

Response surface methodology and its applications in agricultural and food sciences

Abstract

The purpose of this article is to provide an overview of response surface methodology (RSM) which includes the modeling of a response function, the corresponding choice of design, and the determination of optimum conditions. In addition, the use of RSM in agricultural and food sciences is highlighted by citing several examples taken from a variety of applied journals.

Volume 5 Issue 5 - 2017

Andre I. Khuri

Department of Statistics, University of Florida, USA

Correspondence: Andre I Khuri, Department of Statistics, University of Florida, USA, Email ufakhuri@stat.ufl.edu

Received: September 15, 2016 | **Published:** April 11, 2017

Introduction

Response surface methodology (RSM) consists of a group of mathematical and statistical techniques concerned with

a. The selection and construction of an appropriate design that can provide adequate and reliable information concerning a certain response variable, denoted by y .

b. The determination of a suitable model that best fits the data that can be generated from using the design chosen in (1). Such a model gives an approximate functional relationship between the response variable y and a set of control variables believed (by the experimenter) to have an effect on the response y . These variables are denoted by x_1, x_2, \dots, x_k .

c. The finding of optimal settings on the control variables that produce maximum (or minimum) response values within a certain region of interest.

From the historical point of view, the article by Box and Wilson¹ is considered to be the first for having laid the foundations for RSM. Some of the early papers that also contributed to the development of RSM include those by Box and Hunter² and Box and Draper.³ Several review papers on RSM were subsequently published starting with the article by Hill and Hunter⁴ that emphasized practical applications of RSM in the chemical and processing fields. This was followed by more recent reviews by Myers et al.,⁵⁻⁷ A review of RSM from a biometric viewpoint was given by⁸ who emphasized biological applications rather than applications in the physical and engineering sciences. In addition, several books were written on the subject of RSM by Myers,⁹ Khuri and Cornell,¹⁵ Box and Draper¹¹ and Myers et al.,¹²

In the early development of RSM, from 1951 up through the 1970s, the main focus of attention was on controlled experiments that can, for the most part, be performed in a laboratory. This was particularly suited in an industrial setting with possible applications in the physical and engineering sciences. The review paper by Hill and Hunter⁴ made reference to several applications in the chemical

industry. It should be pointed out here that the seminal work of Box and Wilson¹ occurred at a major chemical company, and Box himself was initially trained as a chemist in England.

One of the basic characteristics of a response surface investigation is its sequential nature whereby experiments are performed in stages. Information acquired from one set of experiments is used to plan the strategy for a follow-up set of experiments. This sequential pattern of experimentation was suggested by Box and Youle.¹³ Such an approach works well in an industrial setting since the response values in a given stage can be obtained in a relatively short time. However, this approach may not be feasible in an agricultural setting where experiments typically require long periods between stages. Furthermore, it is quite common in an agricultural experiment that the results of a single experiment may be the only ones available rather than having a series of experiments obtained sequentially. Another main difference between an industrial experiment and an agricultural one has to do with factors' levels. In an industrial experiment, the levels of quantitative factors can be accurately controlled and measured. This, however, may be difficult to do in an agricultural experiment. Edmondson¹⁴ Outlined several other major distinctions between industrial and agricultural experiments.

Mead and Pike⁸ were the first to bring attention to the role of RSM in agriculture and biometric research in general. They also provided a survey of biological and agricultural journals that used response surface methods. However, they emphasized the use of nonlinear models to describe the behavior of biological data. Such models are usually referred to as *Mechanistic*. The traditional modeling scheme used in the early work on RSM was based on the so-called *Empirical modeling* where low-degree polynomials are used to fit the response data. These polynomials are chosen and tested on the basis of the observed data, but are not selected on the basis of information that pertain to certain chemical or physical laws, as is the case with mechanistic models.

While certain approaches used in RSM may not be quite suitable in an agricultural setting, there is a lot to be gained from using certain response surface techniques in an agricultural experiment. The

purpose of this chapter is to provide an expose of some basic methods and models used in RSM. Several applications of RSM in agricultural and food sciences will also be mentioned using examples taken from the corresponding literature.

Response surface models

Let y be a response variable of interest, and let x_1, x_2, \dots, x_k denote control variables believed to have an effect on y . As was mentioned earlier, one of the objectives of RSM is to establish a functional relationship between y and its associated control variables. Such a relationship is, in general, unknown, but can be approximated by a low-degree polynomial model of the form

$$y = f'(x)\beta + \epsilon \tag{2.1}$$

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 $d (\geq 1)$, β is a vector of p unknown constant coefficients referred to as parameters, and ϵ is a random experimental error assumed to have a zero mean. If model (2.1) provides an adequate representation of the response, then the quantity $f(x)\beta$ represents the so-called *Mean response*, or the *Expected Value* of y , and is denoted by $\mu(x)$.

Two models are commonly used in RSM; they are the first-degree model ($d = 1$),

$$y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \epsilon; \tag{2.2}$$

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

$$y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i < j} \beta_{ij} x_i x_j + \sum_{i=1}^k \beta_{ii} x_i^2 + \epsilon; \tag{2.3}$$

Models (2.2) and (2.3) are special cases of model (2.1). The first model is usually used in the initial phase of the experiment representing part of an exploratory process to assess the important factors. The second model is subsequently developed after a series of experiments have been performed resulting in the identification of the important factors to be considered in the experiment. At this stage, the experimenter will be ready to use the model in doing data analysis to determine significance of the model's parameters, estimate the mean response, and to arrive at optimum operating conditions on the control variables that result in a maximum or a minimum response over a certain region of interest which we denote by R .

