carried out till convergence is achieved with respect to some optimality criterion, for example, D-optimality. Sequential designs were proposed by Wu,<sup>112</sup> Sitter and Forbes,<sup>113</sup> Sitter and Wu,<sup>114</sup> among others.

- Bayesian designs: in the Bayesian approach, a prior distribution is assumed on the parameter vector,  $\boldsymbol{\beta}$  ( $\boldsymbol{\beta}$  as in the linear predictor), which is then incorporated into an appropriate design criterion by integrating it over the prior distribution. For example, one criterion maximizes the average over the prior distribution of the logarithm of the determinant of Fisher's information matrix. This criterion is equivalent to D-optimality in linear models. Bayesian versions of other alphabetic optimality criteria can also be used such as A-optimality. One of the early papers on Bayesian D-optimality criterion is the one by Zacks.<sup>115</sup> Later, the Bayesian approach was discussed by several authors including Chaloner,<sup>116</sup> Chaloner and Larntz,<sup>117,118</sup> and Chaloner and Verdinelli.<sup>119</sup> Designs for a family of exponential models were presented by Dette and Sperlich<sup>120</sup> and Mukhopadhyay and Haines<sup>121</sup> (see Ref 122). Atkinson et al.<sup>123</sup> developed D- and  $D_s$ - (optimal for a subset of parameters) optimal Bayesian designs for a compartmental model.
- Quantile dispersion graphs (QDGs) approach: this approach was recently introduced by Robinson and Khuri<sup>124</sup> in a logistic regression situation. In this graphical technique, designs were compared on the basis of their quantile dispersion profiles. Since in small samples, the parameter estimates are often biased, Robinson and Khuri<sup>124</sup> considered the mean-squared error of prediction (MSEP) as a criterion for comparing designs. Khuri and Mukhopadhyay<sup>104</sup> later applied the QDG approach to compare designs for log-linear models representing Poisson-distributed data.
- Robust design approach: in this approach, a minmax procedure is used to obtain designs robust to poor initial parameter estimates. Sitter<sup>125</sup> applied the minmax procedure to binary data and used the D-optimality and the Fieller optimality criteria to select designs. It was shown by Sitter<sup>125</sup> that his D-optimal designs were more robust to poor initial parameters than locally D-optimal designs for binary data. An extension of Sitter's work was given by King and Wong.<sup>126</sup>

Some recent works on designs for GLMs include those by Dror and Steinberg,<sup>127</sup> Woods et al.,<sup>128</sup> and Russell et al.<sup>129</sup> All of these papers are focused on GLMs with several independent variables. Dror and Steinberg<sup>127</sup> and Russell et al.<sup>129</sup> used clustering techniques for constructing optimal designs.

All the above references discuss design issues for GLMs with a single response. Very little work has been done on multiresponse or multivariate GLMs, particularly in the design area. Such models are considered whenever several response variables can be measured for each setting of a group of control variables, and the response variables are adequately represented by GLMs. Books by Fahrmeir and Tutz<sup>130</sup> and McCullagh and Nelder<sup>99</sup> discuss the analysis of multivariate GLMs.

In multivariate generalized linear models (GLMs), the  $q$ -dimensional vector of responses,  $\mathbf{y}$ , is assumed to follow a distribution from the exponential family. The mean response  $\boldsymbol{\mu}(\mathbf{x}) = [\mu_1(\mathbf{x}), \dots, \mu_q(\mathbf{x})]'$  at a given point  $\mathbf{x}$  in the region of interest,  $R$ , is related to the linear predictor  $\boldsymbol{\eta}(\mathbf{x}) = [\eta_1(\mathbf{x}), \dots, \eta_q(\mathbf{x})]'$  by the link function  $\mathbf{g} : R^q \rightarrow R^q$ ,

$$\boldsymbol{\eta}(\mathbf{x}) = \mathbf{Z}'(\mathbf{x})\boldsymbol{\beta} = \mathbf{g}[\boldsymbol{\mu}(\mathbf{x})] \quad (54)$$

where  $\mathbf{x} = (x_1, \dots, x_k)'$ ,  $\mathbf{Z}(\mathbf{x}) = \bigoplus_{i=1}^q \mathbf{f}_i(\mathbf{x})$ ,  $\mathbf{f}_i(\mathbf{x})$  is a known vector function of  $\mathbf{x}$ ,  $\boldsymbol{\beta}$  is a vector of unknown parameters. If the inverse of  $\mathbf{g}$ , denoted by  $\mathbf{h}$ , exists, where  $\mathbf{h} : R^q \rightarrow R^q$ , then

$$\boldsymbol{\mu}(\mathbf{x}) = \mathbf{h}[\boldsymbol{\eta}(\mathbf{x})] = \mathbf{h}[\mathbf{Z}'(\mathbf{x})\boldsymbol{\beta}]. \quad (55)$$

Estimation of  $\boldsymbol{\beta}$  is based on the method of maximum likelihood using an iterative weighted least-squares procedure (see Ref 130, p. 106). The variance-covariance matrix,  $\text{Var}(\hat{\boldsymbol{\beta}})$ , is dependent on the unknown parameter vector  $\boldsymbol{\beta}$ . This causes the design dependence problem in multivariate GLMs. Some of the key references for optimal designs in multivariate GLMs are Refs 131 and 132. Heise and Myers<sup>131</sup> studied optimal designs for bivariate logistic regression, whereas Zocchi and Atkinson's<sup>132</sup> work was based on optimal designs for multinomial logistic models. A recent work by Mukhopadhyay and Khuri<sup>133</sup> compares designs for multivariate GLMs using the technique of quantile dispersion graphs.

The problem of optimization in a GLM environment is not as well developed as in the case of linear models. In single-response GLMs, Paul and Khuri<sup>25</sup> used modified ridge analysis to carry out optimization of the response. Instead of optimizing the mean response directly, Paul and Khuri<sup>25</sup> optimized the linear predictor. Mukhopadhyay and Khuri<sup>134</sup>

used the generalized distance approach, initially developed for the simultaneous optimization of several linear response surface models, for optimization in a multivariate GLM situation.

