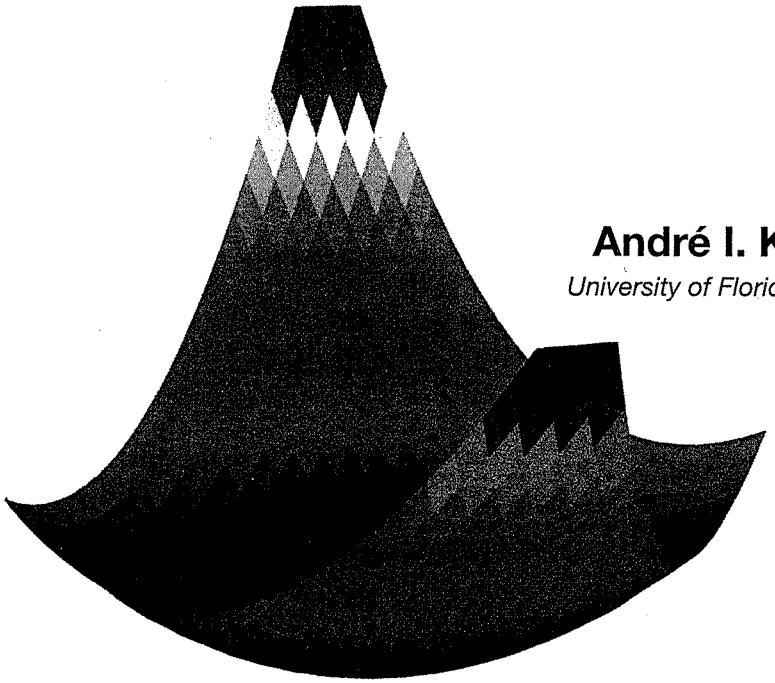
RESPONSE SURFACE METHODOLOGY AND RELATED TOPICS



editor

André I. Khuri

University of Florida, USA

World Scientific

NEW JERSEY • LONDON • SINGAPORE • BEIJING • SHANGHAI • HONG KONG • TAIPEI • CHENNAI

CHAPTER 4

STATISTICAL INFERENCE FOR RESPONSE SURFACE OPTIMA

Dennis K. J. Lin

Department of Supply Chain & Information Systems
The Pennsylvania State University
University Park, PA 16802-3005

John J. Peterson

GlaxoSmithKline Pharmaceuticals, R&D
709 Swedeland Road
King of Prussia, PA 19406-0939

This chapter is a review of research to date on statistical inference for response surface optima and related parameters. The need for such a review stems from the fact that the majority of literature on response surface optimization addresses only optimization of a fitted response surface. However, replication of the experiment will not produce the exact same response surface. As such, it is important to be able to assign some measures of risk or regions of uncertainty to our response surface optimizations. This chapter is an attempt to bring together and review most of the research that has been done on statistical inference for response surface optima and related parameters, such as the eigenvalues of the symmetric matrix of regression coefficients for the popular second-order polynomial model. This review will also include a review of research done on statistical inference for multiple response surfaces.

1. Introduction

Response surface methodology consists of a group of techniques used in the empirical study of the relationship between the response and a

number of input variables. Typically, the experimenter attempts to find the optimal setting for the input variables that maximizes (or minimizes) the predicted response. Suppose we have a set of data containing observations on a variable y and k predictor variables $\xi_1, \xi_2, \ldots, \xi_k$. A response surface model is a mathematical model fitted to y as a function of the ξ_i's in order to provide a summary representation of the behavior of the response, as the predictor variables are changed. This might be done in order to (a) optimize the response (minimize a cost, maximize a percentage yield, or minimize an impurity, for example), (b) find what regions of the ξ-space lead to a desirable product (viscosity within stated bounds, transparency not worse than a standard, appropriate color maintained, for example), or (c) gain knowledge of the general form of the underlying relationship with a view to describing options such as (a) and (b) to customers.

When the mechanism that produced the data is either unknown or poorly understood, so that the mathematical form of the true response surface is unknown, an *empirical model* is often fit to the data. An empirical model is usually linear in the parameters and often of polynomial form, either in the basic predictor variables or in transformed entities constructed from these basic predictors. The purpose of fitting empirical models is to provide a smooth curve that will summarize the data.

There is another useful type of model, however, the mechanistic model. If knowledge of the underlying mechanism that produced the data is available, it is sometimes possible to construct a model that represents the mechanism reasonably well. A mechanistic model usually contains fewer parameters, fits the data better, and extrapolates more sensibly. (Polynomial models often extrapolate poorly.) However, mechanistic models are often nonlinear in the parameters, and more difficult to formulate, to fit, and to evaluate. For information on this topic, see Bates and Watts[1] and Seber and Wilde[2].

When little is known of the nature of the true underlying relationship, the model fitted will usually be a polynomial in the ξ_i's. The philosophy applied here is that we are approximating the true but unknown surface by low-order (equivalently: low degree) terms in its Taylor's series

expansion. Most used in practice are polynomials of first and second order. The first-order model is

$$y_u = \alpha_0 + \alpha_1\xi_{1u} + \cdots + \alpha_2\xi_{2u} + \cdots + \alpha_k\xi_{ku} + \varepsilon_u \qquad (1)$$

where, $(y_u, \xi_{1u}, \ldots, \xi_{ku})$, $u = 1,2, \ldots, n$, are the available data, and where it is usually tentatively assumed that the error term $\varepsilon_u \sim N(0, \sigma^2)$, with independent errors for each observation. Such assumptions should always be carefully checked by examining the residuals (the differences between observed and predicted values of y) for possible contradictory patterns. The second-order model containing $k(k+1)/2$ additional terms is:

$$\begin{aligned} y_u = {} & \alpha_0 + \alpha_1\xi_{1u} + \alpha_2\xi_{2u} + \cdots + \alpha_k\xi_{ku} \\ & + \alpha_{11}\xi_{1u}^2 + \alpha_{22}\xi_{2u}^2 + \cdots + \alpha_{kk}\xi_{ku}^2 \\ & + \alpha_{12}\xi_{1u}\xi_{2u} + \alpha_{13}\xi_{1u}\xi_{3u} \cdots + \alpha_{k-1,k}\xi_{k-1,u}\xi_{ku} + \varepsilon_u \ . \end{aligned} \qquad (2)$$

Polynomial models of order higher than two are not typically used in practice. (However, an exception occurs for mixture experiments where 3rd order models are common. See, for example, Cornell[3]). This is partly because of the difficulty of interpreting the form of the fitted surface, and partly because the region of interest is usually chosen small enough for a first- or second-order model to be a reasonable choice. When a second-order polynomial is not adequate, and often even when it is, the possibility of making a simplifying transformation in y or in one or more of the ξ_i's would usually be explored before proceeding to higher order polynomial. A more parsimonious representation involving fewer terms is generally more desirable.