In order to estimate the unknown parameters in model (2.1), a series on $n (> p)$ experiments are performed in each of which the response y is measured 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 D , of order $n \times k$ called the design matrix,

$$D = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix}; \tag{2.4}$$

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 D represents a point, referred to as a design point, in a k -dimensional Euclidean space. The response value obtained at the u^{th} setting of x , namely $x_u = (x_{u1}, x_{u2}, \dots, x_{uk})'$ ($u=1,2,\dots,n$), is denoted by y_u . From (2.1) we thus have

$$y_u = f'(x_u)\beta + \epsilon_u, \quad u = 1, 2, \dots, n; \tag{2.5}$$

where ϵ_u denotes the error term at the u^{th} experimental run. Model (2.5) can be expressed in matrix form as

$$y = X\beta + \epsilon; \tag{2.6}$$

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

A common method for estimating the parameter vector in (2.6) is the one based on the method of ordinary least squares. This method requires that ϵ has a zero mean and a variance-covariance matrix given by $\sigma^2 I_n$ [see, for example, Khuri and Cornell,¹⁰ (Section 2.3)]. In this case, the least-squares estimator of β , denoted by $\hat{\beta}$, is given by

$$\hat{\beta} = (X'X)^{-1} X'y; \tag{2.7}$$

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

$$\begin{aligned} Var(\hat{\beta}) &= (X'X)^{-1} X'(\sigma^2 I_n) X (X'X)^{-1} \\ &= \sigma^2 (X'X)^{-1}; \end{aligned} \tag{2.8}$$

An estimate $\hat{\mu}(x_u)$, of the mean response at x_u , is obtained by replacing β by $\hat{\beta}$, that is,

$$\hat{\mu}(x_u) = f'(x_u)\hat{\beta}, \quad u = 1, 2, \dots, n; \tag{2.9}$$

The quantity $f'(x_u)\hat{\beta}$ also gives the value of the predicted response, $\hat{y}(x_u)$, at the u^{th} design point ($u=1,2,\dots,n$). In general, at any point, x , in the region R , the predicted response $\hat{y}(x)$ is

$$\hat{y}(x) = f'(x)\hat{\beta}, \quad x \in R; \tag{2.10}$$

Using formula (2.8), the variance of $\hat{y}(x)$ is then given by [Khuri and Cornell,¹⁵ (Section 6.1) or Khuri,⁴³ (Section 12.4)

$$\text{Var}[\hat{y}(x)] = \sigma^2 f'(x) (X'X)^{-1} f'(x) \quad (2.11)$$

The expression on the right-hand side of (2.11) is called the *Prediction variance*. This is an important quantity since the quality of prediction depends on the size of this variance. Also, the determination of optimum operating conditions on the control Variables requires that the prediction variance be small over the region of interest, R , in order to arrive at reliable information about the true optimum of the response over R . This is of course dependent on the assumption that the postulated model in (2.1) does not suffer from lack of fit (LOF). For a study of LOF of a fitted response surface model, see, for example, Khuri and Cornell,¹⁵ (Section 2.6).

Response surface designs

The choice of the design matrix, D , is quite important in any response surface investigation since the prediction variance depends on D as can be seen from formula (2.11). Some common design properties are

I. Orthogonality

A design D is said to be orthogonal if the matrix $X'X$ is diagonal, where X is the model matrix in (2.6). In this case, the elements of $\hat{\beta}$ will be uncorrelated since the off-diagonal elements of $\text{Var}(\hat{\beta})$ in (2.8) will be zero. If the error vector ϵ in (2.6) is assumed to be normally distributed as $N(0, \sigma^2 I_n)$, then these elements will be also stochastically independent. This makes it easier to test the significance of the unknown parameters in the model.

II. Rotatability

A design D is said to be rotatable if the prediction variance in (2.11) 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. This causes the prediction variance to remain 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 will have the same variance.

The necessary and sufficient condition for a design to be rotatable was given by Box and Hunter.² See also Appendix 2B in Khuri¹⁶ introduced a measure that quantifies the amount of rotatability in a response surface design. This measure can be helpful in comparing designs on the basis of rotatability, assessing the extent of departure from rotatability, and in improving rotatability by a proper augmentation of a nonrotatable design.

III. Optimal designs

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, R . The design criteria that address the minimization of the variance associated with the estimation of model (2.1)'s unknown parameters are called *Variance-related criteria*. The

most prominent of such criteria is the D-Optimality criterion which maximizes the determinant of the matrix $X'X$. This amounts to the minimization of the size of the confidence region on the vector β in model (2.6). Another variance-related criterion that is closely related to D-optimality is the G-Optimality criterion which requires the minimization of the maximum over R of the prediction variance in (2.11).