Application of GLMs to the robust parameter design problem has been discussed by several authors including Nelder and Lee,<sup>135</sup> Engel and Huele,<sup>136</sup> Brinkley et al.,<sup>137</sup> Hamada and Nelder,<sup>138</sup> Nelder and Lee,<sup>139</sup> Lee and Nelder,<sup>140</sup> and Myers et al.<sup>141</sup> Nelder and Lee<sup>135</sup> modeled the mean and the variance separately using GLMs. Both mean and variance were functions of the control factors. Nelder and Lee<sup>139</sup> modeled the mean as a function of both the control and the noise variables, whereas the variance was a function of the control variables only. Engel and Huele<sup>136</sup> adopted the single response model of Myers et al.<sup>90</sup> and assumed nonconstant error variances. In their paper, the process variance depended on the noise variables as well as the residual variance. They modeled the residual variance using an exponential model which guaranteed positive variance estimates. This exponential model was previously used by Box and Meyer,<sup>142</sup> Grego,<sup>143</sup> and Chan and Mak.<sup>144</sup> A recent paper by Robinson et al.<sup>145</sup> proposed the use of generalized mixed models in a situation where the response was nonnormal and the noise variable was a random effect.

### Graphical Procedures for Assessing the Prediction Capability of a Response Surface Design

Standard design optimality criteria usually base their evaluations on a single number, like D-efficiency, but do not consider the quality of prediction throughout the experimental region. However, the prediction capability of any response surface design does not remain constant throughout the experimental region. Thus, rather than relying on a single-number design criterion, a study of the prediction capability of the design throughout the design region should give more information about the design's performance. Three graphical methods have been proposed to study the performance of a design throughout the experimental region.

1. *Variance dispersion graphs*: Giovannitti-Jensen and Myers<sup>146</sup> and Myers et al.<sup>147</sup> proposed the graphical technique of variance dispersion graphs (VDGs). VDGs are two-dimensional plots displaying the maximum, minimum, and average of the prediction variance on concentric spheres, chosen within the experimental region,

against their radii. The prediction variance is the variance of the predicted response,

$$\hat{y}(\mathbf{x}) = \mathbf{f}'(\mathbf{x})\hat{\beta} \quad (56)$$

as given in formula (10), where  $\mathbf{f}(\mathbf{x})$  is a known vector function of  $\mathbf{x} = (x_1, \dots, x_k)'$  and  $\hat{\beta}$  is the least-squares estimate of  $\beta$  in Eq. (7). As in Eq. (11), the prediction variance is

$$\text{Var}[\hat{y}(\mathbf{x})] = \mathbf{f}'(\mathbf{x})(\mathbf{X}'\mathbf{X})^{-1}\mathbf{f}(\mathbf{x})\sigma^2. \quad (57)$$

A scaled version of the prediction variance (SPV) is given by

$$\frac{N \text{Var}[\hat{y}(\mathbf{x})]}{\sigma^2} = N\mathbf{f}'(\mathbf{x})(\mathbf{X}'\mathbf{X})^{-1}\mathbf{f}(\mathbf{x}) \quad (58)$$

where  $N$  is the number of experimental runs. The average prediction variance,  $V_r$ , over the surface,  $U_r$ , of a sphere of radius  $r$  is

$$V_r = N \frac{\psi}{\sigma^2} \int_{U_r} \text{Var}[\hat{y}(\mathbf{x})] d\mathbf{x} \quad (59)$$

and the maximum and minimum prediction variances, respectively, are given by

$$\max_{\mathbf{x} \in U_r} \frac{N \text{Var}[\hat{y}(\mathbf{x})]}{\sigma^2}, \min_{\mathbf{x} \in U_r} \frac{N \text{Var}[\hat{y}(\mathbf{x})]}{\sigma^2} \quad (60)$$

where  $U_r = \{\mathbf{x} : \sum_{i=1}^k x_i^2 = r^2\}$  and  $\psi^{-1} = \int_{U_r} d\mathbf{x}$  is the surface area of  $U_r$ . A design is considered to be good if it has low and stable values of  $V_r$  throughout the experimental region. The maximum and minimum prediction variance reflect the extent of variability in the prediction variance values over  $U_r$ . A big gap between the maximum and minimum values implies that the variance function is not stable over the region. The software for constructing VDGs was discussed by Vining.<sup>148</sup> Borkowski<sup>149</sup> determined the maximum, minimum, and average prediction variances for central composite and Box–Behnken designs analytically and showed that they were functions of the radius and the design parameters only. Trinca and Gilmour<sup>150</sup> further extended the VDG approach and applied it to blocked response surface designs. Borrer et al.<sup>151</sup> used the VDG approach to compare robust parameter designs. Similar to the VDG approach, Vining and Myers<sup>152</sup> proposed a graphical approach to evaluate and compare response surface designs on the basis of the mean-squared error of

prediction. Vining et al.<sup>153</sup> made use of VDGs in mixture experiments.

2. *Fraction of design space plots:* Zahran et al.<sup>154</sup> proposed fraction of design space (FDS) plots where the prediction variance is plotted against the fraction of the design space that has prediction variance at or below the given value. They argued that VDGs provide some but not all information on SPV of a design, because two designs may have the same VDG pattern but different SPV distributions. This happens because VDGs fail to weight the information in each sphere by the proportion of design space it represents. Zahran et al.<sup>154</sup> defined the FDS criterion as

$$\text{FDS} = \frac{1}{\phi} \int_A d\mathbf{x} \quad (61)$$

where  $A = \{\mathbf{x} : V(\mathbf{x}) < Q\}$ ,  $V(\mathbf{x}) = \frac{N \text{Var}[\hat{y}(\mathbf{x})]}{\sigma^2}$ ,  $\phi$  is the total volume of the experimental region,  $Q$  is some specified quantity.