In actual applications, it is common practice to code the ξ_i's via, $x_i = (\xi_i - \xi_{i0})/c_i$, $i = 1, 2, \ldots, k$, where ξ_{i0} is some selected central value of the ξ_i range to be explored, and c_i is a selected scale factor. For example, if a temperature ($\xi = T$) range of 150-170^{0}C is to be covered using three levels, 150, 160, 170^{0}C, the coding $x = (T - 160)/10$ will code these levels to $x = -1,0,1$, respectively. The first- and second-order model would then be recast respectively as

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + \varepsilon \tag{3}$$

and

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + \beta_{11} x_1^2 + \ldots + \beta_{kk} x_k^2 + \beta_{12} x_1 x_2 + \ldots + \beta_{k-1,k} x_{k-1} x_k + \varepsilon . \tag{4}$$

However, it is often more useful to write the second-order model in (4) in a more compact form using vector and matrix notation as shown in the model in (5) below.

$$y = \beta_0 + \boldsymbol{x}'\boldsymbol{\beta} + \boldsymbol{x}'\boldsymbol{B}\boldsymbol{x} + \varepsilon , \tag{5}$$

where $\boldsymbol{x} = (x_1,\ldots,x_k)'$, $\boldsymbol{\beta} = (\beta_1,\ldots,\beta_k)'$, and $\boldsymbol{B}$ is a kxk symmetric matrix with diagonal elements equal to β_{ii} (i = 1,..., k) and off-diagonal elements equal to $\frac{1}{2}\beta_{ij}$ ($i < j$, j = 2,...,k). This would usually be fitted by least squares in the coded form. Substitution of the coding formulas into (3) or (4) enables the α's to be expressed in terms of the β's, if desired.

This chapter is organized as follows. First we consider response surface statistical inference for first-order models (3) by reviewing work on confidence regions for the path of steepest ascent (descent). Following this we consider quadratic models as in (5), and more general models as well. We discuss statistical inference for these models to address: the shape of the response surface in (5) (including inference about the $\boldsymbol{B}$ matrix), optimal mean response analysis (and in particular "ridge analysis"), a confidence region for the optimizing $\boldsymbol{x}$ value, multiple comparisons for response surface contrasts, and optima for multiple-response surfaces. We conclude with a brief discussion of some recent research work (with noise variables) and some considerations for future work related to statistical inference for nonstandard response surface models.

2. Statistical Inference for the Path of Steepest Ascent

For the first-order model as in (3), the path of steepest ascent (descent) is the path which produces a maximum (minimum) response subject to the constraint, $\boldsymbol{x}'\boldsymbol{x} = r^2$, as one increases r from zero onwards. More generally, this is also called the ridge trace path for a first order model. (In the next section we discuss the ridge trace path for a second-order model.) From now on we consider only the path (i.e. direction) of steepest ascent. The direction cosines for the true direction of steepest ascent are given by $\delta_i = \beta_i / \{\boldsymbol{\beta}'\boldsymbol{\beta}\}^{1/2}$ (i = 1,..., k), where $\boldsymbol{\beta} = (\beta_1,\ldots,\beta_k)'$. Note that $E(b_i) = \gamma\delta_i$ for each i, where $\gamma = \{\boldsymbol{\beta}'\boldsymbol{\beta}\}^{1/2}$ and b_i is the least squares estimator of the i-th regression coefficient. The direction cosines can easily be converted into direction angles of steepest ascent.

It is important to take into account sampling variation in assessing the direction of the path of steepest ascent. This is because that the estimated path is based on the regression coefficient estimates; and hence the estimated path itself has sampling variation. This sampling variation can lead to a confidence region for the underlying path based upon the true regression coefficients, the β_i's (i = 1,..., k). The value of the confidence region may be illustrated by plots, say in the case of two or three variables. A graphical analysis may indicate the amount of flexibility the practitioner has in experiments along the path. Furthermore, the set of angles of directions formed by this confidence region forms a confidence cone. The angles of directions excluded by such cones can be helpful in avoiding sub-optimal directions when an experimenter is trying to improve the mean responses for one or more first-order response surfaces.

Box and Draper[4] show how to obtain a 100(1−α)% confidence region for the δ_i's as follows. Suppose there are k design variables. The coefficients b_1, b_2, ..., b_k provide estimates of the relative movement of variables along the path. For the first-order model, the true path is defined by parameters β_1, β_2, ..., β_k,. Furthermore, $E(b_i) = \beta_i$, for i=1, ..., k, and the true coefficients are proportional to the relative movement along the path, which implies $\beta_i = \gamma\delta_i$, where the δ_i's represent the direction cosines of the path. In other words, δ_1, δ_2, ..., δ_k

are the constants which, if known, could be used to compute any coordinates on the true path. Now, if we think of this relationship as a regression model in which the b_i are responses and the δ_i are the levels of a single-predictor variables, then γ is the "regression coefficient" of the $b_1, b_2, \ldots, b_k$ on the $\delta_1, \delta_2, \ldots, \delta_k$. The required region is supplied by those elements $\delta_1, \delta_2, \ldots, \delta_k$ which just fail to make the residual mean square significant compared with $Var(b_i) = \sigma_b^2$ at some desired level α. Here, the variance of each b_i will be equal to some common value, σ_b^2, say, if a two-level orthogonal design is used.

Another way to view this problem is to consider the confidence region to be derived by inverting a hypothesis test about the direction cosines. If $\delta_{10}, \delta_{20}, \ldots, \delta_{k0}$ are the specified direction cosines, which equal the true direction cosines, then it follows that the null hypothesis, $H_0 : \boldsymbol{\beta} = \gamma \boldsymbol{\delta}_0$, is true, where $\boldsymbol{\delta}_0 = (\delta_{10}, \ldots, \delta_{k0})'$. Hence the confidence region for the true $\boldsymbol{\delta}$ value can be taken as the set of all $\boldsymbol{\delta}_0$ values such that $H_0 : \boldsymbol{\beta} = \gamma \boldsymbol{\delta}_0$ is not rejected at level α. That is (assuming a two-level orthogonal design) for those δ_i's which satisfy

$$\frac{1}{(k-1)} \sum_{i=1}^{n} \frac{(b_i - \hat{\gamma}\delta_i)^2}{s_b^2} \leq F_{1-\alpha}(k-1, \upsilon_b) \text{ and } \boldsymbol{\delta}'\boldsymbol{\delta} = 1,$$

where s_b^2 is an estimate of σ_b^2 and υ_b is the number of degrees of freedom on which this estimate s_b^2 is based. Here,

$$\hat{\gamma} = \sum_{i=1}^{n} b_i \delta_i \Big/ \sum_{i=1}^{n} \delta_i^2 = \sum_{i=1}^{n} b_i \delta_i .$$

Because all the quantities in the foregoing inequality are known except for the values of the δ_i's, this expression defines a set of acceptable δ_i's, hence a confidence region for the direction of steepest ascent. For details, see Box and Draper[4] or Myers and Montgomery[5]. Recently, Sztendur and Diamond[6] extend these results to cover the cases of heterogeneous error variances, nonorthogonal designs, and generalized linear models.