Such variance-related criteria are often referred to as *Alphabetic optimality*. They are meaningful when the fitted model in (2.1) represents the true relationship connecting y to its control variables. There are, however, many situations where this is not the case due to fitting the "wrong model". This results in the So-called model bias. For example, a first-degree model may be fitted to a data set when in reality a second-degree model would be a better representative of the response data. Box and Draper^{3,42} placed a great emphasis on the role of bias in the choice of a response surface design and advocated choosing designs that minimized model bias. They considered the minimization of bias to be a very important design criterion, and in certain cases, even more important than the variance-related criteria.

Designs for first-degree models

Designs for fitting first-degree models as in (2.2) are called first-order designs. The most common of such designs are the 2^k factorial (k is the number of control variables), and the Plackett-Burman design.

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 D in (2.4) consists of all 1's, all -1's, or a combination thereof. It therefore 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 an agricultural experiment, two levels of fertilizer A are combined with two levels of fertilizer B in order to study their effects on the yield of a certain vegetable crop over a certain period of time. This results in a 2^2 factorial experiment with four treatment combinations.

If k is large ($k \geq 5$), the 2^k design requires a large number of design points. In this case, fractions of 2^k can be considered. For example, we can consider a one-half-fraction design which consists of one-half the number of points of a 2^k design, or a one-fourth-fraction design which 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 (2.2) 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 Box, Hunter, and Hunter,¹⁷ Raktoc, Hedayat, and Federer,¹⁸ and Montgomery.¹⁹ See also Chapter 3 in Khuri and Cornell.¹⁵

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 (2.2). 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 1's 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, D , has 8 points whose coordinates are x_1, x_2, \dots, x_7 , and is of the form

$$D = \begin{pmatrix} 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ -1 & 1 & 1 & 1 & -1 & 1 & -1 \\ -1 & -1 & 1 & 1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 & 1 & 1 & -1 \\ -1 & 1 & -1 & -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 & -1 & -1 & 1 \\ -1 & -1 & -1 & -1 & -1 & -1 & -1 \end{pmatrix} ;$$

Design arrangements for $k=3, 7, 11, \dots, 99$ factors can be found in Plackett and Burman.⁴⁴

Designs for second-degree models

These are designs for fitting second-degree models as in (2.3), which has $p = 1 + 2k + \frac{1}{2}k(k-1)$ parameters (they are also referred to as second-order designs). The number of distinct design points of such design must therefore be at least equal to p . The design settings are sometimes 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 ($u=1, 2, \dots, n, i=1, 2, \dots, k$).

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. For example, for $k=2$, the design matrix, in coded form, consists of 9 points as shown below:

$$D = \begin{pmatrix} -1 & -1 \\ -1 & 0 \\ -1 & 1 \\ 0 & -1 \\ 0 & 0 \\ 0 & 1 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix} ;$$

The number of design points for a 3^k design is 3^k , which can be very large for a large k . The use of full factorial designs is therefore not feasible if the number of experimental units is limited. 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, for example, in¹⁹ Chapter 9. See also McLean and Anderson,²⁰ Appendix 2.

The central composite design: The central composite design (CCD) is perhaps the most popular of all second-order designs. It was first introduced in¹ as an alternative to the 3^k design. This design consists of the following three parts:

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 of the design.

4. 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 α from the design center (chosen as the point at the origin of the coordinates system). We refer to α as the axial parameter.

5. A certain number, n_0 , of replications at the design center ($n_0 \geq 1$). This is called the center-point portion.

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

$$D = \begin{pmatrix} -1 & -1 \\ 1 & -1 \\ -1 & 1 \\ 1 & 1 \\ -\sqrt{2} & 0 \\ \sqrt{2} & 0 \\ 0 & -\sqrt{2} \\ 0 & \sqrt{2} \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

We note that the CCD is obtained by augmenting a first-order design, namely, the $2k$ axial points and the n_0 center-point replications. This design is usually developed in a manner consistent with the sequential nature of a response surface investigation in starting with a first-order design to fit 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 estimating all the p parameters in model (2.3). The fitted model is then used in the determination of optimum operating conditions on the control variables over the region of experimentation.

When k is large ($k \geq 5$), the factorial portion can be replaced by a fraction of a 2^k design. For example, for $k=5$, a one-half fraction

of 2^5 can be used giving a total of 16 points in the factorial portion instead of 32 (for more details about fractionating in the factorial portion) see Khuri and Cornell,¹⁵ Section 4.5.3.

A CCD can become rotatable by assigning the value $F^{1/4}$ to the axial parameter, α , where F denotes the number of points in its factorial portion, that is, $\alpha = F^{1/4}$. In addition, the number of center-point replications, n_0 , can be chosen so that a rotatable CCD will have the additional property of orthogonality (see Section 3). Note that orthogonality of a second-order design is attainable only after expressing model (2.3) in terms of orthogonal polynomials as explained in,² pp. 200-201. See also 15, 4.3. In particular, Table 4.3 in Khuri and Cornell's book can be used to determine the value of n_0 in order for a rotatable CCD to have the additional orthogonality property.

The box-behnken design: This design was developed by Box and Behnken.²¹ 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,¹¹ Khuri and Cornell,¹⁵ (Section 4.5.2), and.¹² Some Box-Behnken designs are rotatable, but this design is not always rotatable. Box and Behnken²¹ list a number of design arrangements for $k = 3, 4, 5, 6, 7, 9, 10, 11, 12,$ and 16 control variables.