3. *Quantile plots:* Khuri et al.<sup>155</sup> proposed the quantile plots (QPs) approach for evaluating and comparing several response surface designs based on linear models. In this graphical technique, the distribution of SPV on a given sphere was studied in terms of its quantiles. Khuri et al.<sup>155</sup> showed through examples that two designs can have the same VDG pattern but may display very different distributions of SPV. They stated that this occurred because VDGs provide information only on the extreme values and average value of the SPV but not on the distribution of SPV on a given sphere. To obtain the QPs on the surface,  $U_r$ , of a sphere of radius  $r$ , several points were first generated randomly on  $U_r$ . Spherical coordinates were used for the random generation of points. The values of the SPV function,  $V(\mathbf{x})$ , at points  $\mathbf{x}$  on  $U(r)$  were computed. All these values formed a sample  $T(r)$ . The quantiles of  $T(r)$  were obtained and the  $p$ th quantile of  $T(r)$  was denoted by  $Q_r(p)$ . Plots of  $Q_r(p)$  against  $p$  gave the QPs on  $U(r)$ . Khuri et al.<sup>156</sup> used the QPs to compare designs for mixture models on constrained regions.

## Graphical Methods for Comparing Response Surface Designs

Quantile dispersion graphs (QDGs) were proposed by Khuri<sup>157</sup> for comparing designs for estimating variance components in an analysis of variance (ANOVA) situation. The exact distribution of the

variance component estimator was determined in terms of its quantiles. These quantiles were dependent on the unknown variance components of the ANOVA model. Plots of the maxima and the minima of the quantiles over a subset of the parameter space produced the QDGs. These graphs assessed the quality of the ANOVA estimators and allowed comparison of designs with respect to their estimation capabilities. Lee and Khuri<sup>158,159</sup> extended the use of QDGs to unbalanced random one-way and two-way models, respectively. They used QDGs to compare designs based on ANOVA and maximum likelihood estimation procedures. QDGs were used by Khuri and Lee<sup>160</sup> to evaluate and compare the prediction capabilities of nonlinear designs with one control variable throughout the region of interest  $R$ . More recently, Saha and Khuri<sup>161</sup> used QDGs to compare designs for response surface models with random block effects.

Robinson and Khuri<sup>124</sup> generalized the work of Khuri and Lee<sup>160</sup> by addressing nonnormality and nonconstant error variance. They considered the problem of discriminating among designs for logistic regression models using QDGs based on the MSEP. Khuri and Mukhopadhyay<sup>104</sup> later applied the QDG approach to compare the prediction capabilities of designs for Poisson regression models, also using the MSEP criterion. The MSEP incorporates both the prediction variance and the prediction bias, which results from using maximum likelihood estimates (MLEs) of the parameters of the fitted model. As in any design criterion for GLMs, the MSEP and its quantiles depend on the unknown parameters of the model. For a given design, quantiles of the MSEP were obtained within a region of interest. To compare designs using QDGs in a region  $R$ , several points were generated on concentric surfaces, denoted by  $R_\nu$ , which were obtained by shrinking the boundary of  $R$  by a factor  $\nu$ . The value of the MSEP was computed for each  $\mathbf{x}$  on  $R_\nu$  and  $\beta$  in a parameter space,  $\mathcal{C}$ . An ‘initial’ data set on the response was used to construct the parameter space  $\mathcal{C}$ . Quantiles of MSEP for a given design  $D$  were computed, and the  $p$ th quantile was denoted by  $Q_D(p, \beta, \nu)$  for  $0 \leq p \leq 1$ . These quantiles provided a description of the distribution of MSEP for values of  $\mathbf{x}$  on  $R_\nu$ . The dependence of the quantiles on  $\beta$  was investigated by computing  $Q_D(p, \beta, \nu)$  for several values of  $\beta$  that formed a grid,  $C$ , inside  $\mathcal{C}$ . Subsequently, the minimum and maximum values of  $Q_D(p, \beta, \nu)$  over the values of  $\beta$  in  $C$  were obtained

$$Q_D^{\min}(p, \nu) = \min_{\beta \in C} \{Q_D(p, \beta, \nu)\} \quad (62)$$

$$Q_D^{\max}(p, \nu) = \max_{\beta \in C} \{Q_D(p, \beta, \nu)\}. \quad (63)$$

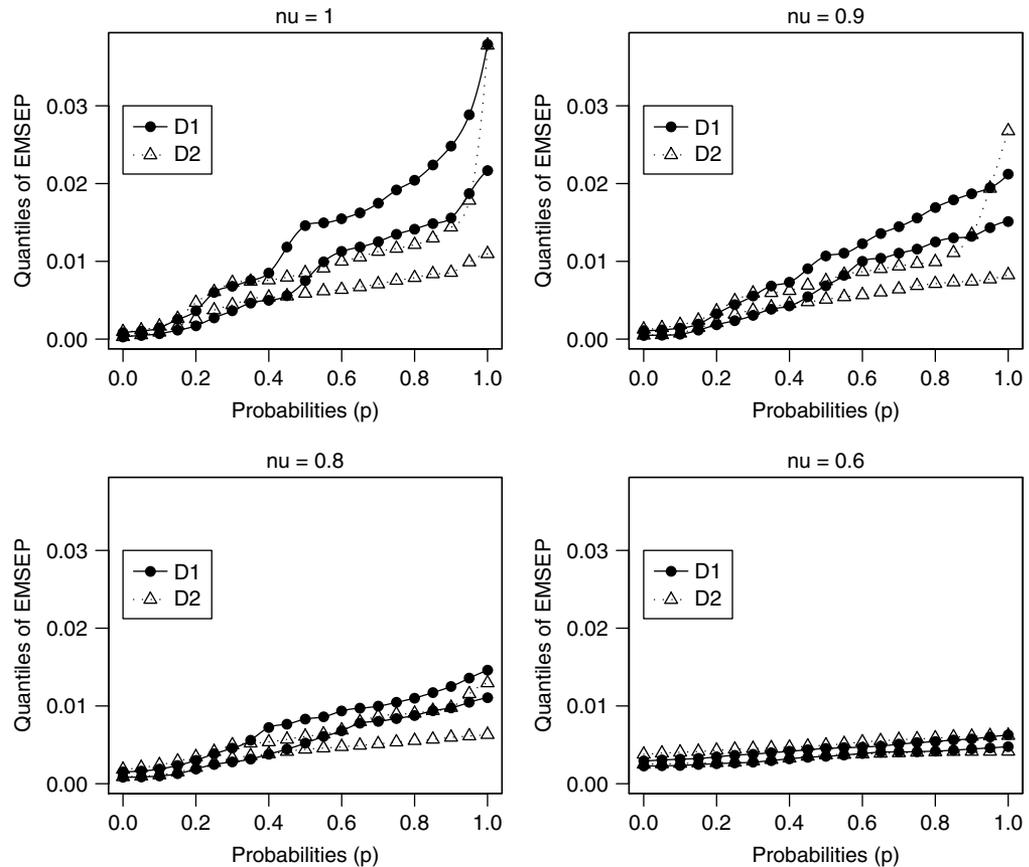
Plotting these values against  $p$  resulted in the QDGs of the MSEP over  $R_\nu$ . Using several values of  $\nu$ , the entire region  $R$  was covered. Given several designs to compare, a design that displays close and small values of  $Q_D^{\max}$  and  $Q_D^{\min}$  over the range of  $p$  is considered desirable because it has good prediction capability and is robust to the changes in  $\beta$ . Khuri and Mukhopadhyay<sup>104</sup> also studied the effect of the choice of the link function and/or the nature of the response distribution on the shape of the QDGs for a given design.

