A GRAPHICAL COMPARISON OF DESIGNS FOR RESPONSE OPTIMIZATION BASED ON SLOPE ESTIMATION

by

Kimberly Kearns Hockman

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Statistics

APPROVED:

R. H. Myers, Chairman

K. H. Hinkelmann

M. R. Reynolds

M. Lentner

D. B. Birch

January 27, 1989

Blacksburg, Virginia

A GRAPHICAL COMPARISON OF DESIGNS FOR RESPONSE OPTIMIZATION BASED ON SLOPE ESTIMATION

by

Kimberly Kearns Hockman

Committee Chairman: Raymond H. Myers, Statistics

(ABSTRACT)

The response surface problem is two-fold: to predict values of the response, and to optimize the response. Slope estimation criteria are well suited for the optimization problem. Response prediction capability has been assessed by plotting the average, maximum, and minimum prediction variances on the surface of spheres with radii ranging across the region of interest. Average and maximum prediction bias plots have recently been added to the spherical criteria. Combined with the prediction variance, a graphical MSE criterion results. This research extends these ideas to the slope estimation objective.

A direct relationship between precise slope estimation and the ability to pinpoint the location of the optimum is developed, resulting in a general slope variance measure related to E-optimality in slope estimation. A more specific slope variance measure is defined and analyzed for use in evaluating standard response surface (RS) designs, where slopes parallel to the factor axes are estimated with equal precision. Standard second order RS designs are then studied in light

of the prediction and optimization goal distinction. Designs which perform well for prediction of the response do not necessarily estimate the slope precisely.

A spherical measure of bias in slope estimation is developed and used to measure slope bias due to model misspecification and due to the presence of outliers. A study of augmenting saturated orthogonal arrays of strength two to detect lack of fit is included as an application of a combined squared bias and variance measure of MSE in slope.

A study of the designs recommended for precise slope estimation in their robustness to outliers and to missing observations is conducted using the slope bias and general slope variance measures, respectively.

DEDICATION

To my husband, for his loving support, and to the memory of my mother for her silent inspiration.

ACKNOWLEDGEMENTS

I wish to express my sincere appreciation to the following:

Dr. Raymond H. Myers for his patient guidence through the dissertation research process, for his determined moral support and tolerance for an out-of-town student, and for his words of wisdom and encouragement concerning my personal and professional development.

Drs. K. H. Hinkelmann, M. R. Reynolds, J. B. Birch, and M. Lentner for graciously serving as committee members.

Dr. James M. Lucas for providing me with fresh ideas, for presenting so many professional opportunities, and for his continued friendship.

The Statistics Department at Virginia Tech for providing an outstanding Statistics program, and for financial support during my time in residence.

Dr. Chuck Pfeifer for providing employment enabling me to finance the completion of my Ph. D. research.

Dr. Max Morris for his assistance with the DETMAX algorithm.

Dr. Geoff Vining for his assistance with his search algorithm.

and I for their gracious hospitality.

My family, especially my father, for their loving support and encouragement.

My loving husband, for his expertise and patient assistance with computing facilities for word processing, graphics, and printing this document, as well as for his never ending love, support (financial and emotional), and encouragement.

TABLE OF CONTENTS

ABSTRACT..... ii

DEDICATION..... iv

ACKNOWLEDGEMENTS..... v

TABLE OF CONTENTS..... vi

CHAPTER 1: INTRODUCTION AND BACKGROUND..... 1

1.1 Introduction to the Response Surface Problem..... 1

1.2 The Relationship Between the Slope Variance and the Location of the Optimum..... 4

1.3 Existing Design Criteria for Response Prediction..... 8

1.3.1 Criteria Based on the Precision of $\underline{b}$..... 8

1.3.2 Prediction Variance Criteria..... 9

1.3.3 Rotatability..... 10

1.3.4 Spherical Variance and its Dispersion..... 11

1.3.5 Bias in Response Prediction on Spheres..... 12

1.4 Slope Estimation Literature..... 14

1.4.1 MSE of the Slope for a First Order Model..... 14

1.4.2 Variance of the Slope in a One Factor Model..... 14

1.4.3 Integrated Variance, Bias, MSE in Slope Estimation..... 15

1.4.4 Slope Rotatability..... 16

1.5 Notation 19

1.6 Families of Second Order Response Surface Designs..... 22

1.6.1 Central Composite Designs..... 22

1.6.2 Box-Behnken Designs..... 24

1.6.3 Hybrid Designs..... 25

1.6.4 Notz Designs..... 26

1.6.5 Hoke Designs..... 27

1.6.6 Box-Draper Designs..... 27

1.6.7 Saturated Orthogonal Arrays of Strength Two..... 28

CHAPTER 2: SPHERICAL SLOPE VARIANCE MEASURES..... 30

2.1 Extending the Relationship Between the Slope Variance and the Location of the Optimum..... 31

2.2 A Slope Estimation Measure Based on E-Optimality..... 32

2.3 Development of the Slope Variance Envelope for One Slope..... 38

2.4 Comparing Standard RS Designs using V^{*r}..... 40

2.5 V^{*r} for the Central Composite Designs..... 40

2.5.1 The Parameter α..... 41

2.5.2 The Number of Center Runs..... 43

2.5.3 Replicating Points in a CCD..... 44

2.5.4 The Slope Variance Envelope for the CCD..... 48

2.6 The Slope Variance on Spheres using Other Second Order Designs..... 50

2.6.1 Dual Axial CCD..... 50

2.6.2 Box-Behnken Design and Two CCDs..... 50

2.6.3 Hybrid Comparison..... 50

2.6.4 Saturated D-Efficient Designs..... 50

2.7 The "Best" Designs for Precise Slope Estimation..... 57

CHAPTER 3: SPHERICAL SQUARED BIAS AND MSE MEASURES..... 61

3.1 Bias in the Predicted Response..... 61

3.1.1 The Hazards of an Underfit Model..... 61

3.1.2 Prediction Bias..... 62

3.2 Bias in Slope Estimation..... 63

3.2.1 Maximizing Slope Bias on Spheres..... 64

3.2.2 Averaging Slope Bias Over Spheres..... 65

3.2.3 Mean Squared Error in Slope Estimation..... 66

3.3 Bias in Slope Estimation Using Augmented Saturated Orthogonal Arrays..... 70

3.3.1 The Alternative Models..... 71

3.3.2 Criterion for Augmentation..... 72

3.3.3 Detection Does Not Guarantee Protection..... 74

CHAPTER 4: A STUDY OF ROBUSTNESS IN SLOPE ESTIMATION..... 83

4.1 Robustness to Outliers in the Data..... 83

4.1.1 The Mean Shift Outlier Model for Slope Estimation..... 83

4.1.2 Maximizing and Averaging Outlier Bias on Spheres..... 86

4.1.3 Conclusions Concerning Robustness to Outliers..... 91

4.2 Robustness to Missing Observations..... 94

4.2.1 Robustness of the CCD to Missing Observations..... 94

4.2.2 Robustness of the BBD to Missing Observations..... 99

4.2.3 Robustness of the Hybrid 311B to Missing Observations.....101

4.2.4 Robustness of the Dual Axial CCD to Missing Observations.....102

4.2.5 Conclusions Concerning Robustness to Missing Observations.....106

CHAPTER 5: SUMMARY AND FUTURE RESEARCH AREAS.....107

5.1 Summary of Significant Contributions.....107

5.1.1 Precise Slope Estimation.....107

5.1.2 Bias and MSE in Slope Estimation.....108

5.1.3 Robustness to Outliers in the Data.....109

5.1.4 Robustness to Missing Observations.....110

5.2 Future Research Areas.....110

REFERENCES.....112

APPENDIX A.0.....117

APPENDIX A.1.....121

APPENDIX A.2.....124

APPENDIX A.3.....128

APPENDIX A.4.....132

VITA.....134

CHAPTER 1

INTRODUCTION AND BACKGROUND

1.1 INTRODUCTION TO THE RESPONSE SURFACE PROBLEM AND ITS GOALS

Response surface methodology (RSM) is a set of procedures for developing models which describe systems in a meaningful and useful way. It is a subject which is demanding increased attention as American industry becomes more quality conscious. Three texts which have been written on RSM are Khuri and Cornell (1987), Box and Draper (1987), and Myers (1976). In addition, a historical review paper addressing RSM techniques will appear in *Technometrics* in May of 1989 (Myers, Khuri and Carter).

A response surface model expresses a continuous, dependent variable, y, as a function of several continuous independent variables, $x_1, x_2, \ldots, x_k$. The true relationships among these variables may be unknown or quite complex, so a working model in the form of a low order polynomial is developed. There is obviously some error involved in describing a complicated system with a simple polynomial, so the analyst is tasked with finding the "best" model in some sense.

A first order model in k variables is written as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \epsilon \qquad (1.1.1)$$

where the β's are the model parameters usually estimated using the method of least squares, and ϵ represents the standard error. In other

cases, a higher order polynomial may be more appropriate. A second order model in k variables is written as

$$y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + \sum_{j=1}^{k} \beta_{jj} x_j^2 + \sum_{i<j}^{k} \beta_{ij} x_i x_j + \epsilon^*. \qquad (1.1.2)$$

First or second order models are more easily interpreted than higher order models, so they are seen more often in practice. For this reason, procedures developed in this dissertation will be limited to first and second order response models.

To fit the response surface model, that is, to estimate its parameters, a set of N experimental runs is performed. Either model can be written as

$$\underline{y} = X\underline{\beta} + \underline{\epsilon}. \qquad (1.1.3)$$

In this expression, $\underline{y}$ is the vector of response values; X is an (Nxp) matrix where p is the number of terms in the model, each column corresponding to a term in the model and each row representing an experimental run; $\underline{\beta}$ is the (px1) vector of model parameters to be estimated, (Graybill, 1976). The least squares estimate of $\underline{\beta}$ is

$$\underline{b} = (X'X)^{-1}X'\underline{y} \qquad (1.1.4)$$

and the variance-covariance matrix of the estimated parameters is

$$\text{var}(\underline{b}) = \sigma^2(X'X)^{-1}, \quad \text{assuming } \underline{\epsilon} \sim (0,\sigma^2 I).$$

The predicted response at a point $\underline{x}_0$, in the model space, is then

$$\hat{y}(\underline{x}_0) = \underline{x}_0'\underline{b} \qquad (1.1.5)$$

and the variance of this prediction is

$$\text{var}(\hat{y}(\underline{x}_0)) = \sigma^2 \underline{x}_0'(X'X)^{-1}\underline{x}_0. \qquad (1.1.6)$$

There are generally two purposes for building a response surface (RS) model. The first is to predict values of the response. The second is to identify values of the x's which yield a desirable or optimal response. In either case, the experimental design employed is critical to the experimenter's ability to achieve his goal. The proper choice of experimental design will depend on the purpose of the investigation. Although certain classes of designs perform well for either purpose, the member within such a class that is best for predicting the response may not be that which is best for locating the optimum.

Optimization techniques involve setting the partial derivatives $\partial y/\partial x_1$, $\partial y/\partial x_2, \ldots, \partial y/\partial x_k$ equal to zero and solving simultaneously for the x's. Thus, estimation of the partial derivatives, or the slope of the response surface, is more important in this case than estimating the response itself. The distinction is realized through the choice of the experimental, or response surface, design. The procedures developed in this dissertation are used to evaluate designs in their ability to precisely estimate the slope and to examine bias in slope estimation due either to model misspecification, or to outliers in the data. The discussions include several comparisons of designs preferred for response prediction to those recommended for slope estimation.

1.2 THE RELATIONSHIP BETWEEN THE SLOPE VARIANCE AND THE LOCATION OF THE OPTIMUM

To illustrate how use of the slope can lead to better estimation of optimum conditions, consider the one variable case. Two designs for fitting a second order model are compared on the basis of the variance of the estimated slope, and by calculating the probability of placing an arbitrary point inside the confidence interval (CI) around the optimum. It can be shown that the design which provides the more precise slope estimation is also that which yields the more attractive CI around the location of the optimum.

For k=1, consider two experimental designs with N=20 design points at -1, 0, or +1. The first design, D1, has eight points at each extreme, with four in the center. The second design, D2, has five points at each extreme, and ten in the center.

$$\text{D1:}\quad (n_{-1},\ n_0,\ n_{+1}) = (8,\ 4,\ 8)$$

$$\text{D2:}\quad (n_{-1},\ n_0,\ n_{+1}) = (5,\ 10,\ 5).$$

The model is $$y = \beta_0 + \beta_1 u + \beta_2 u^2 + \epsilon,$$

where $\epsilon \sim N(0,1)$, iid. The slope is given by

$$s = \beta_1 + 2\beta_2 u,$$

and the optimum, u_0, is the solution obtained by setting the slope equal to zero and solving for u,

$$u_0 = -\beta_1/2\beta_2.$$

The variance of the slope at a point u in the design space, for these designs, is

$$\text{var}(\hat{s}) = \text{var}(b_1) + 4u^2\text{var}(b_2)$$

$$= \sigma^2(c^{11} + 4u^2c^{22}),$$

where the c's denote elements of the $(U'U)^{-1}$ matrix,

$$(U'U)^{-1} = \begin{bmatrix} c^{00} & c^{01} & c^{02} \\ & c^{11} & c^{12} \\ \text{sym.} & & c^{22} \end{bmatrix},$$

and $U = [\underline{1}\ \underline{u}\ \underline{u}^2]$.

Plotted in Figure 1.2.1 are variances of the slope for each design. In the region $-.3<u<.3$, design D1 has the smaller slope variance, but outside (-.3, .3), D2 provides the more precise slope estimation.

Theorem 1.2.1: In the case of one independent variable (u) in a second order model, and symmetric experimental designs Di and Dj,

$$\text{if } \text{var}[\hat{s}(u),Di] > \text{var}[\hat{s}(u),Dj],$$

where u is an arbitrary location in the region, then the probability of placing the point u inside a $(1-\alpha)100\%$ confidence interval for u_0 using Di will be larger than that using Dj. This implies a higher probability of misclassifying non-optimal points as optimal. The proof and calculation of the probabilities are in Appendix A.0.

For example, consider the model

$$y = 1 + 4u - 2u^2.$$

The optimum is located at $u_0 = 1$. Observe in Figure 1.2.1 that D2 yields more precise slope estimation near the optimum. To illustrate that D2 also provides a more attractive CI around x_0, it is shown that the probability that an arbitrary $u \neq u_0$ value lies inside the CI, is less using D2 than for D1. Table 1.2.1 lists these probabilities for several points across the region. Keep in mind that the optimum is at 1.0.

Note that design D2 is the better design, since an arbitrary u has less chance of being in the CI. There is indeed a portion of the region in which D1 has slightly smaller slope variance. D1 would give a

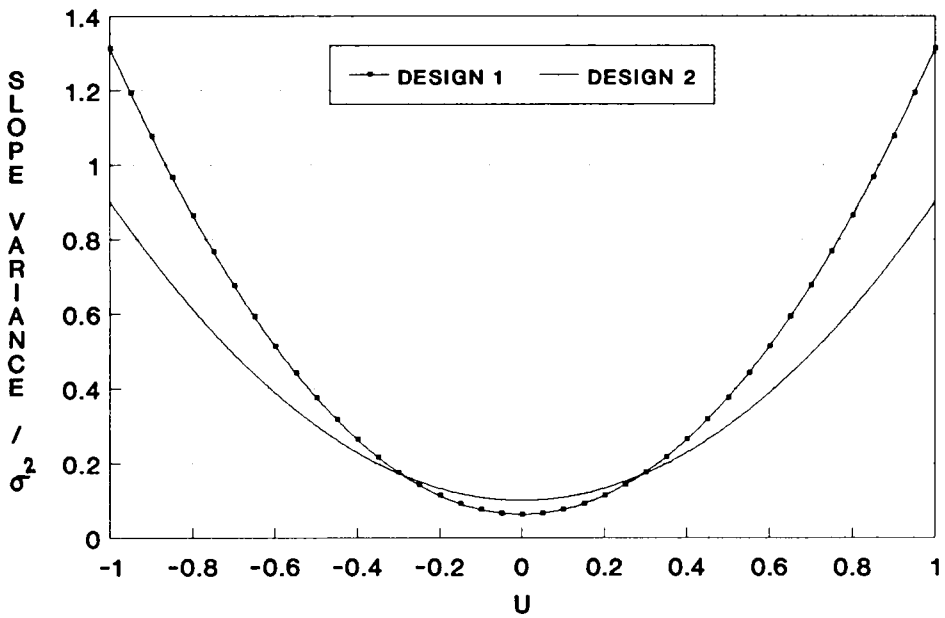

FIGURE 1.2.1
SLOPE VARIANCE COMPARISON FOR K=1
D1: (8, 4, 8) VS D2: (5, 10, 5)

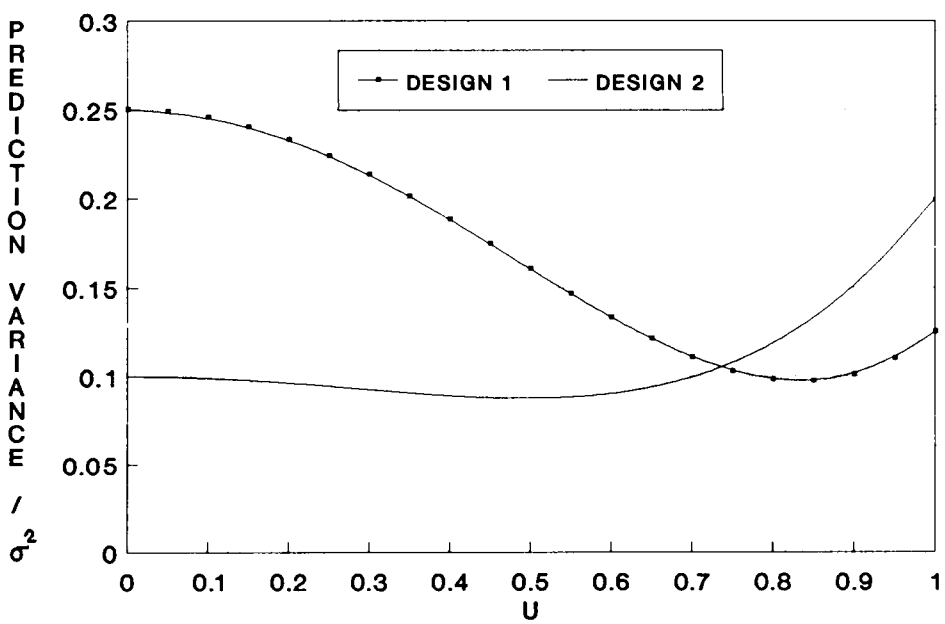

FIGURE 1.2.2
PREDICTION VARIANCE COMPARISON FOR K=1
D1: (8, 4, 8) VS D2: (5, 10, 5)

TABLE 1.2.1

u	DESIGN 1 Pr(u in CI)	DESIGN 2 Pr(u in CI)
.4	0	0
.5	.1	.04
.6	.38	.28
.7	.71	.60
.8	.85	.80
.9	.93	.92
1	1	1

smaller probability of falsely classifying these points ($-.3 \leq u \leq .3$) as optimal, but the probabilities using either design are quite small, so neither is likely to misclassify a point in this portion of the region.

A similar plot involving the variance of response prediction using each of these designs is depicted in Figure 1.2.2. Note that in the region near the optimum ($u_0=1$), design D1 yields more precise response prediction than does design D2, and yet it does not provide for more precisely locating the optimum. Thus, a study of response prediction variance alone could induce a false sense of security when dealing with the optimization problem. An extension of the relationship between slope variance and optimization to k independent variables is given in Chapter 2.

1.3 EXISTING DESIGN CRITERIA FOR RESPONSE PREDICTION

The choice of an experimental design is based on the purpose of the investigation. Many design criteria are available to assist the researcher in this choice. Several criteria will be presented in this section. Each has its merits, but none are specifically aimed at precise estimation of the slope of a surface.

1.3.1 CRITERIA BASED ON THE PRECISION OF $\underline{b}$

Finding the best model, in some sense, is the response surface analyst's goal. Thus, several optimality criteria have been aimed at precise estimation of the model parameters. One variance criterion which has received a lot of attention in the literature is D-optimality.

This seeks to minimize the generalized variance of the estimated coefficients. Specifically,

$$D = |\text{var}(\underline{b})| = |\sigma^2(X'X)^{-1}|,$$

where $|.|$ denotes the determinant. A D-optimal design minimizes D, or equivalently, maximizes $|(X'X)|$. Kiefer and Wolfowitz (1959, 1960) developed the D-optimality criterion by considering a design to be a continuous probability measure. Since real life designs are not truly such, approximate D-optimal designs are used in practice. St. John and Draper (1975) provide an extensive bibliography of D-optimality literature up to 1975 in their review of the criterion for regression designs.

Other criteria aimed at estimation of $\underline{\beta}$ include A-optimality, which minimizes the trace of $(X'X)^{-1}$, and E-optimality, which minimizes the maximum eigenvalue of $(X'X)^{-1}$ (references include Kiefer, 1958, 1959, 1961, 1962, 1974.)

1.3.2 PREDICTION VARIANCE CRIERIA

More directly related to precise estimation of the response is the G-optimality criterion. G-optimal designs minimize the maximun variance of the predicted response, $\sigma^2\underline{x}'(X'X)^{-1}\underline{x}$, over the region of interest in the design variable space. G-optimality, although a prediction variance criterion, is identical to D-optimality when the design is expressed as a probability measure on the design space. (Kiefer and Wolfowitz, 1960)

Another prediction variance criterion, developed by Box and Draper (1959, 1963), is the integrated prediction variance, IV, which seeks to minimize the average prediction variance in a region R. Specifically,

$$IV = \frac{NK}{\sigma^2}\int_R \mathrm{var}[\hat{y}(\underline{x})]d\underline{x}$$

$$= \frac{NK}{\sigma^2}\int_R \sigma^2\underline{x}'(X'X)^{-1}\underline{x}d\underline{x}, \qquad (1.3.1)$$

where $K^{-1} = \int_R d\underline{x}$ is the volume of the region, and N is the number of points in the design used to weight the criterion by the sample size. Box and Draper extended the IV criterion to include an integrated squared bias term due to model misspecification, which resulted in a mean squared error (J) criterion for response prediction. Myers and Lahoda (1975) developed the corresponding criterion for the slope. (See section 1.4.3.)

1.3.3 ROTATABILITY

A design property referred to as rotatability (Box and Hunter, 1957) allows designs with this property to be compared by looking at variance pictures of the predicted response. A rotatable design is one which results in equal precision of response prediction at all locations equidistant to the design center. For a rotatable design, a plot of prediction variance against distance from the center displays, in a simple picture, every prediction variance in the experimental region. This does not guarantee however, that the prediction variance will be small everywhere in the region, and the plot is meaningless outside the class of rotable designs.

1.3.4 SPHERICAL VARIANCE AND ITS DISPERSION

Hussey (1983), and Hussey, Myers and Houck (1988) developed the spherical variance to assess prediction capability on spheres for non-rotatable designs. The spherical variance, V^r, is the average prediction variance over the surface of a k-dimensional hypersphere of radius r, centered at the design origin. With a factor of N, which is inserted for design comparisons, V^r is a special case of the IV of Box and Draper.

$$V^r = \frac{NK}{\sigma^2} \int_{U_r} \text{var}[\hat{y}(\underline{x})]d\underline{x}, \qquad (1.3.2)$$

where $K^{-1} = \int_{U_r} d\underline{x}$ is the surface area of the sphere $U_r=\{\underline{x}: \sum_{i=1}^{k} x_i^2=r^2\}$. V^r is calculated for many concentric hyperspheres of varying radii. The spherical variance is plotted against the radius to provide a picture of the average prediction precision in the region.

Unfortunately, the spherical variance gives no information on the consistency of the prediction variance on spheres. Giovannitti-Jensen and Myers (1989) have since developed prediction variance dispersion measures to provide this information. A range for $\text{var}[\hat{y}(\underline{x})]/\sigma^2$ on spheres is defined

$$\text{RofV}(r) = \max_{\underline{x}\in U_r} \text{var}[\hat{y}(\underline{x})]/\sigma^2 - \min_{\underline{x}\in U_r} \text{var}[\hat{y}(\underline{x})]/\sigma^2. \qquad (1.3.3)$$

But more informative than the range is the plot of V^r, the maxima, and the minima on spheres, appearing on the same graph.

Presenting the spherical prediction variance and the maximum and minimum prediction variances corresponding to each radius in a single plot provides a clear, concise picture of prediction capability on

spheres for a given design. The plot is termed a Variance Dispersion Graph (VDG). Giovannitti-Jensen and Myers inspected many designs in this light, but for comparison to slope estimation designs in the variance sense, only those for fitting a second order model are of interest.

The forms of $var[\hat{y}(\underline{x})]$ and RofV depend on the model. Due to the complexity of the second order case, an optimization algorithm must be used to determine RofV. The authors used a computer program developed by Vining (1988b) to locate the extreme variances on spheres.

Figure 1.3.1 is a VDG for the central composite design (Section 1.6) with α=1.0 and three center runs. Note that the design can be evaluated efficiently in its ability to provide a model useful for response prediction across the region using a VDG.

1.3.5 BIAS IN RESPONSE PREDICTION ON SPHERES

Vining (1988c), develops prediction bias plots as an anologue to the variance dispersion graphs to evaluate designs in their ability to protect against model misspecification and the mean shift outlier model. Combined with the spherical variance plots, a spherical MSE measure results. Designs performing well by this criterion are reasonable compromises between minimum variance and minimum bias designs. More details pertaining to this research will be revealed as the ideas are modified to deal with slope estimation bias.

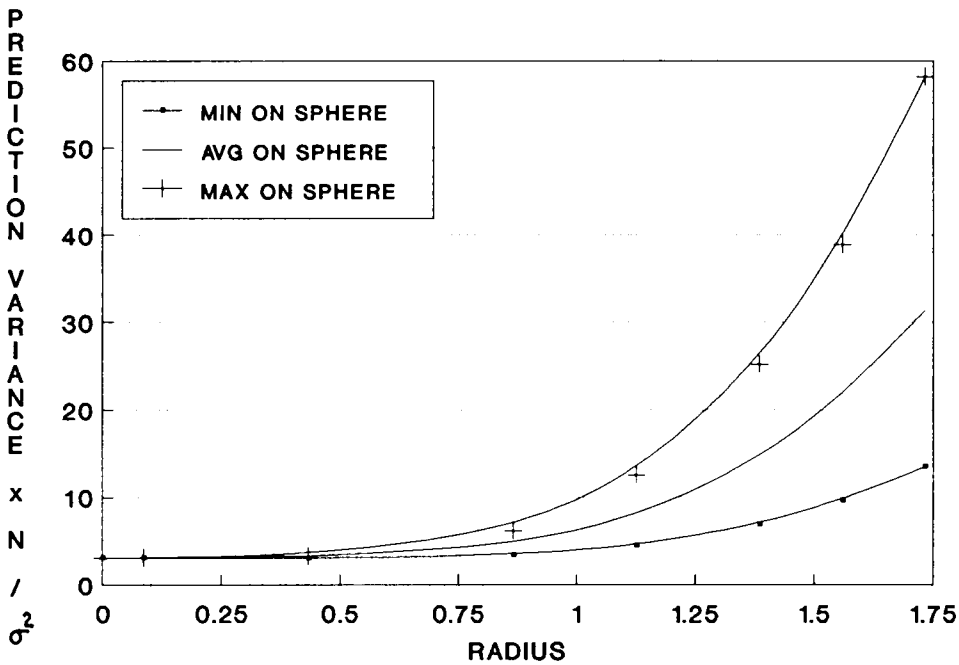

FIGURE 1.3.1
VARIANCE DISPERSION GRAPH
CCD FOR k=3, ALPHA=1.0, 3 C.R.s

1.4 SLOPE ESTIMATION LITERATURE

There is limited literature on slope estimation. Results from the papers discussed here will either be helpful in the development of the present methods for evaluating designs, or the designs found in the literature will be evaluated using the new methods.

1.4.1 MSE OF THE SLOPE FOR A FIRST ORDER MODEL

Atkinson (1970) considers slope estimation by fitting a first order model when the true response is quadratic and uses a minimum squared error criterion. His design criterion assumes the relative magnitudes of bias and variance in order to construct designs which minimize mean squared error at a point $\underline{x}_0$, or equivalently over a region centered at $\underline{x}_0$. It does not restrict $\underline{x}_0$ to the center of the experimental region. An interesting feature of his criterion is that designing experiments to estimate any directional slope is equivalent to designing experiments to estimate slopes along the axes of the design variables.

1.4.2 VARIANCE OF THE SLOPE IN A ONE FACTOR MODEL

Ott and Mendenhall (1972) consider estimating the slope of a second order response curve in the k=1 case, with three design point locations. Their goal is to find locations for, and assign design points to, the three locations, so as to achieve a near minimum variance of the least squares estimate of slope [V(s|u)] over the experimental region. They found that

1) The variance V(s|u) is inversely proportional to r^2, the squared radius of the experimental region;

2) V(s|u) is a parabola in u opening upward;

3) With the locations scaled by the experimental region, the number of points at -1 should equal the number at +1, and the third location should be at u=0, the design center;

4) The proportion of points at the extremes depends on whether the experimenter is more interested in predicting the slope in the center of the region or at the extremes.

These characteristics will be seen in the slope estimation criterion developed for standard designs in more than one independent variable.