3. Statistical Inference Related to the Shape of the Quadratic Model

In process optimization, as the experimenter gets closer to a stationary point he or she can expect to encounter some curvature of the underlying response surface. As such, if one wants to have an adequate model in the neighborhood of a stationary point one should consider a response surface model with curvature. While we may not know the true functional form of the response surface, a second-order model is often a reasonable approximation about the neighborhood of a stationary point. The second-order model, $y = \beta_0 + \boldsymbol{x}'\boldsymbol{\beta} + \boldsymbol{x}'\boldsymbol{B}\boldsymbol{x} + \varepsilon$, can also be thought of as a second-order Taylor series approximation to the true response surface function.

The matrix $\boldsymbol{B}$ is important in that it contains information about the curvature of the second-order response surface. If all of the eigenvalues of $\boldsymbol{B}$ are positive (negative) then the stationary point of (5) corresponds to a minimum (maximum). If some eigenvalues are positive while the rest are negative, this corresponds to a saddle surface. If one or more eigenvalues are zero then this corresponds to a line, plane, or hyperplane of points that may form a rising ridge or a stationary ridge of points. If all of the eigenvalues are nonpositive (nonnegative) with some equal to zero, and there is a stationary point, then there will exist a line, plane, or hyperplane of stationary points that are equivalent maximum (minimum) points. This allows the experimenter some leeway with choosing factors that may satisfy other desirable conditions such as those relating to economic considerations. In any case, it is useful to be able to make statistical inferences about the eigenvalues of $\boldsymbol{B}$.

Box and Draper[4] stated without proof that for a rotatable design the variances of the eigenvalues of $\boldsymbol{B}$ are approximately equal to the variances of the estimates of the β_{ii}, terms in the model. A proof of this is outlined in Peterson[7,8] and also in Carter, Chinchilli, and Campbell[9]. This fact allows one to construct approximate confidence intervals for the eigenvalues of $\boldsymbol{B}$ when the design is rotatable or approximately rotatable.

Carter, Chinchilli, Myers, and Campbell[10] created a confidence interval for an eigenvalue of $\boldsymbol{B}$ for rotatable or nonrotatable

designs. However, their method can be quite conservative and requires solving the following two nonlinear programming problems

$$\left[\min_{\boldsymbol{B}\in C} \lambda_i(\boldsymbol{B}),\ \max_{\boldsymbol{B}\in C} \lambda_i(\boldsymbol{B})\right],$$

where $\lambda_i(\boldsymbol{B})$ is the i^{th} eigenvalue of $\boldsymbol{B}$ and C is the usual elliptically-shaped (normal-theory) confidence region for the elements of $\boldsymbol{B}$.

Carter, Chinchilli, and Campbell[9] sought less conservative confidence intervals for the eigenvalues of $\boldsymbol{B}$ by using the delta-method. They did a simulation study to show that their method works well for most designs occurring in practice, whether they are rotatable or nonrotatable.

Peterson[8] showed how to compute a confidence interval for the maximum (or minimum) eigenvalue of $\boldsymbol{B}$ as a byproduct of a ridge analysis methodology (which will be discussed later on in this chapter). This method is also applicable whether or not the design is rotatable. Bisgaard and Ankenman[11] provide a "double linear regression" method to compute the variances of the eigenvalues of $\hat{\boldsymbol{B}}$ and use them to get approximate confidence intervals. They then show that this approach is equivalent to the delta method of Carter, Chinchilli, and Campbell[9] but it can be done just using a standard statistical package that does regression analysis. However, a transformation of the design matrix using matrix multiplication is also needed. Ankenman[12] extends the approach of Bisgaard and Ankenman[11] to construct a hypothesis test to identify rising ridge behavior in a quadratic response surface. Here, one or more of the eigenvalues are zero and no stationary point exists. A simple example of such a quadratic response surface model with a rising ridge is a two-factor model with $\beta_{12} = \beta_{22} = 0$ and $\beta_2 \neq 0$.

4. Statistical Inference for the Optimal Mean Response

Even if statistical inference for the eigenvalues of $\boldsymbol{B}$ gives the experimenter a good idea of the shape of the response surface, further exploratory work may need to be done to understand how the response surface changes in an optimal fashion as operating conditions depart from the center of the design space. A graphical and analytical method

for making such an assessment is ridge analysis. The idea of ridge analysis was introduced by Hoerl[13,14] and given a more mathematical analysis by Draper[15].

In standard ridge analysis we maximize (or minimize) the mean response on circles, spheres, or hyperspheres centered at the center of the design space. If $\hat{y}(\boldsymbol{x})$ is the predicted mean response, then a ridge analysis (involving maximization) solves the following optimization problems for various radius values, r,

$$\hat{y}(\boldsymbol{x}_r) = \max_{\boldsymbol{x}'\boldsymbol{x}=r^2} \hat{y}(\boldsymbol{x}),$$

where $\boldsymbol{x}_r$ is the value of $\boldsymbol{x}$ that maximizes $\hat{y}(\boldsymbol{x})$ on $\boldsymbol{x}'\boldsymbol{x} = r^2$. The results of a ridge analysis are displayed on two plots. One plot is the "optimal response plot" which is a plot of $\hat{y}(\boldsymbol{x}_r)$ vs. r, while the other plot is called the "optimal coordinate plot", which is an overlay plot of x_{ir} vs. r $(i = 1,\ldots,k)$, where $\boldsymbol{x}_r = (x_{1r},\ldots,x_{kr})'$. Below in Figure 1 is the optimal response plot for a five factor response surface constructed from the data in example 11-5 in Box and Draper[4]. The corresponding optimal coordinate plot is given in Figure 2. In this example, the goal is to increase the percentage yield of a chemical process.

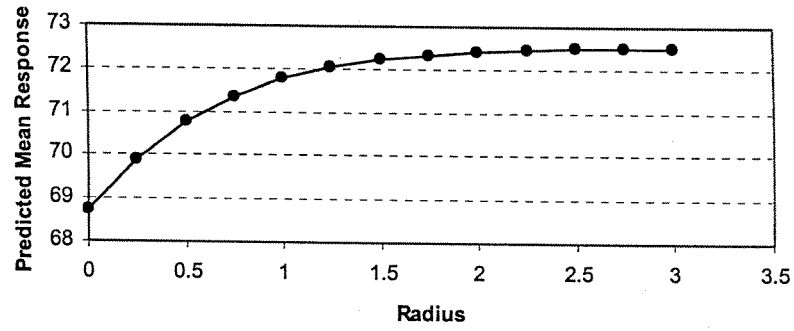


Fig. 1. The optimal mean response plot for the ridge analysis of the five-factor response surface example.

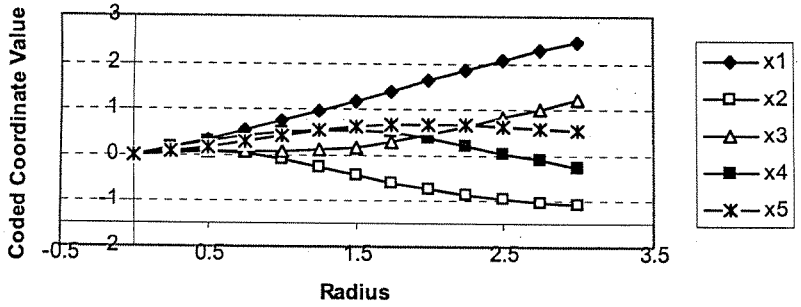


Fig. 2. The optimal coordinate plot for the ridge analysis of the five factor response surface example.