The san cristobal design: Rojas²² introduced a variant of the CCD, called the San Cristobal Design (SCD), for sugar farming experiments. It is utilized in situations where the levels of k control variables are restricted to be nonnegative, as is the case with fertilizers experiments. The SCD consists of 2^k factorial points combined with center and axial points, all contained within the positive orthant. It also includes a control where no fertilizers are applied. More recently, the performance of this design was evaluated by Haines²³ who reviewed some of its properties.

Determination of optimum conditions

Optimization plays a key role in any response surface investigation. One of the main objectives of modeling the response is to use the fitted model in determining optimum conditions on the model's control variables that result in a maximum (or minimum) response over a certain region of interest, R . This, of course, assumes that the model has been screened to determine its suitability for providing an adequate representation of the mean response over the region R .

Quite often, a second-degree model is employed 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. We shall therefore only mention optimization techniques that are applicable to such a model.

Optimization of a second-degree model

Let us consider the second-degree model in (2.3), which can be written as

$$y = \beta_0 + x' \beta_* + x' Bx + \epsilon; \tag{4.1}$$

where $x = (x_1, x_2, \dots, x_k)'$, $\beta_* = (\beta_1, \beta_2, \dots, \beta_k)'$, and B is a symmetric matrix of order $k \times k$ whose i^{th} diagonal element is β_{ii} ($i=1, 2, \dots, k$), and its $(i, j)^{th}$ 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 D as in (2.4), then (4.1) can be written in vector form as in (2.6), where the parameter vector β consists of β_0 and the elements of β_* and B . Assuming that $E(\epsilon) = 0$ and $Var(\epsilon) = \sigma^2 I_n$, the least-squares estimate of β is $\hat{\beta}$ as given in (2.7). The predicted response at a point x in the region R is then of the form

$$\hat{y}(x) = \hat{\beta}_0 + x' \hat{\beta}_* + x' \hat{B}x, \tag{4.2}$$

where $\hat{\beta}_0$ and the elements of $\hat{\beta}_*$ and \hat{B} are the least-squares estimates of β_0 and the corresponding elements of β_* and B , respectively.

An unconstrained optimum is obtained by optimizing $\hat{y}(x)$ unconditionally with respect to x . This is achieved by taking the partial derivatives of $\hat{y}(x)$ with respect to x_1, x_2, \dots, x_k , equating each one to zero and then solving the resulting k equations. The solution to these equations provides the coordinates of the so-called *stationary point* which we denote by x_0 . This point may not necessarily be a point of optimum. For a maximum at x_0 , the matrix \hat{B} must be negative definite, or equivalently, if its eigenvalues are all negative. For a minimum \hat{B} , must be positive definite, or equivalently, if its eigenvalues are all positive (if some of the eigenvalues are positive and some are negative, then x_0 is a saddle point). Of course, an optimum is only meaningful if x_0 falls within the region R . If the location of the optimum falls outside this region, then it will be necessary to use the method of ridge analysis to determine Optimum conditions over R . This is explained in the next section.

The method of ridge analysis: When the location of the stationary point falls outside the region of interest, R , the next step is to determine optimum operating conditions within the boundary of R . For this purpose we use the method of ridge analysis, which was originally introduced by Hoerl²⁴ and later formalized by Draper.²⁵ This method optimizes $\hat{y}(x)$ in (4.2) subject to x being on the surface of a hypersphere of radius r and centered at the origin, namely,

$$\sum_{i=1}^k x_i^2 = r^2 \tag{4.3}$$

This constrained optimization is conducted using several values of r corresponding to hypersphere contained within the region R . The rationale for doing this is to get information about the optimum at various distances from the origin within R . Since this optimization is subject to the equality constraint given by (4.3), the method of Lagrange multipliers can be used to search for the optimum. Let us therefore consider the function

$$H = \hat{\beta}_0 + x' \hat{\beta}_* + x' \hat{B}x - \lambda (x'x - r^2), \tag{4.4}$$

where λ is a Lagrange multiplier. Differentiating H with respect to x and equating the derivative to zero, we get

$$\hat{\beta}_* + 2(\hat{B}x - \lambda x) = 0. \tag{4.5}$$

Solving for x , we obtain

$$x = -\frac{1}{2} (\hat{B}x - \lambda I_n)^{-1} \hat{\beta}_*. \tag{4.6}$$

A maximum (minimum) is achieved at this point if the matrix $\frac{\partial}{\partial x} \left[\frac{\partial H}{\partial x} \right]$ of second-order partial derivatives of H with respect to x is negative definite (positive definite). From (4.5), this matrix is given by

To achieve a maximum, Draper²⁵ suggested that λ be chosen larger than the largest eigenvalue of \hat{B} . Such a choice causes $\hat{B} - \lambda I$ to be negative definite. Choosing λ smaller than the smallest eigenvalue of \hat{B} causes $\hat{B} - \lambda I$ to be positive definite which results in a minimum. Thus, by choosing several values of λ in this fashion, we can, for each λ , find the location of the optimum (maximum or minimum) by using formula (4.6) and hence obtain the value of x $x = r^2$. The solution from (4.6) is feasible provided that r corresponds to a hypersphere that falls entirely within the region R . The optimal value of $\hat{y}(x)$ is computed by substituting x from (4.6) into the right-hand side of (4.2). 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 Khuri and Cornell,¹⁵ Section 5.7, and,¹¹ Chapter 19.