1.4.3 INTEGRATED VARIANCE, BIAS, AND MSE IN SLOPE ESTIMATION

Myers and Lahoda (1975) contribute to the collection of slope estimation criteria by considering both bias and variance as does Atkinson. Their work extends the J criterion (MSE) of Box and Draper to the slope. They seek to minimize

$$J^{**} = \frac{NK}{\sigma^2}\int_R E[\hat{\gamma}(\underline{x})-\gamma_2(\underline{x})]'[\hat{\gamma}(\underline{x})-\gamma_2(\underline{x})]d\underline{x} \qquad (1.4.1)$$

where $K^{-1}=\int_R d\underline{x}$ is the volume of the region of interest, N is the number of design points, $\hat{\gamma}(\underline{x})$ is the vector of partial derivatives of the fitted response function, $\gamma_2(\underline{x})$ is the vector of true partial derivatives, and R is the region of interest. Common to the MSE criteria is the partition into variance, V**, and bias, B**, where

$$V^{**}=\frac{NK}{\sigma^2}\int_R E[\hat{\gamma}(\underline{x})-E(\hat{\gamma}(\underline{x}))]'[\hat{\gamma}(\underline{x})-E(\hat{\gamma}(\underline{x}))]d\underline{x} \qquad (1.4.2a)$$

and

$$B^{**}=\frac{NK}{\sigma^2}\int_R [E(\hat{\gamma}(\underline{x})-\gamma_2(\underline{x})]'[E(\hat{\gamma}(\underline{x})-\gamma_2(\underline{x})]d\underline{x}. \qquad (1.4.2b)$$

When fitting a second order model in the presence of a true cubic response, V** and B** cannot be simultaneously minimized, but B** and (V**|min B**) can be. Designs which meet this criterion are dual axial central composite designs, where the design matrix is written

$$D = \begin{bmatrix} \pm g & \pm g & \cdots & \pm g \\ +\alpha_1 & 0 & \cdots & 0 \\ -\alpha_1 & 0 & \cdots & 0 \\ 0 & +\alpha_1 & \cdots & 0 \\ 0 & -\alpha_1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & +\alpha_1 \\ 0 & 0 & \cdots & -\alpha_1 \\ +\alpha_2 & 0 & \cdots & 0 \\ -\alpha_2 & 0 & \cdots & 0 \\ 0 & +\alpha_2 & \cdots & 0 \\ 0 & -\alpha_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & +\alpha_2 \\ 0 & 0 & \cdots & -\alpha_2 \end{bmatrix} . \qquad (1.4.3)$$

Values for g, α_1, and α_2 are given for k=2 to 7 for both a cuboidal and a spherical region of interest. These designs will be evaluated using the slope oriented methodology in this dissertation.

1.4.4 SLOPE-ROTATABILITY

Hader and Park (1978) determined the parameters of the central composite design (Box and Wilson, 1951) which achieve the slope rotatability property -- the analogue of the Box-Hunter rotatability property for the second order response surface design. Slope rotatability implies that the variance of an estimated slope in the direction of an independent variable axis is a function only of a point's distance from the design origin. They assume designs that have

the property that slopes in the direction of each factor axis have the same variance, and unfortunately, rotatability holds only for each of them separately and not for a slope in any arbitrary direction.

By plotting the variance of the estimated slope at a point, against its distance from the design center (henceforth referred to as its radius) for several radii in the design region, the performance of a slope-rotatable design is evaluated efficiently. The present methods will allow similar plots to be informative even if the experimental design is not slope-rotatable.

Figure 1.4.1 depicts the variance of the slope apart from σ^2, using a slope-rotatable CCD, plotted against radii ranging from zero to the experimental design perimeter. The design matrix is given below, but in the figure it has been scaled so that all design points lie on or within a sphere of radius $\sqrt{k}$ for the k=3 case. Such a plot is meaningful only if the design is slope-rotatable.

$$D = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \\ -1 & 1 & 1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \\ -1 & -1 & -1 \\ -2.4324 & 0 & 0 \\ 2.4324 & 0 & 0 \\ 0 & -2.4324 & 0 \\ 0 & 2.4324 & 0 \\ 0 & 0 & -2.4324 \\ 0 & 0 & 2.4324 \\ 0 & 0 & 0 \end{bmatrix} . \qquad (1.4.4)$$

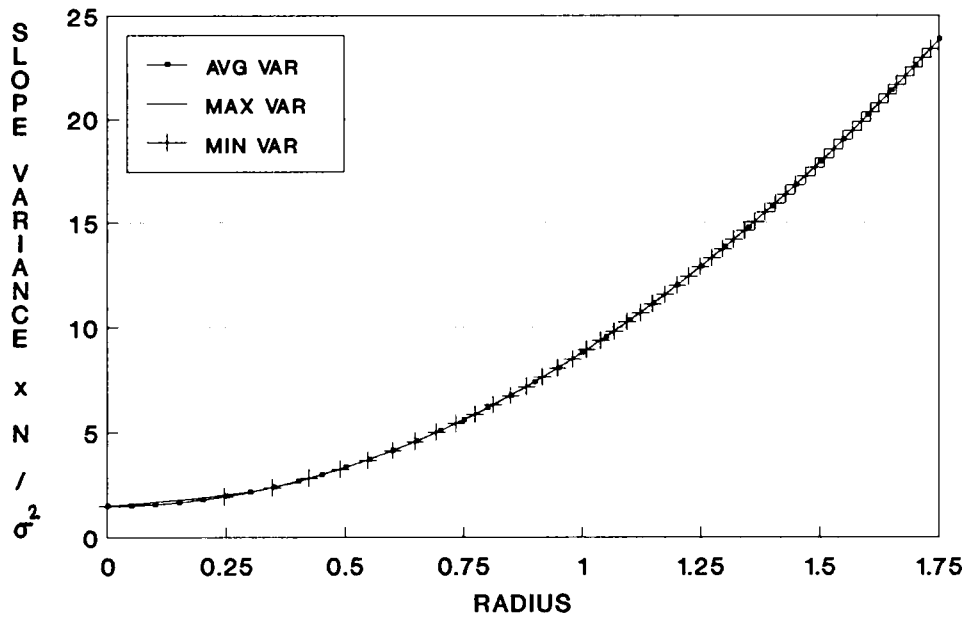

FIGURE 1.4.1
SLOPE-ROTATABLE CCD WITH ONE CENTER RUN;
SLOPE VARIANCE FOR K=3 AND ALPHA=2.4324

1.5 NOTATION

This section will summarize the notation used in the remaining pages. Specific meanings may change as needed and will be duly noted.

Let D denote an (Nxk) design matrix for the investigation of k variables in N experimental runs.

$$D = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{k1} \\ x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & & \vdots \\ x_{1N} & x_{2N} & \cdots & x_{kN} \end{bmatrix}. \qquad (1.5.1)$$

X is the corresponding (Nxp) model matrix, where p is the number of terms in the model. In the first order case, p=k+1; in the second order case, p=k(k+3)/2 +1. A second order model matix is given by

$$X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{k1} & x_{11}^2 & \cdots & x_{k1}^2 & x_{11}x_{21} & \cdots & x_{k-1,1}x_{k,1} \\ 1 & x_{12} & \cdots & x_{k2} & x_{12}^2 & \cdots & x_{k2}^2 & x_{12}x_{22} & \cdots & x_{k-1,2}x_{k,2} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 1 & x_{1N} & \cdots & x_{kN} & x_{1N}^2 & \cdots & x_{kN}^2 & x_{1N}x_{2N} & \cdots & x_{k-1,N}x_{k,N} \end{bmatrix}. \qquad (1.5.2)$$

The (Nx1) vector of responses is denoted $\underline{y}$, and the (px1) vector of coefficients is denoted $\underline{\beta}$. The predicted responses are $\hat{\underline{y}}$, and the estimated coefficients are $\underline{b}$.

The predicted response model is written

$$\hat{\underline{y}} = X\underline{b}. \qquad (1.5.3)$$

In the first order, k=2 case, $X = [\underline{1}\ \underline{x}_1\ \underline{x}_2]$. In the second order, k=2

case, $X = [\underline{1}\ \underline{x}_1\ \underline{x}_2\ \underline{x}_1^2\ \underline{x}_2^2\ \underline{x}_1\underline{x}_2]$.

The estimated slope of the response surface is obtained by taking the partial derivatives of the predicted response function. The estimated slope in the direction of the x_1 axis is given by

$$\hat{s}_1 = \underline{l}'\underline{b}, \quad \text{where } \underline{l}' = [0\ 1\ 0\ 2x_1\ 0\ x_2] \qquad (1.5.4)$$

$$\text{and } \underline{b}' = [b_0\ b_1\ b_2\ b_{11}\ b_{22}\ b_{12}].$$

Denote the design moment matrix by $M=(1/N)(X'X)$. For k=2,

$$M = \begin{bmatrix} 1 & [1] & [2] & [11] & [22] & [12] \\ & [11] & [12] & [111] & [122] & [112] \\ & & [22] & [112] & [222] & [122] \\ & \text{sym.} & & [1111] & [1122] & [1112] \\ & & & & [2222] & [1222] \\ & & & & & [1122] \end{bmatrix} \qquad (1.5.5)$$

$$\text{where } [i] = \sum_{u=1}^{N} x_{iu}/N,$$

$$[ij] = \sum_{u=1}^{N} x_{iu}x_{ju}/N$$

$$[iij]=[jii]=[iji] = \sum_{u=1}^{N} x_{iu}^2x_{ju}/N$$

$$[iijj] = \sum_{u=1}^{N} x_{iu}^2x_{ju}^2/N, \quad \text{etc.}$$

Elements of M^{-1} are denoted by c's with the appropriate superscripts. In the k=2 case,

$$M^{-1} = \begin{bmatrix} c^0 & c^1 & c^2 & c^{11} & c^{22} & c^{12} \\ & c^{11} & c^{12} & c^{111} & c^{122} & c^{112} \\ & & c^{22} & c^{112} & c^{222} & c^{122} \\ & & & c^{1111} & c^{1122} & c^{1112} \\ & \text{sym.} & & & c^{2222} & c^{1222} \\ & & & & & c^{1122} \end{bmatrix} . \qquad (1.5.6)$$

When fitting a model of order d_1 when the true model is of order d_2, subscripts will be added to these quantities. The fitted model matrix, X_1 will contain a column for each term in the fitted model, while X_2 contains one for each term in the true model not represented in X_1. $\underline{\beta}_1$ contains coefficients to be fitted, while $\underline{\beta}_2$ contains ignored coefficients. The corresponding momemt matrices are

$$M_{11}=(1/N)(X_1'X_1), \tag{1.5.7a}$$

$$M_{22}=(1/N)(X_2'X_2), \tag{1.5.7b}$$

and

$$M_{12}=(1/N)(X_1'X_2). \tag{1.5.7c}$$

Spherical region moments are used to develop spherical variances and average biases. A spherical region moment of order g is defined

$$\sigma_{g_1 g_2 \ldots g_k} = K\int_{U_r} x_1^{g_1} x_2^{g_2} \ldots x_k^{g_k} d\underline{x} \tag{1.5.8}$$

where $K^{-1} = \int_{U_r} d\underline{x}$ is the surface area of the hypersphere U_r,

and $\sum_{i=1}^{k} g_i = g$.

If any of the g_i are zero, it does not appear in the region moment. Due to the symmetric nature of the region, (U_r), $\sigma_{g_1 g_2 \ldots g_k}$ is null whenever any g_i is odd. Stroud (1971) provides the details of the integration, but some useful results are presented here.

$$\sigma_2 = K\int_{U_r} x_i^2 d\underline{x} = r^2/k \tag{1.5.9a}$$

$$\sigma_4 = K\int_{U_r} x_i^4 d\underline{x} = 3r^4/(k(k+2)) \tag{1.5.9b}$$

$$\sigma_{22} = K\int_{U_r} x_i^2 x_j^2 d\underline{x} = r^4/(k(k+2)). \tag{1.5.9c}$$

Region moments only depend on the region of interest and not on the design. A matrix of spherical region moments is defined as

$$S = K\int_{U_r} \underline{x}\underline{x}'d\underline{x}, \qquad (1.5.10)$$

where $\underline{x}$ depends on the model.

Further details and additional notation notes will be provided and duly noted throughout the dissertation.

1.6 FAMILIES OF SECOND ORDER RESPONSE SURFACE DESIGNS

In order to fit a second order response surface model in k factors, a design with at least three levels of each variable must be used. A natural choice would be the family of 3^k factorial designs, in which each factor appears at exactly three levels. But, when k is large, say $k > 3$, too many experimental runs are required.

1.6.1 CENTRAL COMPOSITE DESIGNS

A family of designs that requires fewer points than the 3^k factorial is the class of central composite designs (CCD) devised by Box and Wilson. A CCD in k variables is comprised of a first order 2^k factorial or fractional factorial design with a set of axial points. The design matrix is given by

$$\text{CCD} = \begin{bmatrix} \pm\underline{1} & \pm\underline{1} & \cdots & \pm\underline{1} \\ +\alpha & 0 & \cdots & 0 \\ -\alpha & 0 & \cdots & 0 \\ 0 & +\alpha & \cdots & 0 \\ 0 & -\alpha & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & +\alpha \\ 0 & 0 & \cdots & -\alpha \end{bmatrix} . \qquad (1.6.1)$$

Center runs may be added to the design if desired. The locations of the axial points are determined by the parameter α. The experimenter chooses the number of center runs, n_0, and α to achieve desired design properties. This research will investigate choices of n_0 and α, and replicating the set of axial points n_a times, to achieve good spherical variance pictures for predicting the slope. A common criticism of the CCD is that it involves five levels for each of the design variables (unless α=1.0). This may be prohibitive in certain experimental situations.

1.6.2 BOX-BEHNKEN DESIGNS

Box and Behnken (1960) introduced as an alternative, the class of Box-Behnken designs, which use only three levels of each design variable. The designs are constructed from balanced incomplete block structures. For k = 3, 4, or 5, the design is built in $\binom{k}{2}$ parts. Each part contains a 2^2 factorial design in two variables, with the levels of the (k-2) other factors at zero. There is one such part for each pair of variables, resulting in $N=4\binom{k}{2}+n_0$ design points, where n_0 is the number of center runs. The design matrix for a Box-Behnken design (BBD) in three variables with one center run is given by

$$BBD = \begin{bmatrix} -1 & -1 & 0 \\ -1 & 1 & 0 \\ 1 & -1 & 0 \\ 1 & 1 & 0 \\ -1 & 0 & -1 \\ -1 & 0 & 1 \\ 1 & 0 & -1 \\ 1 & 0 & 1 \\ 0 & -1 & -1 \\ 0 & -1 & 1 \\ 0 & 1 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} . \qquad (1.6.2)$$

Both the CCD and the BBD classes of designs require more design points than are necessary to estimate all of the parameters in a second order model. The remaining designs presented are saturated, or nearly saturated designs, in which the number of design points is equal to or close to the minimum required to estimate the parameters.

1.6.3 HYBRID DESIGNS

The Hybrid designs (Roquemore, 1976) are nearly saturated designs that were constructed to be near y-rotatable. A hybrid design for k factors is built from a CCD in k-1 factors with an additional column for the k^{th} factor chosen to provide desirable moment conditions. The factorial and/or axial portions of the design matrix may be rotated or scaled. Three hybrid designs will be studied for the k=3 case. In Roquemore's notation the appropriate designs are the hybrid 310, 311A, and 311B. The corresponding design matrices are given below.

$$D_{310} = \begin{bmatrix} 0 & 0 & 1.2906 \\ 0 & 0 & -.1360 \\ -1 & -1 & .6386 \\ 1 & -1 & .6386 \\ -1 & 1 & .6386 \\ 1 & 1 & .6386 \\ 1.1736 & 0 & -.9273 \\ -1.1736 & 0 & -.9273 \\ 0 & 1.1736 & -.9273 \\ 0 & -1.1736 & -.9273 \\ 0 & 0 & 0 \end{bmatrix} \qquad (1.6.3a)$$

$$D_{311A} = \begin{bmatrix} 0 & 0 & 2 \\ 0 & 0 & -2 \\ -\sqrt{2} & -\sqrt{2} & 1 \\ \sqrt{2} & -\sqrt{2} & 1 \\ -\sqrt{2} & \sqrt{2} & 1 \\ \sqrt{2} & \sqrt{2} & 1 \\ 2 & 0 & -1 \\ -2 & 0 & -1 \\ 0 & 2 & -1 \\ 0 & -2 & -1 \\ 0 & 0 & 0 \end{bmatrix} \qquad (1.6.3b)$$

$$D_{311B} = \begin{bmatrix} 0 & 0 & 2.44949 \\ 0 & 0 & -2.44949 \\ -.7507 & 2.1063 & 1 \\ 2.1063 & .7507 & 1 \\ .7507 & -2.1063 & 1 \\ -2.1063 & -.7507 & 1 \\ .7507 & 2.1063 & -1 \\ 2.1063 & -.7507 & -1 \\ -.7507 & -2.1063 & -1 \\ -2.1063 & .7507 & -1 \\ 0 & 0 & 0 \end{bmatrix} . \qquad (1.6.3c)$$

1.6.4 NOTZ DESIGNS

Notz (1982) devised minimum or nearly-minimum point designs by augmenting a 2^k factorial or fractional factorial with a kxk identity matrix. In three variables, a Notz design with one center run is

$$D = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ -1 & 1 & 1 \\ 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \\ -1 & -1 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} . \qquad (1.6.4)$$

1.6.5 HOKE DESIGNS

Hoke (1974) designs are also saturated (or nearly so) three level designs. Irregular fractions of a 3^k factorial design, generated by sets of partially balanced arrays, are used to build Hoke designs. The Hoke-D_2 design with one center run is illustrated below.

$$D = \begin{bmatrix} -1 & -1 & -1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ -1 & 1 & 1 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \\ -1 & 0 & 0 \\ 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \\ 0 & 0 & 0 \end{bmatrix} . \qquad (1.6.5)$$

1.6.6 BOX-DRAPER DESIGNS

Box and Draper (1971, 1974) introduced designs restricted to be saturated with points on or within the unit cube. Subject to these restrictions, they found the D-optimal design. For three variables,

$$D = \begin{bmatrix} -1 & -1 & -1 \\ 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \\ .1925 & .1925 & -1 \\ .1925 & -1 & .1925 \\ -1 & .1925 & .1925 \\ -.2912 & 1 & 1 \\ 1 & -.2912 & 1 \\ 1 & 1 & -.2912 \end{bmatrix} . \qquad (1.6.6)$$

1.6.7 SATURATED ORTHOGONAL ARRAYS (SOA) OF STRENGTH TWO

In addition to the above families of designs, orthogonal arrays of strength two (Fisher 1945, Plackett and Burman 1946, Rao 1947, Seiden 1954, Addelman and Kempthorne 1961, Addelman 1962a 1962b 1972, and Kackar 1982) will be studied before and after augmentation to detect lack of fit due to the presence of a higher order underlying model. These designs are advocated by Taguchi for off-line quality control purposes (Taguchi, 1986). Such designs do not allow for the estimation of interactions between design factors, when fitting either a main effects model or a second order model without interactions. His claim is that main effects and/or pure quadratic effects far outweigh the importance of interactions, so he does not consider them at all in his design recommendations. Of course, if interactions do exist, the resulting estimated factor effects will be biased. Thus, the slope estimates will also be biased.

An application of the slope bias measure and the MSE measure in slope estimation to be developed will involve two designs typical of those recommended by Taguchi: the L_8 array, which is an eight point SOA of strength two, used to fit a seven factor main effects model; and the L_9 array, which is a nine point SOA of strength two, used to fit a second order model without interactions in four factors. The design matrices are given below. Augmentation of these designs to detect lack of fit will be discussed in Chapter 3.

$$L_8 = \begin{bmatrix} -1 & -1 & -1 & -1 & 1 & -1 & -1 \\ -1 & -1 & -1 & 1 & -1 & 1 & 1 \\ -1 & 1 & 1 & -1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & -1 & 1 \\ 1 & -1 & 1 & 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & -1 & -1 & 1 & -1 \\ 1 & 1 & -1 & 1 & 1 & -1 & 1 \end{bmatrix} . \qquad (1.6.7a)$$

$$L_9 = \begin{bmatrix} -1 & -1 & -1 & -1 \\ -1 & 0 & 0 & 0 \\ -1 & 1 & 1 & 1 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & 1 & -1 \\ 0 & 1 & -1 & 0 \\ 1 & -1 & 1 & 0 \\ 1 & 0 & -1 & 1 \\ 1 & 1 & 0 & -1 \end{bmatrix} . \qquad (1.6.7b)$$

CHAPTER 2

SPHERICAL SLOPE VARIANCE MEASURES

This chapter extends the relationship between slope variance and optimization to the case of k independent variables. It develops a measure of variance in the k jointly estimated slopes parallel to the factor axes, which is general enough in nature to be used to evaluate any response surface design in its ability to efficiently and precisely locate the optimum.

Due to the wide scope of applicability of this slope variance measure, very few analytic results can be developed that hold true in general. So in order to gain more insight about slope estimation precision, a measure based on the variance of one slope is developed. This more specific measure is applicable to the class of standard response surface designs in which slopes parallel to the factor axes are estimated with equal precision. Thus, consideration of one slope is adequate.

The family of central composite designs is studied in some detail. Choices of design parameters are made which facilitate precise slope estimation. These choices are then compared to those recommended for response prediction. Often the best design for slope estimation in a given portion of the region is not necessarily that which is best for response prediction.

Several other families of second order designs are investigated in somewhat less detail, and guidelines are provided as to what designs are "best" for precise slope estimation. These designs will be studied further in terms of robustness to outliers in the data, and to missing observations, in Chapter 4.

2.1 EXTENDING THE RELATIONSHIP BETWEEN THE SLOPE VARIANCE AND THE LOCATION OF THE OPTIMUM TO THE K VARIABLE CASE

Consider using the method of least squares to fit a second order model in k independent factors (equations 1.1.2 and 1.5.3). The estimated kx1 vector of slopes is given by

$$\hat{\underline{s}} = \frac{\partial}{\partial \underline{x}} (\hat{\underline{y}}).$$

In the case of three independent variables, this is specified as

$$\hat{\underline{s}} = \begin{bmatrix} b_1 + 2b_{11}x_1 + (b_{12}x_2 + b_{13}x_3) \\ b_2 + 2b_{22}x_2 + (b_{12}x_1 + b_{23}x_3) \\ b_3 + 2b_{33}x_3 + (b_{13}x_1 + b_{23}x_2) \end{bmatrix} . \qquad (2.1.1)$$

Denote the symmetric kxk variance-covariance matrix of $\hat{\underline{s}}$ as

$$V = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1k} \\ & v_{22} & \cdots & v_{2k} \\ & & \ddots & \vdots \\ \text{sym.} & & & v_{kk} \end{bmatrix} . \qquad (2.1.2)$$

Making the standard normal assumptions about the error vector,

$$\underline{\epsilon} \sim N(\underline{0}, \sigma^2 I),$$

the estimated slope vector at the optimum, $\hat{\underline{s}}_0$, follows the multivariate normal distribution

$$\hat{\underline{s}}_0 \sim N_k(\underline{0}, V),$$

while at an arbitrary location (not the optimum), $\underline{x}$,

$$\hat{\underline{s}}'V^{-1}\hat{\underline{s}} \sim \chi^{2\prime}(k, \mu/2), \qquad \text{where} \quad \mu = \underline{s}'V^{-1}\underline{s} . \qquad (2.1.3)$$

In order to describe the $(1-\alpha)100\%$ confidence region (CR) around the optimum, $\underline{x}_0$, define P to be the probability that an arbitrary $\underline{x} \neq \underline{x}_0$ lies inside the $(1-\alpha)100\%$ CR around $\underline{x}_0$. It can be shown that

$$P = Pr\{\chi^{2\prime}(k, \mu/2) \leq \chi^2_{\alpha,k}\}. \qquad (2.1.4)$$

That is, P is the probability that a noncentral chi-square random variate is less than or equal to a critical point in a central chi-square distribution. If P is large, then one is highly likely to classify a non-optimal point as optimal; if P is small, it is not very probable that a point will be misclassified as optimal. The value of P decreases as μ, the noncentrality parameter, increases, so in terms of its ability to precisely identify the optimum of a resonse surface, a design is desirable if it yields a large value for μ.

2.2 A SLOPE ESTIMATION MEASURE BASED ON E-OPTIMALITY

Recall for response prediction the E-optimality design criterion. An E-optimal design is one which minimizes the maximum eigenvalue of $(X'X)^{-1}$, the variance-covariance matrix of $\underline{b}$. Herein is developed a slope estimation criterion for the optimization problem, which chooses the design yielding the smallest maximum eigenvalue of V,

the variance-covariance matrix of $\hat{\underline{s}}$. It will be shown that using this criterion (E-optimality for slopes) is equivalent to choosing the design yielding the most desirable confidence region around the location of the optimum.

Recall that we wish to choose a design with a large value of μ, as defined in (2.1.3). We will do this by choosing the design with the largest minimum value of μ for a given length of the slope vector.

Theorem 2.2.1: With V defined in (2.1.2),

$$\min_{\underline{s}'\underline{s}=1} \underline{s}'V^{-1}\underline{s} = 1/[\lambda_L(V)],$$

where $\lambda_L(V)$ is the largest eigenvalue of V.

Thus, calculating $\lambda_L(V)$ for several candidate designs, and choosing the design yielding the smallest $\lambda_L(V)$, is equivalent to choosing the design with a large minimum μ value, and thus a small probability of falsely classifying non-optimal points as optimal.

Proof:

Since V is a kxk real symmetric matrix, there exists an orthogonal matrix P such that

$$P'VP = \text{diag}(\lambda_1, \lambda_2, \ldots \lambda_k) = D$$

where the λ_j's are the k eigenvalues of V.

$$(P'VP)^{-1} = \text{diag}(1/\lambda_1, 1/\lambda_2, \ldots 1/\lambda_k) = D^{-1}$$

$$P'V^{-1}P = D^{-1}.$$

The eigenvalues of V^{-1} are the inverses of the eigenvalues of V and the same orthogonal matrix P diagonalizes both V and V^{-1}.

If $a < b$, then $1/a > 1/b$,

so the largest eigenvalue of V is the inverse of the smallest eigenvalue of V^{-1}. Now,

$$\underline{s}'V^{-1}\underline{s} = \underline{s}'PP'V^{-1}PP'\underline{s} \quad \text{since } PP'=I$$

$$= \underline{z}'D^{-1}\underline{z} \quad \text{where } \underline{z}=P'\underline{s} \text{ is an orthogonal transformation of } \underline{s}, \text{ which implies that } \underline{s}'\underline{s}=\underline{s}'PP'\underline{s}=\underline{z}'\underline{z}.$$

$$= \sum_{j=1}^{k} (1/\lambda)_j \, z_j^2$$

To minimize this subject to $\underline{s}'\underline{s}=r^2$, or equivalently $\underline{z}'\underline{z}=r^2$, we must set $z_1=\pm r$ and the other $z_j=0$ for $j\neq 1$, where z_1 corresponds to the smallest eigenvalue of V^{-1}. But note that the smallest eigenvalue of V^{-1} is the inverse of the largest eigenvalue of V, so

$$\min_{\underline{s}'\underline{s}=r^2} \underline{s}'V^{-1}\underline{s} = r^2/[\lambda_L(V)] \, ,$$

and thus the theorem is proved.■

The eigenvalues of the slope estimation variance-covariance matrix V, are complicated functions both of the design used, and of the location of the point at which the slope is to be estimated, $\underline{x}$. For fairness in the comparison of designs with differing sizes, define

$$\lambda_L^*(V) = N \times \max_{\lambda_j} \{ \lambda_j : |V-\lambda_j I|=0 \}. \qquad (2.2.1)$$

Using modern algebra techniques, it is possible to express the λ_j's as functions of $\underline{x}$, but the result is not in a form easily manipulated. It cannot be determined, from the expressions, which eigenvalue is largest, so integrating the largest eigenvalue is not possible. Hence, no average of $\lambda_L^*(V)$ on spheres will be calculated.

A spherical measure of slope variance is developed for a given design by maximizing and minimizing $\lambda_L^*(V)$ over the surface of k-dimensional spheres of varying radii, using the appropriate V matrix at each location. The optimum functional values are found using Vining's search algorithm. The maxima and minima are then plotted against the radius, resulting in a general slope variance envelope (GSVE) for the jointly estimated slope vector. In effect, what is being plotted is the inverse of the minimum noncentrality parameter defined in (2.1.3) maximized and minimized on spheres in the design space. We pick the design with the lowest maximum curve plotted, thus yielding the "smallest" confidence region around the location of the optimum.

To illustrate this measure, Figure 2.2.1 is a general slope variance envelope (GSVE) for $\hat{\underline{s}}$ using the three variable CCD (1.6.1) with α=1.732 and three center runs. Figure 2.2.2 is the GSVE for a Box-Behnken design (1.6.2) with three center runs. It is not appropriate to compare Figure 2.2.1 to Figure 2.2.2 as the designs have not been scaled to lie on the same radius. Comparison of these and other standard RS designs are made with the appropriate scaling, in Section 2.3, using the variance of one slope, since the k slopes are estimated with equal precision. Figures 2.2.3 and 2.2.4 are GSVEs for the hybrid 311B (1.6.3c) and the dual axial CCD (1.4.3) where g=.7816, α_1=.5833, and α_2=1.3015, respectively. It will be shown that these designs are recommended as those four "best" for precise slope estimation. They will then be studied for robustness in Chapter 4.