From Figure 1, we see that going out a distance of 2.5 to 3 from the center of the design space appears to maximize the mean yield of the chemical process. From Figure 2, we see that the corresponding process factors change in a smooth linear fashion. This is a nice example of the power of ridge analysis to capture the important aspects of a response surface optimization problem where the entire response surface itself cannot be easily plotted as we have more than two x-factors.

The optimal response plot allows the experimenter to see how the mean response changes in an optimal fashion regardless of the number of factors. Likewise, the optimal coordinate plot allows the experimenter to see how the factors change in an optimal fashion as we leave the center of the design space. Hence ridge analysis becomes increasingly important, and contour plots less interpretable, as the number of factors increase. Hoerl[16] gives a nice overview of classical ridge analysis. Extensions of ridge analysis to mixture experiments have been given by Hoerl[17], Peterson[8], and Draper and Pukelsheim[18,19]. Draper[15] established conditions to find the appropriate Lagrange multiplier value needed to compute $\hat{y}(\boldsymbol{x}_r)$ for each value of r. See Westlake[20] and Fan[21] for some discussion of the numerical analysis aspects. Khuri and Myers[22] introduce a modification of ridge analysis that takes into consideration the variance of the estimated response, $\hat{y}$. Paul and Khuri[23] extend this modification to weighted least squares regression and to generalized linear models.

Carter Chinchilli, Myers and Campbell[10] first proposed a confidence band for the underlying optimal response trace, $y(\boldsymbol{x})$ vs. r, in ridge analysis. This provides the investigator with a "guidance band" to help with making decisions about how far from the center of the design the new operating conditions should be set. Peterson[8] improved upon this approach and in doing so generalized the statistical model to the form

$$y = z(\boldsymbol{x})'\boldsymbol{\theta} + \varepsilon \ , \tag{6}$$

where z is a general function of the vector of factors, $\boldsymbol{x}$. The more general model in (6) is useful for modeling response surfaces that are not symmetric about their stationary point (such as cubic models) or in modeling some of the more exotic response surface functions useful in mixture experiments[3]. Peterson, Cahya, and del Castillo[24] propose an informal bootstrap approach for graphically assessing the uncertainty in a quadratic model ridge analysis with regard to both the optimal response and optimal coordinate plots. Their approach is particularly useful when there are more than three or four factors. Peterson, Cahya, and del Castillo[24] use example 11-5 of Box and Draper[4] to illustrate their proposed approach.

If it appears that the factor levels corresponding to the overall optimal response are well within the experimental region, then one may decide not to do a ridge analysis but to make a statistical inference about the global optimal response. Let $\eta(\boldsymbol{\theta})$ represent the maximum of the response surface function in (5), where $\boldsymbol{\theta}$ is the vector of all of the regression coefficients. Carter, W. H., Chinchilli, V. M., Campbell, E. D., and Wampler, G. L.[25], proposed a conservative approach to this problem where finding the confidence interval endpoints for $\eta(\boldsymbol{\theta})$ involves minimizing and maximizing $\eta(\boldsymbol{\theta})$ over the confidence region for $\boldsymbol{\theta}$. Unfortunately, this approach can become very conservative if the dimension of $\boldsymbol{\theta}$ is not small.

Using a general, differentiable response surface form, $f(\boldsymbol{x};\boldsymbol{\theta})$, Peterson[26] constructs a large-sample (delta method) confidence interval for the maximum of $f(\boldsymbol{x};\boldsymbol{\theta})$, where the maximum can be unconstrained or subject to a differentiable equality constraint, such as $\boldsymbol{x}'\boldsymbol{x} = r^2$. Chinchilli, Carter, Breen, and Campbell[27] propose a delta method

confidence interval for the maximum of the unconstrained quadratic form in (5); however, their method can be viewed as a special case of the confidence interval given by Peterson[26]. Chinchilli Carter, Breen, and Campbell[27] provide a small simulation study that indicates that about ten replications of a 3^2 design provides close to nominal coverage.

5. A Confidence Region for the Optimizing Factor Levels

In addition to having a confidence interval for the optimal mean response, it is useful to have a confidence region for the factor levels associated with the optimal point. Box and Hunter[28] (hereafter referred to as BH) proposed a confidence region for the stationary point of a response surface for the quadratic model in (5). However, they point out that their methodology can be applied to the more general parametrically linear model in (6). Unfortunately, for the more general model in (6) it can be difficult to characterize which stationary points are global optimal points. For the quadratic model, one can use the statistical inferences about the eigenvalues of $\boldsymbol{B}$ to provide evidence as to whether the response surface is convex, concave, or a saddle surface. Even here, however, one must be cautious as shown by Peterson, Cahya, and del Castillo[29] (hereafter referred to as PCD). They provide an example from Box and Draper[4] where a 90% confidence interval for the maximum eigenvalue of $\boldsymbol{B}$ indicates that it is negative, thereby indicating that $\boldsymbol{B}$ is negative definite (n.d.). A $\boldsymbol{B}$ that is n.d. corresponds to a concave response surface with a unique stationary point that is the global maximum. However, the associated 90% BH confidence region looks rather odd, being unbounded and split into two disjoint regions. PCD provide a theorem which helps explain why this can happen with the BH confidence region.

Sometimes the global response surface optimum may be outside of the region of admissible experimentation, and as such, one may desire a confidence region for constrained optimal factor levels. Stablein, Carter, and Wampler[30] (hereafter referred to as SCW) proposed a modification of the BH confidence region by using a Lagrange multiplier to model a constrained optimum. Böckenholt[31] extends the work of SCW to address

the situation where it is reasonable to have a common optimal $\boldsymbol{x}$-point in a multiple response surface experiment.

PCD propose a methodology for finding a confidence region for optimal factor levels that can be used for constrained or unconstrained situations, and can be used with the more general response surface model in (6). In addition, the constraint region can be completely general and involve equality or inequality constraints. The approach of PCD avoids Lagrange multipliers and is focused on global optimal points, not stationary points. Cahya, del Castillo, and Peterson[32] provide an algorithm that improves the computational speed and accuracy of the of PCD method. For the $\boldsymbol{x}'\boldsymbol{x} = r^2$ constraint in ridge analysis, Gilmour and Draper[33] propose a clever modification of the confidence region of SCW that appears to be conservative, although they do not provide a proof that their approach guarantees at least nominal coverage. Gilmour and Draper[33] prefer a somewhat conservative version of the SCW confidence region to compensate for the fact that the SCW confidence region is approximate.