Khuri and Myers²⁶ provided a modification of the method of ridge analysis by placing an added constraint on the size of the prediction variance associated with the predicted response in (4.2) within the region r . The rationale for the additional constraint stems from the fact that since optimization is based on using $\hat{y}(x)$, which is a random variable, it would be necessary for the prediction variance not to be highly variable within R . This modification can provide better optimization results, especially when the design used to fit model (4.2) is not rotatable.

The results of ridge analysis can be easily obtained using PROC RSREG in SAS²⁷ and adding the "Ridge Max", or "Ridge Min" In this example, the predicted response is

$$\hat{y}(x) = 10.462 - 0.574x_1 + 0.183x_2 + 0.456x_3 - 0.678x_1x_2 + 1.183x_1x_3 + 0.233x_2x_3 - 0.676x_1^2 + 0.563x_2^2 - 0.273x_3^2.$$

The matrix \hat{B} [see formula (4.2)] is given by

$$\hat{B} = \begin{bmatrix} -0.676 & -0.339 & 0.592 \\ -0.339 & 0.563 & 0.117 \\ 0.592 & 0.117 & -0.273 \end{bmatrix} :$$

The stationary point x_0 corresponding to $\hat{y}(x)$ is located at (-.394, -.364, -0.175) is a saddle point since the eigenvalues of \hat{B} are 1.841, 0.367, -3.304 which have mixed signs. Thus \hat{B} is neither positive definite nor negative definite, that is, it is indefinite. To find the maximum of $\hat{y}(x)$ over R , it is necessary here to use the method of ridge analysis.

The needed SAS code to obtain the results of ridge analysis (using PROC RSREG) is given below

```
DATA;
INPUT x1, x2, x3, Y;
CARDS;
(enter here the data from Table 1)
PROC SORT;
BY X1 - X3;
RUN;
PROC RSREG;
MODEL Y = X1 - X3 /LACKFIT;
RIDGE MAX;
RUN;
```

statements, depending on whether it is desired to have a maximum response or a minimum response, respectively, over the region R . It should be noted that regardless of how the control variables are coded, SAS codes the variables so that the boundary of R has a radius equal to 1 (assuming that R is spherical). The next numerical example gives details of the SAS code needed to fit the model, get its parameter estimates, determine the nature of the stationary point, and finally display the results of ridge analysis.

A. Numerical example

A central composite rotatable design with 6 center-point replications was set up to investigate the effects of three fertilizer ingredients on the yield of snap beans. The fertilizer ingredients and actual amounts applied were nitrogen (N), from 0.94 to 6.29 lb/plot; phosphoric acid (P_2O_5), from 0.59 to 2.97 lb/plot; and potash (K_2O), from 0.60 to 4.22 lb/plot. The response of interest, y , is the average yield in pounds per plot of snap beans. The following: coded variables, x_1, x_2, x_3 , are given by

$$x_1 = \frac{N-3.62}{1.59}, x_2 = \frac{P_2O_5-1.78}{0.71}, x_3 = \frac{K_2O-2.42}{1.07}$$

The design settings (in coded form) and corresponding response values are given in Table 1 (same as Table 5.9 in,¹⁵ p. 185): We note that the design is rotatable since the axial parameter value is $\alpha = F^{1/4} = 1.682$, where $F = 8$ is the number of points in the factorial portion of this CCD. The region R is therefore spherical with a radius = 1.682.

The MODEL statement in PROC RSREG fits a second-degree model in the control variables, x_1, x_2, x_3 . Note that the statements, "PROC SORT" and "BY $X_1 - X_3$," are needed to perform a lack-of-fit test [see Section 2.6 in Khuri and Cornell (15)] on the second-degree model. The data are sorted by the variables x_1, x_2, x_3 , so that the replicated observations at the design center are grouped together. The actual lack-of-fit test is performed by adding the option "LACKFIT" to the MODEL statement in PROC RSREG. Using the data in Table 1, the resulting lack-of-fit F test statistic has the value $F = 2.91$ with 5 and 5 degrees of freedom. The corresponding p-value is 0.1333 which is not significant at the 0.10 level.

The results of ridge analysis are shown in Table 2. We note that the maximum response value, 12.886, is attained on the boundary of R (identified by the coded radius 1 which corresponds to the radius $r = 1.682$) at the point $x_1 = -0.544, x_2 = 1.589, x_3 = 0.089$. Expressed in units of pounds per plot, the corresponding levels of the original factors are $N = 2.755, P_2O_5 = 2.908, K_2O = 2.515$.