Mukhopadhyay and Khuri<sup>133</sup> extended the application of GLMs to compare response surface designs for multivariate GLMs. Since the MSEP is a matrix in the multivariate situation, they considered a scalar-valued function of the MSEP, namely the largest eigenvalue of the MSEP matrix (EMSEP), as their comparison criterion. Similar to the MSEP, EMSEP also depends on  $\mathbf{x}$  and  $\beta$ . As in the univariate case, quantiles of EMSEP were computed on concentric regions  $R_\nu$ , for a given design  $D$ . Mukhopadhyay and Khuri<sup>133</sup> chose the parameter space  $\mathcal{C}$  to be the  $(1 - \alpha)100\%$  confidence region of  $\beta$ . Subsequently, the minimum and maximum quantiles were computed over the values of  $\beta$  in a grid of points from  $\mathcal{C}$  and plotted against  $p$  to obtain the QDGs. The authors illustrated their proposed methodology using a data set from a combination drug therapy study on male mice taken from Gennings et al.,<sup>162</sup> pp. 429–451 (see Section 6 in Mukhopadhyay and Khuri.<sup>133</sup>)

*A numerical example.* In a drug therapy experiment on male mice, the pain relieving effects of two drugs, morphine sulfate ( $x_1$ ), and  $\Delta^9$ -tetrahydrocannabinol ( $x_2$ ), on two binary responses, pain relief ( $y_1$ ), and side effect ( $y_2$ ), were studied. The response  $y_1$  takes the value 1 if relief occurs, otherwise it takes the value zero. The response  $y_2$  is equal to one if a harmful side effect develops, otherwise it is equal to zero. A  $5 \times 7$  factorial design with six mice in each run was considered. The experimental region  $R$  was rectangular in shape with  $R : \{0 \leq x_1 \leq 8, 0 \leq x_2 \leq 15\}$ . The following first-degree models were used to fit the data

$$\begin{aligned} \eta_1(\mathbf{x}) &= \beta_1 + \beta_2 x_1 + \beta_3 x_2, \\ \eta_2(\mathbf{x}) &= \beta_4 + \beta_5 x_1 + \beta_6 x_2, \\ \eta_3(\mathbf{x}) &= \beta_7 + \beta_8 x_1 + \beta_9 x_2. \end{aligned} \quad (64)$$

Here,  $\boldsymbol{\eta}(\mathbf{x}) = [\eta_1(\mathbf{x}), \eta_2(\mathbf{x}), \eta_3(\mathbf{x})]'$  is the three-dimensional linear predictor as explained in Eq. (54).



**FIGURE 1** | Comparison of the QDGs for designs  $D_1$  ( $5 \times 7$  factorial) and  $D_2$  ( $3^2$  factorial), for  $p = 0(0.05)1$  and  $\nu = 1, 0.9, 0.8, 0.6$ .

The corresponding link function used was

$$\hat{\mu}_i(\mathbf{x}) = \frac{\exp[\hat{\eta}_i(\mathbf{x})]}{1 + \sum_{l=1}^3 \exp[\hat{\eta}_l(\mathbf{x})]}, \quad i = 1, 2, 3 \quad (65)$$

where  $\hat{\eta}_i(\mathbf{x})$  is the estimate of  $\eta_i(\mathbf{x})$  for  $i = 1, 2, 3$ . Note that  $\mu_i(\mathbf{x})$  is the  $i$ th element of  $\boldsymbol{\mu}(\mathbf{x}) = [\mu_1(\mathbf{x}), \mu_2(\mathbf{x}), \mu_3(\mathbf{x})]'$ , where for a given  $\mathbf{x}$ ,  $\mu_1(\mathbf{x})$  is the probability that  $y_1 = 1$  and  $y_2 = 1$ ,  $\mu_2(\mathbf{x})$  is the probability that  $y_1 = 1$  and  $y_2 = 0$ , and  $\mu_3(\mathbf{x})$  is the probability that  $y_1 = 0$  and  $y_2 = 1$ . The original  $5 \times 7$  factorial design ( $D_1$ ) was compared with  $D_2$ , a  $3^2$  factorial design with the center point (4, 7.5) replicated

three times and all other points replicated four times. The EMSEP values were computed for  $\mathbf{x}$  on concentric rectangles  $R_\nu$  and  $\beta \in \mathcal{C}$ . The 95% confidence region of  $\beta$  was chosen to be  $\mathcal{C}$  and  $\mathcal{C}$  was a set of 500 randomly chosen points from  $\mathcal{C}$ . The QDGs comparing designs  $D_1$  and  $D_2$  on the basis of the quantiles of EMSEP values are shown in Figure 1. From Figure 1, it can be noted that both designs are robust to the changes in the parameter values, and overall,  $D_2$  has better prediction capability than  $D_1$  for almost all values of  $\nu$  and  $p$ . More details concerning this example can be found in Sections 5 and 6 in Ref 133.