A MatLab® program for executing the algorithm in Cahya, del Castillo, and Peterson[32] is available for free download at: http://www.ie.psu.edu/researchlabs/Engineeringstatistics/software.htm. The computer code of a Maple™ program called BH.mws for computing the BH confidence region is also available at the same address as above. This program is discussed in del Castillo and Cahya[34].

6. Multiple Comparison for Response Surface Contrasts

One area of statistical inference for response surface models that has received only modest attention is multiple comparisons. Sa and Edwards[35] introduced some special-case solutions for the "multiple comparisons with a control" (MCC) problem applied to quadratic response surfaces. Here statistical inference centers on the function,

$$\delta(\boldsymbol{x}) = E(Y \mid \boldsymbol{x}) - E(Y \mid \boldsymbol{x} = \boldsymbol{0}),$$

where $\boldsymbol{x} = \boldsymbol{0}$ is considered a combination of factor levels associated with a control. If the optimization goal is maximization, it is desired to find

simultaneous lower confidence bounds for $\delta(\boldsymbol{x})$. Such sets are typically displayed to show a region of $\boldsymbol{x}$-points where $\delta(\boldsymbol{x}) > 0$ with simultaneous confidence. Sa and Edwards[35] achieve this using existing regression methodology applied to response surface models. Sa and Edwards[35] consider the quadratic regression model in (4) and obtain an exact result for $k = 1$. They obtain a conservative solution for rotatable designs using a quadratic model.

Later, Merchant, McCann, and Edwards[36] improved upon this method by providing improved bounds for more widely applicable designs. Both of these solutions employ sophisticated probability inequalities to obtain the necessary critical values. However, the general Monte Carlo approaches discussed in Hsu[37] can also be used to compute MCC intervals. This approach is easy to understand and widely applicable for the parametrically linear response surface model, as in (4). A somewhat related problem has been discussed by Gilmour and Mead[38].

Moore and Sa[39] introduced "multiple comparisons with the best" (MCB) in response surface methodology. For MCB, they are considering instead the function,

$$\delta(\boldsymbol{x}) = E(Y \mid \boldsymbol{x}_0) - E(Y \mid \boldsymbol{x}),$$

where $\boldsymbol{x}_0$ corresponds to a stationary point of a quadratic response surface model. They assume that the stationary point for their model is a global optimum. They apply two different approaches, the delta method and an F-projection approach. Their delta method approach can be easily derived as a straightforward consequence of Peterson[26]. Both their delta method and their F-projection approaches use the Scheffe' critical value, which is very conservative. Using the delta method and the approach of Merchant, McCann, and Edwards[36], Miller and Sa[40] improve upon the methods of Moore and Sa[39] by obtaining approximate but much less conservative results. These MCB problems may have an intimate connection with the confidence region of PCD since their confidence region is composed of the $\boldsymbol{x}$-points for which the estimated optimal point is not statistically significantly better. A special set of multiple comparisons useful for combination-drug response surfaces has been proposed by Hung[41]. These simultaneous confidence intervals identify

treatment combinations that give better responses than either respective treatment component given alone. However, the methods in Hsu[37] can be used as well to compute these simultaneous confidence intervals, but in a less conservative manner.

7. Statistical Inference for Multiple-Response Surfaces

Statistical inference dealing with multiple response surface optimization has received more attention in recently. Overviews of multiple response surface methodology can be found in the books by Khuri and Cornell[42] and Myers and Montgomery[5]. The standard regression model for multiple response surface modelling is the classical multivariate multiple regression model,

$$\boldsymbol{Y} = \boldsymbol{B}\boldsymbol{z}(\boldsymbol{x}) + \boldsymbol{\varepsilon}, \tag{7}$$

where $\boldsymbol{Y}$ is a qx1 vector of q response types, $\boldsymbol{B}$ is now a qxp matrix of regression coefficients and $\boldsymbol{z}(\boldsymbol{x})$ is a px1 vector-valued function of $\boldsymbol{x}$. The vector $\boldsymbol{\varepsilon}$ has a multivariate normal distribution with mean vector $\boldsymbol{0}$ and variance-covariance matrix, $\boldsymbol{\Sigma}$. Typically in response surface analysis, the model in (7) takes the form of q quadratic models for each mean response.

The methods used most often for optimizing multiple response surfaces are "overlapping contour plots" and "desirability functions". The overlapping contour plot approach simply finds $\boldsymbol{x}$-points associated with estimated mean response values that satisfy some simultaneous (desirable) set of conditions, C, such as a set of $\boldsymbol{x}$-points corresponding to $\hat{\boldsymbol{y}} \in C$, where $\hat{\boldsymbol{y}} = \hat{\boldsymbol{B}}\boldsymbol{z}(\boldsymbol{x})$ and $\hat{\boldsymbol{B}}$ is the least squares estimate of $\boldsymbol{B}$.

A desirability function, $D(\hat{\boldsymbol{y}})$, is typically a (weighted) geometric mean of q individual desirability functions, $d_i(\hat{y}_i)$, one for each element, $\hat{y}_i$ of $\hat{\boldsymbol{y}}$. Each $d_i(\hat{y}_i)$ function is scaled to be between 0 and 1, with 0 indicating unacceptable quality and 1 indicating optimal quality with regard to the response associated with y_i. Since $D(\hat{\boldsymbol{y}})$ is a geometric mean of the $d_i(\hat{y}_i)$'s, it can only be close to 1 if *all* of the $d_i(\hat{y}_i)$'s are. Likewise, $D(\hat{\boldsymbol{y}})$ will be small if *any* of the $d_i(\hat{y}_i)$'s are sufficiently close to zero. One then seeks to find values of $\boldsymbol{x}$ to maximize

$D(\hat{y})$. For other approaches to desirability functions, see for example, Kim and Lin[43] and Kim and Lin[44].

Cahya, del Castillo, and Peterson[45] have proposed a large sample approach to a confidence region for the optimal factor levels, $\boldsymbol{x}_0$, for a desirability function. Here, they take a log or logit transformation of the desirability function to obtain a function, $g(\boldsymbol{B};\boldsymbol{x})$, of the regression model parameters in (7) and the factor levels, $\boldsymbol{x}$. They then approximate $g(\boldsymbol{B};\boldsymbol{x})$ by a first-order Taylor series about $\boldsymbol{B}$ and apply the methodology of Peterson, Cahya, and del Castillo[29] to obtain the confidence region for $\boldsymbol{x}_0$.