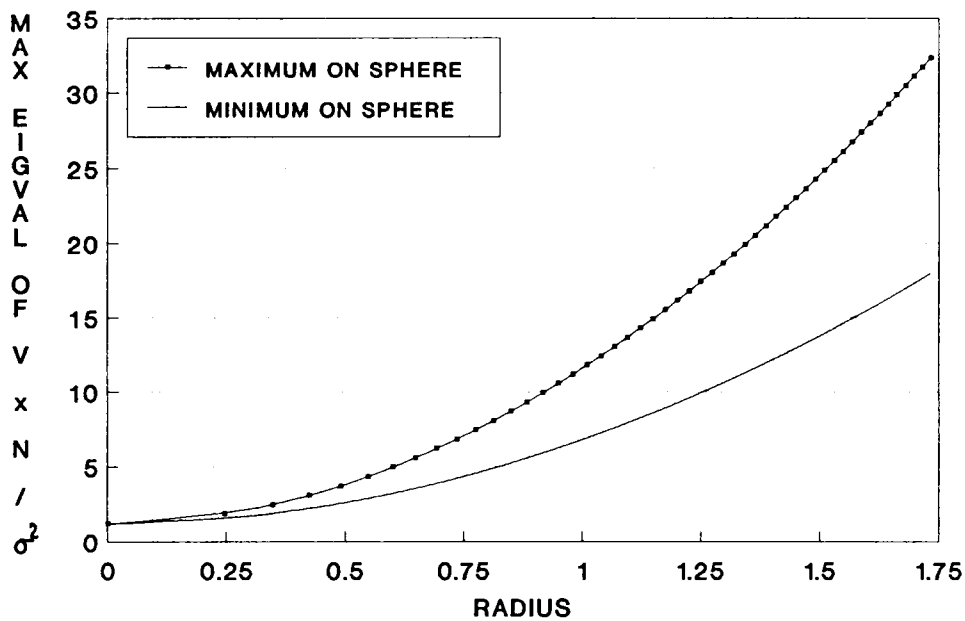

FIGURE 2.2.1
GENERAL SLOPE VARIANCE ENVELOPE
CCD FOR K=3: ALPHA=1.732, 3 CENTER RUNS

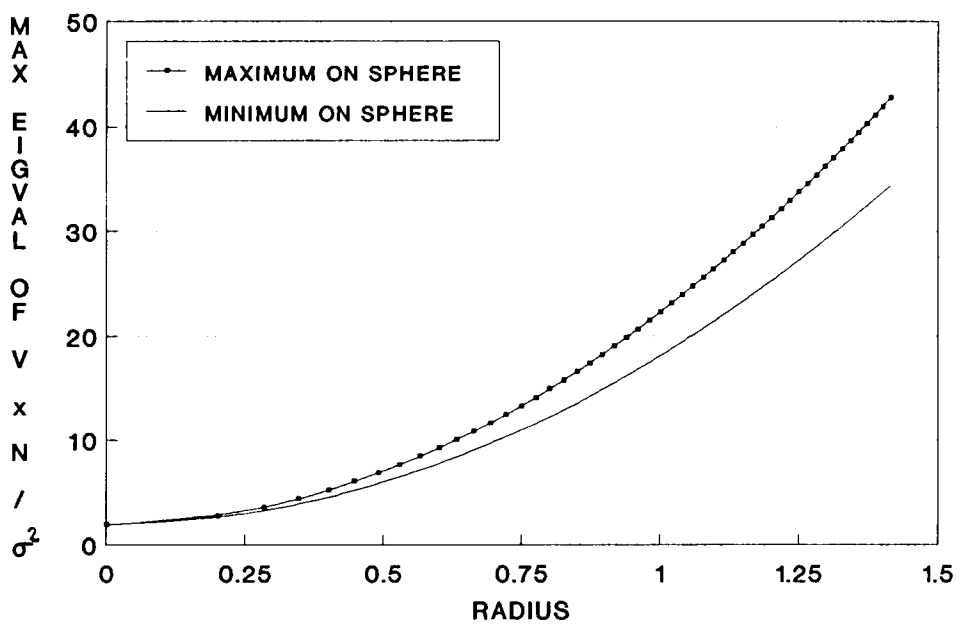

FIGURE 2.2.2
GENERAL SLOPE VARIANCE ENVELOPE
BOX-BEHNKEN DESIGN WITH 3 CENTER RUNS

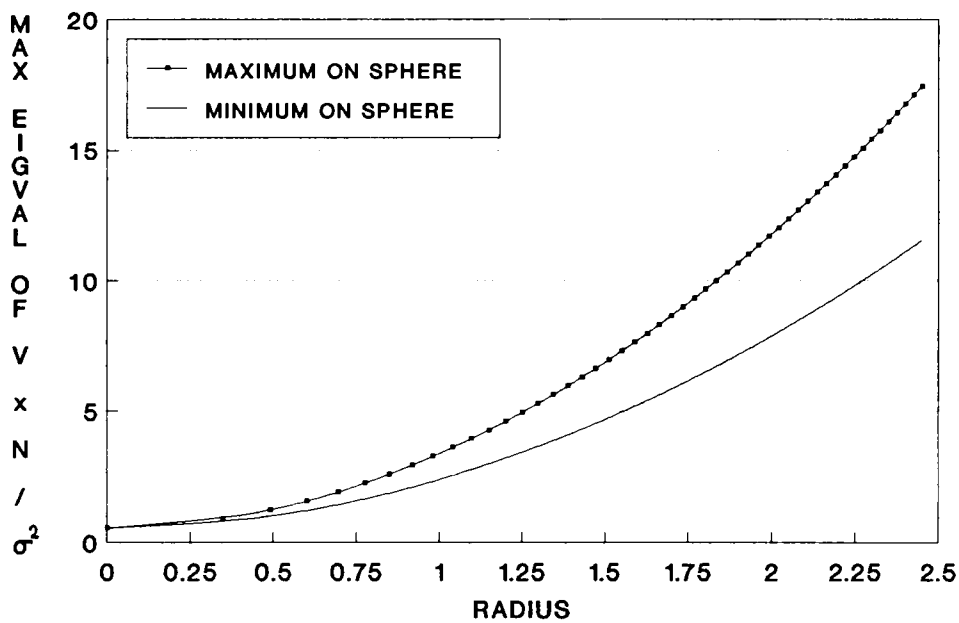

FIGURE 2.2.3
GENERAL SLOPE VARIANCE ENVELOPE
THE HYBRID 311B DESIGN

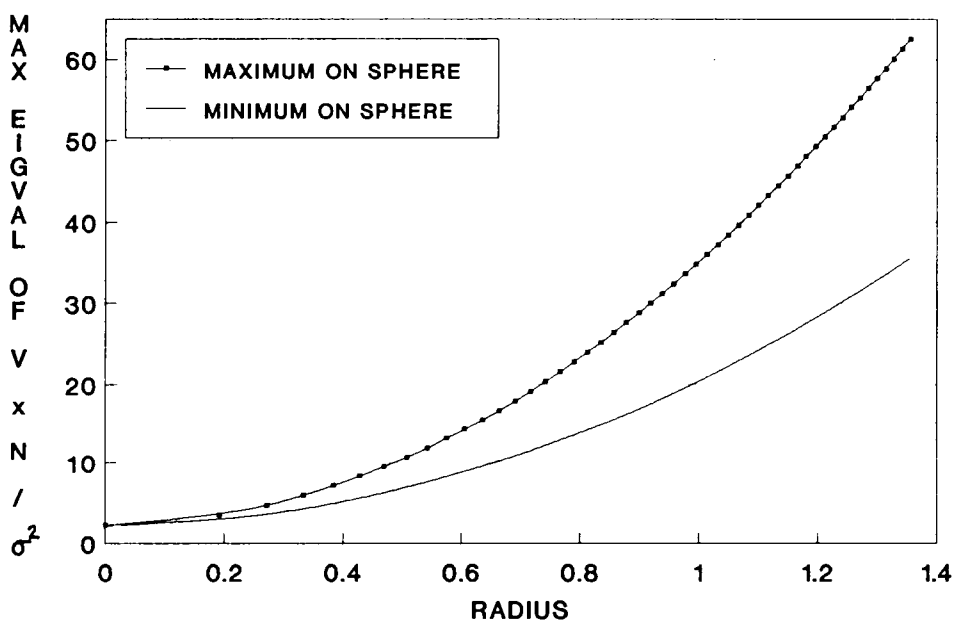

FIGURE 2.2.4
GENERAL SLOPE VARIANCE ENVELOPE
DUAL AXIAL CCD

2.3 DEVELOPMENT OF THE SLOPE VARIANCE ENVELOPE FOR ONE SLOPE

To gain more insight into precise slope estimation, consider the estimation of a single slope parallel to any one of the factor axes. The slope of the surface in a direction parallel to the i^{th} factor axis is obtained by taking the partial derivative of the response function with respect to x_i. For three factors, the slope in the direction of x_1 is

$$\begin{aligned} s_1 &= \partial y(\underline{x})/\partial x_1 \\ &= [0\ 1\ 0\ 0\ 2x_1\ 0\ 0\ x_2\ x_3\ 0]\underline{\beta} \\ &= \underline{l}_1'\underline{\beta}. \end{aligned} \qquad (2.3.1)$$

For standard response surface designs and least squares estimation of $\underline{\beta}$, the k slopes are estimated with equal precision. Thus, without loss of generality and for ease of notation, the slope criterion will be derived for s_1. Note however that the estimated slopes are not independent. Due to this fact, the results will only apply to the estimation of each slope separately and not to joint estimation. Recall the general variance measure for the kxl vector of jointly estimated slopes developed for use with non-standard RS designs. The dependence among the slopes when using standard RS designs is small enough that both criteria yield similar results. The present method is less complex, so it is used for the standard RS designs to gain information about slope estimation precision.

We seek to assess slope estimation capability in a given region. Recall the spherical variance for the response, given in equation 1.3.2. The sample size is used in the expression to punish a design for using

many points when compared with a smaller design.

To provide a similar measure for the average slope prediction variance (henceforth called slope variance) on spheres, the same averaging convention is used. With K and U_r defined in (1.3.2), define

$$V^{*r} = \frac{NK}{\sigma^2}\int_{U_r} \mathrm{var}(\hat{s}_i(\underline{x}))d\underline{x} \qquad (2.3.2)$$

$$= \mathrm{tr}\{\mu M^{-1}\}$$

$$= c^{ii} + \frac{r^2}{k}(4c^{iiii} + \sum_j c^{iijj}), \qquad (2.3.3)$$

where $\mu = K\int_{U_r} \underline{l}_i\underline{l}_i'd\underline{x}$,

$M^{-1} = N(X'X)^{-1}$ with elements denoted as c's,

$\sum_j$ is over the (k-1) interactions involving factor x_j,

c^{ii} is the variance of b_i weighted by N,

c^{iiii} is the variance of b_{ii} weighted by N,

and c^{iijj} is the variance of b_{ij} weighted by N.

Details of the development and the form of M^{-1} are found in Appendix A.1. The calculation of V^{*r} for a given design is straightforward once $(X'X)^{-1}$ has been calculated. Optimization on spheres, however, remains quite complex. The search algorithm developed by Vining, modified for use on the slope, is used to calculate the maxima and minima on spheres. A graph containing curves for the maxima, minima, and average slope variances, is a slope variance envelope (SVE) for one slope. Every slope variance for a slope parallel to a factor axis lies somewhere inside the envelope. The slope variance envelope (SVE) describes a standard RS design's ability to precisely estimate the slope.

2.4 COMPARING STANDARD RS DESIGNS USING V^{*r}

In order to provide a fair comparison of standard response surface designs in their ability to precisely estimate the slope in a region of interest, each design is scaled so that the design radius is equal to $\sqrt{k}$. In addition, V^{*r} includes a factor of N to punish a larger design for requiring so many points. By comparing graphs for designs adjusted as such, where V^{*r} is plotted against the radius of several concentric spheres for each design, a simple, yet fair, comparison of slope estimation precision inside the design region results.

2.5 V^{*r} FOR THE CENTRAL COMPOSITE DESIGN

The design matrix for the CCD is given in (1.6.1). The three parameters which specify the CCD are α, the location of the axial points, n_0, the number of center runs, and n_a, the number of runs taken at each axial. In this section, choices for these parameters will be considered in order to facilitate precise slope estimation.

Hussey (1983) showed for the CCD that

$$(X'X)^{-1} = \begin{bmatrix} a & \underline{0}' & b\underline{1}_k' & \underline{0}' \\ \underline{0} & cI_k & \underline{0}' & \underline{0}' \\ b\underline{1}_k & \underline{0} & (d-e)I_k+e\underline{1}\underline{1}_k' & \underline{0}' \\ \underline{0} & \underline{0} & \underline{0} & fI_i \end{bmatrix} \qquad (2.5.1)$$

where $i=k(k-1)/2$ is the number of interaction terms in the model,

$$a = (kF+2\alpha^4)/D$$

$$b = -(F+2\alpha^2)/D$$

$$c = 1/(F+2\alpha^2)$$

$$d = (D+J)/2\alpha^4 D$$

$$e = J/2\alpha^4 D$$

$$f = 1/F$$

$$D = N(kF+2\alpha^4)-k(F+2\alpha^2)^2$$

$$J = (F+2\alpha^2)^2-NF$$

F = # of factorial points

N = # of points in the design

α = location of the axial points.

Making the appropriate substitutions for c^{ii}, c^{iiii}, and c^{iijj},

$$V^*r_{CCD}(\hat{s}_i) = \frac{N}{(F+2\alpha^2)} + \frac{4r^2\{N^2(kF+2\alpha^4)-N(k-1)(F+2\alpha^2)^2-N^2F\}}{2\alpha^4kN(kF+2\alpha^4)-2\alpha^4k^2(F+2\alpha^2)^2} + \frac{r^2N(k-1)}{kF} . \qquad (2.5.2)$$

This will be used to determine optimal values for the parameters in a CCD.

2.5.1 THE PARAMETER α

The location of the axial points in a CCD affects the slope estimation capabilities of the design.

Theorem 2.5.1: When the central composite design is scaled so that the design radius is $\sqrt{k}$, V^*r as defined in (2.3.2) and (2.3.3) is minimized for all radii, $0 \le r \le \sqrt{k}$, with respect to α, when $\alpha = \sqrt{k}$. The proof is in Appendix A.2.

Note that this agrees with Lucas's (1976) study of D-efficiencies for the CCD.

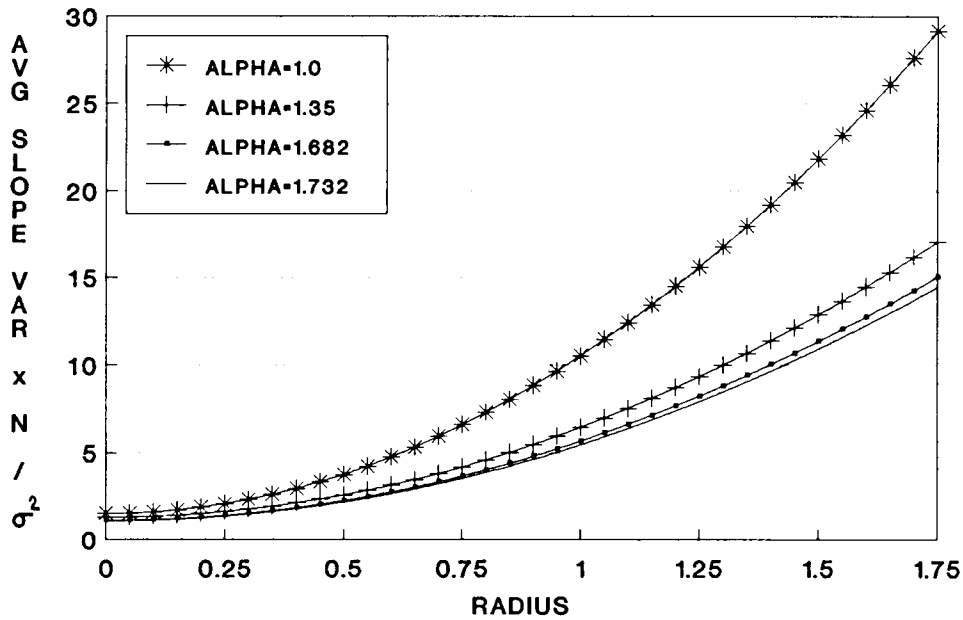

FIGURE 2.5.1a
AVERAGE SLOPE VARIANCE COMPARISON
CCDs FOR K=3, 1 C.R., VARYING ALPHA

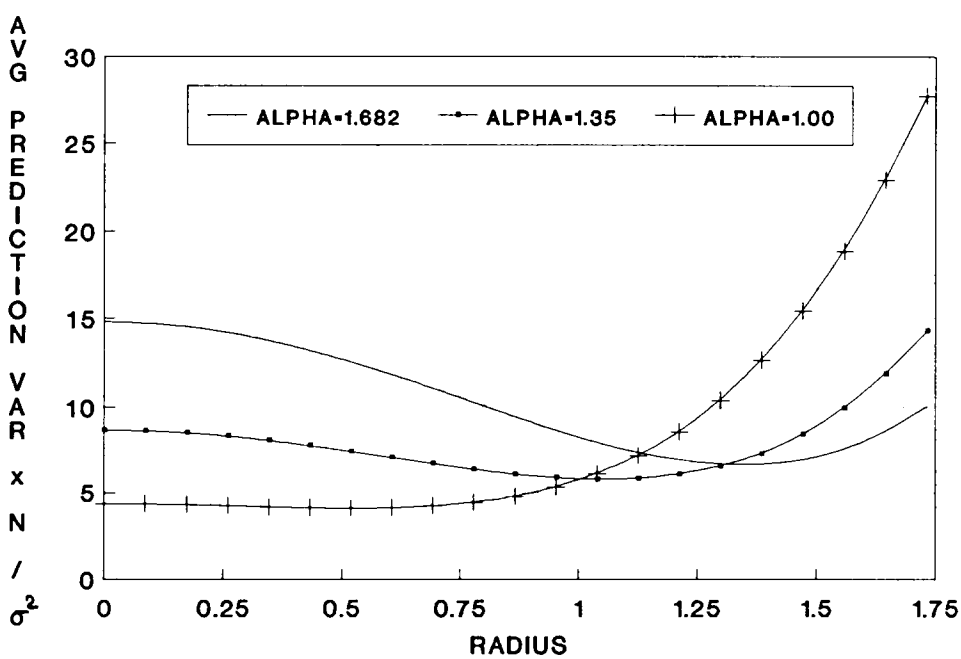

FIGURE 2.5.1b
SPHERICAL VARIANCE COMPARISON
CCDs FOR K=3, 1 C.R., VARYING ALPHA

Figure 2.5.1a illustrates the spherical slope variance for CCD designs, varying α. Four axial locations are compared for designs with one center run. Clearly the CCD with $\alpha=\sqrt{k}$ is uniformly optimal in V^{*r}. Figure 2.5.1b portrays a comparison of three CCDs with α ranging from 1.0 to 1.682 (the response rotatable value) in prediction variance. By comparing Figure 2.5.1a to Figure 2.5.1b, it becomes clear that designs uniformly better for slope estimation are not uniformly better for response prediction.

2.5.2 THE NUMBER OF CENTER RUNS

It can be shown using differencing methods, that there is no clear answer as to the number of center runs and axials to use that holds across all values of α and over the experimental region. As Ott and Mendenhall found for the k=1 case and three experimental locations, it depends on where in the region one wishes to predict. This is why pictures such as the VDG, GSVE and SVE are necessary and useful. Results are stated below, but details are given in Appendix A.3.

If α is sufficiently small, say up to 1.0, center runs worsen slope prediction across the region. If α is larger, $1 \leq \alpha \leq \sqrt{k}$, slope estimation is improved with the addition of center runs (1, 2, or 3 only) in the outer portion of the region. Slope prediction precision is actually lessened in the center of the region, but by a negligible amount. More than three center runs adversely affects slope estimation capabilities. Recall that V^{*r} is weighted by N. The effect of this is seen in these results. A decrease in performance as judged by this

measure may indicate that there is simply not enough reduction in variance obtained at the cost of the additional points. Note that adding points in one location may reduce performance at that spot since it takes two locations to determine a slope. It is understandable then, that center runs can actually damage slope prediction at the design origin, though not appreciably. Figure 2.5.2 illustrates the case α=1. As the analytic results revealed, when α is small, center runs are undesirable. Figure 2.5.3a shows the improvement obtained by adding center runs when α is near $\sqrt{k}$. Note that there is greater improvement at the perimeter than there is damage in the center. These results for slope estimation are opposite to those for response prediction as seen in Figure 2.5.3b.

2.5.3 REPLICATING POINTS IN A CCD

Since it takes more than one location to determine a slope, one might expect that replication efforts should be concentrated at the axial points. A study of V^{*r} performance for the CCD with n_a=1 and the CCD with n_a=2 revealed that indeed, sometimes this is the case. But once again, it depends on the location of the axial points, and the use of pictures is again necessary.

If α is small, replicating axials improves performance in the outer portion of the region. If α is close to, but less than $C=\sqrt{kF/(F+n_0)}$, then replicating axials worsens prediction everywhere. When α becomes larger than C, there is slightly improved prediction in the center, but not enough to offset the damage done to the rest of the

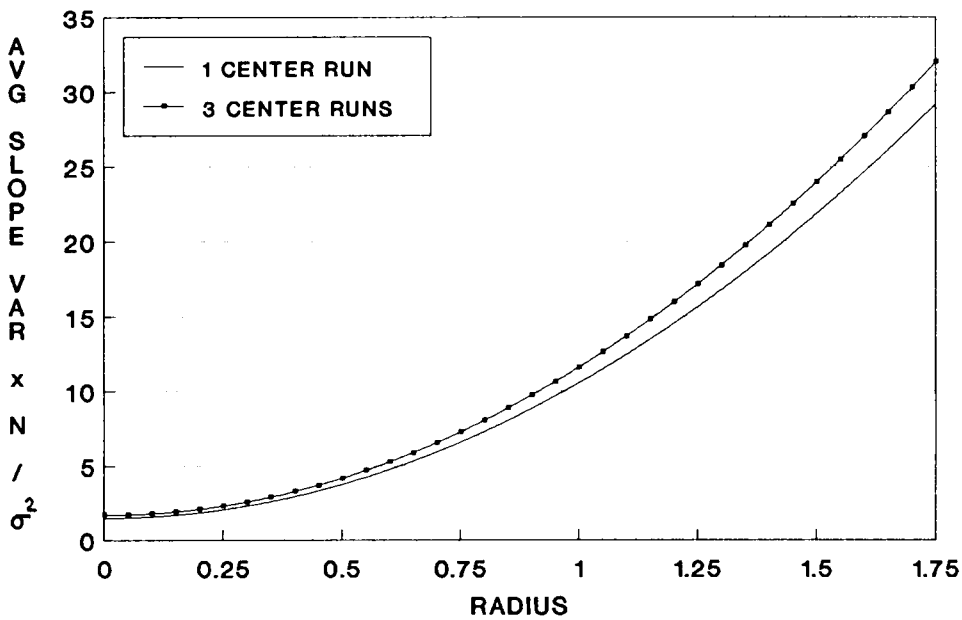

FIGURE 2.5.2
AVERAGE SLOPE VARIANCE COMPARISON
CCDs FOR K=3, ALPHA=1, 1 VS 3 C.R.s

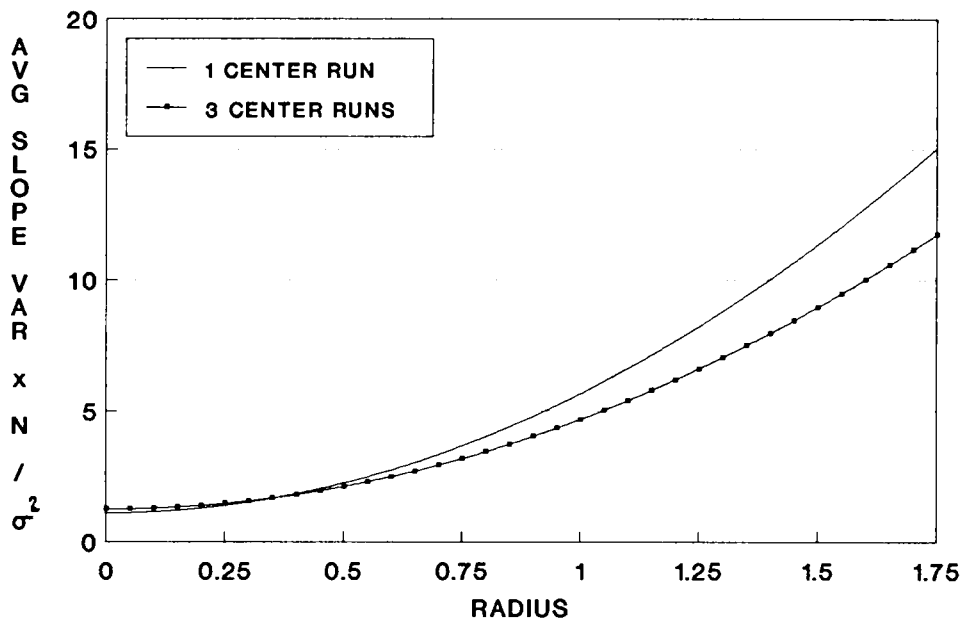

FIGURE 2.5.3a
AVERAGE SLOPE VARIANCE COMPARISON
CCDs FOR K=3, ALPHA=1.682, 1 VS 3 C.R.s

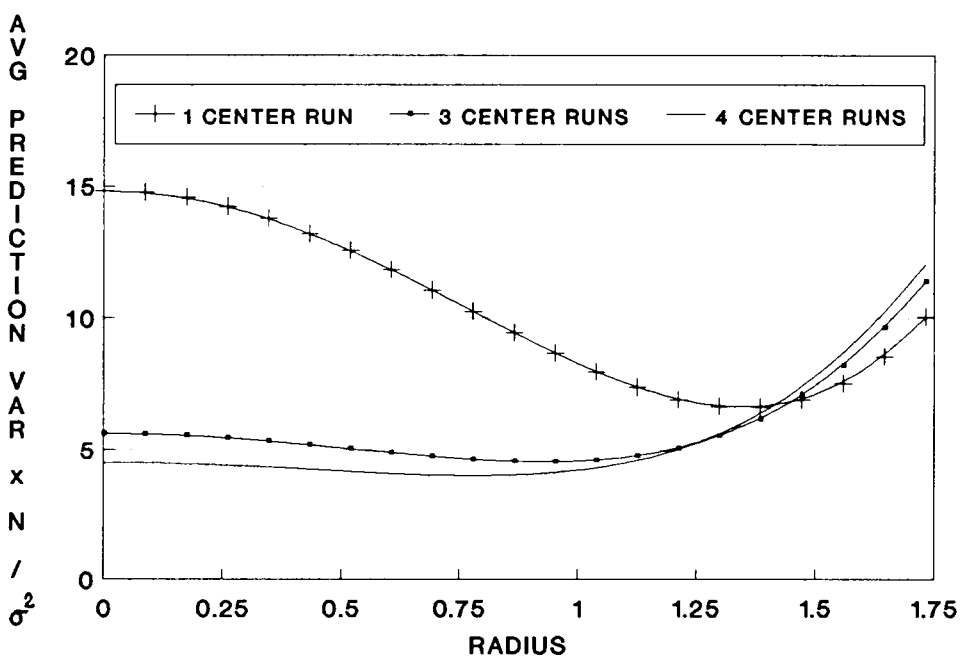

FIGURE 2.5.3b
SPHERICAL VARIANCE COMPARISON
CCDs FOR K=3, ALPHA=1.682, VARYING #C.R.

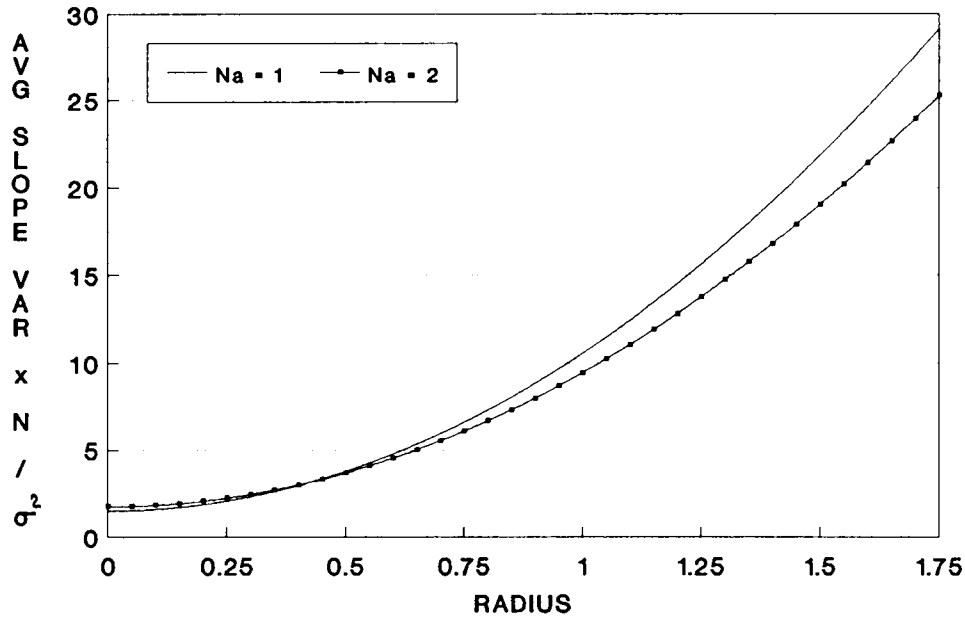

FIGURE 2.5.4a: REPLICATING AXIALS
AVERAGE SLOPE VARIANCE COMPARISON
CCDs FOR K=3, 1 C.R., ALPHA=1.0

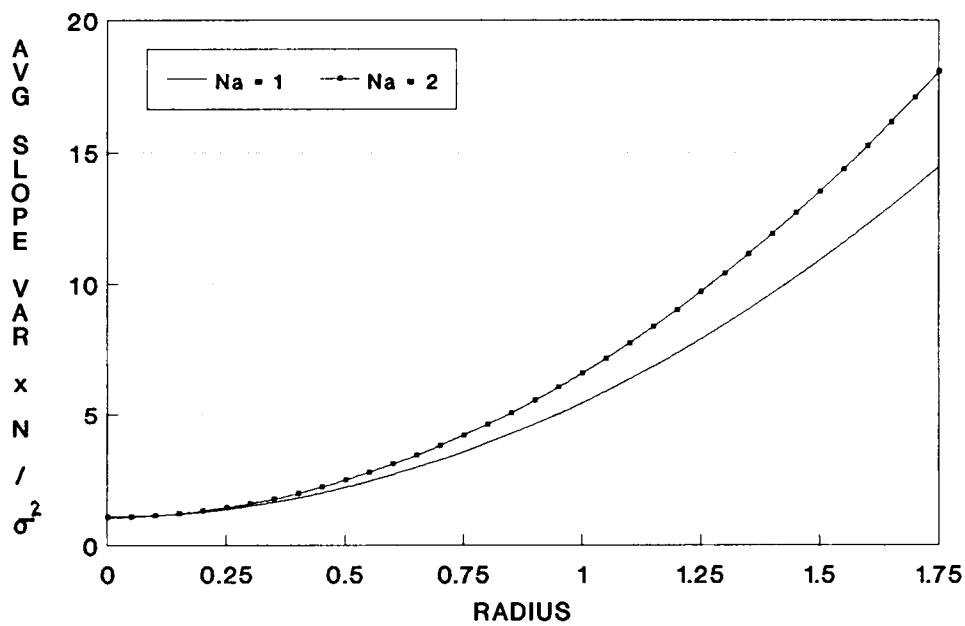

FIGURE 2.5.4b: REPLICATING AXIALS
AVERAGE SLOPE VARIANCE COMPARISON
CCDs FOR K=3, 1 C.R., ALPHA=1.732

region. Figures 2.5.4a and 2.5.4b illustrate these results. The results indicate that for precise slope estimation, if using a face-centered-cube (α=1.0) then it is better to replicate at the axials than in the center, but if using an $\alpha \approx \sqrt{k}$ CCD, center runs should be replicated, not axials. Details of the investigations of center runs and axial points are in Appendix A.3.