Table 1 Design settings and yield values

x_1	x_2	x_3	N	P_2O_5	K_2O	Yield
-1	-1	-1	2.03	1.07	1.35	11.28
1	-1	-1	5.21	1.07	1.35	8.44
-1	1	-1	2.03	2.49	1.35	13.19
1	1	-1	5.21	2.49	1.35	7.71
-1	-1	1	2.03	1.07	3.49	8.94
1	-1	1	5.21	1.07	3.49	10.9
-1	1	1	2.03	2.49	3.49	11.85
1	1	1	5.21	2.49	3.49	11.03
-1.682	0	0	0.94	1.78	2.42	8.26
1.682	0	0	6.29	1.78	2.42	7.87
0	-1.682	0	3.62	0.59	2.42	12.08
0	1.682	0	3.62	2.97	2.42	11.06
0	0	-1.682	3.62	1.78	0.6	7.98
0	0	1.682	3.62	1.78	4.22	10.43
0	0	0	3.62	1.78	2.42	10.14
0	0	0	3.62	1.78	2.42	10.22
0	0	0	3.62	1.78	2.42	10.53
0	0	0	3.62	1.78	2.42	9.5
0	0	0	3.62	1.78	2.42	11.53
0	0	0	3.62	1.78	2.42	11.02

Applications in agricultural and food sciences

Mead and Pike⁸ were among the first authors to explore the use of RSM in biological research. They examined a large number of papers in biological journals to determine the extent of using RSM ideas.

They reported that "not much awareness of current RSM methods was shown." They proposed a "joint development by the biologist and the statistician of particular biologically reasonable models for particular practical research problems." This is a good advice since the practical research worker will be more interested in methods that pertain to his (her) particular field of application rather than pursuing general results.

Fortunately, RSM has since become more applicable to a wide spectrum of applied research areas, including those with biological and agricultural applications. The development of new statistical software and the introduction of fast computers have made it a lot easier for practitioners to attempt more advanced RSM technique than was possible before. The food industry, in particular, has been a prime user of RSM since the early 1970s.⁵ devoted two sections to review various applications of RSM in the food and biological sciences. I myself was involved in one such application in determining the optimum combination of the levels of washing temperature, washing ratio of water volume to sample weight, and washing time on the quality of minced mullet flesh.²⁸

Table 2 Details of ridge analysis

Coded radius	Estimated response	x_1	x_2	x_3
0	10.462	0	0	0
0.1	10.575	-0.106	0.102	0.081
0.2	10.693	-0.17	0.269	0.11
0.3	10.841	-0.221	0.438	0.118
0.4	11.024	-0.269	0.605	0.12
0.5	11.243	-0.316	0.771	0.117
0.6	11.499	-0.362	0.935	0.113
0.7	11.79	-0.408	1.099	0.108
0.8	12.119	-0.453	1.263	0.102
0.9	12.484	-0.499	1.426	0.096
1	12.886	-0.544	1.589	0.089

In the remainder of this section, several papers will be cited to highlight some applications of RSM. These papers represent only a small sample since the actual number of papers with RSM applications is very large.

Edmondson¹⁴ provided an interesting application of RSM to greenhouse experiments and presented some valuable insights into the use of RSM in an agricultural setting versus an industrial one, as was mentioned earlier. Schmidt et al.,²⁹ investigated the effects of cysteine and calcium chloride on the textural and water-holding characteristics of dialyzed whey protein concentrates gel systems. These characteristics were measured by hardness (y_1), cohesiveness (y_2), springiness (y_3), and compressible water (y_4). A central composite design with five center-point replications was used and a second-degree model was fitted to each of the four responses. This experiment involved four response variables and is therefore labeled

as a multiresponse experiment. This is an important and a relatively recent area in RSM. It has attracted a lot of attention, especially in the context of simultaneous optimization of the various responses considered in the experiment. By this it is meant the determination of operating conditions on the control variables that result in optimum, or near optimum, values for all the responses. Khuri and Cornell¹⁵ applied a multiresponse optimization technique introduced by³⁰ to the simultaneous maximization of the four responses, namely y_1 , y_2 , y_3 , y_4 , in²⁹ experiment. A detailed review of multiresponse experiments can be found in Khuri.^{31,10} See also Chapter 7 in Khuri and Cornell.¹⁵ Another example concerning a multiresponse experiment was described in Evans et al.,³² who considered data of seed-germination percentages after four weeks incubation of four plant species in response to 55 alternating and constant-temperature regimes in dark laboratory germinators. A second-degree model was fitted to the data from each of the four species. This could have been treated as a multiresponse experiment involving the responses from the four species. Evans et al., however, chose to fit the models individually to their respective data, which is not advisable since the responses can be correlated and such individual fits ignore any interrelationships that may exist among the responses. Instead, multiresponse techniques should be used to fit the models in the multiresponse system [see, for example, Section 7.2 in¹⁵]

Keisling et al.,³³ utilized response surface techniques to predict weed age and future weed size from weed height. The objectives of their study were to: (a) utilize response models to generate data for describing weed interference in soybeans, (b) present strategies for estimating multispecies interference, and (c) project yield loss from existing data. Such a study was designed to produce information to assist soybean producers in recognizing economically detrimental threshold levels of weed infestations which require the initiation of control measures. Broudiscou et al.,³⁴ investigated the effects of several mineral compounds on feed degradation and microbial growth in a continuous culture system using RSM. The models considered were of the second degree fitted to data generated by a nonstandard design that consisted of 16 points giving seven levels to each of the four factors in the experiment. The design had good characteristics (by comparison to a CCD with 25 experimental runs, as shown in their Table II on page 257), was close to being orthogonal, and almost rotatable. The authors used Khuri's¹⁶ measure of rotatability to assess the percent rotatability of their design, which turned out to be 99.6 as compared to 89.2 for the CCD. Furthermore, the design was also more G-efficient.