Using the model in (7), Ding, Lin, and Peterson[46] have constructed a large-sample confidence band about an optimal ridge trace for a desirability function. The optimal ridge trace here is defined as the function

$$\eta(\boldsymbol{B};r) = \max_{\boldsymbol{x}'\boldsymbol{x}=r^2} g(\boldsymbol{B};\boldsymbol{x}), \tag{8}$$

where $g(\boldsymbol{B};\boldsymbol{x})$ is defined using a logistic transformation of the overall desirability function. They construct $100(1-\alpha)\%$ asymptotic simultaneous confidence intervals for $\eta(\boldsymbol{B};r)$ for various values of the radius, r. Two versions of these simultaneous confidence intervals are obtained, respectively, by means of two different critical values. One critical value obtained using the Bonferroni adjustment is, $z_{\alpha/2h}$, where $z_{\alpha/2h}$ is the standard normal distribution percentile corresponding to the $\alpha/2h$ upper-tail-area and h is the number of radius values (i.e. intervals) used. The other critical value used is $\chi^2_{1-\alpha}(2)$. This is the asymptotic limit of the critical value used in Peterson[8] which was $2F_{1-\alpha}(2,\upsilon)$, where υ is the degrees of freedom associated with tests about the regression coefficients. Peterson's simulations had showed that $2F_{1-\alpha}(2,\upsilon)$ can be a good approximate critical value for a confidence band about a ridge trace for a univariate mean response surface. A simulation based upon parameter estimates from a realistic data set show that both critical values give good coverage with the Bonferroni adjustment being somewhat conservative and the $\chi^2_{1-\alpha}(2)$ critical value being close to the nominal 95% coverage rate. However, more simulation work needs to be done here to examine more examples and to see how small the sample sizes can be and still provide reasonable asymptotic results.

For the standard multivariate multiple regression model, Peterson[47] has proposed a Bayesian approach to multiple response surface optimization that uses a posterior predictive distribution to compute the probability that a vector of responses, $\boldsymbol{Y}$, lies in a desired region, A. This approach can also be easily modified to compute the posterior reliability that a desirability, $D(\boldsymbol{Y})$, or quadratic loss function, $Q(\boldsymbol{Y})$, will be within a desired range. In addition, a modification is possible to compute a Bayesian credible region for the factor levels that correspond to a good Bayesian reliability. Peterson[47] finds values of $\boldsymbol{x}$ to maximize the posterior probabilities,

$$Pr(\boldsymbol{Y} \in A \mid \boldsymbol{x}),\ Pr\left(D(\boldsymbol{Y}) \geq D^* \mid \boldsymbol{x}\right),\ \text{or}\ Pr\left(Q(\boldsymbol{Y}) \leq Q^* \mid \boldsymbol{x}\right), \tag{9}$$

where A, D^*, or Q^* are associated with desirable values specified by the experimenter. He shows that such posterior probabilities can be noticeably less than one might expect. This is due in part to the fact that the events within the probability measures in (9) may require the individual response-elements of $\boldsymbol{Y}$ to be within a small region of the joint distribution of $\boldsymbol{Y}$. The probabilities in (9) can also reduced noticeably due to model parameter uncertainty, for which the Bayesian approach takes account. Such reduction is more noticeable for small sample sizes.

A few extensions to Peterson[47] have been recently proposed. Miró-Quesada, del Castillo, and Peterson[48] have proposed a modification of Peterson[47] to include noise variables (see section 8). Peterson, Miró-Quesada, and del Castillo[49] have proposed a generalization of Peterson[47] to address seemingly unrelated regression models[50] and t-distribution, as well as normal distribution, error terms. Here they use Gibbs Sampling[51] to simulate the necessary posterior predictive distributions. Rajagopal and del Castillo[52] have proposed a modification to assess the uncertainty of the model form itself using "Bayesian model averaging"[15]. Rajagopal, del Castillo, and Peterson[53] extend this further with the incorporation of noise variables and t-distributed residual errors into the Bayesian model-

averaging framework. A further extension worth considering involves constructing a posterior predictive distribution for the multiple response surface model proposed by Chiao and Hamada[54] . Their model is nice in that it allows the variance-covariance matrix of the residual error vector to be a function of the covariates. However, it appears that some sophisticated Markov Chain Monte Carlo work would be needed to enable one to sample from the posterior predictive distributions.

8. Recent Results and Some Future Directions for Research

We wrap up this chapter with a discussion of some more recent work and possible future directions on statistical inference for response surface optima. Some recent work with regard to statistical inference for response surface optima involves modeling some of the factors as noise variables. A noise variable is a factor that cannot be completely controlled when the process is in actual operation, even though it can be controlled in an experimental setting. For example, this might involve the baking temperature of a food product that is cooked in a consumer's oven.

The quadratic response surface model often used for incorporating noise variables with controllable factors was initially proposed by Box and Jones[55], and appears in well-known response surface texts by Myers and Montgomery[5] and Khuri and Cornell[42]. It has the form

$$Y = \beta_0 + \boldsymbol{\beta}'\boldsymbol{x} + \boldsymbol{x}'\boldsymbol{B}\boldsymbol{x} + \boldsymbol{\gamma}'\boldsymbol{z} + \boldsymbol{x}'\boldsymbol{\Delta}\boldsymbol{z} + \varepsilon\,, \qquad (10)$$

where $\boldsymbol{x}$ is a kx1 vector of control factors, $\boldsymbol{z}$ is an lx1 vector of noise variables, and ε is a random normal error term with mean zero and variance σ^2 . The parameters $\boldsymbol{\beta}$ and $\boldsymbol{B}$ are as defined in (5), $\boldsymbol{\gamma} = (\gamma_1,\ldots,\gamma_l)'$, and $\boldsymbol{\Delta}$ is a kxl matrix composed of elements δ_{ij} ($i = 1,\ldots, k; j = 1,\ldots l$). In the model in (10), it is assumed that the noise variables have a multivariate normal distribution, and have been scaled so that they have a mean vector equal to $\mathbf{0}$ and a variance-covariance matrix equal to $\sigma_z^2\boldsymbol{\Omega}$. It is also often assumed that σ_z^2 and $\boldsymbol{\Omega}$ are known, based upon experience with the noise factors.[56]

In the presence of random noise variables, z, we wish to do robust optimization, that is we wish to find factor levels of $\boldsymbol{x}$ (the controllable factors) that minimize the "loss function" of the mean square error of the predictive model about a target value, T, i.e. minimize

$$MSE(\boldsymbol{x}) = E_z\left[m(\boldsymbol{x}, z; \boldsymbol{\phi}) - T\right]^2, \qquad (11)$$

where $m(\boldsymbol{x}, z; \boldsymbol{\phi})$ is the predictive mean model conditional on z. Here, $m(\boldsymbol{x}, z; \boldsymbol{\phi})$ is given by the first five terms in the model in (10). In (11) the expectation is done with respect to z, and $\boldsymbol{\phi}$ is the vector of all of the regression coefficient parameters in (10).

Statistical inference for response surface optima involving noise variables was first introduced by Kuhn[57]. This Kuhn modified the confidence region approach of SCW to consider a noise variable situation where one desires to minimize the variance of a response when the mean is subject to an equality constraint. It is important to point out that this problem has two sources of variability: the random replication error and the error due to the noise variables. It is the goal of robust process optimization to find controllable factor levels such that they minimize the influence of the noise variables on the variability of the response. This is possible if the some of the noise variables in the response surface regression model appear in interaction terms with some of the controllable factors[5].