2.5.4 THE SLOPE VARIANCE ENVELOPE FOR THE CCD

The average slope variance on a sphere does not give any description of the dispersion of the slope variance over the sphere. It turns out that the design parameter values which are favorable for V^{*r} are also usually favorable for RofV*, where

$$\text{RofV}^*(r) = \max_{\underline{x} \in Ur} \text{var}(\hat{s}(\underline{x})) - \min_{\underline{x} \in Ur} \text{var}(\hat{s}(\underline{x})). \qquad (2.5.3)$$

When V^{*r} increases, in most cases, RofV* increases as well. This is described by Hockman and Myers (1988). But when RofV* is small, it does not guarantee in every case, that V^{*r} is small. Figure 2.5.5 portrays the slope variance envelope for the three variable CCD with $\alpha = \sqrt{3}$ and three center runs. This is the best CCD by the measure V^{*r}. Clearly RofV* is not as small as for the slope rotatable CCD, but note that even the maximum slope variance curve for the $\alpha=\sqrt{3}$ CCD is smaller than the V^{*r} curve for the slope rotatable CCD shown in Figure 1.4.1. In three variables, the recommended single-axial CCD for precise slope estimation is that with $\alpha=\sqrt{3}$ and three center runs.

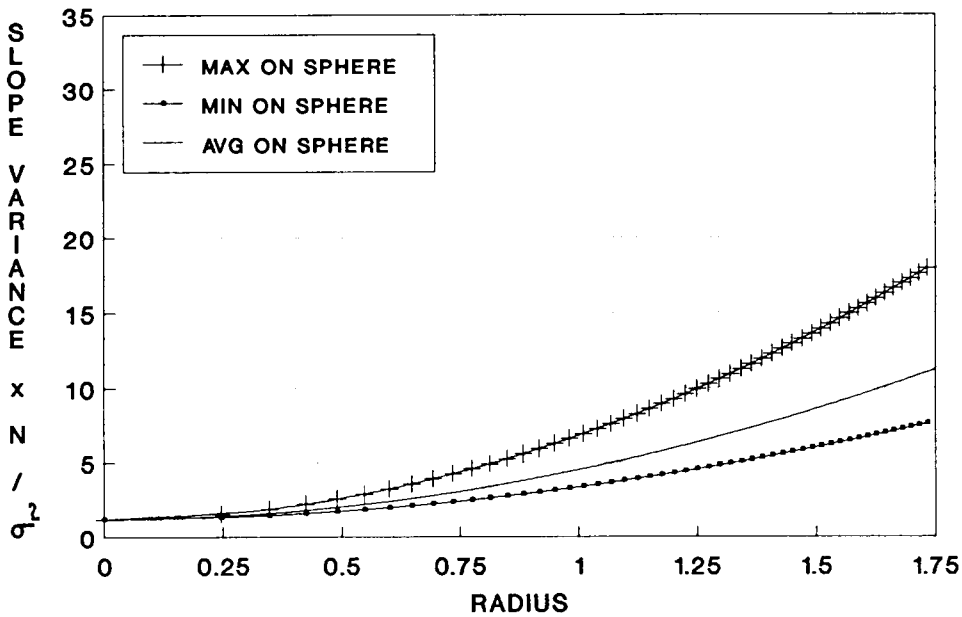

FIGURE 2.5.5
SLOPE VARIANCE ENVELOPE FOR ONE SLOPE
CCD FOR K=3, ALPHA=1.732, 3 C.R.'s

2.6 SLOPE VARIANCE ON SPHERES USING OTHER SECOND ORDER DESIGNS

The prediction capability of the dual axial CCD (satisfying the min $V^{**}|\min B^{**}$ criterion of Myers and Lahoda), the Box-Behnken design (BBD), three Hybrid designs, a Notz design and the Hoke-D_2 design are presented and compared in this section for the three factor case. Recall that all designs have been scaled (if necessary) to lie on and within a radius of $\sqrt{k}$.

2.6.1 DUAL AXIAL CCD

The dual axial CCD is a good design by the V^{*r} measure. Figure 2.6.1 provides an assessment of its slope prediction precision. The dual axial CCD has properties of V^{*r} and RofV* that are comparable to the single axial CCD where $\alpha < \sqrt{k}$.

2.6.2 BOX-BEHNKEN DESIGN

The prediction picture for a Box-Behnken design is given in Figure 2.6.2. Its performance is very similar to that of the $\alpha=\sqrt{k}$ CCD. Box-Behnken designs are nearly $\hat{y}$-rotatable out to a radius of about 1.0, (see Giovannitti-Jensen, 1987), but they are not nearly slope-rotatable.

2.6.3 HYBRID COMPARISON

Three Hybrid designs are assessed in Figures 2.6.3-2.6.6. Clearly 311B is the best for slope estimation. This class of designs is very promising for use in estimating the slope of a response surface.

2.6.4 SATURATED D-EFFICIENT DESIGNS

Saturated D-efficient designs (Atwood, 1977) are generally poor designs by this criterion. The Box-Draper, the Notz, and the Hoke-D_2

designs, each with one center run, are evaluated in Figures 2.6.7, 2.6.8, and 2.6.9 respectively. All three have large V^{*r} and $RofV^{*}$ across the region, and thus do not represent desirable alternatives to the CCD, the BBD, or the Hybrid classes of designs.

Clearly, in many cases, the designs best for response prediction are not those best for slope estimation. If the response surface analyst is more interested in the optimization problem than in prediction, the slope variance envelope is at least as important a criterion for choosing designs as the response prediction criteria.

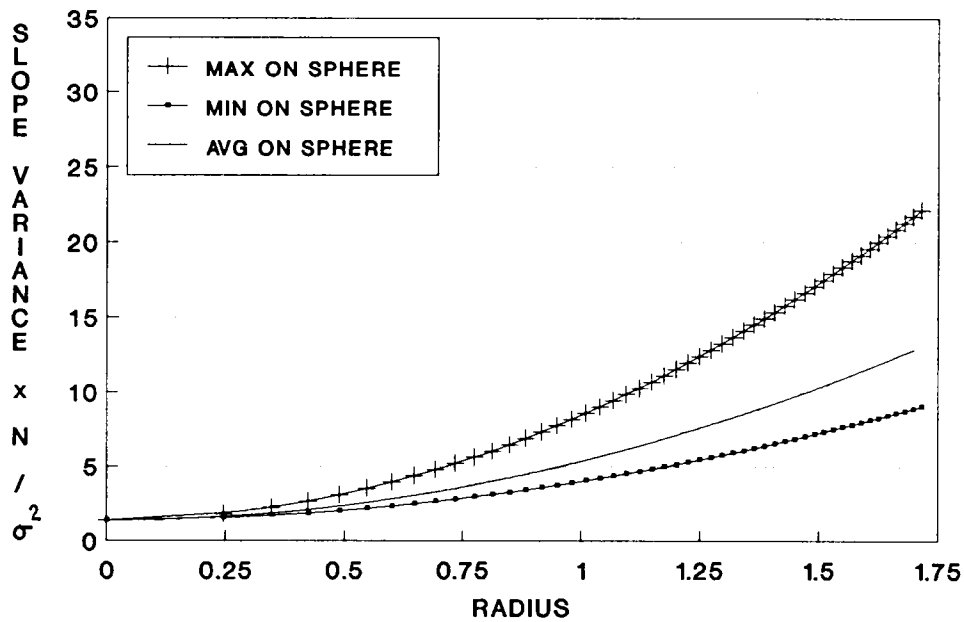

FIGURE 2.6.1: SLOPE VARIANCE ENVELOPE
DUAL AXIAL CCD FOR K=3 WITH 0 C.R.s
ALPHA1=.5833, ALPHA2=1.3015, FACT=.7816

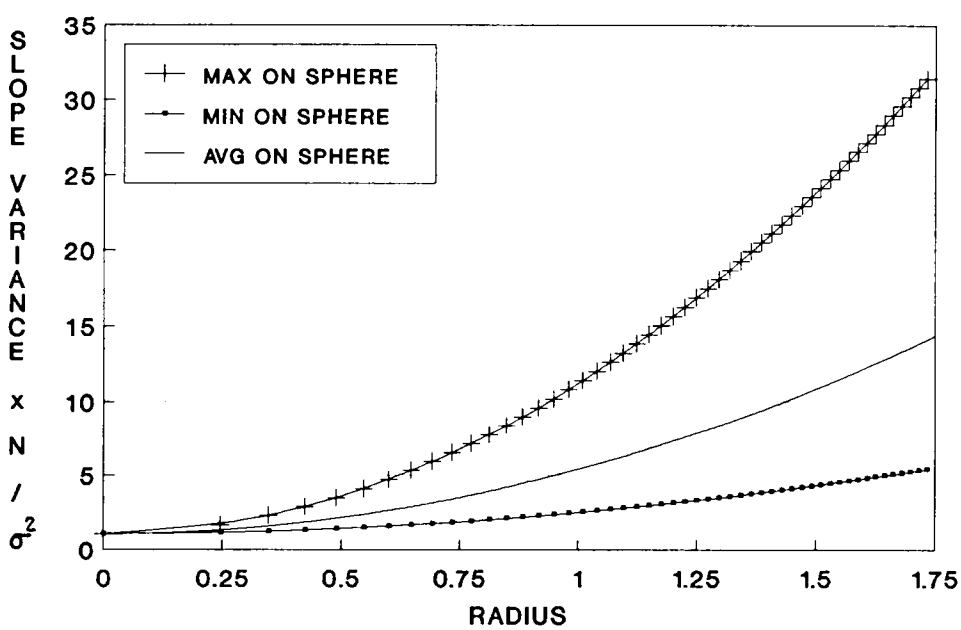

FIGURE 2.6.2
SLOPE VARIANCE ENVELOPE FOR ONE SLOPE
BOX-BEHNKEN DESIGN FOR K=3 WITH 1 C.R.

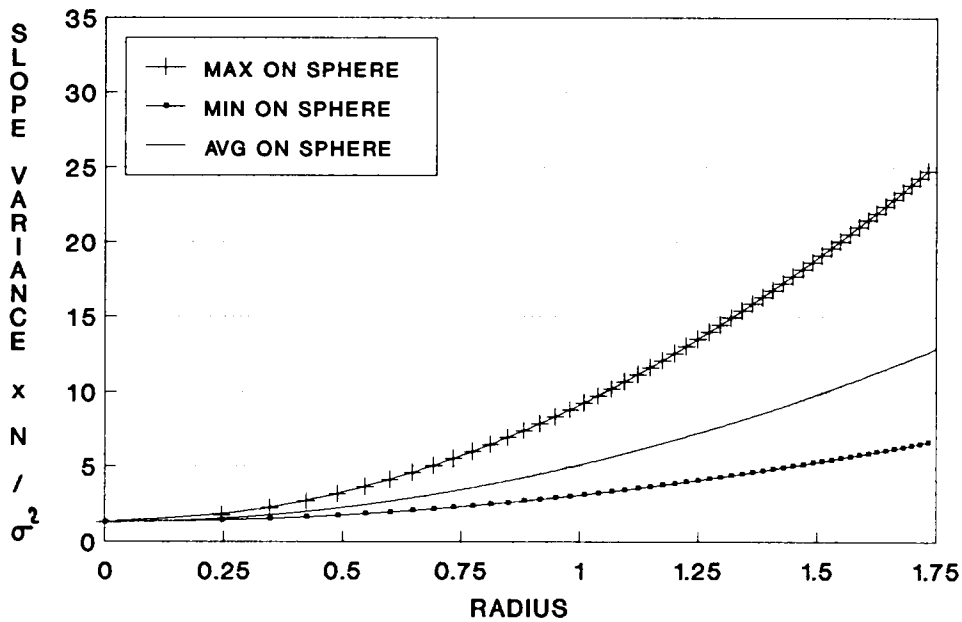

FIGURE 2.6.3
SLOPE VARIANCE ENVELOPE FOR ONE SLOPE
HYBRID 310 DESIGN FOR K=3

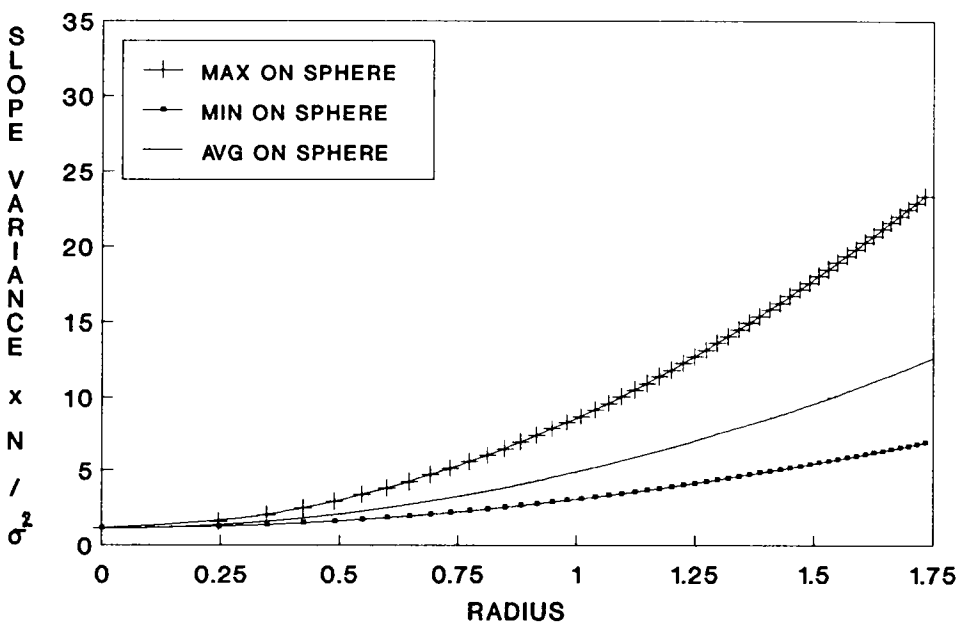

FIGURE 2.6.4
SLOPE VARIANCE ENVELOPE FOR ONE SLOPE
HYBRID 311A DESIGN FOR K=3

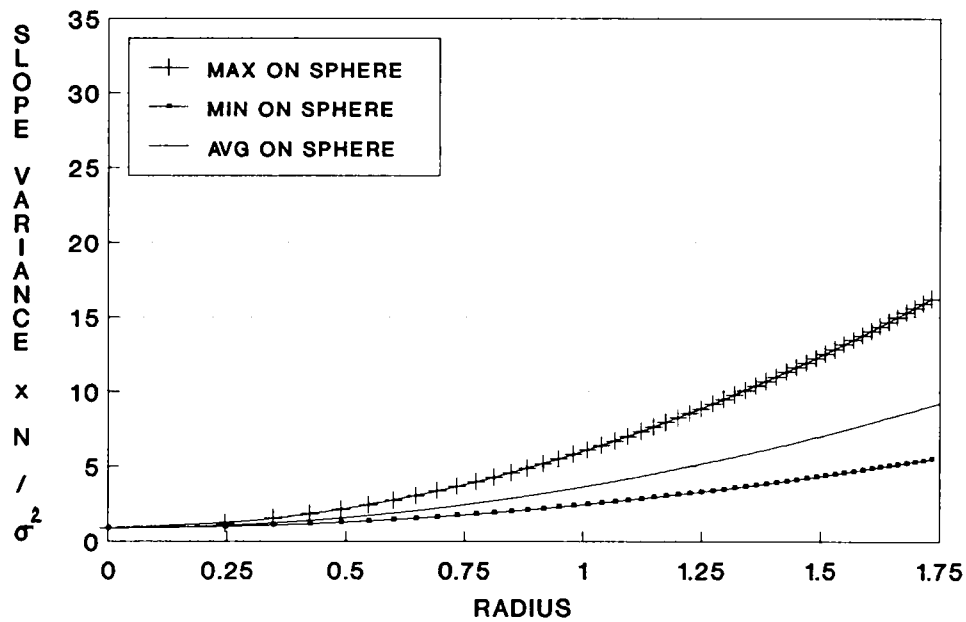

FIGURE 2.6.5
SLOPE VARIANCE ENVELOPE FOR ONE SLOPE
HYBRID 311B DESIGN FOR K=3

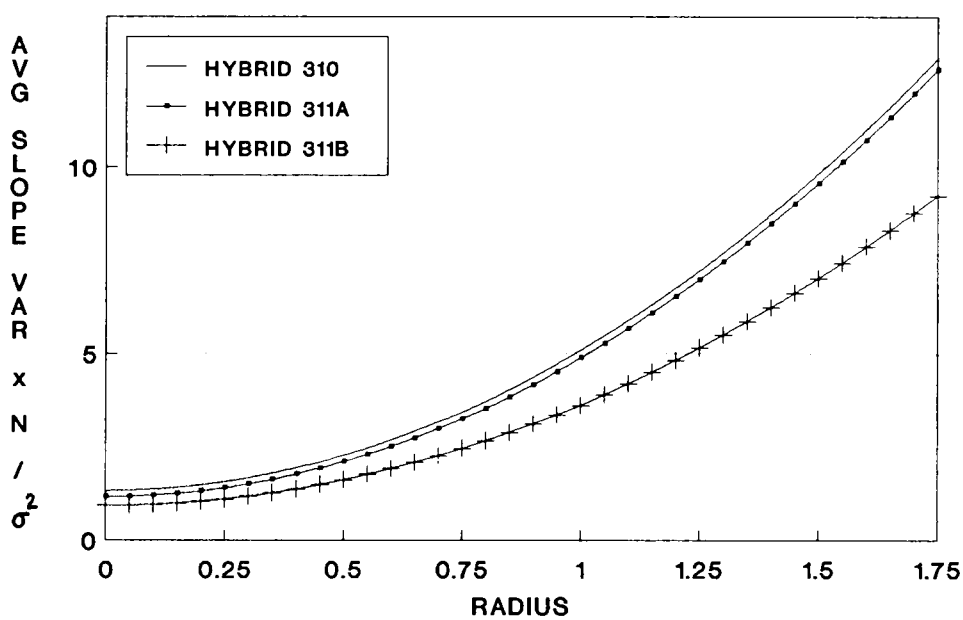

FIGURE 2.6.6
AVERAGE SLOPE VARIANCE COMPARISON
HYBRID 310 VS 311A VS 311B

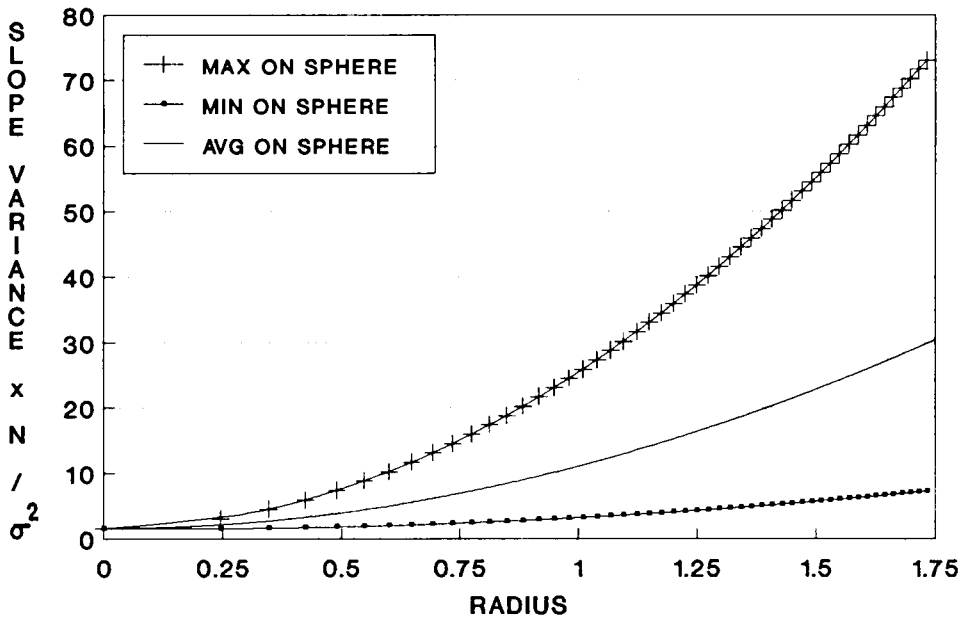

FIGURE 2.6.7
SLOPE VARIANCE ENVELOPE FOR ONE SLOPE
BOX-DRAPER DESIGN FOR K=3 WITH 1 C.R.

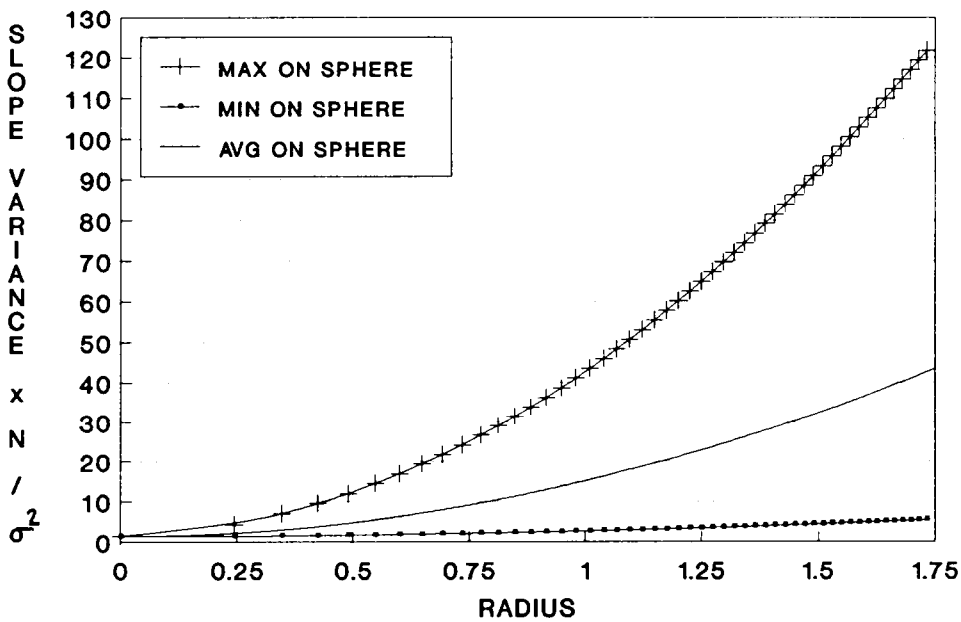

FIGURE 2.6.8
SLOPE VARIANCE ENVELOPE FOR ONE SLOPE
NOTZ DESIGN FOR K=3 WITHOUT C.R.'s

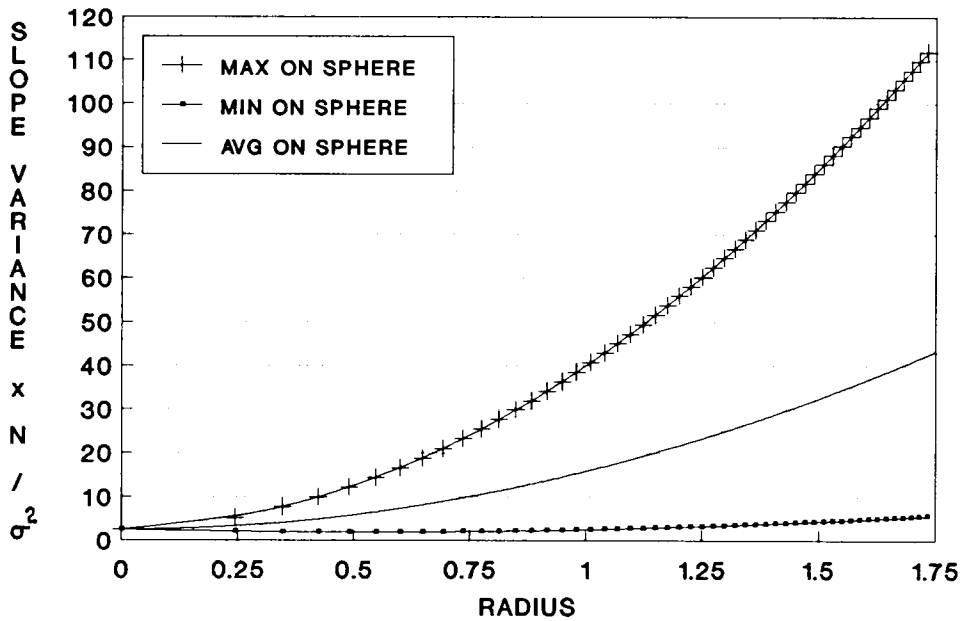

FIGURE 2.6.9
SLOPE VARIANCE ENVELOPE FOR ONE SLOPE
HOKE D2 DESIGN FOR K=3 WITHOUT C.R.'s

2.7 THE "BEST" DESIGNS FOR PRECISE SLOPE ESTIMATION

Four specific response surface designs have been chosen as recommended for slope estimation in the k=3 case. The methods developed to make such choices can easily be applied to more variables, resulting in similar recommendations.

The four designs are:

1. The single-axial CCD with $\alpha = \sqrt{k}$ and 3 center runs,
2. The Hybrid 311B with an additional center run (N=12),
3. The Box-Behnken design with three center runs,

and 4. The dual-axial CCD with no center runs, where g=.7816, α_1 = .5833, and α_2 =1.3015.

They are all very close in their capability to estimate the slope precisely, as is shown in Figure 2.7.1. Note that for all four designs, their maximum variance curves are below the slope variance curve for the slope-rotatable CCD, which is seen in Figure 2.7.2. This implies that even at their worst, these designs estimate the slope more precisely than does the s-rotatable CCD at its best (except for a very small area in the center).

Plots such as the VDG, GSVE and SVE are more informative than measures of efficiency such as G-efficiency (Lucas, 1976). Giovannitti-Jensen and Myers (1989) show that G-efficiencies alone do not adequately describe response prediction variance across the region. It is recommended to compare designs for optimization purposes by comparing slope variance envelopes (or GSVEs), but when many designs are considered, it would be useful to have some kind of efficiency

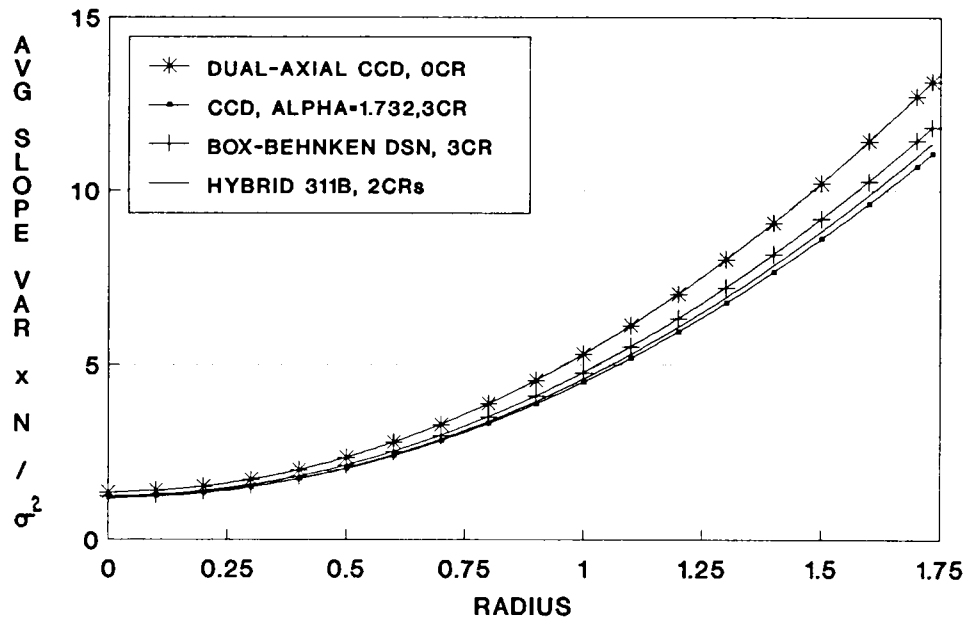

FIGURE 2.7.1
AVERAGE SLOPE VARIANCE COMPARISON
BEST DESIGNS FOR SLOPE ESTIMATION, K=3

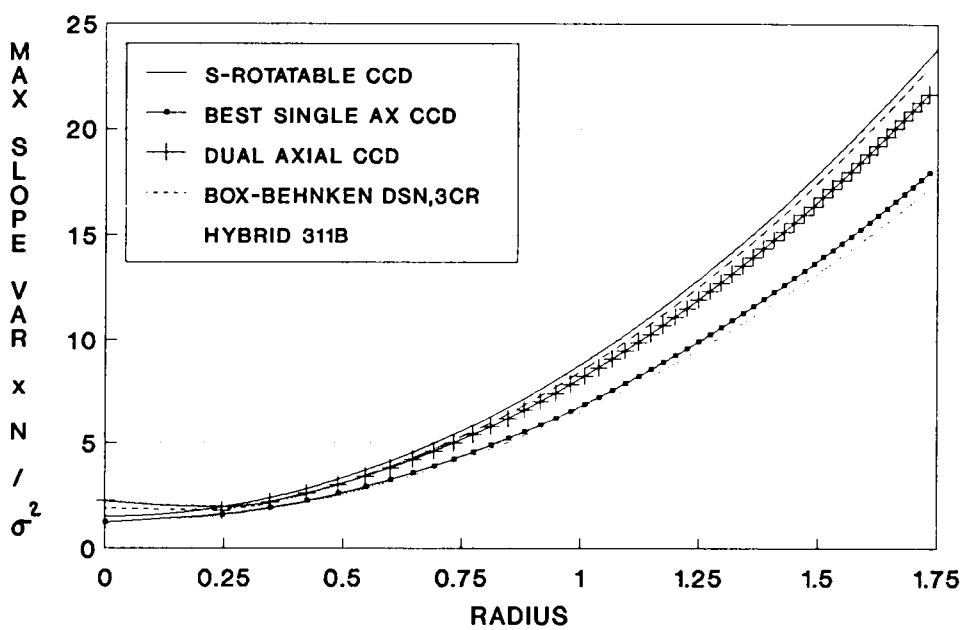

FIGURE 2.7.2
MAXIMUM SLOPE VARIANCE COMPARISON
BEST DESIGNS FOR SLOPE ESTIMATION, K=3

measure. There is no measure of G-efficiency for slope estimation, but an efficiency measure can be developed by measuring a design's performance relative to the "best" design for precise slope estimation. Although several designs have been recommended for precise slope estimation, an efficiency measure will be defined relative to the single axial CCD with $\alpha=\sqrt{k}$ and 3 center runs. Unlike the case of response prediction, for most standard response surface designs, the maximum difference in slope estimation precision is found at a specific location in the region, that is, at the perimeter.