## REFERENCES

1. Khuri, AI, Cornell, JA. *Response Surfaces*. 2nd ed. New York: Dekker; 1996.
2. Box GEP, Hunter JS. Multifactor experimental designs for exploring response surfaces. *Ann Math Stat* 1957, 28:195–241.
3. Khuri AI. A measure of rotatability for response surface designs. *Technometrics* 1988, 30:95–104.
4. Box GEP, Draper NR. Robust designs. *Biometrika* 1975, 62:347–352.
5. Kiefer J, Wolfowitz J. The equivalence of two extremum problems. *Can J Math* 1960, 12:363–366.
6. Pazman A. *Foundations of Optimum Experimental Design*. Dordrecht, Holland: Reidel; 1986.

7. Box GEP, Hunter WG, Hunter JS. *Statistics for Experimenters*. New York: John Wiley & Sons; 1978.
8. Ratkoe BL, Hedayat A, Federer WT. *Factorial Designs*. New York: John Wiley & Sons; 1981.
9. Montgomery DC. *Design and Analysis of Experiments*. 6th ed. New York: John Wiley & Sons; 2005.
10. Plackett RL, Burman JP. The design of optimum multifactorial experiments. *Biometrika* 1946, 33:305–325.
11. Box GEP. Multi-factor designs of first order. *Biometrika* 1952, 39:49–57.
12. McLean RA, Anderson VL. *Applied Factorial and Fractional Designs*. New York: Dekker; 1984.
13. Box GEP, Wilson KB. On the experimental attainment of optimum conditions. *J R Stat Soc., Ser B* 1951, 13:1–45 (with discussion).
14. Box GEP, Behnken DW. Some new three-level designs for the study of quantitative variables. *Technometrics* 1960, 2:455–475.
15. Box GEP, Draper NR. *Response Surfaces, Mixtures, and Ridge Analyses*. 2nd ed. Hoboken, New Jersey: John Wiley & Sons; 2007.
16. Myers RH, Montgomery DC. *Response Surface Methodology*. New York: John Wiley & Sons; 1995.
17. Hoke AT. Economical second-order designs based on irregular fractions of the  $3^n$  factorial. *Technometrics* 1974, 16:375–384.
18. Box MJ, Draper NR. Factorial designs, the  $|X'X|$  criterion and some related matters. *Technometrics* 1971, 13:731–742.
19. Doehlert DH. Uniform shell designs. *J R Stat Soc., Ser C* 1970, 19:231–239.
20. Roquemore KG. Hybrid designs for quadratic response surfaces. *Technometrics* 1976, 18:419–423.
21. Myers RH, Khuri AI. A new procedure for steepest ascent. *Commun Stat: Theory Methods* 1979, 8:1359–1376.
22. Hoerl AE. Optimum solutions of many variable equations. *Chem Eng Prog* 1959, 55:69–78.
23. Draper NR. Ridge analysis of response surfaces. *Technometrics* 1963, 5:469–479.
24. Khuri AI, Myers RH. Modified ridge analysis. *Technometrics* 1979, 21:467–473.
25. Paul S, Khuri AI. Modified ridge analysis under non-standard conditions. *Commun Stat: Theory Methods* 2000, 29:2181–2200.
26. Hill WJ, Hunter WG. A review of response surface methodology: a literature review. *Technometrics* 1966, 8:571–590.
27. Myers RH, Khuri AI, Carter WH Jr. Response surface methodology: 1966–1988. *Technometrics* 1989, 31:137–157.
28. Khuri AI. Multiresponse surface methodology. In: Ghosh S, Rao CR eds. *Handbook of Statistics*, vol. 13. Amsterdam: Elsevier Science B.V.; 1996, 377–406.
29. Zellner A. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *J Am Stat Assoc* 1962, 57:348–368.
30. Draper NR, Hunter WG. Design of experiments for parameter estimation in multiresponse situations. *Biometrika* 1966, 53:525–533.
31. Box MJ, Draper NR. Estimation and design criteria for multiresponse non-linear models with non-homogeneous variance. *J R Stat Soc., Ser C* 1972, 21:13–24.
32. Fedorov V. *Theory of Optimal Experiments*. New York: Academic Press; 1972.
33. Wijesinha M, Khuri AI. The sequential generation of multiresponse D-optimal designs when the variance-covariance matrix is not known. *Commun Stat: Simulation Comput* 1987, 16:239–259.
34. Krafft O, Schaefer M. D-optimal designs for a multivariate regression model. *J Multivariate Anal* 1992, 42:130–140.
35. Bischoff W. On D-optimal designs for linear models under correlated observations with an application to a linear model with multiple response. *J Stat Plan Infer* 1993, 37:69–80.
36. Chang SI. Some properties of multiresponse D-optimal designs. *J Math Anal Appl* 1994, 184:256–262.
37. Chang S. An algorithm to generate near D-optimal designs for multiple response surface models. *IIE Trans* 1997, 29:1073–1081.
38. Imhof L. Optimum designs for a multiresponse regression model. *J Multivariate Anal* 2000, 72:120–131.
39. Atashgah AB, Seifi A. Application of semi-definite programming to the design of multi-response experiments. *IIE Trans* 2007, 39:763–769.
40. Hatzis C, Larntz K. Optimal design in nonlinear multiresponse estimation: Poisson model for filter feeding. *Biometrics* 1992, 48:1235–1248.
41. Wijesinha MC, Khuri AI. Construction of optimal designs to increase the power of the multiresponse lack of fit test. *J Stat Plan Infer* 1987, 16:179–192.
42. Wijesinha MC, Khuri AI. Robust designs for first-order multiple design multivariate models. *Commun Stat Theory Methods* 1991, 41:2987–2999.
43. Yue R. Model-robust designs in multiresponse situations. *Stat Probab Lett* 2002, 58:369–379.
44. Khuri AI. A test for lack of fit of a linear multiresponse model. *Technometrics* 1985, 27:213–218.
45. Morrison DF. *Multivariate Statistical Methods*. 2nd ed. New York: McGraw-Hill; 1976.
46. Levy MS, Neill JW. Testing for lack of fit in linear multiresponse models based on exact or near