Peterson and Kuhn[58] (hereafter referred to as PK) have introduced an approach to ridge analysis that allows the experimenter to incorporate noise variables into his/her analysis. Instead of doing a ridge analysis on the predicted mean response, $\hat{y}(\boldsymbol{x})$, PK propose minimizing the mean square about a target value in (11) subject to constraints of the form, $\boldsymbol{x}'\boldsymbol{x} = r^2$. PK also provide a method to compute a confidence band about the optimal ridge trace. They also provide a modification of their approach to handle "larger the better" (LTB) and "smaller the better" (STB) optimization problems. This is done by replacing the target, T, in (11) by an estimate of a maximum or minimum, respectively, over the experimental region. It is also possible to adapt the approach of PK to the dual response surface optimization problem as posed by Lin and

Tu[59]. Here, the mean is as in (4) or (5), but the standard deviation of the residual error is a parametrically linear model in the factors as well.

One computational issue with the ridge analysis approaches of Peterson[8] and PK is that if the number of factors is not small, then some kind of efficient global optimization procedure is needed to do the ridge analysis and to compute the confidence band about the optimal ridge trace. It has been suggested by Peterson[8] and PK that a global optimization algorithm may be a useful for doing the necessary computations. Therefore it is interesting to consider if a genetic algorithm[60] might be an efficient way to do the computations when the experimenter has, say, more than three or four (controllable) factors.

These ridge analysis procedures involve maximizing nonlinear functions of the factors. For each r, these functions may have several local optima on the constraint set, $S = \{\boldsymbol{x}: \boldsymbol{x}'\boldsymbol{x} = r^2\}$. For $\boldsymbol{x} = (x_1,...,x_k)'$, the spherical constraint set, S, can be transformed into a $(k\text{-}1)$ dimensional rectangle by the use of polar coordinates. Here, $\boldsymbol{x} = r\boldsymbol{t}(\boldsymbol{a})$, where $\boldsymbol{a}$ is a $(k\text{-}1)$x1 vector of (radian) angles in the set

$$A = \left\{\boldsymbol{a}: -\frac{\pi}{2} < a_i \le \frac{\pi}{2}(i = 1,...,k-2),\ -\pi < a_{k-1} \le \pi\right\}$$

The form of the vector-valued function, $\boldsymbol{t}$, can be found in Peterson[8] ((A.1)-(A.3)). Using this polar coordinate transformation, optimizations can now be done over A for various fixed values of r. Using genetic algorithms applied to the functions of angles on A, it may be possible to have reasonably large values of k when doing the necessary optimizations. But future work is needed to assess this computational approach to ridge analysis.

An additional area for future research work deals with the issue of obtaining reliable inferences with nonstandard regression models. Based on the empirical work of Lewis, Montgomery, and Myers[61], PCD have suggested that their confidence region method for the optimal $\boldsymbol{x}$-point ought to work reasonably well for generalized linear models but the simulation work still needs to be done. Likewise, one could ask if some of the methodologies above could be adapted to rank-based regression[62] or nonparametric regression[63]. It would not be surprising to be able to make such adaptations successful with large sample sizes, but the real challenge would be to see how small the experimental designs could be and still provide useful inferences about response surface optima.

References

1. Bates, D.M. and Watts, D.G. (1988), *Nonlinear Regression Analysis and Its Applications*. Wiley: New York.
2. Seber, G.A.F. and Wild, C.J. (1989), *Nonlinear Regression*, Wiley:New York.
3. Cornell, J. A. (2002), *Experiments with Mixtures*, 3rd ed., John Wiley and Sons, Inc., New York.
4. Box, G. E. P. and Draper, N. R. (1987) *Empirical Model Building and Response Surfaces*, New York: John Wiley.
5. Myers, R. H. and Montgomery, D.C. (2002), *Response Surface Methodology*, 2nd ed., John Wiley and Sons Inc., New York.
6. Sztendur, E. M. and Diamond, N. T. (2002) "Extensions to Confidence Region Calculations for the Path of Steepest Ascent", *Journal of Quality Technology*, **34**, 289-296.
7. Peterson, J. J. (1988), "A General Approach to Ridge Analysis with Confidence Intervals", Technical Report S-44, Syracuse University, Department of Mathematics.
8. Peterson, J. J. (1993) "A General Approach to Ridge Analysis with Confidence Intervals", *Technometrics*, **35**, 204-214.
9. Carter, W. H., Chinchilli, V. M., and Campbell, E. D. (1990), "A Large Sample Confidence Region Useful in the Characterization of the Stationary Point of a Response Surface", *Technometrics*, **32**, 425-435.
10. Carter, W. H., Chinchilli, V. M., Myers, R. H., and Campbell, E. D. (1986), "Confidence Intervals and an Improved Ridge Analysis of Response Surfaces", *Technometrics*, **28**, 339-346.
11. Bisgaard, S. and Ankenman, B. (1996), "Standard Errors for the Eigenvalues in Second-Order Response Surface Models", *Technometrics*, **38**, 238-246.
12. Ankenman, B. E. (2003). "Identifying Rising Ridge Behavior in Quadratic Response Surfaces," *IIE Transactions*, **35**, 493-502.
13. Hoerl, A. E. (1959), "Optimum Solution to Many Variables Equations", *Chemical Engineering Progress*, **55**, 69-78.
14. Hoerl, A. E. (1964), "Ridge Analysis", *Chemical Engineering Progress Symposium Series*, **60**, 67-77.
15. Draper, N. R. (1963), "Ridge Analysis of Response Surfaces", *Technometrics*, **5**, 469-479.
16. Hoerl, R. W. (1985), "Ridge Analysis 25 Years Later", *The American Statistician*, **39**, 186-192.
17. Hoerl, R. W. (1987), "The Application of Ridge Techniques to Mixture Data: Ridge Analysis", *Technometrics*, 29, 161-172.