Theorem 2.7.1: If V^{*r} as defined in (2.3.2) and (2.3.3) is used to measure and compare the slope estimation precision capabilities of two standard RS designs, the maximum difference in V^{*r} from one design to the other is located at the design perimeter, that is, at $r = \sqrt{k}$, provided that

$$\left| \frac{2(c_1^{ii}-c_2^{ii})}{4(c_1^{iiii}-c_2^{iiii})+(\sum_{i\neq j} c_1^{iijj}-\sum_{i\neq j} c_2^{iijj})} \right| \leq 1.$$

The proof is given in Appendix A.4.

Although it has not been proved, it appears that the biggest difference in the maximum slope variance on spheres occurs at the perimeter as well. Based on this conjecture, an efficiency measure for slope estimation is defined to be the ratio of $\max [\mathrm{var}(\hat{s}_i)]$ for a standard response surface design, to $\max [\mathrm{var}(\hat{s}_i)]$ for the stated "best" CCD. That is, define the slope variance efficiency measure as

$$\mathrm{SVEM(DSN)} = \frac{\max_{\underline{x}'\underline{x}=k} [\mathrm{var}(\hat{s}_i|\mathrm{CCD})]}{\max_{\underline{x}'\underline{x}=k} [\mathrm{var}(\hat{s}_i|\mathrm{DSN})]} . \qquad (2.7.1)$$

If SVEM(DSN) is less than 1.0, then DSN does not provide as precise slope estimation as does the CCD (with $\alpha=\sqrt{k}$ and three center runs) at the perimeter, which is where the difference in designs is conjectured to be greatest. It is hoped that in future research, a precise measure comparable to G-efficiency in slope estimation can be developed.

The SVEMs for the four recommended designs are given in Table 2.7.1.

TABLE 2.7.1

Design for k=3	SVEM
CCD, α=1.732, 3 C.R.s	1.00
Hybrid 311B, 2 C.R.s	1.04
Dual axial CCD, g=.7816, α_1=.5833, α_2=1.3015	0.83
BBD, 3 C.R.s	0.78

CHAPTER 3

SPHERICAL SQUARED BIAS AND MSE MEASURES

In some experimental situations, especially in the initial phase of a response surface investigation, the experimenter may not specify the complete or correct model form. Thus, he may wish to afford himself protection against bias in either response prediction or in slope estimation. In this chapter, a measure of slope bias due to an underspecified model will be developed. It is a spherical squared bias measure analogous to that developed by Vining for response prediction. The spherical slope bias, combined with the spherical slope variance, into an MSE measure, will be applied in an investigation of saturated orthogonal arrays. The squared slope bias is then used in Chapter 4 to evaluate the designs recommended for precise slope estimation in their robustness to outliers in the data.

3.1 BIAS IN THE PREDICTED RESPONSE

3.1.1 THE HAZARDS OF AN UNDERFIT MODEL

Suppose a model is fit which is of order d_1.

$$\underline{y} = X_1\underline{\beta}_1 + \underline{\epsilon}, \tag{3.1.1}$$

when the correct model is of order $d_2 > d_1$,

$$\underline{y} = X_1\underline{\beta}_1 + X_2\underline{\beta}_2 + \underline{\epsilon}, \tag{3.1.2}$$

where X_1 is the Nxp_1 model matrix for the fitted model, $\underline{\beta}_1$ is the p_1x1 vector of corresponding coefficients, X_2 is the Nxp_2 matrix of

independent variable terms not in the fitted model, and $\underline{\beta}_2$ is the corresponding p_2x1 vector of parameters not present in the presumed model. The model has been underfit.

There are hazards associated with underfitting. The estimate of σ^2 is biased upward. That is,

$$E(s^2)= \sigma^2 + \underline{\beta}_2'[X_2'X_2 - X_2'X_1(X_1'X_1)^{-1}X_1'X_2]\underline{\beta}_2. \quad (3.1.3)$$

The estimated model coefficients, $\underline{b}_1$, are biased:

$$E(\underline{b}_1) = \underline{\beta}_1 + (X_1'X_1)^{-1}X_1'X_2\underline{\beta}_2. \quad (3.1.4)$$

In addition, there is bias in response estimation:

$$E(\hat{y}(\underline{x}_{0,1})) = \underline{x}_{0,1}'[\underline{\beta}_1 + (X_1'X_1)^{-1}X_1'X_2\underline{\beta}_2]. \quad (3.1.5)$$

Thus, the prediction bias at a point $\underline{x}_0'=(\underline{x}_{0,1}' \ \underline{x}_{0,2}')$, in the true model space, is written

$$E(\hat{y}(\underline{x}_{0,1})) - E(y(\underline{x}_0)) = (\underline{x}_{0,1}'A - \underline{x}_{0,2}')\underline{\beta}_2, \quad (3.1.6)$$

where $A=(X_1'X_1)^{-1}X_1'X_2$ is the alias matrix.

3.1.2 PREDICTION BIAS

Vining (1988a, 1988c) has developed the analogue to the variance dispersion graph of Giovannitti-Jensen and Myers, assuming that the experimenter has fit a polynomial of order d_1, when a polynomial model of order $d_2>d_1$ is appropriate. The squared prediction bias at a point $\underline{x}_0$ in the true model space is given by

$$B(\underline{x}_0) = \underline{\beta}_2'[\underline{x}_{0,1}'A - \underline{x}_{0,2}']'[\underline{x}_{0,1}'A - \underline{x}_{0,2}']\underline{\beta}_2. \quad (3.1.7a)$$

For notational simplicity, Vining defines

$$A^{**}(\underline{x}_0) = [\underline{x}_{0,1}'A - \underline{x}_{0,2}]'[\underline{x}_{0,1}'A - \underline{x}_{0,2}]. \quad (3.1.8)$$

Hence, $$B(\underline{x}_0) = \underline{\beta}_2' A^{**}(\underline{x}_0)\underline{\beta}_2. \qquad (3.1.7b)$$

By maximizing and averaging $B(\underline{x}_0)$ over the surface of several concentric hyperspheres in the design space, centered at the design center, bias plots similar to the variance dispersion graphs for response prediction result.

3.2 BIAS IN SLOPE ESTIMATION

When a researcher assumes a polynomial model of order d_1 when the true model is of order $d_2 > d_1$, there is bias in the estimated slopes of the fitted surface as well. The presumed response model of order d_1 is given by (3.1.1), while the true model of order d_2 is given by (3.1.2). The slope of the surface in the direction of the x_j axis using the presumed model is obtained through partial differentiation of the presumed model with respect to x_j, and is written

$$s_j = \underline{l}_{j1}' \underline{\beta}_1, \qquad (3.2.1)$$

but the true slope is

$$s_j = \underline{l}_{j1}' \underline{\beta}_1 + \underline{l}_{j2}' \underline{\beta}_2, \qquad (3.2.2)$$

where $\underline{l}_{j1}$ is the vector such that $\underline{l}_{j1}' \underline{\beta}_1$ is the partial derivative of the presumed model, and $\underline{l}_{j2}$ is such that $\underline{l}_{j2}' \underline{\beta}_2$ contains the terms in the partial derivative of the true model not contained in $\underline{l}_{j1}' \underline{\beta}_1$. It follows that

$$\begin{aligned} E(\hat{s}_j) &= E(\underline{l}_{j1}' \underline{b}_1) \\ &= \underline{l}_{j1}' (X_1' X_1)^{-1} X_1' E(\hat{\underline{y}}) \\ &= \underline{l}_{j1}' (X_1' X_1)^{-1} X_1' [X_1 \underline{\beta}_1 + X_2 \underline{\beta}_2] \end{aligned}$$

$$= \underline{l}_{j1}'[\underline{\beta}_1 + A\underline{\beta}_2], \qquad (3.2.3)$$

where $A=(X_1'X_1)^{-1}X_1'X_2$ is the alias matrix. Hence, the bias in the slope is

$$E(\hat{s}_j) - E(s_j) = [\underline{l}_{j1}'A - \underline{l}_{j2}']\underline{\beta}_2 \qquad (3.2.4)$$

and the squared slope estimation bias in the estimated slope parallel to the x_j axis, at a point $\underline{x}_0$ in the true model space, is

$$B_j^*(\underline{x}_0) = \underline{\beta}_2'[\underline{l}_{j1}'A - \underline{l}_{j2}']'[\underline{l}_{j1}'A - \underline{l}_{j2}']\underline{\beta}_2$$
$$= \underline{\beta}_2'A_j^*(\underline{x}_0)\underline{\beta}_2, \qquad (3.2.5)$$

where $A_j^*(\underline{x}_0) = [\underline{l}_{j1}'A - \underline{l}_{j2}']'[\underline{l}_{j1}'A - \underline{l}_{j2}']$.

3.2.1 MAXIMIZING SLOPE BIAS ON SPHERES

Analogous to the slope variance plots, the maximum and average squared slope biases on hyperspheres of radius r are plotted against the radius. However, $B_j^*(\underline{x}_0)$ is a function of $\underline{\beta}_2$, which is unknown. A result from linear algebra (Noble and Daniel, 1977) states that since $B_j^*(\underline{x}_0)$ is a quadratic form in $\underline{\beta}_2$,

$$\max_{\underline{\beta}_2'\underline{\beta}_2=1} \underline{\beta}_2'A_j^*(\underline{x}_0)\underline{\beta}_2 = \lambda_k,$$

where λ_k is the largest eigenvalue of $A_j^*(\underline{x}_0)$. Noting that $A_j^*(\underline{x}_0)$ is positive semidefinite and of rank one, it has exactly one positive eigenvalue. Since the trace of a matrix is equal to the sum of its eigenvalues, $tr\{A_j^*(\underline{x}_0)\}=\lambda_k$. By plotting the slope bias in units of $\underline{\beta}_2'\underline{\beta}_2$,

$$\max_{\underline{\beta}_2'\underline{\beta}_2=1} \underline{\beta}_2'A_j^*(\underline{x}_0)\underline{\beta}_2 = tr\{A_j^*(\underline{x}_0)\} \qquad (3.2.6)$$

and

$$\max_{\underline{x}_0'\underline{x}_0=r^2} \underline{\beta}_2'A_j^*(\underline{x}_0)\underline{\beta}_2 = \max_{\underline{x}_0'\underline{x}_0=r^2} tr\{A_j^*(\underline{x}_0)\}. \qquad (3.2.7)$$

Note that the bias in the estimated slopes parallel to the k factor axes are equal for most standard response surface designs. To account for those cases where they are not equal however, we will maximize the maximum of the k slope biases over the surface of the hypersphere. Define

$$B^*(\underline{x}_0) = \max_j B_j^*(\underline{x}_0).$$

The computer algorithm developed by Vining is used to search the surface of spheres to locate optimum functional values.

$$B^*_{max}(r) = \max_{\underline{x}_0'\underline{x}_0=r^2} \{ \max_j [tr(A_j^*(\underline{x}_0))] \}.$$

3.2.2 AVERAGING SLOPE BIAS OVER SPHERES

The average slope bias on a sphere of radius r, $B^*_{avg}(r)$, is also a function of $\underline{\beta}_2$. A double integration approach is used to calculate the average slope bias. The first average is over the surface of a hypersphere in the $\underline{\beta}_2$ space, (U_β), where $|\underline{\beta}_2|=1$ and $\underline{\beta}_2$ is the vector of parameters in the true model that are not in the fitted model. Note that this averaging convention simply implies that the plot of slope estimation bias will be in units of $|\beta_2|$. The second average is over the surface of the hypersphere of radius r, (U_r), in the design variable space. That is,

$$B^*_{avg}(r) = K\int_{U_\beta} \int_{U_r} B^*(\underline{x}_0) d\underline{x} d\underline{\beta}_2$$

$$= (1/p_2)tr[A'\sigma^*_{11}A - 2\sigma^*_{12}A + \sigma^*_{22}], \qquad (3.2.8)$$

where $K^{-1} = \int_{U_\beta} \int_{U_r} d\underline{x} d\underline{\beta}_2,$

$$\overset{*}{\sigma}_{11} = K\int_{U_r} \underline{l}_{j1}\underline{l}_{j1}'d\underline{x},$$

$$\overset{*}{\sigma}_{12} = K\int_{U_r} \underline{l}_{j1}\underline{l}_{j2}'d\underline{x},$$

$$\overset{*}{\sigma}_{22} = K\int_{U_r} \underline{l}_{j2}\underline{l}_{j2}'d\underline{x}$$

and j identifies the slope for which $B^*(\underline{x}_0)=\max_j \overset{*}{B}_j(\underline{x}_0)$. The quantities $\overset{*}{\sigma}_{11}$, $\overset{*}{\sigma}_{12}$, and $\overset{*}{\sigma}_{22}$ are spherical region moment matrices and are calculable. Stroud (1971) provides the details of the integration but the useful results are given in Section 1.5.

3.2.3 MEAN SQUARED ERROR IN SLOPE ESTIMATION

When conducting a response surface investigation for optimization purposes, the interest is in finding a model with both small slope variance and adequate protection against slope bias. Bias is of particular importance when a first order model is fit. Consider combining the slope variance and slope bias information into one measure of slope estimation capability. By assessing the relative importance of bias to variance, we will define the MSE measures.

Recall that the mean squared error in slope estimation is given by

$$MSE[\hat{s}_j(\underline{x}_0)] = \sigma^2\underline{l}_{j1}'(X_1'X_1)^{-1}\underline{l}_{j1} + \underline{\beta}_2'A^*_j(\underline{x}_0)\underline{\beta}_2.$$

A result from linear algebra (Noble and Daniel 1977) states that

$$\max_{|\underline{\beta}_2|=q} \underline{\beta}_2'A_j^*(\underline{x}_0)\underline{\beta}_2 = q^2\lambda_L$$

where λ_L is the largest eigenvalue of $A_j^*(\underline{x}_0)$. It has been shown that $\lambda_L = tr[A_j^*(\underline{x}_0)]$. Hence, define

$$MSE_{max}(r) = \max_{\underline{x}_0'\underline{x}_0=r^2} \max_j \{\sigma^2 \underline{l}_{j1}'(X_1'X_1)^{-1}\underline{l}_{j1} + q^2 tr[A_j^*(\underline{x}_0)]\}$$

Note that q is the length of the unknown $\underline{\beta}_2$ vector, which is the source of the bias.

Before dealing with the unknown σ^2 and q^2, consider the average mean squared error.

$$MSE_{avg}(r) = \operatorname{avg}_{\underline{x}_0'\underline{x}_0=r^2} \operatorname{avg}_{|\underline{\beta}_2|=q} [\sigma^2 \underline{l}_{j1}'(X_1'X_1)^{-1}\underline{l}_{j1} + \underline{\beta}_2' A^*(\underline{x}_0)\underline{\beta}_2]$$

where $A^*(\underline{x}_0)$ is chosen from the $A_j^*(\underline{x}_0)$ so that $B^*(\underline{x}_0)=\max_j B_j^*(\underline{x}_0)$. From results in Chapter 2 and Section 3.2.2, this can be written

$$MSE_{avg}(r) = \sigma^2 K \int_{U_r} \underline{l}_{j1}'(X_1'X_1)^{-1}\underline{l}_{j1} d\underline{x}_0 + \frac{q^2}{p_2} tr[A'\overset{*}{\sigma}_{11}A - 2\overset{*}{\sigma}_{12}A + \overset{*}{\sigma}_{22}].$$

Now the unknown σ^2 and q^2 must be addressed. Dealing with these quantities effectively is critical in that they describe the state of nature, that is, the amount of bias present relative to the variance. Consider

$$MSE^*_{max}(r) = \frac{MSE_{max}(r)}{\sigma^2 + q^2}$$

$$= \max_{\underline{x}_0'\underline{x}_0=r^2} \max_j \left[\frac{\sigma^2}{\sigma^2+q^2} \underline{l}_{j1}'(X_1X_1)^{-1}\underline{l}_{j1} + \frac{q^2}{\sigma^2+q^2} tr A_j^*(\underline{x}_0)\right]$$

$$= \max_{\underline{x}_0'\underline{x}_0=r^2} \max_j [(1-w_B)\underline{l}_{j1}'(X_1'X_1)^{-1}\underline{l}_{j1} + w_B tr A_j^*(\underline{x}_0)],$$

and $$MSE^*_{avg}(r) = \frac{MSE_{avg}(r)}{\sigma^2 + q^2}$$

$$= (1-w_B)K \int_{U_r} \underline{l}_{j1}'(X_1'X_1)^{-1}\underline{l}_{j1} d\underline{x}_0 + \frac{w_B}{p_2} tr[A'\overset{*}{\sigma}_{11}A - 2\overset{*}{\sigma}_{12}A + \overset{*}{\sigma}_{22}],$$

where w_B is the relative importance of slope bias in the mean squared error of slope estimation. This division by (σ^2+q^2) simply states that the $MSE^*(r)$ plots will be in units of (σ^2+q^2).

Naturally the relative weight, w_B, must be chosen to provide a meaningful balance between variance and bias. One must address the amount of bias which is important, and thus should be protected against, relative to the variance, and choose w_B to reflect that amount. In practice, one is not smart to use a design to fit a model known apriori to be underspecified. Typically, it is good to build into the design the ability to detect lack of fit. The choice of w_B will be made through consideration of the lack of fit test.

It can be shown that the noncentrality parameter (n.c.p.) in a lack of fit test can be written as

$$\nu = \frac{1}{2\sigma^2}\underline{\beta}_2'[X_2'X_2 - X_2'X_1(X_1'X_1)^{-1}X_1'X_2]\underline{\beta}_2.$$

The n.c.p. is not constant for all $\underline{\beta}_2$ of length q, so consider an average noncentrality parameter,

$$\nu_{avg} = \underset{|\underline{\beta}_2|=q}{avg}\ \nu = \frac{q^2}{2\sigma^2 p_2}.$$

Given the average n.c.p., the average power of the lack of fit test can be calculated.

So, the experimenter determines the q^2/σ^2 value which yields the appropriate average power to detect lack of fit, thus defining the amount of bias he considers to be important. He then calculates

$$w_B = \frac{(q^2/\sigma^2)}{1+(q^2/\sigma^2)} \quad ,$$

which is used in the measures of mean squared error in slope estimation.

The above determination of w_B is carried out before any experimentation, so regardless of the design chosen for experimentation, a design good for detecting lack of fit must be used in the calculation of w_B. The chosen design for practice is then evaluated using $MSE^*_{max}(r)$ and $MSE^*_{avg}(r)$ in its ability to adequately protect against a predetermined amount of bias. This bias is that which would have been detected with a known power if the ideal design for detecting lack of fit had been used.

In the following section, saturated orthogonal arrays (SOAs) will be evaluated using the MSE^* measures in their ability to protect against bias that would cause a lack of fit test to have power .5, that is, the amount that would cause a lack of fit test to reject the hypothesis of model misspecification 50% of the time. The designs used to calculate w_B are those determined to be good for detecting lack of fit, and will be given in each case studied. The SOAs will then be augmented with points added to detect lack of fit, and the resulting designs will be evaluated in their ability to protect against bias using the MSE^* measures. In order to compare designs of differring sizes (SOAs versus augmented SOAs), $MSE^*_{avg}(r)$ and $MSE^*_{max}(r)$ will include a factor of N, the number of points in the design. Adjusted as such, they represent a measure of efficiency in slope bias

protection. Note that power to detect lack of fit does not guarantee protection from slope bias.

3.3 BIAS IN SLOPE ESTIMATION USING AUGMENTED SATURATED ORTHOGONAL ARRAYS

Quality improvement is receiving increased attention in the United States today. American industry is being forced to change its views on quality. One reason for this is the success of the Japanese. They have been able to produce high quality products at low cost. Taguchi credits off-line quality control measures as the most efficient way to assure product performance. Off-line quality control refers to efforts to build quality into a product/process at the design stage. (Taguchi, 1986.)

Taguchi uses experimental designs composed of orthogonal arrays of strength two with a marginal means analysis to choose the best level for each of many design factors investigated in what he calls parameter design. This section will apply the MSE* measures to evaluate designs like those recommended by Taguchi.

Users of Taguchi methods claim that the goal of parameter design is not to estimate coefficients or even to fit a model, but simply to identify design factors which can be used to control the response, whether it be a signal-to-noise ratio or a product characteristic. But once these factors are identified, a level for each is chosen to optimize the response. This choice is made by looking at graphs, but the technique is equivalent to estimating linear coefficients in a

sometimes oversimplified model.

When a model is underspecified, it is appropriate to investigate induced biases. In this section, the slope estimation biases are investigated before and after augmenting an orthogonal array with points which maximize the power to detect lack of fit due to a higher order underlying model. The use of MSE* plots will be illustrated.

3.3.1 THE ALTERNATIVE MODELS

The matrix $A_j^*(\underline{x}_0)$ depends on the choice of the higher order alternative model. We consider three alternatives:

1) fit a first order model and protect against only two-factor interactions,

2) fit a first order model and protect against bias due to a full quadratic,

and 3) fit the second order model without two-factor interactions, yet protect against them.

These three situations are of particular interest in the evaluation of screening designs like those recommended by Taguchi. The spherical MSE* measures for slope estimation will be applied to evaluate saturated, or near-saturated, orthogonal arrays before and after augmentation to detect lack of fit. It is hoped that we can determine whether or not power to detect lack of fit translates into

protection from bias due to model misspecification.

3.3.2 CRITERION FOR AUGMENTATION

Morris and Mitchell (1980, 1983) present designs for detecting the presence of interactions which are good in the sense of tr(L)-optimality and augment these designs according to the D-optimality criterion. The result is a compromise between tr(L)-and D-optimal designs. Basically, tr(L)-optimality is an attempt to maximize the power of detection of the lack of fit due to an underspecified model. We will start with the D-optimal orthogonal arrays of Taguchi, and augment them according to tr(L)-optimality.

Suppose that a model is underspecified. For example, say the experimenter fits the main effects model

$$\underline{y} = X_1\underline{\beta}_1,$$

$$\text{where } \underline{\beta}_1' = (\beta_0\ \beta_1\ \ldots\ \beta_k),$$

$$\text{and } X_1 = [\underline{1}\ \underline{x}_1\ \underline{x}_2\ \ldots\ \underline{x}_k],$$

when the true model includes two-factor interactions,

$$\underline{y} = X_1\underline{\beta}_1 + X_2\underline{\beta}_2,$$

$$\text{where } \underline{\beta}_2' = (\beta_{12}\ \beta_{13}\ \ldots\ \beta_{k-1,k}),$$

$$\text{and } X_2 = [\underline{x}_1\underline{x}_2\ \underline{x}_1\underline{x}_3\ \ldots\ \underline{x}_{k-1}\underline{x}_k].$$

Suppose further that the method of least squares is used to fit the presumed model. It can be shown that the expected residual sum of squares can be written

$$E(RRS) = (N - r(X_1))\sigma^2 + \underline{\beta}_2'L\underline{\beta}_2, \qquad (3.3.1a)$$

where $r(X_1)$ denotes the rank of the matrix X_1, and

$$L = X_2'[I - X_1(X_1'X_1)^{-}X_1']X_2 \qquad (3.3.1b)$$

is the lack-of-fit matrix. Morris and Mitchell (1980) showed that maximizing the trace of the lack-of-fit matrix, tr(L), also maximizes

$$E(\underline{\beta}_2'L\underline{\beta}_2) = \beta^2_{INT}tr(L), \qquad (3.3.2)$$

where β^2_{INT} is the average of the squared interaction coefficients.

For any alternative model, to use the tr(L) criterion, one picks a design that has large tr(L), thus large E(RSS), when $|\underline{\beta}_2| > 0$. The authors modified the DETMAX algorithm of Mitchell (1974) to search a set of given candidate points for those which maximize tr(L) for a given number of design points. In practice, they actually maximize $tr(T^{-1}L)$, where T is of the same form as L, but X_2 and X_1 are replaced by their candidate points counterpart. This is similar to the idea of a region moment matrix in RS terminology, where the region is defined by the candidates. They use this form of the tr(L) criterion because it has some nice invariance properties.

The value of $tr(T^{-1}L)$ increases with each added point. This means that the power to detect lack-of-fit is increasing. It seems intuitive that the increased detection power may also, in these cases, provide protection against bias in response prediction and/or in slope estimation. We will use the spherical MSE^* plots to measure the difference in MSE^* attained by augmenting according to the tr(L)-optimality criterion.

3.3.3 DETECTION DOES NOT GUARANTEE PROTECTION

Two of the designs typical of those recommended by Taguchi are the L_8 and L_9 SOAs given by (1.6.7a) and (1.6.7b). The first, L_8, is used to fit the seven factor main effects model. The second, L_9, is recommended for fitting a second order model without interactions in four factors. Each design, unaugmented, is saturated, and thus provides no power to detect lack of fit. Each array has been augmented, according to tr(L)-optimality, one point at a time, until the design size is doubled. With each added point, the power to detect lack of fit is increased. The alternative models considered as the sources of lack of fit are those listed in Section 3.3.1, with L_8 augmented to detect lack of fit due to the first alternative model, and L_9 augmented for the third case. The augmented designs are summarized in Tables 3.1 and 3.2.

The MSE^* measures are used to measure protection against bias (and variance) obtained from the added points. Augmented L_8 arrays are investigated in their ability to protect against slope bias from alternative models 1 and 2, while the L_9 augmented arrays are evaluated in their ability to protect against slope bias from alternative model 3. The designs used to calculate w_B in the MSE^* measures in each case are 1) the resolution IV fractional factorial in seven variables, 2) the resolution IV fractional factorial with center runs, 3) a central composite design in four variables.

The results of the investigations are portrayed in Figure 3.3.1 through Figure 3.3.5.

TABLE 3.1

AUGMENTED L_8 DESIGNS

	FACTORS							POINTS* TO BE ADDED TO MAXIMIZE $TR(T^{-1}L)$								
RUN	1	2	3	4	5	6	7	9	10	11	12	13	14	15	16	j N
1	-	-	-	-	+	-	-							X	X	
2	-	-	-	+	-	+	+					X	X	X	X	
3	-	+	+	-	+	+	+		X	X	X	X	X	X	X	
4	-	+	+	+	-	-	-								X	
5	+	-	+	-	-	-	+	X	X	X	X	X	X	X	X	
6	+	-	+	+	+	+	-				X	X	X	X	X	
7	+	+	-	-	-	+	-						X	X	X	
8	+	+	-	+	+	-	+			X	X	X	X	X	X	
			$tr(T^{-1}L)/N$:					4.7	8.4	11.5	14	16.2	18	19.6	21	

*Points checked "X" are foldovers of the points across from the check, in the L_8 design. e.g. the 11-point design contains L_8 and foldovers of points 3, 5, and 8.

A foldover of a point is the point which reverses all of the signs in the original. For example, the foldover of (+1,-1,+1,-1,-1,-1,+1) is (-1,+1,-1,+1,+1,+1,-1).

TABLE 3.2

AUGMENTED L_9 DESIGNS

N	$tr(T^{-1}L)/N$	ADDED RUNS*
10	5.0	(+ - - -)
11	5.9	(- - + +) (+ + + +)
12	6.4	(- - + +) (+ + + +) (- + - +)
13	6.9	(+ - - -) (- + - -) (- - + -) (- - - +)
14	8.1	(+ + + +) (- - + +) (- + - +) (- + + -) (+ - - -)
15	8.6	(+ + + +) (- - + +) (- + - +) (- + + -) (+ - - -) (+ - - +)
16	8.6	(+ + + +) (- - + +) (- + - +) (- + + -) (+ - + -) (+ - - +) (+ + - -)
17	9.3	(+ + + +) (- - - +) (+ - + +) (- + - -) (+ + - +) (- - + -) (+ - - -) (+ + + -)
18	9.4	(- - - +) (+ - + +) (- + - -) (- + + -) (- - + -) (+ - - -) (+ + + +) (+ + - +) (+ + + -)

* The basic L_9 array is contained in all of the resulting designs.

The design matrix for L_9 is given by (1.6.7b).

Figure 3.3.1 contains the $MSE^*_{avg}(r)$ against radius plots for five L_8 arrays when protecting against the model with two factor interactions: L_8 unaugmented, L_8 plus two points, L_8 plus four points, L_8 plus six points, and the folded over L_8 design. Recall that $MSE^*_{avg}(r)$ has been adjusted by a factor of N for each design. Without this adjustment, each added point does provide some additional protection against slope bias uniformly for all radii, but the measure assumes a cost of zero for the additional points. It is more realistic to think in terms of maximizing efficiency in bias protection by including the sample size factor. By studying Figure 3.3.1, we see that in the perimeter of the region (which is most important when searching for an optimum using a first order model), the added points are not worth their cost in providing slope bias protection, on the average, against a model with interactions. At the perimeter, the unaugmented L_8 array has the smallest MSE^*_{avg}, and the measure gets larger with each additional point. Thus we must conclude that when protecting against slope bias in the presence of interactions, added points providing detection power do not efficiently provide protection.