- replicates. *Commun Stat: Theory Methods* 1990, 19:1987–2002.
47. Lind EE, Goldin J, Hickman JB. Fitting yield and cost response surfaces. *Chem Eng Prog* 1960, 56:62–68.
48. Harrington ECJ. The desirability functions. *Ind Quality Control* 1965, 12:494–498.
49. Derringer G, Suich R. Simultaneous optimization of several response variables. *J Quality Technol* 1980, 12:214–219.
50. Khuri AI, Conlon M. Simultaneous optimization for multiple responses represented by polynomial regression functions. *Technometrics* 1981, 23:363–375.
51. Myers RH, Carter WH. Response surface techniques for dual response systems. *Technometrics* 1973, 15:301–317.
52. Biles WE. A response surface method for experimental optimization of multiple response processes. *Ind Eng Chem Process Design Dev* 1975, 14:152–158.
53. Del Castillo E. Multiresponse process optimization via constrained confidence regions. *J Quality Technol* 1996, 28:61–70.
54. Del Castillo E, Fan SK, Semple J. Optimization of dual response systems: a comprehensive procedure for degenerate and nondegenerate problems. *Eur J Oper Res* 1999, 112:174–186.
55. Fan SK. A generalized global optimization algorithm for dual response systems. *J Quality Technol* 2000, 32:444–456.
56. Del Castillo E, Montgomery DC, McCarville DR. Modified desirability functions for multiple response optimization. *J Quality Technol* 1996, 28:337–345.
57. Wu FC. Optimization of correlated multiple quality characteristics using desirability function. *Quality Eng* 2005, 17:119–126.
58. Vining G. A compromise approach to multiresponse optimization. *J Quality Technol* 1998, 30:309–313.
59. Pignatiello JJ. Strategies for robust multiresponse quality engineering. *IIE Trans* 1993, 25:5–15.
60. Ames A, Mattucci N, McDonald S, Szonyi G, Hawkins D. Quality loss function for optimization across multiple response surfaces. *J Quality Technol* 1997, 29:339–346.
61. Khuri AI, Valeroso ES. Optimization methods in multiresponse surface methodology. In: Park SH, Vining GG, eds. *Statistical Process Monitoring and Optimization*. New York: Dekker; 2000, 411–433.
62. Kacker RN. Off-line quality control, parameter design, and the Taguchi method. *J Quality Technol* 1985, 17:176–209.
63. Taguchi G, Wu Y. *Introduction to Off-Line Quality Control*. Nagoya, Japan: Central Japan Quality Control Association; 1985.
64. Taguchi G. *System of Experimental Design: Engineering Methods to Optimize Quality and Minimize Cost*. White Plains: UNIPUB/Kraus International; 1987.
65. Nair VN, Shoemaker AC. The role of experimentation in quality engineering: a review of Taguchi's contributions. In: Ghosh S, ed. *Statistical Design and Analysis of Industrial Experiments*. New York: Marcel Dekker; 1990, 247–277.
66. Taguchi G. *Introduction to Quality Engineering*. White Plains: UNIPUB/Kraus International; 1986.
67. Phadke MS. *Quality Engineering Using Robust Design*. Englewood Cliffs: Prentice-Hall; 1989.
68. Wu CFJ, Hamada M. *Experiments: Planning, Analysis, And Parameter Design Optimization*. New York: Wiley-Interscience; 2000.
69. Myers RH, Montgomery DC. *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. 2nd ed. New York: John Wiley & Sons; 2002.
70. Myers RH, Montgomery DC, Vining GG, Borror CM, Kowalski SM. Response surface methodology: a retrospective and literature survey. *J Quality Technol* 2004, 36:53–77.
71. Robinson TJ, Borror CM, Myers RH. Robust parameter design: a review. *Quality Reliab Eng Int* 2004, 20:81–101.
72. Box GEP. Discussion of off-line quality control, parameter design, and the Taguchi methods. *J Quality Technol* 1985, 17:198–206.
73. Box GEP. Signal-to-noise ratios, performance criteria, and transformations. *Technometrics* 1988, 30:1–17.
74. Easterling RB. Discussion of off-line quality control, parameter design, and the Taguchi methods. *J Quality Technol* 1985, 17:198–206.
75. Pignatiello JJ, Ramberg JS. Discussion of off-line quality control, parameter design, and the Taguchi methods. *J Quality Technol* 1985, 17:198–206.
76. Nair VN, Pregibon D. Analyzing dispersion effects from replicated factorial experiments. *Technometrics* 1988, 30:247–258.
77. Welch WJ, Yu TK, Kang SM, Sacks J. Computer experiments for quality control by parameter design. *J Quality Technol* 1990, 22:15–22.
78. Nair VN. Taguchi's parameter design: a panel discussion. *Technometrics* 1992, 34:127–161.
79. Vining GG, Myers RH. Combining Taguchi and response surface philosophies: a dual response approach. *J Quality Technol* 1990, 22:38–45.
80. Del Castillo E, Montgomery DC. A nonlinear programming solution to the dual response problem. *J Quality Technol* 1993, 25:199–204.
81. Lin DKJ, Tu W. Dual response surface optimization. *J Quality Technol* 1995, 27:34–39.