RSM has received attention for modeling the performance of agronomic experiments. For example,³⁵ used inverse polynomials to model the yield of maize against three control variables, namely levels of nitrogen, phosphorous, and potassium. A 3^2 factorial design was used and the experiment was conducted in a randomized complete block layout with two replications per treatment combination. The inverse polynomial model³⁶ provided a better fit than the traditional second-degree model. The latter model may produce negative estimates of the yield response, which, of course, must be positive. This shows that taking into account any physical knowledge known about the response can be very beneficial when choosing an appropriate model.

Food science has also benefited from the application of RSM to its various areas of research. Diniz and Martin³⁷ used RSM to study the effects of pH, temperature, and enzyme-substrate ratio (E/S) on the degree of hydrolysis of dogfish muscle protein. The effect of the

hydrolysis variables was described using a Box-Behnken design. This design was also utilized by Jiang et al.,³⁹ to investigate the effects of x_1 = ultrasonic temperature (30-70 °C), x_2 = power (120-300 W), and x_3 = time (10-50 min) on ultrasonic assisted extraction for oligosaccharides from longan fruit pericarp (OLFP). Their fitted second-degree model was then used to obtain optimum conditions on x_1, x_2, x_3 that maximize the OLFP response. Optimization was also the goal of a study conducted by Cao et al.,³⁸ to study the effects of temperature, pH, and enzyme concentration/substrate concentration (E/S) ratio on the response, degree of hydrolysis (DH) for a marine shrimp called *acetes chinensis* that was harvested in China. The design used was a CCD for three control variables with $n_0 = 6$ center-point replications and an axial parameter $\alpha = 1.682$. This causes the design to be rotatable. Also, since $n_0 = 6$, the design has the additional uniform precision property.¹⁵ By definition, a rotatable design has the uniform precision property if the prediction variance at the design center is equal to the prediction variance at a distance of one (in the coded space) from the center. Such a property for a rotatable design maintains approximate uniform distribution of precision (in estimating the response) in the vicinity of the design center.¹⁵ The results of Cao et al.'s study indicated that hydrolysis of shrimp (*acetes chinensis*) resulted in a maximum DH value of about 26.33 % under the optimal conditions on temperature, pH, and E/S ratio.

Another optimization experiment was carried out by Zhang, et al.,⁴⁰ in a study concerning pyridoxine (PN), which is one of the three members of the vitamin B6 group. It has broad applications in the food industry, cosmetics, and medical supplies. RSM was successfully applied to determine optimum operating conditions for maximum conversion of PN. The control variables were reaction temperature, reaction time, enzyme loading, molar substrate ratio, and water activity. The response was the conversion of PN. The design used was a CCD whose factorial portion consisted of a one-half fraction of a 2^5 factorial, its axial portion contained 10 points with an axial parameter $\alpha = 2$, and $n_0 = 6$ center-point replications. This design is rotatable since $\alpha = F^{1/4}$ where $F = 16$ is the number of factorial points. It also has the uniform precision property since $n_0 = 6$ (Table 4.3 in¹⁵.) A listing of several applications of RSM in the optimization of chemical and biochemical processes was given by.⁴¹ In addition to their review of the recent literature on RSM applications to the aforementioned areas, they also provided a critique concerning the misuses of RSM in some of the reviewed articles.

Acknowledgements

None.

Conflicts of interest

None.

References

1. Box GEP, Wilson KB. On the experimental attainment of optimum conditions. *Journal of the Royal Statistical Society*. 1951;B 13:1–45.
2. Box GEP, Hunter JS. Multifactor experimental designs for exploring response surfaces. *Ann Math Statist*. 1957;28(1):195–241.
3. Box GEP, Draper NR. A basis for the selection of a response surface design. *J Amer Stat Assoc*. 1959;54:622–654.
4. Hill WJ, Hunter WG. A review of response surface methodology: A literature survey. *Technometrics*. 1966;8(4):571–590.