Figure 3.3.2 is more interesting. This plot contains the $MSE^*_{max}(r)$ versus radius curves for the same five L_8 arrays. Continuing to focus on the perimeter of the region, again we see that the unaugmented L_8 array is most efficient in protecting against bias in a minimax sense. If we are to augment at all however, according to tr(L)-optimality, the next best design is the completely folded over array with eight additional points. Keep in

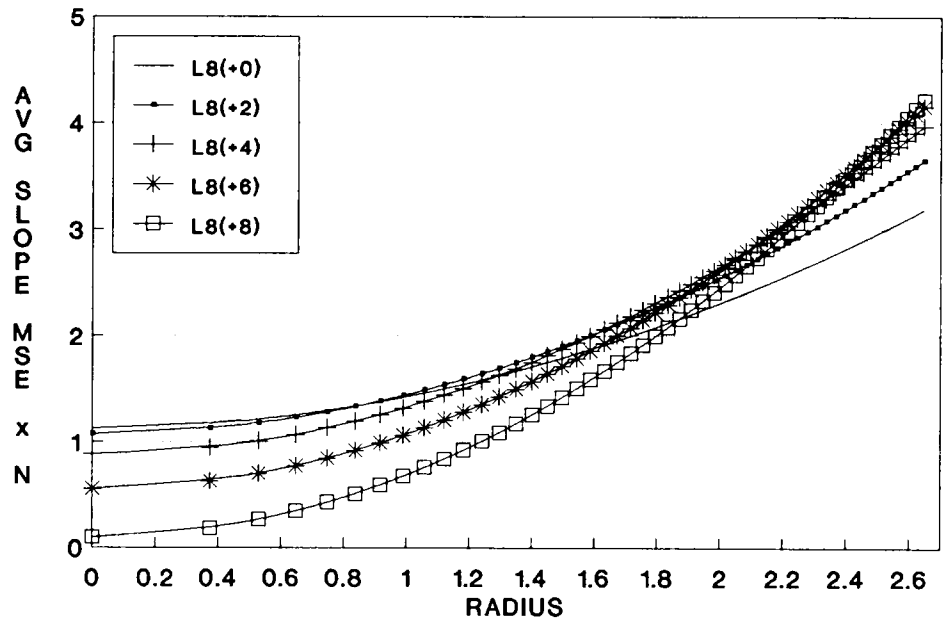

FIGURE 3.3.1: AVG SLOPE MSE COMPARISON
L8 ARRAYS AUGMENTED BY TR(L)-OPTIMALITY
PROTECT AGAINST INTERACTIONS

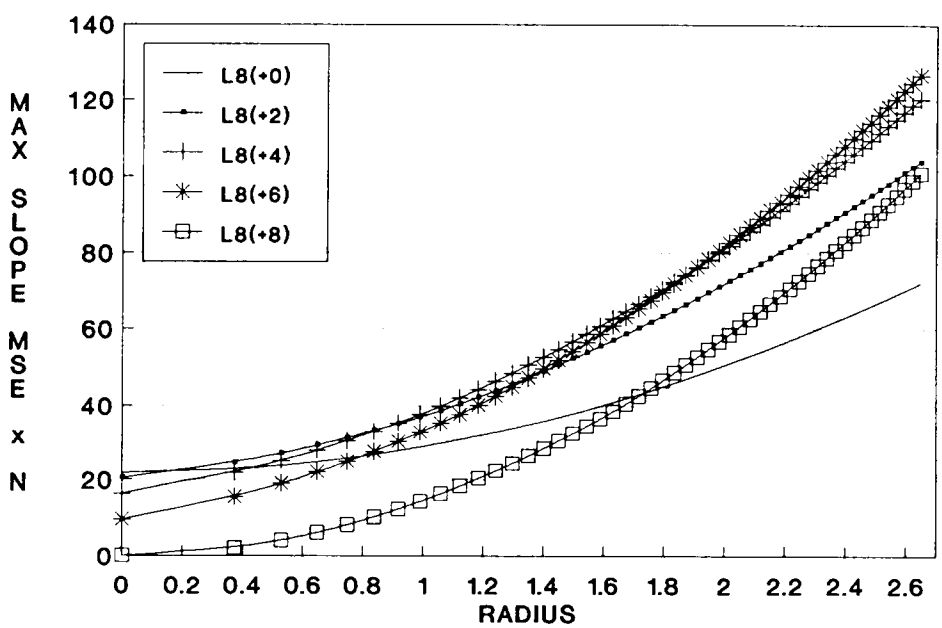

FIGURE 3.3.2: MAX SLOPE MSE COMPARISON
L8 ARRAYS AUGMENTED BY TR(L)-OPTIMALITY
PROTECT AGAINST INTERACTIONS

mind that the MSE^* measure includes a factor of N for each design, punishing the larger design for requiring the extra points. But the conclusion is still the same: points added to detect lack of fit due to interactions do not efficiently provide slope bias protection in the portion of the region where it is most needed.

Figures 3.3.3 and 3.3.4 are illustrations of the $MSE^*_{avg}(r)$ and $MSE^*_{max}(r)$ measures respectively, where the same five L_8 arrays are evaluated in their ability to provide protection from bias induced by the full quadratic model. The same conclusion is stated even more strongly in this case.

Finally, Figure 3.3.5 portrays the results of the L_9, second order, investigation. In the figure are plots of $MSE^*_{avg}(r)$ including the sample size factor, against the radius r, for four L_9 arrays: the unaugmented L_9 array, L_9 plus three points added to maximize the power to detect lack of fit due to interactions, L_9 plus the six best points, and L_9 plus the nine best points. Clearly, the added points for detecting lack of fit are not efficiently protecting against slope bias due to interactions. It can be shown that even if the factor N were omitted, the added points provide negligible amounts of added protection. A plot containing the $MSE^*_{max}(r)$ information has the same appearance and thus has not been included.

The conclusion then, is that augmenting saturated orthogonal arrays to detect the lack of fit due to an underspecified model, will not necessarily provide efficient protection against slope

estimation bias in the portion of the region where it matters most. It is only fair to note here, that the points considered as candidates for augmentation of L_8 and L_9 are those in the complete 2^7 factorial with center points, and the complete 3^4 factorial respectively. These are quite reasonable sets of candidate points, but the results may have been different if candidate points were chosen say from a uniform distribution over the region. The intent here is simply to illustrate the usefulness of the MSE* measures in such an investigation.

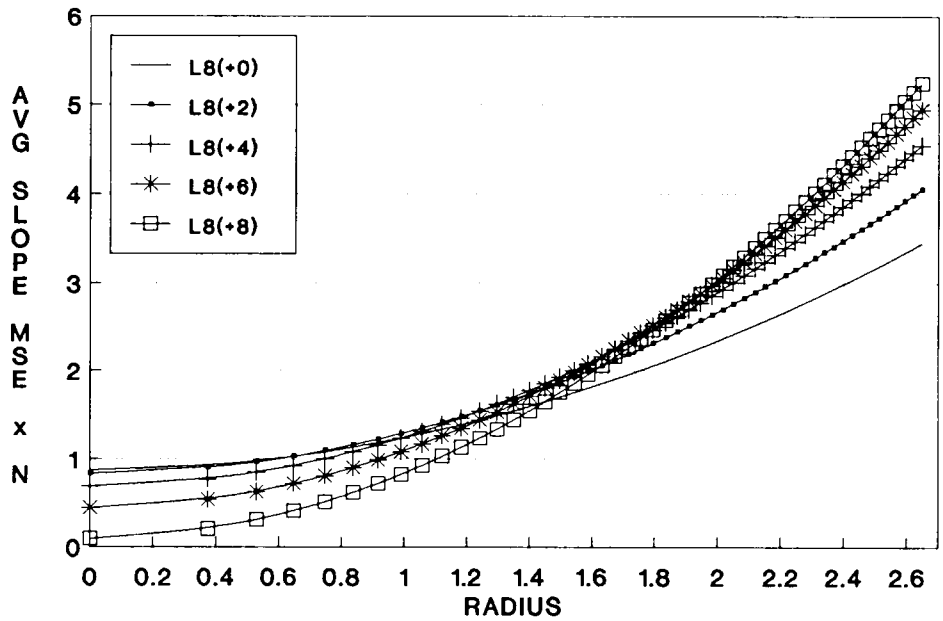

FIGURE 3.3.3: AVG SLOPE MSE COMPARISON
L8 ARRAYS AUGMENTED BY TR(L)-OPTIMALITY
PROTECT AGAINST FULL 2nd ORDER

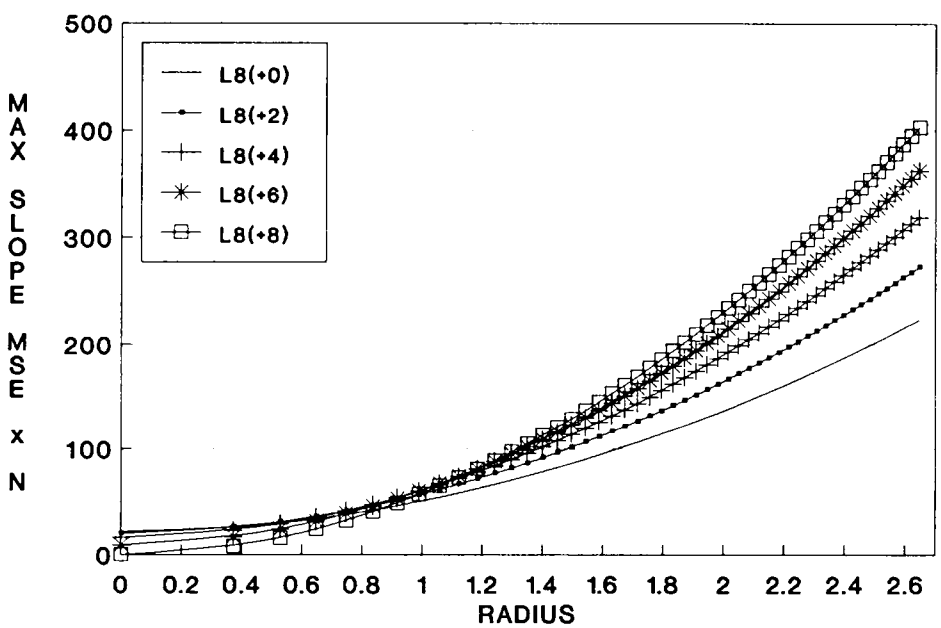

FIGURE 3.3.4: MAX SLOPE MSE COMPARISON
L8 ARRAYS AUGMENTED BY TR(L)-OPTIMALITY
PROTECT AGAINST FULL 2nd ORDER

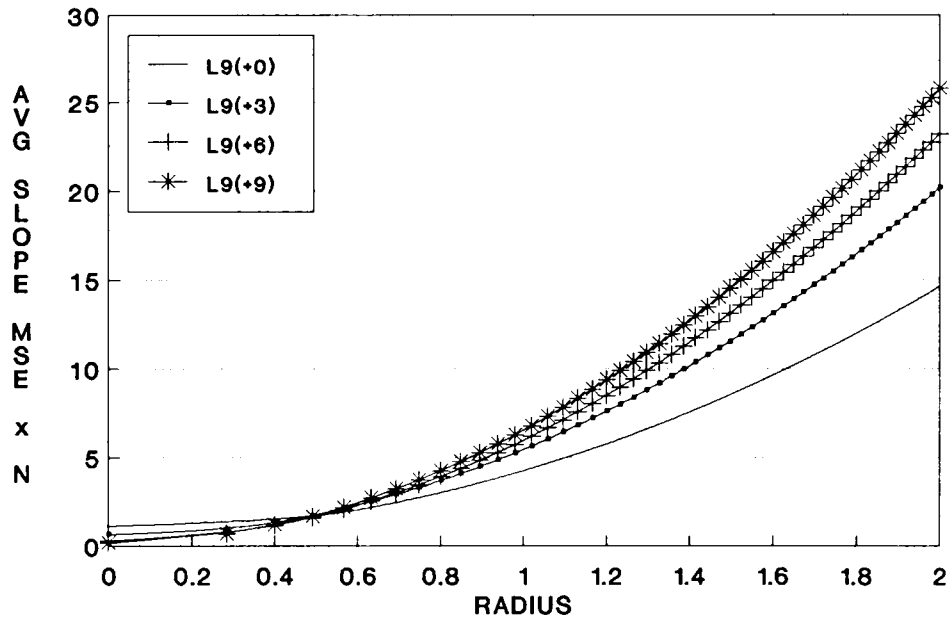

FIGURE 3.3.5: AVG SLOPE MSE COMPARISON
L9 ARRAYS AUGMENTED BY TR(L) OPTIMALITY
PROTECT AGAINST INTERACTIONS

CHAPTER 4

A STUDY OF ROBUSTNESS IN SLOPE ESTIMATION

Recommendations have been stated regarding designs for precise slope estimation. It is important to study the performance of such designs in other than ideal conditions. This chapter addresses the presence of outliers in the data, and the problem of missing observations. The first problem results in biased slope estimation, while the second changes the variance of $\hat{\underline{s}}$.

4.1 ROBUSTNESS TO OUTLIERS IN THE DATA

There are sources of bias in slope estimation other than model misspecification. One such source is the presence of outliers in the data. In this section the slope bias measures of Section 3.2 are applied to the mean shift outlier model to develop maximum and average measures of slope bias due to outliers. These measures are then used in an evaluation of the response surface designs recommended in Section 2.7 for precise slope estimation, in their robustness to outliers.

4.1.1 THE MEAN SHIFT OUTLIER MODEL FOR SLOPE ESTIMATION

Consider the Nx1 response vector $\underline{y}$. Suppose there are p_2 suspected outliers among the N responses. Rearrange the elements of $\underline{y}$ so that the outliers are the last p_2 elements of the vector. Correspondingly, rearrange the fitted model matrix X_1 so as to

maintain the proper associations between response and factor levels. That is, fit the model

$$\underline{y} = X_1\underline{\beta}_1 + \underline{\epsilon},$$

where $\underline{y} = \begin{bmatrix} \underline{y}_{-I} \\ \underline{y}_I \end{bmatrix}$ $(n-p_2)x1$ vector of non-outliers,
$(p_2 \times 1)$ vector of outliers,

and $X_1 = \begin{bmatrix} X_{1,-I} \\ X_{1,I} \end{bmatrix}$ $(n-p_2)xp$: corresponding to non-outliers
$(p_2 \times p)$: corresponding to outliers.

If there are indeed p_2 mean shifts or outliers in the data, then the true model form can be written as

$$\underline{y} = X_1\underline{\beta}_1 + \underline{X}_2\beta_2 + \underline{\epsilon},$$

where $\underline{X}_2$ is an Nx1 column vector of zeros and ones, with unity occupying the last p_2 positions. β_2 is a scalor representing the size of the mean shift at those points. The slope bias due to outliers will be developed for the second order fitted model with a single outlier in the data. Thus,

$$\underline{X}_2' = [\ 0\ \ 0\ \ 0\ \ .\ .\ .\ 0\ \ 1\ ].$$

The slope parallel to the x_j axis is $s_j = \underline{l}_{j1}'\underline{\beta}_1$ but the estimate of $\underline{\beta}_1$ will be biased.

$$\begin{aligned} E(\hat{s}_j) &= \underline{l}_{j1}'(X_1'X_1)^{-1}X_1'E(\hat{\underline{y}}) \\ &= \underline{l}_{j1}'(X_1'X_1)^{-1}X_1'[X_1\underline{\beta}_1 + \underline{X}_2\beta_2] \\ &= \underline{l}_{j1}'[\underline{\beta}_1 + \underline{A}\beta_2] \end{aligned}$$

where $\underline{A} = (X_1'X_1)^{-1}X_1'\underline{X}_2$.

The slope estimation bias is then

$$E(\hat{s}_j) - E(s_j) = \underline{l}_{j1}'\underline{A}\beta_2,$$

and the squared slope bias due to an outlier at a point x_i in the design D, is

$$B^*_{j,out,i}(\underline{x}_0) = \beta_2^2[\underline{A}_i'\underline{l}_{j1}]^2.$$

For standard response surface designs and least squares estimation, the oulier biases in slopes parallel to the k factor axes are equal. So without loss of generality, the measures will be developed for one slope and the subscript j will be dropped from the notation.

Note that an outlier could potentially occur at any one of the N design points (thus the subscript i on $\underline{A}$). We will consider two approaches to dealing with this. First, if the experimenter wishes to provide himself with overall protection from outlier bias when an outlier at any point is equally likely, he could pick a design with small average outlier bias, where the average is taken over the design points. The second is a minimax approach, where the design is chosen with small maximum bias, where the maximum is over the design points. In the first case, we are providing protection against bias from an outlier at any design point, whereas in the second case, we are protecting against the worst case bias. In either case however, we still need to maximize and average the measure over the surface of spheres of varying radii to generate a plot describing outlier bias in slope estimation across the region.

4.1.2 MAXIMIZING AND AVERAGING OUTLIER BIAS ON SPHERES

In developing measures of oulier bias in slope estimation, several cases could be considered, but they are not all unique. When previously dealing with bias, we either averaged or maximized bias over the sphere $|\underline{\beta}_2|=1$ in the $\underline{\beta}_2$ space, which means we plotted the bias in units of $|\underline{\beta}_2|$. With outlier bias, the two cases are identical since β_2 is a scalor, and no averaging or maximizing is needed in the β_2 space. We also permitted the possibility that the bias in slopes parallel to the k factor axes may differ, and we considered the maximum of the k biases. Since these measures of outlier bias are to be used on standard RS designs, this is no longer necessary since the k biases are equal. Two items remain for consideration: averaging versus maximizing over the design points, and averaging versus maximizing over spheres of radius r in the design space.

In order to develop a plot analogous to the slope variance envelope, we will first average the slope outlier bias over the design points.

$$B^{*}_{out,avg}(r) = \underset{\underline{x}_i \in D}{\text{avg}} \; B^{*}_{out,i}(r) \text{ such that } |\beta_2| = 1$$

$$= \underset{\underline{x}_i \in D}{\text{avg}} \; [\underline{A}_i' \underline{l}_1]^2$$

$$= (1/N) \sum_{i=1}^{N} [\underline{A}_i' \underline{l}_1]^2$$

Then the optimum functional values on spheres are located with Vining's algorithm, and a slope outlier bias envelope results.

$$\max_{\underline{x}_0'\underline{x}_0=r^2} B^*_{out,avg}(r) = \max_{\underline{x}_0'\underline{x}_0=r^2} (1/N) \sum_{i=1}^{N} [\underline{A}_i'\underline{l}_1]^2$$

$$\min_{\underline{x}_0'\underline{x}_0=r^2} B^*_{out,avg}(r) = \min_{\underline{x}_0'\underline{x}_0=r^2} (1/N) \sum_{i=1}^{N} [\underline{A}_i'\underline{l}_1]^2 .$$

Also on this graph is plotted the average taken over the surface of the sphere.

$$\operatorname{avg}_{\underline{x}_0'\underline{x}_0=r^2} B^*_{j,out,avg}(r) = K \int_{U_r} (1/N) \sum_{i=1}^{N} [\underline{A}_i'\underline{l}_{j1}]^2$$

$$= (1/N) \sum_{i=1}^{N} [a^2_{(j+1),i} + (r^2/k)(4a^2_{(k+1+j),i} + \sum_{m} a^2_{m,i}]$$

where $a_{z,i}$ denotes the z^{th} element of the $\underline{A}_i$ alias vector resulting from an outlier in position $\underline{x}_i$. The subscript j is included for clarity in the calculation, but the average bias will have the same value for all k values of j. $\sum_{m}$ denotes the sum over all m positions in the $\underline{A}_i$ alias vector corresponding to an interaction term involving x_j. (e.g. if j=1 then $\sum_{m} = \sum_{m=8}^{9}$; if j=3 then $\sum_{m} = \sum_{m=9}^{10}$, etc.)

Picking a design with an attractive slope outlier bias envelope constructed as such, is choosing that which provides adequate protection from outlier bias when an outlier is equally likely to occur at any of the design points. When we include the factor of N for comparing design of differing sizes, we see that here we are interested in minimizing the total slope estimation bias of all possible outlier positions in the data. Figures 4.1.1 through 4.1.4 evaluate the four designs recommended for precise slope estimation using this slope outlier bias envelope measure.

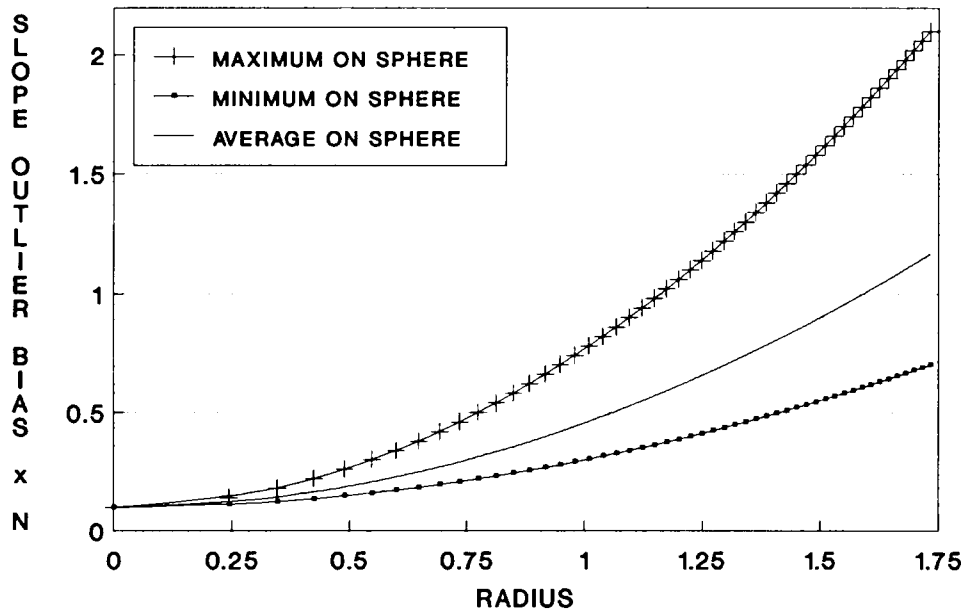

FIGURE 4.1.1: BIAS DUE TO AN OUTLIER
HYBRID 311B, 1 C.R., N=11
SLOPE BIAS AVERAGED OVER DESIGN POINTS

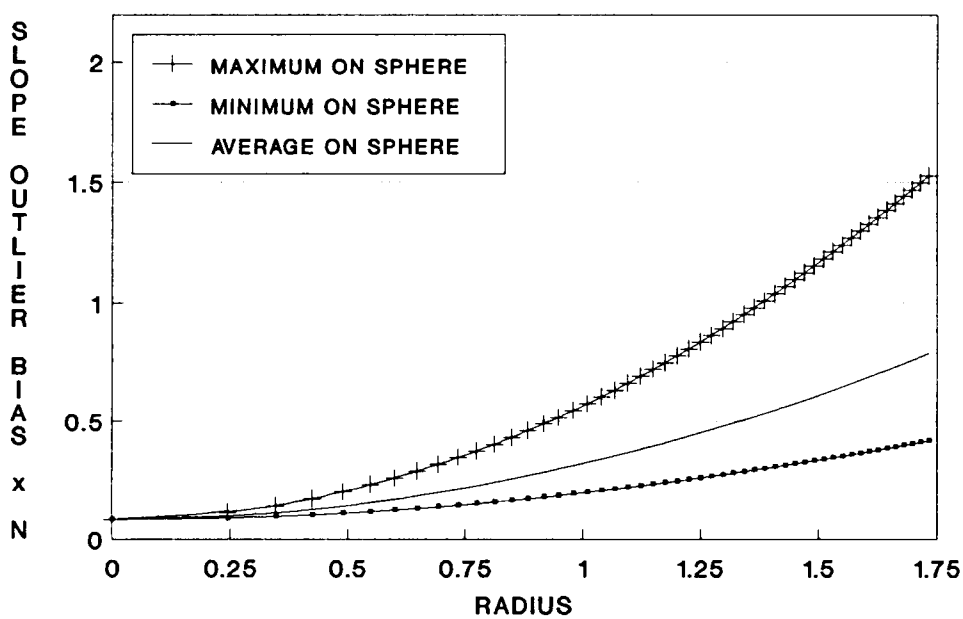

FIGURE 4.1.2: BIAS DUE TO AN OUTLIER
BOX-BEHNKEN DSN FOR K=3 WITH 3 C.R.s
SLOPE BIAS AVERAGED OVER DESIGN POINTS

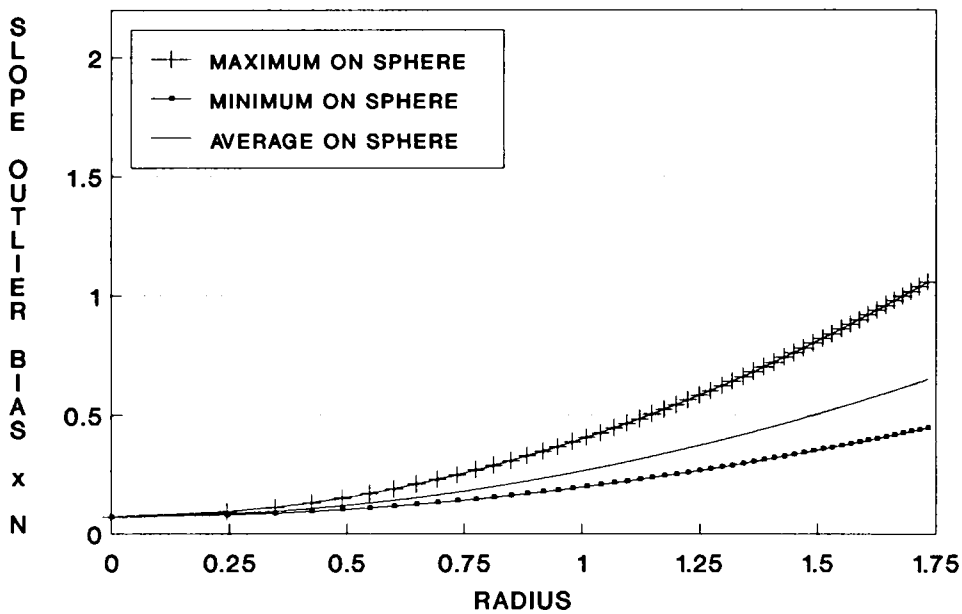

FIGURE 4.1.3: BIAS DUE TO AN OUTLIER
CCD FOR K=3, A=1.732, 3 C.R.s, Na=1
SLOPE BIAS AVERAGED OVER DESIGN POINTS

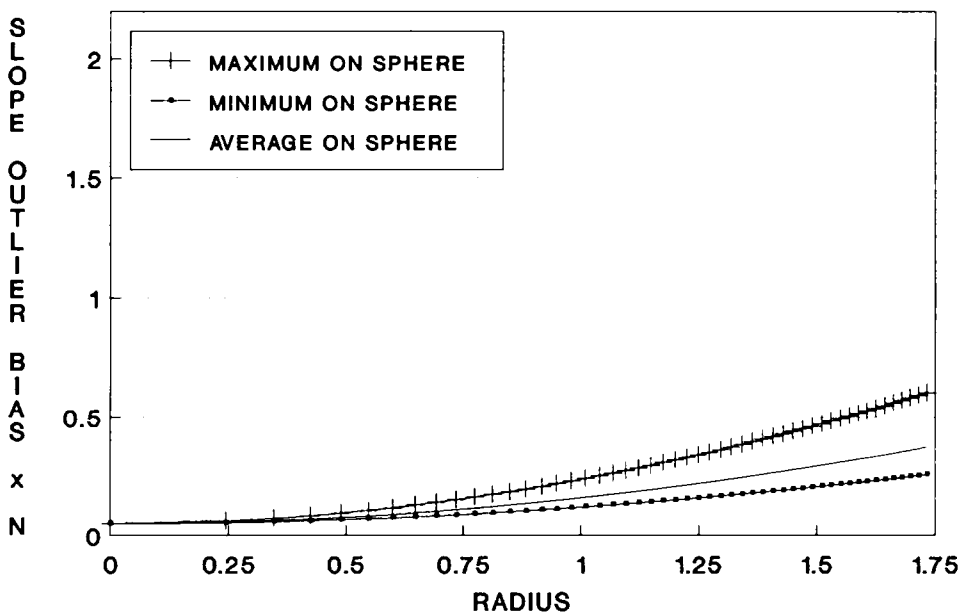

FIGURE 4.1.4: BIAS DUE TO AN OUTLIER
DUAL AXIAL CCD FOR K=3
SLOPE BIAS AVERAGED OVER DESIGN POINTS

The design most robust to overall outlier bias in slope estimation is the previously recommended dual axial CCD (Figure 4.1.4). The design least robust is the Hybrid 311B (Figure 4.1.1). The CCD and BBD rank in the middle. Note that the ranking of robustness matches the sizes of the designs (largest is most robust), but this is merely a coincidence. For example, using the CCD or the dual axial CCD, a large bias results in the slope s_j estimate when the outlier is at an axial point in x_j (an outer axial for the dual axial CCD). This is because these points are used to estimate the pure quadratic effects, and the bias in the pure quadratic coefficients is heavily weighted in this measure of slope estimation bias. The biases resulting from outliers at other points are quite small however, so the designs evaluated using the average measures look robust.

Using the BBD, the worst place for an outlier to occur when estimating s_j is at a "factorial" point where $x_j=\pm 1$. There are eight such points, so even though the bias at each is not very large, the sum of the eight results in a large overall bias measure. When using the Hybrid 311B, the largest bias ocurrs at a center run, but half of the remaining points contribute a large amount of bias as well, so the overall robustness is poor by this measure.