82. Copeland KAF, Nelson PR. Dual response optimization via direct function minimization. *J Quality Technol* 1996, 28:331–336.
83. Kim K, Lin DKJ. Dual response surface optimization: a fuzzy modeling approach. *J Quality Technol* 1998, 30:1–10.
84. Del Castillo E, Fan SK, Semple J. Computation of global optima in dual response systems. *J Quality Technol* 1997, 29:347–353.
85. Koksoy O, Doganaksoy N. Joint optimization of mean and standard deviation using response surface methods. *J Quality Technol* 2003, 35:239–252.
86. Kim YJ, Cho BR. Development of priority-based robust design. *Quality Eng* 2002, 14:355–363.
87. Tang LC, Xu K. A unified approach for dual response optimization. *J Quality Technol* 2002, 34:437–447.
88. Lucas JM. Achieving a robust process using response surface methodology. Paper presented at the Joint Statistical Meetings, Washington, DC, August 6–10, 1989.
89. Shoemaker AC, Tsui KL, Wu CFJ. Economical experimentation methods for robust parameter design. Paper presented at the Fall Technical Conference, American Society of Quality Control, Houston, TX, 1989.
90. Myers RH, Khuri AI, Vining GG. Response surface alternatives to the Taguchi robust parameter design approach. *Am Stat* 1992, 46:131–139.
91. Aggarwal ML, Bansal A. Robust response surface designs for quantitative and qualitative factors. *Commun Stat Theory Methods* 1998, 27:89–106.
92. Aggarwal ML, Gupta BC, Bansal A. Small robust response surface designs for quantitative and qualitative factors. *Am J Math Manage Sci* 2000, 39:103–130.
93. Brenneman WA, Myers WR. Robust parameter design with categorical noise variables. *J Quality Technol* 2003, 35:335–341.
94. SAS Institute Inc. *Version 9.1.3*. Cary, NC: SAS; 2008. [OnlineDoc]. <http://support.sas.com/onlinedoc/913/docMainpage.jsp>. Accessed 2008.
95. Kenward MG, Roger JH. Small-sample inference for fixed effects from restricted maximum likelihood. *Biometrics* 1997, 53:983–997.
96. Khuri AI. Response surface models with mixed effects. *J Quality Technol* 1996, 28:177–186.
97. Khuri AI. Mixed response surface models with heterogeneous within-block error variances. *Technometrics* 2006, 48:206–218.
98. Nelder JA, Wedderburn RWM. Generalized linear models. *J R Stat Soc., Ser A* 1972, 135:370–384.
99. McCullagh P, Nelder JA. *Generalized Linear Models*. 2nd ed. London: Chapman and Hall; 1989.
100. Lindsey JK. *Applying Generalized Linear Models*. New York: Springer; 1997.
101. Dobson AJ. *An Introduction to Generalized Linear Models*. 2nd ed. Boca Raton: Chapman and Hall; 2001.
102. McCulloch CE, Searle SR. *Generalized, Linear, and Mixed Models*. New York: John Wiley & Sons; 2001.
103. Myers RH, Montgomery DC, Vining GG. *Generalized Linear Models with Applications in Engineering and the Sciences*. New York: John Wiley & Sons; 2002.
104. Khuri AI, Mukhopadhyay S. GLM designs: the dependence on unknown parameters dilemma. In: Khuri AI, ed. *Response Surface Methodology and Related Topics*. Singapore: World Scientific; 2006, 203–223.
105. Khuri AI, Mukherjee B, Sinha BK, Ghosh M. Design issues for generalized linear models: a review. *Stat Sci* 2006, 21:376–399.
106. Mathew T, Sinha BK. Optimal designs for binary data under logistic regression. *J Stat Plan Infer* 2001, 93:295–307.
107. Abdelbasit KM, Plackett RL. Experimental design for binary data. *J Am Stat Assoc* 1983, 78:90–98.
108. Minkin S. Optimal design for binary data. *J Am Stat Assoc* 1987, 82:1098–1103.
109. Khan MK, Yazdi AA. On D-optimal designs for binary data. *J Stat Plan Infer* 1988, 18:83–91.
110. Wu CFJ. Optimal design for percentile estimation of a quantal response curve. In: Dodge Y, Fedorov VV, Wynn HP, eds. *Optimal Design and Analysis of Experiments*. Amsterdam: Elsevier Science; 1988, 213–223.
111. Sitter RR, Wu CFJ. Optimal designs for binary response experiments: Fieller, D, and A criteria. *Scand J Stat* 1993, 20:329–341.
112. Wu CFJ. Efficient sequential designs with binary data. *J Am Stat Assoc* 1985, 80:974–984.
113. Sitter RR, Forbes B. Optimal two-stage designs for binary response experiments. *Stat Sin* 1997, 7:941–956.
114. Sitter RR, Wu CFJ. Two-stage design of quantal response studies. *Biometrics* 1999, 55:396–402.
115. Zacks S. Problems and approaches in design of experiments for estimation and testing in non-linear models. In: Krishnaiah PR, ed. *Multivariate Analysis*, vol. 4. Amsterdam: North-Holland; 1977, 209–223.
116. Chaloner K. An approach to design for generalized linear models. *Proceedings of the Workshop on Model-Oriented Data Analysis*, Wartburg. Lecture Notes in Economics and Mathematical Systems. Berlin: Springer; 1987, 3–12.
117. Chaloner K, Larntz K. Optimal Bayesian designs applied to logistic regression experiments. *J Stat Plan Infer* 1989, 21:191–208.