18. Draper, N. R. and Pukelsheim, F. (2000), "Ridge Analysis of Mixture Response Surfaces", *Statistics & Probability Letters*, **48**, 131-140.
19. Draper, N. R. and Pukelsheim, F. (2002), "Generalized Ridge Analysis Under Linear Restrictions, with Particular Applications to Mixture Experiments Problems", *Technometrics*, **44**, 250-255.
20. Westlake, W. J. (1962), "A Numerical Analysis Problem in Constrained Quadratic Regression Analysis", *Technometrics*, **4**, 426-430.
21. Fan, S. S. (1999), "Computational Schema on Ridge Analysis", *Communications in Statistics – Simulations and Computation*, **28**, 767-783.
22. Khuri, A. I. and Cornell, J. A. (1996), *Response Surfaces*, 2nd ed., Marcel-Dekker, New York.
23. Paul, S. and Khuri, A. I. (2000) "Modified Ridge Analysis Under Nonstandard Conditions", *Communications in Statistics – Theory and Methods*, **29**, 2181-2200.
24. Peterson, J.J., Cahya, S., and del Castillo, E. (2004), Letter to the Editor, *Technometrics*, **46**, 355-357.
25. Carter, W. H., Chinchilli, V. M., Campbell, E. D, and Wampler, G. L. (1984), "Confidence Interval about the Response at the Stationary Point of a Response Surface, with an Application to Preclinical Cancer Therapy", *Biometrics*, **40**, 1125-1130.
26. Peterson, J. J. (1989) "First and Second Order Derivatives Having Applications to Estimation of Response Surface Optima", *Statistics & Probability Letters*, **8**, 29-34.
27. Chinchilli, V. M., Carter, W. H., Breen, T. J., Campbell, E. D. (1991), "An Asymptotic Confidence Interval for the Response at the Stationary Point of a Quadratic Response Surface", *Journal of Biopharmaceutical Statistics*, **1**, 287-301.
28. Box, G. E. P. and Hunter, J. S. (1954) "A Confidence Region for the Solution of a Set of Simultaneous Equations with an Application to Experimental Design", *Biometrika*, **41**, 109-199.
29. Peterson, J. J., Cahya, S. and del Castillo, E. (2002), "A General Approach to Confidence Regions for Optimal Factor Levels of Response Surfaces", *Biometrics*, **58**, 422-431.
30. Stablein, D. L., Carter, W. H., and Wampler, G. L., (1983) "Confidence Regions for Constrained Optima in Response Surface Experiments", *Biometrics*, **39**, 759-763.
31. Böckenholt, U. (1989) "Analyzing Optima in the Exploration of Multiple Response Surfaces", *Biometrics*, **45**, 1001-1008.
32. Cahya, S., del Castillo, E., and Peterson, J.J. (2004) "Computation of Confidence Regions for Optimal Factor Levels in Constrained Response Surface Problems", *Journal of Computational and Graphical Statistics*, **13** (2).
33. Gilmour, S. G. and Draper, N. R. (2004), "Confidence Regions Around the Ridge of Optimal Response on Fitted Second-Order Response Surfaces", *Technometrics*, **45**, 333-339.
34. Del Castillo, E. and Cahya, S. (2001), "A Tool for Computing Confidence Regions on the Stationary Point of a Response Surface", *The American Statistician*, **55**, 358-365.
35. Sa, P. and Edwards, D. (1993), "Multiple Comparisons with a Control in Response Surface Methodology", *Technometrics*, **35**, 436-445.

36. Merchant, A., McCann, M., and Edwards, D. (1998), "Improved Multiple Comparisons With a Control in Response Surface Analysis", *Technometrics*, **40**, 297-303.
37. Hsu, J. C. (1996), *Multiple Comparisons Theory and Methods*, Chapman & Hall, Suffolk, UK.
38. Gilmour, S. G. and Mead, R. (1995), "Stopping Rules for Sequences of Factorial Designs", *Applied Statistics*, **44**, 343-355.
39. Moore, L. J. and Sa, P. (1999), "Comparisons With the Best in Response Surface Methodology", *Statistics & Probability Letters*, **44**, 189-194.
40. Miller, L. K. and Sa, P. (2003), "Improved Multiple Comparisons With the Best in Response Surface Methodology", *Journal of Modern Applied Statistics*, **2**, 247-255.
41. Hung, H. M. J. (1992), "On Identifying a Positive Dose-Response Surface for Combination Agents", *Statistics in Medicine*, **11**, 703-711.
42. Kim, K. and Lin, Dennis K.J. (1998) "Dual Response Surface Optimization: A Fuzzy Modeling Approach," *Journal of Quality Technology*, **30**, 1-10.
43. Kim, K. and Lin, Dennis K.J. (2000) "Simultaneous Optimization of Mechanical Properties of Steel by Maximizing Exponential Desirability Functions," *Journal of Royal Statistical Society, Ser. C (Applied Statistics)*, **49**, 311-325.
44. Khuri, A. I. and Myers, R. H. (1979), "Modified Ridge Analysis", *Technometrics*, **21**, 467-473.
45. Cahya, S., del Castillo, E., and Peterson, J.J. (2002), "A Confidence Region for the Most Desirable Settings in a Multiple Response System", paper presented at the 46th Annual Fall Technical Conference of the American Society for Quality, Valley Forge, PA.
46. Ding, R., Lin, Dennis K.J. and Peterson, J.J. (2004), "A Note on Large-Sample Confidence Band for Multi-Response Ridge Path," *Quality and Reliability Engineering, International*, forthcoming.
47. Peterson, J. J. (2004) "A Posterior Predictive Approach to Multiple Response Surface Optimization", *Journal of Quality Technology*, **36**, 139-153.
48. Miró-Quesada, G., del Castillo, E. , and Peterson, J.J. (2004), "A Bayesian Approach to Multiple Response Surface Optimization in the Presence of Noise Variables", *Journal of Applied Statistics*, **31**, 251-270.
49. Peterson, J.J., Miró-Quesada, G., del Castillo, E. (2003) "A General Approach to Multiple Response Surface Optimization Based upon Posterior Predictive Distributions", paper presented at the 47th Annual Fall Technical Conference of the American Society for Quality, El Paso, TX.
50. Zellner, A. (1971), *An Introduction to Bayesian Inference in Econometrics*, John Wiley, New York, NY.
51. Press, S. J. (2003), *Subjective and Objective Bayesian Statistics: Principles, Models, and Applications*, 2nd edition. John Wiley, New York.
52. Rajagopal, R. and del Castillo, E. (2005), "Model–Robust Process Optimization using Bayesian Model Averaging", *Technometrics* (to appear).
53. Rajagopal, R., del Castillo, E., and Peterson J. J. (2005), "Model and Distribution-Robust Process Optimization with Noise Factors", *Journal of Quality Technology* (to appear).

54. Chiao, C. and Hamada, M. (2001) "Analyzing Experiments with Correlated Multiple Responses", *Journal of Quality Technology*, **33**, pp. 451-465.
55. Box, G. E. P. and Jones, S. (1990), "Designing Products that are Robust to the Environment", *Report Series in Quality and Productivity*, CPQI, University of Wisconsin, Number 56.
56. Borror, C. M., Montgomery, D. C., and Myers, R. (2002), "Evaluation of statistical designs for experiments involving noise variables", *Journal of Quality Technology*, **34**, 54-70.
57. Kuhn, A. M. (2003), "Optimizing Response Surface Experiments with Noise Factors Using Confidence Regions", *Quality Engineering*, **15**, 419-426.
58. Peterson, J. J. and Kuhn, A. M., (2005) "A Ridge Analysis with Noise Variables", *Technometrics* (to appear).
59. Lin, D. K. J. and Tu, W. (1995), "Dual Response Surface Optimization", *Journal of Quality Technology*, **27**, 34-39.
60. Michalewicz, Z. (1996), *Genetic Algorithms + Data Structures = Evolution Programs*, (3rd ed.) Springer-Verlag, New York.
61. Lewis, S. L., Montgomery, D. G., and Myers, R. H. (2001), Confidence Interval Coverage for Designed Experiments Analyzed with GLM's", *Journal of Quality Technology*, **33**, 279-292.
62. McKean, J. W. and Hettmasperger, T. P. (1998) *Robust Nonparametric Statistical Methods*, Oxford University Press, Oxford, UK.
63. Härdle, W. (1992) *Applied Nonparametric Regression*, Cambridge University Press, Cambridge, UK.