A true minimax approach to outlier bias involves choosing the design with the smallest "worst case" outlier bias in slope estimation.

$$\max_{\underline{x}_0'\underline{x}_0=r^2} B^*_{out,max}(r) = \max_{\underline{x}_0'\underline{x}_0=r^2} \max_{\underline{x}_i\in D} B^*_{out}(\underline{x}_0) \text{ such that } |\beta_2|=1$$

$$= \max_{\underline{x}_0'\underline{x}_0=r^2} \max_{\underline{x}_i \in D} [\underline{A}_i'\underline{l}_1]^2.$$

Figure 4.1.5 illustrates the result of evaluating the four RS designs using this minimax approach. Clearly the ranking of robustness has changed. This measure plots the worst amount of bias possible at a distance r from the design center, if the outlier occurs at the worst possible design point. Now the BBD and Hybrid 311B rank better than the CCD or the dual axial CCD.

4.1.3 CONCLUSIONS CONCERNING ROBUSTNESS TO OUTLIERS

Since the two approaches rank the four designs differently in robustness to outliers, it is difficult to choose the RS design best for all situations. The situation must dictate which of the two approaches to use. For example, if the experimenter knows which portion of the region is likely to contain outliers, a design can be recommended based on the minimax approach, where the maximum bias is taken over the design points located in that portion of the design region. If however, the experimenter expects outliers to occur at any location with equal probability, the bias averaged over design points is more appropriate, and the dual axial CCD or the single axial CCD should be used.

It is important when using the CCD or the dual axial CCD, to be extra careful in executing the axial point block. With the Hybrid 311B or the BBD however, a single outlier is not nearly as capable of badly biasing the slope estimation results. The robustness of these

four designs will be studied further in the next section, in relation to missing observations.

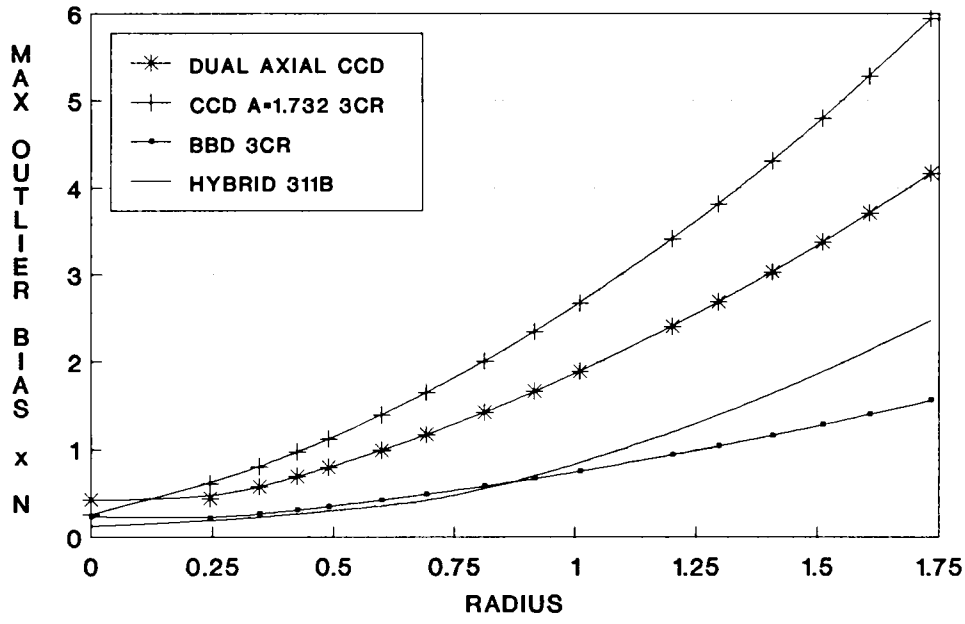

FIGURE 4.1.5: BIAS DUE TO OUTLIERS
A COMPARISON: CCD, H311B, BBD, DUAL CCD
SLOPE BIAS MAXIMIZED OVER DESIGN POINTS

4.2 ROBUSTNESS TO MISSING OBSERVATIONS

Sections 2.1 and 2.2 developed a general measure of variance in the estimated slope vector. The measure is applicable to any response surface design, not just those which provide equally precise estimates of slopes parallel to factor axes. The general measure will be used in this section to examine the robustness of those designs recommended in Section 2.7 to missing observations.

For comparison of designs of differing sizes, $\lambda_L^*(V)$ as defined in (2.2.1), includes a factor of N. In order to compare designs across design classes, in terms of robustness to missing observations, define the missing data loss for a given design, relative to the best full design in its class, as

$$MDL(DSN1) = \frac{\lambda_L^*[V(rmax),DSN1] - \lambda_L^*[V(rmax),FULL\ DSN]}{\lambda_L^*[V(rmax),FULL\ DSN]} .$$

The index considers $\lambda_L^*[V(r)]$ at the perimeter of each design (rmax=radius of design) since this is where the designs differ the most in $\lambda_L^*(V)$. Note that the scaling of designs to a radius of $\sqrt{k}$ is not necessary with the use of this index. While $\lambda_L^*(V)$, maximized on spheres, will be used to compare designs within a class for robustness, the MDL for a design is used when comparing across design classes. The design with the smaller MDL will be more robust.

4.2.1 ROBUSTNESS OF THE CCD TO MISSING OBSERVATIONS

The design matrix for the central composite design is given by (1.6.1) and the parameter values chosen for precise slope estimation in

three variables, are $\alpha=\sqrt{3}$, $n_0=3$, and $n_a=1$. To verify that the same parameter choices result from the current, more general slope variance measure, see Figures 2.2.1 and 4.2.1. The first graph is the general slope variance envelope for the recommended CCD, and Figure 4.2.1 gives the plots of $\lambda_L^*(V)$ maximized over spheres in the design space for several other CCDs. The recommendations made in Section 2.5 are strengthened here. The investigation of robustness to missing observations is conducted relative to the full CCD depicted in Figure 2.2.1.

There are three different kinds of points in a CCD: center runs, axial points, and factorial points. This investigation considers dropping one or more observations of each kind, and measuring the increase in slope estimation variance, relative to the full CCD, by plotting $\max_{\underline{x}'\underline{x}=r^2} \lambda_L^*(V)$ for the competing designs on the same graph.

Note that by comparing Figure 2.2.1 to 4.2.1, we see that dropping two center runs does not result in much deterioration in slope estimation precision. Figure 4.2.2 investigates dropping either a factorial or an axial point from a CCD. Dropping any one axial point results in the same maximum $\lambda_L^*(V)$, as dropping any one of the eight factorial points gives the same $\lambda_L^*(V)$ plot. Clearly, one missing axial point is not nearly as damaging as a missing factorial point.

Figure 4.2.3 illustrates the results of dropping up to four factorial points. The four chosen for deletion are those in the full CCD not present in a small composite design (Hartley, 1959). Label

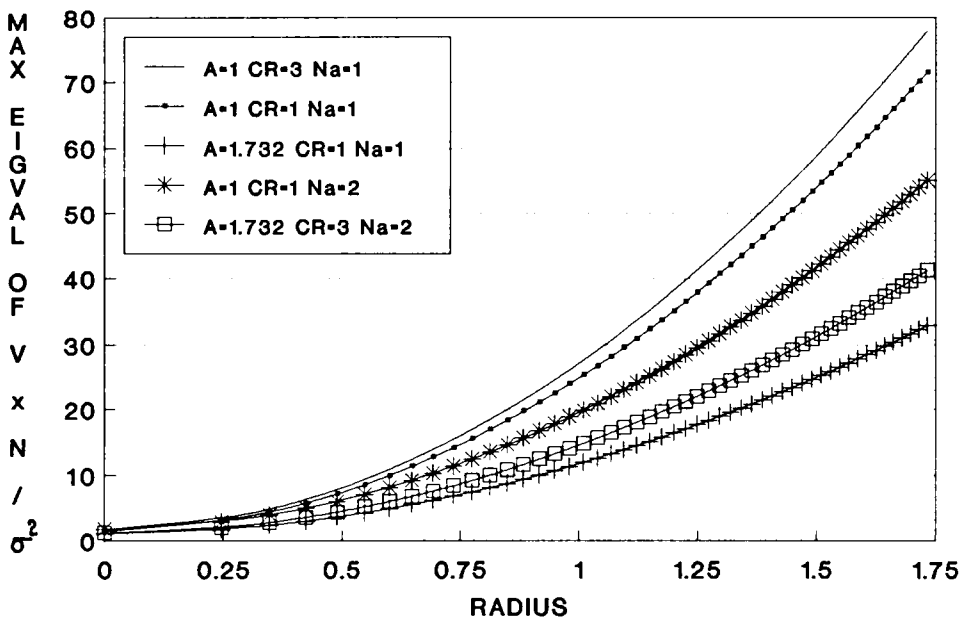

FIGURE 4.2.1: PARAMETERS IN A CCD
USING THE GENERAL SLOPE VARIANCE MEASURE
MAXIMA ON SPHERES

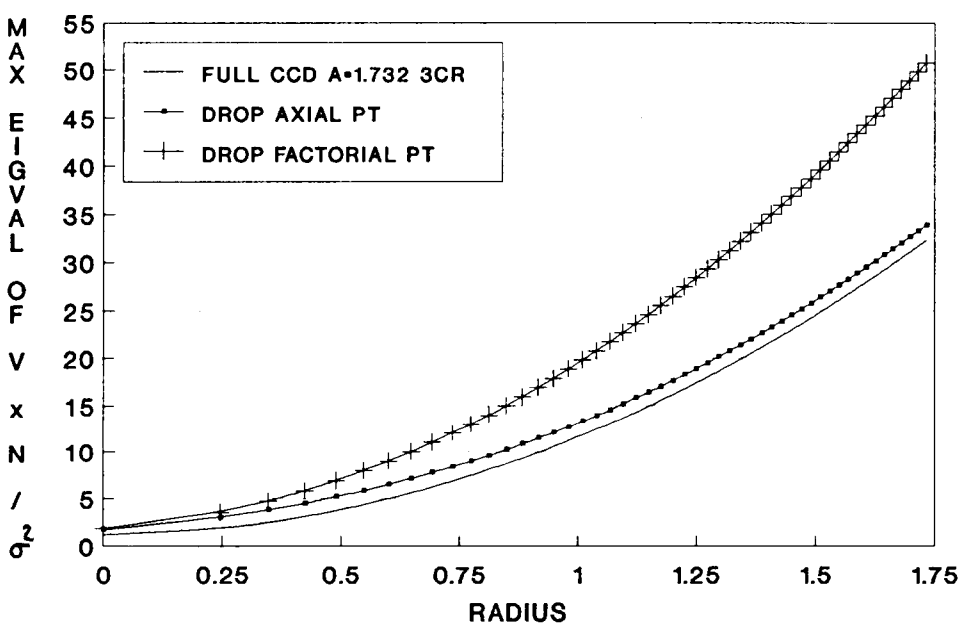

FIGURE 4.2.2: MISSING ONE POINT IN A CCD
DROPPING AXIALS VS. DROPPING FACTORIALS
(-A, 0, 0) VS. (-1, -1, -1)

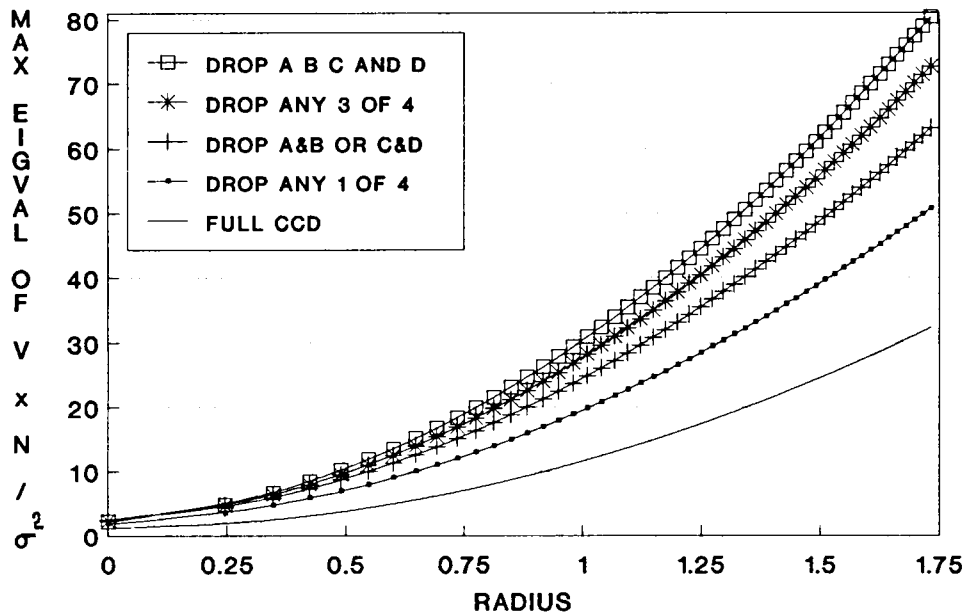

FIGURE 4.2.3
MISSING DATA IN A CCD
DROPPING UP TO FOUR FACTORIAL POINTS

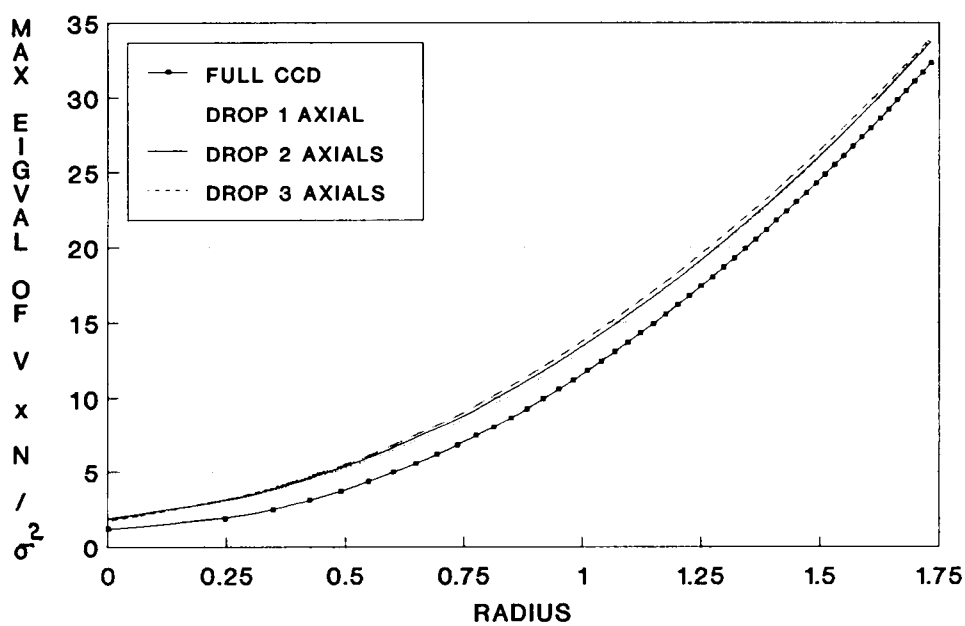

FIGURE 4.2.4
MISSING DATA IN A CCD
DROPPING UP TO THREE AXIAL POINTS

those dropped as

A: (-1, -1, 1)
B: (-1, 1, -1)
C: (1, -1, -1)
and D: (1, 1, 1).

As stated, dropping any one of these will yield the same resulting maximum $\lambda_L^*(V)$ plot. Similarly, there is a unique result for dropping any three of the four. When only two are dropped at a time, there are two unique results: one if either A and B or C and D are dropped, and another if A and D or B and C are dropped. The results are quite close, but the worst case is plotted in Figure 4.2.3. The ability of the CCD to precisely locate the optimum is diminished when there are missing factorial points.

Lost axial points, on the other hand, are not as devastating. Figure 4.2.4 depicts the maximum $\lambda_L^*(V)$ results for CCDs with up to three axial points missing. Note that two axial points in the same variable cannot be dropped simultaneously and maintain estimability of the quadratic coefficients in the model. For example, if $(-\alpha, 0, 0)$ is missing, then $(\alpha, 0, 0)$ must be present for model estimation.

To lose two center runs and/or up to three axial points is better than to lose a single factorial point. This is understandable and perhaps even expected, when the form of V is considered. Each element of V, v_{ij}, is highly dependent on the precision of the estimated interaction coefficients, and they are estimated through the factorial portion of the design.

For purposes of future comparisons, the MDLs have been calculated for several CCDs with missing observations. They are tabulated in Table

4.2 at the end of this section, as the percentage loss in slope estimation precision.

4.2.2 ROBUSTNESS OF THE BBD TO MISSING OBSERVATIONS

The design matrix for the three variable Box-Behken design is given in (1.6.2). Three center runs were since determined as optimal. That is the design evaluated for robustness to missing observations. The general slope variance envelope for the full BBD is shown in Figure 2.2.2. It can be shown that dropping two of the three center runs does not have much of an effect on the plot. Note that it is not scaled to a radius of $\sqrt{k}$, since the MDL is used for comparisons across design classes.

In the investigation, up to three points will be dropped from the full BBD, and $\lambda_L^*(V)$ will be plotted to determine robustness. Consider the design as center runs and the collection of three "blocks" of four points each: the first four listed in (1.6.2), The second set of four in (1.6.2), and the third set of four listed. Up to one point from each "block" will be dropped. More than one point missing from a "block" could make coefficients inestimable.

The results of dropping one, two, and three points from the full BBD are displayed in Figure 4.2.5. Dropping any single point (of the twelve) yields the same plot of $\lambda_L^*(V)$ maximized on spheres. The loss of a single point appears quite damaging to slope estimation. Dropping a second point yields two distinct, but close, results, the worst of which is plotted in Figure 4.2.5. Three missing points is not much

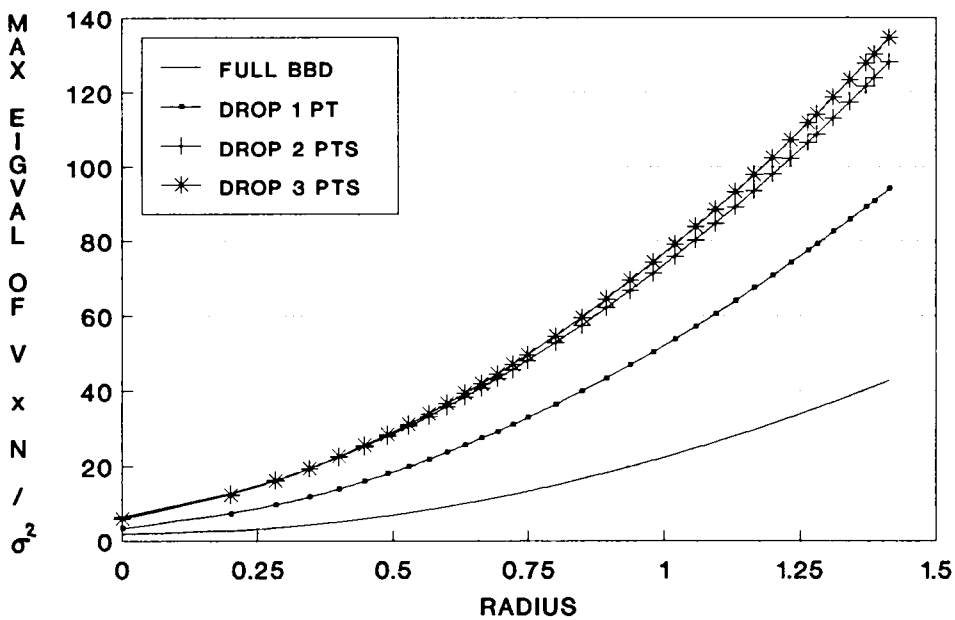

FIGURE 4.2.5: MISSING DATA IN A BBD
DROPPING UP TO THREE POINTS
MAXIMA ON SPHERES

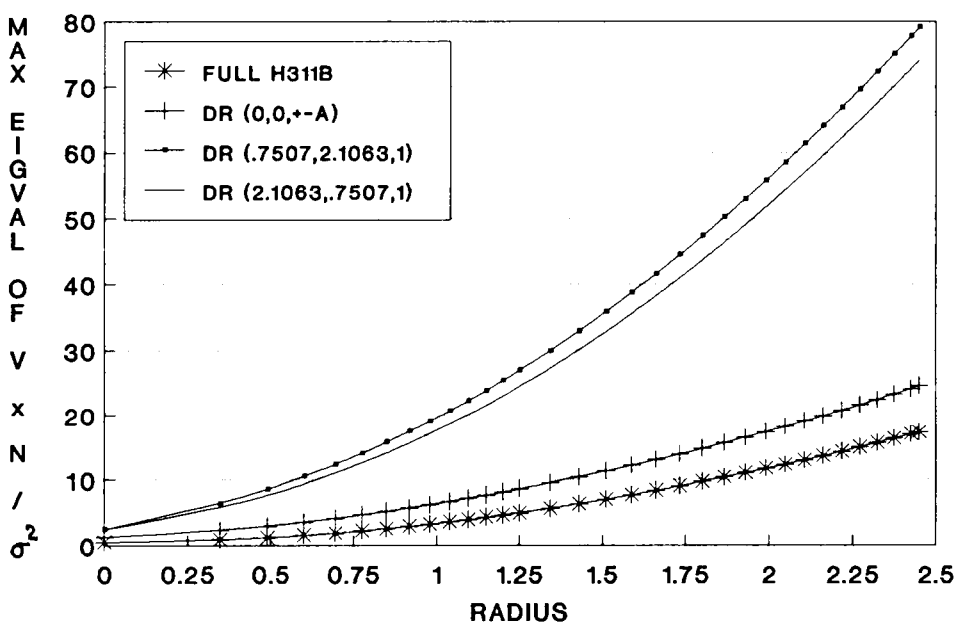

FIGURE 4.2.6
DROPPING ONE POINT FROM A HYBRID 311B
3 UNIQUE RESULTS IN MAXIMA ON SPHERES

worse than dropping two observations, and yields four distinct, yet close, results, the worst of which is shown on the graph.

The missing data relative indices for the BBD are given in Table 4.2 for comparison of robustness to missing data of the BBDs to other classes of designs. It appears that the class of BBDs is not as robust to lost points as the class of CCDs.

4.2.3 ROBUSTNESS OF THE HYBRID 311B TO MISSING OBSERVATIONS

The hybrid 331B design to be evaluated using $\lambda_L^*(V)$ plots for robustness to missing data is given in (1.6.3c). Note that only one point can be dropped in order for all terms in the model to be estimable, and that the center run cannot be lost. Figure 2.2.3 depicts the general slope variance envelope for the full hybrid 311B. Keep in mind that the designs in this chapter are not scaled to lie on a radius of $\sqrt{k}$ since the MDL measure used to compare designs across design classes is invariant to scale.

Dropping one point from the design yields three unique results: one for the "axials", or the first two points given in (1.6.3c); one for the points where $(x_1\ x_2\ x_3) = (\pm.7507, \pm2.1063, \pm1)$; and one for the points where $(x_1\ x_2\ x_3) = (\pm2.1063, \pm.7507, \pm1)$. Figure 4.2.6 illustrates each of the three cases relative to the full hybrid 311B. Dropping either the first or second design point is damaging to slope estimation, but not nearly as devastating as dropping any other point in the design. The missing data relative indices are listed in Table 4.2. Dropping an "axial" from the hybrid 311B is not as bad as dropping

a point from a BBD, nor as bad as a missing factorial in a CCD, but dropping one of the other points from the 311B is worse than any lost point in either of the other classes.

4.2.4 ROBUSTNESS OF THE DUAL AXIAL CCD TO MISSING OBSERVATIONS

The design matrix for the dual axial CCD is given by (1.4.3). The general slope variance envelope for this design is given in Figure 2.2.4. Keep in mind that $\lambda_L^*(V)$ is not scaled for comparison across design classes.

There are many ways to drop points from the dual axial CCD, since it contains many more points than there are parameters in the model. A few of the interesting cases are considered here to further illustrate the usefulness of the general spherical slope variance measures, and to give an overall idea of the robustness of the dual axial CCD, but a detailed study is not included for the sake of brevity.

There are three kinds of points in the dual axial CCD: the factorial points, the inner axial points, and the outer axial points. Figure 4.2.7 contains the results of dropping one point of each of the three types. It is interesting, yet not surprising, to note that the inner axial points do not contribute much to precise slope estimation, so dropping one inner axial point actually improves this measure of efficiency in precise slope estimation. Dropping one outer axial point causes slight damage to slope estimation precision, but the loss of a factorial point is much more troublesome. This is similar to the single axial CCD case.

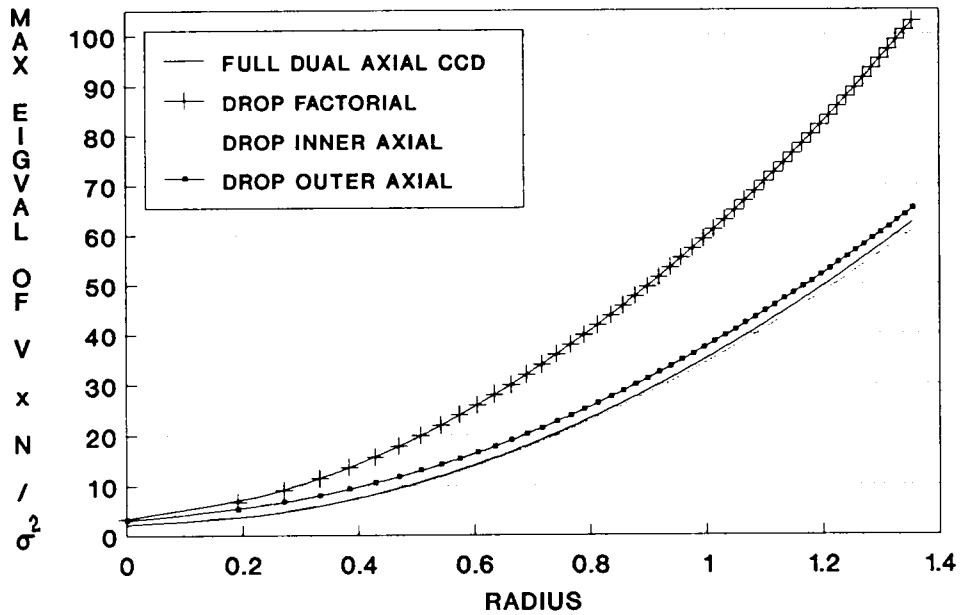

FIGURE 4.2.7
DROPPING ONE POINT FROM A DUAL AXIAL CCD
MAXIMA ON SPHERES

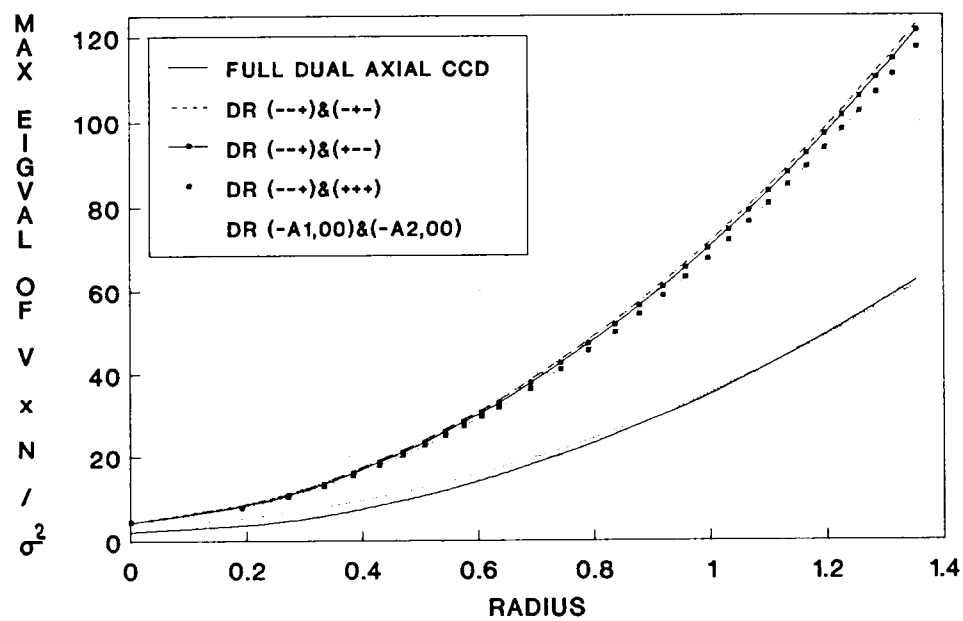

FIGURE 4.2.8
DROPPING 2 POINTS FROM A DUAL AXIAL CCD
FOUR UNIQUE RESULTS OF MAXIMA ON SPHERES

Figure 4.2.8 investigates dropping two points at a time. Dropping any two factorial points brings marked reduction in slope estimation precision. The three distinct cases are plotted in Figure 4.2.8. Dropping one inner axial point with the corresponding outer axial point still provides increased efficiency in slope estimation. The inner point is not necessary and by removing it, we offset the reduction in efficiency attained through the loss of the outer axial point. Many more cases could be considered, but we can say that the loss of axial points may or may not reduce slope estimaiton variance, while missing factorial points worsen slope estimation precision considerably. Some "missing data loss" indices are given in Table 4.2 for comparison to other design classes.

TABLE 4.2

MISSING DATA LOSS INDICES

DESIGN	MDL
CCD - 2 center runs	2.00%
CCD - 1 axial point	4.81%
CCD - 2 axial points	4.40%
CCD - 3 axial points	5.23%
CCD - 1 factorial pt	57.16%
CCD - 2 factorial pts	95.59%
CCD - 3 factorial pts	124.85%
CCD - 4 factorial pts	148.57%
BBD - 1 point	119.90%
BBD - 2 points	199.55%
BBD - 3 points	215.01%
H311B - "axial point"	41.16%
H311B - pt with $x_1=\pm.7507$	353.89%
H311B - pt with $x_1=\pm2.1063$	325.02%
DCCD - 1 inner axial pt	-2.51%
DCCD - 1 outer axial pt	4.29%
DCCD - 1 factorial pt	64.47%
DCCD - 2 factorial pts	96.64%
DCCD - 2 corresponding axial pts	-1.56%

4.2.5 CONCLUSIONS CONCERNING ROBUSTNESS TO MISSING POINTS

The dual axial and single axial central composite designs are clearly the most robust designs in terms of missing data points. With these designs however, factorial points are more expensive to lose than axial points. The Box-Behnken design is not robust to missing points. This is evidenced by the fact that the loss of one point more than doubles $\overset{*}{\lambda}_L(V)$ at the perimeter.