118. Chaloner K, Larntz K. Optimal Bayesian design for accelerated life testing experiments. *J Stat Plan Infer* 1991, 33:245–259.
119. Chaloner K, Verdinelli I. Bayesian experimental design: a review. *Stat Sci* 1995, 10:273–304.
120. Dette H, Sperlich S. A note on Bayesian D-optimal designs for generalization of the simple exponential growth model. *South African Stat J* 1994, 28:103–117.
121. Mukhopadhyay S, Haines L. Bayesian D-optimal designs for the exponential growth model. *J Stat Plan Infer* 1995, 44:385–397.
122. Atkinson AC, Haines LM. Designs for nonlinear and generalized linear models. In: Ghosh S, Rao CR, eds. *Handbook of Statistics*, vol. 13. Amsterdam: Elsevier Science; 1996, 437–475.
123. Atkinson AC, Chaloner K, Herzberg AM, Juritz J. Optimum experimental designs for properties of a compartmental model. *Biometrics* 1993, 49:325–337.
124. Robinson KS, Khuri AI. Quantile dispersion graphs for evaluating and comparing designs for logistic regression models. *Comput Stat Data Anal* 2003, 43:47–62.
125. Sitter RR. Robust designs for binary data. *Biometrics* 1992, 48:1145–1155.
126. King J, Wong W. Minimax D-optimal designs for the logistic model. *Biometrics* 2000, 56:1263–1267.
127. Dror HA, Steinberg DM. Robust experimental design for multivariate generalized linear models. *Technometrics* 2006, 48:520–529.
128. Woods DC, Lewis SM, Eccleston JA, Russell KG. Designs for generalized linear models with several variables and model uncertainty. *Technometrics* 2008, 48:284–292.
129. Russell KG, Woods DC, Lewis SM, Eccleston JA. D-optimal designs for Poisson regression models. *Stat Sin* 2009, 19:721–730.
130. Fahrmeir L, Tutz G. *Multivariate Statistical Modelling Based on Generalized Linear Models*. 2nd ed: New York: Springer; 2001.
131. Heise MA, Myers RH. Optimal designs for bivariate logistic regression. *Biometrics* 1996, 52:613–624.
132. Zocchi SS, Atkinson AC. Optimum experimental designs for multinomial logistics models. *Biometrics* 1999, 55:437–443.
133. Mukhopadhyay S, Khuri AI. Comparison of designs for multivariate generalized linear models. *J Stat Plan Infer* 2008, 138:169–183.
134. Mukhopadhyay S, Khuri AI. Optimization in a multivariate generalized linear model situation. *Comput Stat Data Anal* 2008, 52:4625–4634.
135. Nelder JA, Lee Y. Generalized linear models for the analysis of Taguchi-type experiments. *Appl Stoch Models Data Anal* 1991, 7:107–120.
136. Engel J, Huele AF. A generalized linear modeling approach to robust design. *Technometrics* 1996, 38:365–373.
137. Brinkley PA, Meyer KP, Lu JC. Combined generalized linear modelling/non-linear programming approach to robust process design: a case-study in circuit board quality improvement. *Appl Stat* 1996, 45:99–110.
138. Hamada M, Nelder JA. Generalized linear models for quality-improvement experiments. *J Quality Technol* 1997, 29:292–304.
139. Nelder JA, Lee Y. Letter to the editor. *Technometrics* 1998, 40:168–171.
140. Lee Y, Nelder JA. Robust design via generalized linear models. *J Quality Technol* 2003, 35:2–12.
141. Myers WR, Brenneman WA, Myers RH. A dual-response approach to robust parameter design for a generalized linear model. *Technometrics* 2005, 37:130–138.
142. Box GEP, Meyer RD. Dispersion effects from fractional designs. *Technometrics* 1986, 28:19–27.
143. Grego JM. Generalized linear models and process variation. *J Quality Technol* 1993, 25:288–295.
144. Chan LK, Mak TK. A regression approach for discovering small variation around a target. *Appl Stat* 1995, 44:369–377.
145. Robinson TJ, Wulff SS, Montgomery DC, Khuri AI. Robust parameter design using generalized linear mixed models. *J Quality Technol* 2006, 38:65–75.
146. Giovannitti-Jensen A, Myers RH. Graphical assessment of the prediction capability of response surface designs. *Technometrics* 1989, 31:159–171.
147. Myers RH, Vining GG, Giovannitti-Jensen A, Myers SL. Variance dispersion properties of second order response surface designs. *J Quality Technol* 1992, 24:1–11.
148. Vining GG. A computer program for generating variance dispersion graphs. *J Quality Technol* 1993, 25:45–58.
149. Borkowski JJ. Spherical prediction-variance properties of central composite and Box-Behnken designs. *Technometrics* 1995, 37:399–410.
150. Trinca LA, Gilmour SG. Variance dispersion graphs for comparing blocked response surface designs. *J Quality Technol* 1998, 30:349–364.
151. Borrer CM, Montgomery DC, Myers RH. Evaluation of statistical designs for experiments involving noise variables. *J Quality Technol* 2002, 34:54–70.
152. Vining GG, Myers RH. A graphical approach for evaluating response surface designs in terms of the mean squared error of prediction. *Technometrics* 1991, 33:315–326.

153. Vining GG, Cornell JA, Myers RH. A graphical approach for evaluating mixture designs. *Appl Stat* 1993, 42:127–138.
154. Zahran A, Anderson-Cook CM, Myers RH. Fraction of design space to assess prediction capability of response surface designs. *J Quality Technol* 2003, 35:377–386.
155. Khuri AI, Kim HJ, Um Y. Quantile plots of the prediction variance for response surface designs. *Comput Stat Data Anal* 1996, 22:395–407.
156. Khuri AI, Harrison JM, Cornell JA. Using quantile plots of the prediction variance for comparing designs for a constrained mixture region: an application involving a fertilizer experiment. *Appl Stat* 1999, 48:521–532.
157. Khuri AI. Quantile dispersion graphs for analysis of variance estimates of variance components. *J Appl Stat* 1997, 24:711–722.
158. Lee J, Khuri AI. Graphical technique for comparing designs for random models. *J Appl Stat* 1999, 26:933–947.
159. Lee J, Khuri AI. Quantile dispersion graphs for the comparison of designs for a random two-way model. *J Stat Planning Infer* 2000, 91:123–137.
160. Khuri AI, Lee J. A graphical approach for evaluating and comparing designs for nonlinear models. *Comput Stat Data Anal* 1998, 27:433–443.
161. Saha S, Khuri AI. Comparison of designs for response surface models with random block effects. *Quality Tech Quantitative Manage* 2009, 6:219–234.
162. Gennings C, Carter WH Jr, Martin BR. Drug interactions between morphine and marijuana. In: Lange N, Ryan L, Billard L, Brillinger D, Conquest L, et al. eds. *Case Studies in Biometry*. New York: John Wiley & Sons; 1994, 429–451.