5. Myers RH, Khuri AI, Carter WH. Response surface methodology: 1966-1988. 1989;31:137-157.
6. Myers RH. Response surface methodology - Current status and future directions. *J Qual Technology*. 1999;31:30-44.
7. Myers RH, Montgomery DC, Vining GG, et al. Response surface methodology: A retrospective and literature survey. *J Qual Technology*. 2004;36:53-77.
8. Mead, R, Pike DJ. A review of response surface methodology from a biometric viewpoint. *Biometrics*. 1975;31:803-851.
9. Myers RH. Response Surface Methodology. Allyn and Bacon, Boston; 1971.
10. Khuri, AI. Multiresponse surface methodology. In: Handbook of Statistics, Vol. 13, Subir Ghosh, CR Rao, editors. Elsevier Science B. V, Amsterdam; 1996. p. 377-406.
11. Box GEP, Draper NR. Response Surfaces, Mixtures, and Ridge Analyses, 2nd ed. 2007.
12. Myers RH, Montgomery DC, Anderson-Cook CM. Response Surface Methodology. Wiley, Hoboken, New Jersey; 2009.
13. Box GEP, Youle PV. The exploration and exploitation of response surfaces: An example in the link between the fitted surface and the basic mechanism of the system *Biometrics*. 1955;11(3):287-323.
14. Edmondson RN. Agricultural response surface experiments based on four-level factorial designs. *Biometrics*. 1991;47(4):1435-1448.
15. Khuri AI, Cornell JA. Response Surfaces, 2nd ed. Dekker, New York, USA; 1996.
16. Khuri AI. A measure of rotatability for response surface designs, *Technometrics*. 1988;30:95-104.
17. Box GEP, Hunter WG, Hunter JS. Statistics for Experimenters; 1978.
18. Ratkoe BL, Hedayat A, Federer WT. Factorial Designs. Wiley, New York, USA; 1981.
19. Montgomery DC. Design and Analysis of Experiments. 6th ed. Wiley, New York, USA; 2005.
20. McLean RA, Anderson VL. Applied Factorial and Fractional Designs. Dekker, New York; 1984.
21. Box GEP, Behnken DW. Some new three-level designs for the study of quantitative variables. *Technometrics*. 1960;2(4):455-475.
22. Rojas BA. The San Cristobal design for fertilizer experiments, Proceedings of the 11th Congress of International Society for Sugar Cane Technology. 1962. p. 197-203.
23. Haines LM. Evaluating the performance of non-standard designs: The San Cristobal design. In: Response Surface Methodology and Related Topics. In: Andre I Khuri, editor. World Scientific, Singapore; 2006. p. 251-281.
24. Hoerl AE. Optimum solutions of many variable equations. *Chem Eng Prog*. 1953;55:69-78.
25. Draper NR. Ridge analysis of response surfaces. *Technometrics*. 1963;5(4):469-479.
26. Khuri AI and Myers RH. Modified ridge analysis. *Technometrics*. 1979;21:467-473.
27. SAS Institute, Inc. SAS Online Doc, Version 9.1.3. Author, Cary, North Carolina; 2008.
28. Tseo, CL, Deng JC, Cornell JA, et al. Effect of washing treatment on quality of minced mullet flesh. *J Food Sci*. 1983;48:163-167.
29. Schmidt RH, Illingworth BL, Deng JC, et al. Multiple regression and response surface analysis of the effects of calcium chloride and cysteine on heat-induced whey protein gelation. *J Agric Food Chem*. 1979;27:529-532.
30. Khuri AI, Conlon M. Simultaneous optimization for multiple responses represented by polynomial regression functions. *Technometrics*. 1981;23:363-375.
31. Khuri AI. The analysis of multiresponse experiments: A review. In: Statistical Design and Analysis of Industrial Experiments. Subir Ghosh, editor. Dekker, New York, USA; 1990. p. 231-246.
32. Evans RA, Easi DA, Book DN, et al. Quadratic response surface analysis of seed-germination trials. *Weed Science*. 1982;30(4):411-416.
33. Keisling TC, Oliver LR, Crowley RH, et al. Potential use of response surface analyses for weed management in soybeans (*Glycine max*), *Weed Sci*. 1984;32:552-557.
34. Broudiscou LP, Papon Y, Broudiscou AF. Effects of minerals on feed degradation and protein synthesis by rumen micro-organisms in a dual e uent fermenter. *Reprod Nutr Dev*. 1999;39:255-268.
35. Salawu IS, Adeyemi RA, Aremu TA. Modified inverse polynomial and ordinary polynomial as a response surface model: A case study of nitrogen, phosphate and potassium levels on the yield of maize. *Inter J Pure Appld Sciences*. 2007;1:18-24.
36. Nelder JA. Inverse polynomials, a useful group of multi-factor response function. *Biometrics*. 1962;22:128-141.
37. Diniz FM, Martin AM. Use of response surface methodology to describe the combined effects of pH, temperature, and E/S ratio on the hydrolysis of dog sh (*Squalus acanthias*) muscle. *Inter J Food Sci Tech*. 2003;31:419-426.
38. Cao W, Zhang C, Hong P, et al. Optimization of the production of shrimp waste protein hydrolysate using microbial proteases adopting response surface methodology. *Inter J Food Sci Tech*. 2009;44:1602-1608.
39. Jiang G, Jiang Y, Yang B, et al. Structural characteristics and antioxidant activities of oligosaccharides from longan fruit pericarp. *J Agric Food Chem*. 2009;57(19):9293-9298.
40. Zhang DH, Bai S, Dong XY, Sun Y. Optimization of lipase-catalyzed regioselective acylation of pyridoxine (vitamin B6). *J Agric Food Chem*. 2007;55:4526-4531.
41. Bas D, Boyaci IH. Modeling and optimization I: Usability of response surface methodology. *Journal of Food Engineering*. 2007;78(3):836-845.
42. Box GEP, Draper NR. The choice of a second-order rotatable design. *Biometrika*. 1965;50(1-2):335-352.
43. Khuri, AI. Linear Model Methodology. Chapman & Hall/CRC, Boca Raton, Florida; 2009.
44. Plackett RL, Burman JP. The design of optimum multifactorial experiments. *Biometrika*. 1946;33:305-325.