If a point is lost from the nearly saturated hybrid 311B, the result is usually devastating. The design requires fewer points, but each point is indeed necessary to provide precise slope estimation.

CHAPTER 5

SUMMARY AND FUTURE RESEARCH AREAS

5.1 SUMMARY OF SIGNIFICANT CONTRIBUTIONS

5.1.1 PRECISE SLOPE ESTIMATION

When identifying an optimal set of operating conditions for a system is the goal of the response surface analysis, then slope estimation is more important than response prediction. A tool has been developed to measure the variance in the jointly estimated slope vector. The tool measures and depicts the probability of incorrectly placing a non-optimal point (location in the region) inside a $(1-\alpha)100\%$ confidence region around the location of the optimum. One then picks the design with the smallest probability of such a misclassification.

The criterion involves plotting the largest eigenvalue of the variance-covariance matrix of the estimated slope vector, weighted by the design size, N. The matrix is a function of both the design and of the location where one wishes to estimate the slope. The measure is maximized and minimized over the surfaces of concentric spheres centered at the design center, and these are plotted against the radii of the spheres. The result is a general slope variance envelope (GSVE). It is appropriate for use in the evaluation of any response surface (RS) design in its ability to precisely locate the optimum, regardless of balance in the design. Due to its wide scope of applicability, it is quite complex mathematically, and perhaps more complex than is necessary when evaluating standard RS designs, where the slopes

parallel to the k factor axes are estimated with equal precision.

For evaluating the standard RS designs, a simplified slope variance envelope for one slope was developed, which is analagous to the variance dispersion graph (VDG) of Giovannitti-Jensen and Myers. A detailed study of central composite designs (CCDs) was conducted to determine the optimal design parameters for precise slope estimation. It was found that the axial points should be placed on the same radius as the factorial points, that is, $\alpha=\sqrt{k}$; when $0\leq\alpha\leq 1$ it is better to replicate axial points than center runs; and when $1<\alpha$ replication efforts should be concentrated at the center and not the axial points. The opposite is true when considering precise response prediction. Many other designs were studied, and four were chosen as "best" for precise slope estimation:

1) the CCD with $\alpha=\sqrt{k}$ and k center runs,
2) the dual axial CCD,
3) the hybrid designs, and
4) the Box-Behnken designs with k center runs.

These designs were investigated further in a robustness study.

5.1.2 BIAS AND MSE IN SLOPE ESTIMATION

A measure of squared bias in slope estimation was developed, which is analagous to that developed by Vining for response prediction. It measures the maximum bias of the k slope estimates, then depicts it in a picture through maximizing and averaging it over the surfaces of concentric spheres centered at the design center, and plotting the

results against the radii of the spheres. Combined with the measure of variance in one slope, a measure of MSE in slope estimation results. This was used to evaluate the protection from bias due to an underspecified model, obtained through augmenting saturated orthogonal arrays to maximize the detection of lack of fit. It turned out that points added to detect lack of fit do not efficiently provide protection from bias due to an underspecified model.

5.1.3 ROBUSTNESS TO OUTLIERS IN THE DATA

The four designs recommended for precise slope estimation in the k=3 case were evaluated in their robustness to outliers. The squared bias measure was applied to the mean shift outlier model to define a measure of bias due to outliers. Two approaches were taken to deal with the fact that an outlier could occur at any one of the design points. The first assumes that an outlier is equally likely to occur at any design point, and averages the resulting bias over the design points. The second is a minimax approach, where the criterion measures the bias occurring at the worst possible location in the design. The CCD and dual axial CCD are most robust to outliers, noting that extreme care must be taken when executing the axial points in either design.

The tools developed for conducting this study are best for use in such studies involving standard RS designs, where the bias in slopes parallel to the k factor axes are equal, and the dependence among the slopes is not large. The measure of bias is limited in that it does not consider the dependence among the slopes.

5.1.4 ROBUSTNESS TO MISSING OBSERVATIONS

The general slope variance envelope developed to measure the variance in the vector of jointly estimated slopes was used in a study of the four recommended designs in their robustness to missing data. The tool is extremely useful, in that it is appropriate for use in evaluating any RS design, whether or not it is balanced in the design space. Again, the CCD and the dual axial CCD are recommended. Each design studied contains a few different "types" of points (e.g. factorial, axial, and center points in a CCD). At least one of each type was dropped from each design, and a summary plot was given to depict the importance each has with respect to precise slope estimation. Some points can be lost without the variance results changing drastically, while the loss of other points yields devastating variance results. It is interesting to note that in the case of the CCD, if an axial point is a suspected outlier, it may be better simply to drop it from the analysis. It can cause extreme outlier bias, yet the loss in slope estimation precision is not great.

5.2 FUTURE RESEARCH AREAS

The slope variance envelopes, either in general or for one slope, are useful as tools for comparing designs. Unfortunately, a single plot is not very meaningful. It would be useful to develop an analogue to G-efficiency for application to slope estimation, so that a single plot could be compared to a theoretical optimum value for slope variance.

It is planned to evaluate design performance in a cuboidal region of interest.

The general slope variance envelope is related to E-optimality in slope estimation. One seeks to minimize the maximum eigenvalue of the variance-covariance matrix of the slope vector. Further development of the E-optimality idea as applied to slope estimation could lead to designs constructed to be slope-E-optimal.

A measure of variance in the jointly estimated slope vector has been developed, but bias has been measured in a single slope, ignoring the dependence among slopes. It would be useful to develop a measure of bias in the vector of jointly estimated slopes. Once completed, it could be combined with the general slope variance into a measure of MSE in the vector of estimated slopes.

REFERENCES

Addelman, S. and Kempthorne, O. (1961). "Some Main-Effect Plans and Orthogonal Arrays of Strength Two," *Annals of Mathematical Statistics*, 32, 1167-1176.

Addelman, S. (1962a). "Orthogonal Main-Effect Plans for Asymmetrical Factorial Experiments," *Technometrics*, 4, 21-46.

_____ (1962b). "Symmetrical and Asymmetrical Fractional Factorial Plans," *Technometrics*, 4, 47-58.

_____ (1972). "Recent Developments in the Design of Factorial Experiments," *Journal of the American Statistical Association*, 67-337, 103-111.

Atkinson, A. C. (1970). "The Design of Experiments to Estimate the Slope of a Response Surface," *Biometrika*, 57, 319-328.

Atwood, C. L. (1969). "Optimal and Efficient Designs of Experiments", *Annals of Mathematical Statistics*, 40, 1570-1602.

Box, G. E. P. and Behnken, D. W. (1960). "Some New Three-Level Designs for the Study of Quantitative Variables," *Technometrics*, 2, 455-475.

Box, G. E. P. and Draper, N. R. (1959). "A Basis for the Selection of a Response Surface Design," *Journal of the American Statistical Association*, 54, 622-652.

_____ (1963). "The Choice of a Second Order Rotatable Design," *Biometrika*, 50, 335-352.

_____ (1987). *Empirical Model-Building and Response Surfaces*, New York: John Wiley and Sons, Inc.

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, 190-199.

_____ (1957). "Multifactor Designs for Exploring Resonse Surfaces," *Annals of Mathematical Statistics*, 28, 195-241.

Box, G. E. P., Hunter, W. G. and Hunter, J. S. (1978). *Statistics for Experiments*, New York: John Wiley & Sons.

Box, G. E. P. and Wilson, K. B. (1951). "On the Experimental Attainment of Optimum Conditions," *Journal of the Royal Statistical Society*, B-13, 1-45.

Box, M. J. and Draper, N. R. (1971). "Factorial Designs, The $|X'X|$ Criterion, and Some Related Matters," *Technometrics*, 13, 731-742.

_____ (1974). "On Minimum-Point Second-Order Designs," *Technometrics*, 16-4, 613-616.

Carter, W. H. Jr., Chinchilli, V. M., Myers, R.H. and Campbell, E. D. (1986). "Confidence Intervals and an Improved Ridge Analysis of Response Surfaces," *Technometrics*, 28-4, 339-346.

Draper, N. R. and Guttman, I. (1988). "An Index of Rotatability," *Technometrics*, 30-1, 105-112.

Fisher, R. A. (1945). "A System of Confounding for Factors with More Than Two Alternatives Giving Completely Orthogonal Cubes and Higher Powers," *Annals of Eugenics*, 12, 283-290.

Giovannitti-Jensen, A. (1987). "Graphical Assessment of the Prediction Capability of Response Surface Designs," unpublished doctoral dissertation, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.

Giovannitti-Jensen, A. and Myers, R. H. (to appear May, 1989). "Graphical Comparison of Response Surface Designs", *Technometrics*.

Graybill, F. A. (1976). *Theory and Application of the Linear Model*, Belmont, CA: Wadsworth Publishing Co., Inc.

Hader, R. J. and Park, S. H. (1978). "Slope-Rotatable Central Composite Designs," *Technometrics*, 20-4, 413-417.

Hartley, H. O. (1959). "Smallest Composite Designs for Quadratic Response Surfaces", *Biometrics*, 15, 611-624.

Hockman, K. K. and Myers, R. H. (1988). "Graphical Comparison of Designs for Response Optimization Based on Slope Estimation", Department of Statistics, Virginia Polytechnic Institute and State University, Technical Report No. 88-23.

Hoerl, A. E. (1959). "Optimum Solutions of Many Variable Equations," *Chemical Engineering Progress*, 55, 69-78.

Hoke, A. T. (1974). "Economical Second-Order Designs Based on Irregular Fractions of the 3^n Factorial," *Technometrics*, 16-3, 375-384.

Hussey, J. R. (1983). "Effects of Correlation Induction Schemes on Variance Criteria in Experimentally Designed Computer Simulations", unpublished doctoral dissertation, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.

Hussey, J. R., Myers, R. H., and Houck, E. (1987). "Pseudorandom Number Assignment in Quadratic Response Surface Designs", *Institute of Industrial Engineering Transactions*, 19-4.

_____ (1988). "Correlated Simulation Experiments in First Order Response Surface Designs," *Operations Research*, 35-5, 744-758.

Kackar, R. N. (1982). "Some Orthogonal Arrays for Screening Designs," Technical Memorandum (unpublished but released), AT&T Bell Laboratories, Holmdel, NJ.

_____ (1985). "Off-Line Quality Control, Parameter Design, and the Taguchi Method," and discussions, *Journal of Quality Technology*, 17-4, 176-209.

Khuri, A. I. (1988). "A Measure of Rotatability for Response-Surface Designs," *Technometrics*, 30-1, 95-104.

Khuri, A. I. and Cornell, J. A. (1987). *Response Surfaces*, New York: Marcel-Dekker, Inc.

Kiefer, J. (1958). "On the Nonrandomized Optimality and Randomized Nonoptimality of Symmetrical Designs", *Annals of Mathematical Statistics*, 20, 675-699.

_____ (1959). "Optimal Experimental Designs", *Journal of Royal Statistical Society*, B, 21, 272-319.

_____ (1961). "Optimal Designs in Regression Problems,II", *Annals of Mathematical Statistics*, 32, 298-325.

_____ (1962). "Two More Criteria Equivalent to D-Optimality of Designs", *Annals of Mathematical Statistics*, 33, 792-796.

_____ (1974). "General Equivalence Theory for Optimal Designs (Approximate Theory)", *Annals of Statistics*, 2, 842-879.

Kiefer, J. and Wolfowitz, J. (1959). "Optimum Designs in Regression Problems," *Annals of Mathematical Statistics*, 30-2,271-294.

_____ (1960). "The Equivalence of Two Extremum Problems," *Canadian Journal of Mathematics*, 12, 363-366.

Lucas, J. M. (1976). "Which Response Surface Design is Best," *Technometrics*, 18-4, 411-417.

Mitchell, T. J. (1974). "An Algorithm for the Construction of 'D-Optimal' Experimental Designs," *Technometrics*, 16-2, 203-210.

_____ (1974). "Computer Construction of D-Optimal First-Order Designs," *Technometrics*, 16-2, 211-220.

Morris, M. P. and Mitchell, T. J. (1980). "Two-Level Multifactor Experiment Designs for Detecting the Presence of Interactions," Technical Report No. 2112, available from the Mathematics Research Center, University of Wisconsin-Madison, 610 Walnut St., Madison, WI 53706.

_____ (1983). "Two-Level Multifactor Designs for Detecting the Presence of Interactions," *Technometrics*, 25, 345-355.

Myers, R. H. (1976). *Response Surface Methodology*, Ann Arbor, MI: Edwards Brothers.

Myers, R. H., Khuri, A. I. and Carter, W. H. (to appear May, 1989). "Response Surface Methodology: Historical Review 1966-1988", *Technometrics*.

Myers, R. H. and Lahoda, S. J. (1975). "A Generalization of the Response Surface Mean Square Error Criterion with a Specific Application to the Slope," *Technometrics*, 17-4, 481-486.

Noble, B. and Daniel, J. W. (1977). *Applied Linear Algebra*, Englewood Cliffs, NJ: Prentice Hall, Inc.

Notz, W. (1982). "Minimum Point Second Order Designs," *Journal of Statistical Planning and Inference*, 6, 47-58.

Ott, L. and Mendenhall, W. (1972). "Designs for Estimating the Slope of a Second Order Linear Model", *Technometrics*, 14-2, 341-353.

Plackett, R. L. and Burman, J. P. (1946). "The Design of Optimum Factorial Experiments," *Biometrika*, 33, 305-325.

Rao, C. R. (1947). "Factorial Experiments Derivable from Combinatorial Arrangements of Arrays," *Journal of the Royal Statistical Society*, Supplement, 9, 128-139.

Roquemore, K. G. (1976). "Hybrid Designs for Quadratic Response Surfaces," *Technometrics*, 18, 419-423.

Seiden, E. (1954). "On the Problem of Construction of Orthogonal Arrays," *Annals of Mathematical Statistics*, 25, 151-156.

St. John, R. C. and Draper, N. R. (1975). "D-Optimality for Regression Designs: A Review," *Technometrics*, 17-1, 15-23.

Stroud, A. H. (1971). *Approximate Calculation of Multiple Integrals*, Englewood Cliffs, New Jersey: Prentice-Hall.

Taguchi, G. (1986). *Introduction to Quality Engineering*, Asian Productivity Organization, Tokyo, Japan.

Vining, G. G. (1988a). "A Graphical Approach for Evaluating the Potential Impact of Bias due to Model Misspecification in Response Surface Designs," unpublished doctoral dissertation, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.

_____ (1988b). "An Algorithm for Graphical Assessment of Experimental Designs", Department of Statistics, Virginia Polytechnic Institute and State University, Technical Report No. 88-x.

_____ (1988c). "An Analysis of the Performance of Some Three Variable Second Order Response Surface Designs in Terms of Prediction Bias", Department of Statistics, Virginia Polytechnic Institute and State University, Technical Report No. 88-17.

Wardrop, D. M. and Myers, R. H. (to appear, 1989). "Some Response Surface Designs for Finding Optimal Conditions," *Journal of Statistical Planning and Inference*.

APPENDIX A.0

Theorem 1.2.1: In the case of one independent variable (u) in a second order model, and symmetric experimental designs Di and Dj,

$$\text{if } \text{var}[\hat{s}(u),Di] > \text{var}[\hat{s}(u),Dj]$$

where u is an arbitrary location of the region, then the probability of placing the point u inside the $(1-\alpha)100\%$ confidence interval for u_0 using Di will be larger than that using Dj.

Proof: Fit the following model using the method of least squares.

$$\hat{y} = b_0 + b_1 u + b_2 u^2.$$

The estimated slope of the response curve at the optimum, u_0, is

$$\hat{s}(u_0) = \hat{s}_0 = b_1 + 2b_2 u_0.$$

$$E(\hat{s}_0) = \beta_1 + 2\beta_2 u_0 = 0, \quad \text{and}$$

$$\text{var}(\hat{s}_0) = \text{var}(b_1) + 4u_0^2 \text{var}(b_2),$$

since b_1 and b_2 are uncorrelated for symmetric designs. Hence,

$$\hat{s}_0 \sim N(0,\ \sigma_{b1}^2 + 4u_0^2 \sigma_{b2}^2).$$

It follows that,

$$\frac{[\hat{s}(u_0)]^2}{\text{var}[s(u_0)]} = \frac{(b_1 + 2b_2 u_0)^2}{\sigma^2(c^{11} + 4u_0^2 c^{22})} \sim \chi^2_1,$$

where c^{ii} are diagonal elements of

$$(U'U)^{-1} = \begin{bmatrix} c^{00} & c^{01} & c^{02} \\ & c^{11} & c^{12} \\ \text{sym.} & & c^{22} \end{bmatrix}$$

and $U=[\underline{1}\ \underline{u}\ \underline{u}^2]$ is the model matrix corresponding to the experimental design. For an arbitrary, fixed $u \neq u_0$,

$$\frac{[\hat{s}(u)]^2}{\text{var}[s(u)]} \sim \chi^{2'}_{1,\nu}, \quad \text{where } \nu = \frac{(\beta_1 + 2\beta_2 u)^2}{\sigma^2(c^{11} + 4u^2 c^{22})}$$

is the noncentrality parameter. To show that if $\text{var}[\hat{s}(u),Di] > \text{var}[\hat{s}(u),Dj]$, where D is the experimental design used to estimate the slope, then the $(1-\alpha)100\%$CI for u_0 using Di is "better" than that from using Dj, it is shown that the probability that an arbitrary $u \neq u_0$ lies inside the CI is larger for Di than for Dj. That is, show that if

$$\text{var}[\hat{s}(u),Di] > \text{var}[\hat{s}(u),Dj],$$

then

$$P_i = \Pr\left\{ \frac{(b_1+2b_2u)_i^2}{c_i^{11}+4u^2c_i^{22}} \leq \chi^2_{\alpha,1} \right\}$$

$$> \Pr\left\{ \frac{(b_1+2b_2u)_j^2}{c_j^{11}+4u^2c_j^{22}} \leq \chi^2_{\alpha,1} \right\} = P_j,$$

where P_m is the probability that an arbitrary u lies inside the $(1-\alpha)100\%$CI around the optimum, using design Dm. For either design,

$$P = \Pr\{ \chi^{2'}_{1,\nu} \leq \chi^2_{\alpha,1} \}$$

$$= 1 - \Pr\{ \chi^{2'}_{1,\nu} > \chi^2_{\alpha,1} \}.$$

The value of P decreases as the noncentrality parameter, ν, increases. So, it must be shown that the noncentrality parameter using design Di, ν_i, is smaller than that for using Dj, ν_j.

Denote the noncentrality parameters as

$$\nu_i = \frac{(\beta + 2\beta_2 u)^2}{\sigma^2(c_i^{11}+4u^2c_i^{22})}$$

$$\text{and} \quad \nu_j = \frac{(\beta_1 + 2\beta_2 u)^2}{\sigma^2(c_j^{11}+4u^2c_j^{22})} \qquad \text{(A.0.1)}$$

The denominator of the noncentrality parameter is the variance of the estimated slope. If $\text{var}[\hat{s}(u),Di] > \text{var}[\hat{s}(u),Dj]$, then $\nu_1 < \nu_2$, and hence, $P_i > P_j$, and the theorem is proved.∎

Estimating the slope precisely does, in fact, guarantee a smaller probability of misclassifying a non-optimal point as optimal.

To calculate the probabilities in Table 1.2.1:
The noncentrality parameters are calculated for each design, and P_1 and P_2 are obtained from tables of the noncentral Chi-square distribution. Table A.0.1 lists values of ν_1, ν_2, P_1 and P_2 for values of u across the experimental region, using the model, D1 and D2 as defined in section 1.2:

$$\hat{y} = 1 + 4u - 2u^2$$

D1: $(n_{-1}, n_0, n_{+1}) = (8, 4, 8)$

D2: $(n_{-1}, n_0, n_{+1}) = (5, 10, 5)$.

Note that for values of u where $\text{var}[\hat{s}(u),Di] > \text{var}[\hat{s}(u),Dj]$, then $\nu_i < \nu_j$. When u is sufficiently far from u_0, however, neither design would put u inside the CI, and when u is close to 1 (the optimum in this case), the variance of the slope is so large for both designs that either would place u inside the CI.

TABLE A.0.1

u	ν_1	ν_2	P_1	P_2
.1	172	120	0	0
.2	91	78	0	0
.3	44.8	45	0	0
.4	21.9	25.3	0	0
.5	10.7	13.3	.1	.04
.6	5.0	6.6	.38	.28
.7	2.13	2.9	.71	.60
.8	.742	1.05	.85	.80
.9	.149	.214	.93	.92
1	0	0	1	1

APPENDIX A.1

DEVELOPMENT OF THE SPHERICAL VARIANCE OF SLOPE ESTIMATION

Define the average slope variance on a hypersphere of radius r as

$$V^{*r} = \frac{NK}{\sigma^2} \int_{U_r} \text{var}(s(\underline{x}))d\underline{x}, \quad \text{where} \quad K^{-1} = \int_{U_r} d\underline{x} \tag{2.3.2}$$

$$\text{and} \quad U_r=\{\underline{x}: \sum_{i=1}^{k} x_i^2=r^2\}.$$

For standard RS designs and least squares estimation, the slopes of the surface in directions parallel to the k factor axes are estimated with equal precision. Without loss of generality, consider the slope in the direction of the x_1 axis. The development is given for k=3 for ease in notation, but the results generalize directly to the k variable case. The second order response model is written

$$\hat{y} = b_0 + \sum_{j=1}^{3} b_j x_j + \sum_{j=1}^{3} b_{jj} x_j^2 + \sum_{i<j}^{3} b_{ij} x_i x_j \tag{1.1.2}$$

The slope in the direction of the x_i axis is given by (2.3.1). Hence,

$$\begin{aligned} V^{*r} &= \frac{NK}{\sigma^2} \int_{U_r} \text{var}(\underline{l}'\underline{b})d\underline{x} \\ &= \frac{NK}{\sigma^2} \int_{U_r} \underline{l}'\text{var}(\underline{b})\underline{l}\, d\underline{x} \\ &= \frac{NK}{\sigma^2} \int_{U_r} \underline{l}'\sigma^2(X'X)^{-1}\underline{l}\, d\underline{x} \\ &= \frac{NK}{\sigma^2} \int_{U_r} \sigma^2 \text{tr}(\underline{l}'(X'X)^{-1}\underline{l})\, d\underline{x} \\ &= \frac{NK}{\sigma^2} \int_{U_r} \sigma^2 \text{tr}(\underline{l}\underline{l}'(X'X)^{-1})d\underline{x} \end{aligned}$$

$$= \frac{NK}{\sigma^2}\{\sigma^2 tr[(\int_{U_r} \underline{l}\underline{l}'d\underline{x})(X'X)^{-1}]d\underline{x}\}$$

$$= tr[(K\int_{U_r} \underline{l}\underline{l}'d\underline{x})M^{-1}]$$

$$= tr\,[\,\mu M^{-1}], \qquad (A.1.1)$$

where $\mu = K\int_{U_r} \underline{l}\underline{l}'d\underline{x}$ and $M^{-1} = N(X'X)^{-1}$.

The spherical moment matrix for the slope s_1 is

$$\mu = K\int_{U_r} \underline{l}\underline{l}'d\underline{x}$$

$$= K\int_{U_r} \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 2x_1 \\ 0 \\ 0 \\ x_2 \\ x_3 \\ 0 \end{bmatrix} [0\ 1\ 0\ 0\ 2x_1\ 0\ 0\ x_2\ x_3\ 0]\, d\underline{x}$$

$$= K\int_{U_r} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 2x_1 & 0 & 0 & x_2 & x_3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2x_1 & 0 & 0 & 4x_1^2 & 0 & 0 & 2x_1x_2 & 2x_1x_3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & x_2 & 0 & 0 & 2x_1x_2 & 0 & 0 & x_2^2 & x_2x_3 & 0 \\ 0 & x_3 & 0 & 0 & 2x_1x_3 & 0 & 0 & x_2x_3 & x_3^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} d\underline{x}$$

$$= \begin{bmatrix} 0 & & & & & & & & & \\ & 1 & & & & & & & & \\ & & 0 & & & & & [0] & & \\ & & & 0 & & & & & & \\ & & & & 4\sigma_2 & & & & & \\ & & & & & 0 & & & & \\ & & & & & & 0 & & & \\ & & [0] & & & & & \sigma_2 & & \\ & & & & & & & & \sigma_2 & \\ & & & & & & & & & 0 \end{bmatrix} \qquad (A.1.2)$$

where $\sigma_2 = K\int_{U_r} x_i^2 d\underline{x} = r^2/k$ (1.5.9a).

The weighted variance-covariance matrix for k=3 is given by

$$M^{-1} = N(X'X)^{-1}$$

$$= \begin{bmatrix} c^0 & c^1 & c^2 & c^3 & c^{11} & c^{22} & c^{33} & c^{12} & c^{13} & c^{23} \\ & c^{11} & c^{12} & c^{13} & c^{111} & c^{122} & c^{133} & c^{112} & c^{113} & c^{123} \\ & & c^{22} & c^{23} & c^{112} & c^{222} & c^{233} & c^{122} & c^{123} & c^{223} \\ & & & c^{33} & c^{113} & c^{223} & c^{333} & c^{123} & c^{133} & c^{233} \\ & & & & c^{1111} & c^{1122} & c^{1133} & c^{1112} & c^{1113} & c^{1123} \\ & & & & & c^{2222} & c^{2233} & c^{1222} & c^{1223} & c^{2223} \\ & & & & & & c^{3333} & c^{1233} & c^{1333} & c^{2333} \\ & & \text{sym.} & & & & & c^{1122} & c^{1123} & c^{1223} \\ & & & & & & & & c^{1133} & c^{1233} \\ & & & & & & & & & c^{2233} \end{bmatrix} \quad (A.1.3)$$

Substituting (A.1.2) and (A.1.3) into (A.1.1), we obtain

$$\begin{aligned} V^*r &= tr(\mu M^{-1}) \\ &= c^{11} + 4\sigma_2 c^{1111} + \sigma_2(c^{1122} + c^{1133}) \\ &= c^{11} + \frac{r^2}{3}[4c^{1111} + c^{1122} + c^{1133}] \end{aligned} \quad (A.1.4)$$

In general, for the k variable case, the spherical variance of a slope in the direction of the i^{th} factor axis is written

$$V^*r(\hat{s}_i) = c^{ii} + \frac{r^2}{k}(4c^{iiii} + \sum_{i\neq j} c^{iijj}) \quad (2.3.3)$$

where $\sum_{i\neq j}$ is over the (k-1) interactions involving factor x_i.

APPENDIX A.2

Theorem 2.5.1: When the CCD is scaled so that the design radius is $\sqrt{k}$, V^{*r} as defined in (2.3.2) and (2.3.3) is minimized for all radii $0 \le r \le \sqrt{k}$, with respect to α, when $\alpha = \sqrt{k}$.

Proof: Two cases will be treated separately and then the results will be combined:

1. CCD with $\alpha \le \sqrt{k}$
2. CCd with $\alpha \ge \sqrt{k}$.

Case 1:

Hussey provided the form of the $(X'X)^{-1}$ matrix for the CCD, (2.5.1). Making the appropriate substitutions, V^{*r}_{CCD} is given in (2.5.2). Differentiating (2.5.2) with respect to α provides information on desirable and undesirable locations for the axial points, in terms of the spherical slope variance.

$$V^{*r}_{CCD} = \frac{N}{F+2\alpha^2} + \frac{N(k-1)r^2}{kF} + \frac{4r^2\{N^2(kF-F+2\alpha^4)-N(k-1)(F+2\alpha^2)^2\}}{2\alpha^4 kN(kF+2\alpha^4)-2\alpha^4k^2(F+2\alpha^2)^2} .$$

$$\frac{\partial V^{*r}}{\partial \alpha} = \frac{-4\alpha N}{(F+2\alpha^2)^2} + \frac{\partial(3^{rd}\ term)}{\partial \alpha} .$$

$$(3^{rd}\ term) = \frac{r^2(A+B\alpha^2+C\alpha^4)}{(D\alpha^4+E\alpha^6+G\alpha^8)} ,$$

where $A = 4NF(k-1)(N-F)$

$B = -16NF(k-1)$

$C = 8N(N-2(k-1))$

$$D = 2k^2F(N-1)$$

$$E = -8k^2F$$

$$G = 4k(N-2k).$$

$$\frac{\partial V^{*r}}{\partial \alpha} = \frac{-4\alpha N}{(F+2\alpha^2)^2}$$

$$+ \frac{r^2[(D\alpha^4+E\alpha^6+G\alpha^8)(2B\alpha+4C\alpha^3)-(A+B\alpha^2+C\alpha^4)(4D\alpha^3+6E\alpha^5+8G\alpha^7)]}{[D\alpha^4+E\alpha^6+G\alpha^8]^2}$$

$$= -D1 + r^2\left(\frac{NUM}{DEN}\right),$$

where $D1 = \frac{4\alpha N}{(F+2\alpha^2)^2} \geq 0$, $DEN = [D\alpha^4+E\alpha^6+G\alpha^8]^2 \geq 0$,

and NUM is the numerator of the second term in $\partial V^{*r}/\partial\alpha$ apart from r^2.

To determine for which radii V^{*r} is decreasing as α becomes larger, consider the following.

$$\frac{\partial V^{*r}}{\partial \alpha} < 0 \quad \text{when} \quad r^2 < D1(DEN/NUM) \text{ if } NUM>0$$

$$r^2 > D1(DEN/NUM) \text{ if } NUM<0.$$

Since $D1 \geq 0$ and $DEN \geq 0$, V^{*r} decreases at all r when $NUM < 0$. Due to the complex nature of NUM, a computer was used to calculate its value for $0<\alpha\leq \lfloor k$. NUM is less than zero for all $\alpha \leq \lfloor k$, so V^{*r} is decreasing on every radius in the design region when α is increasing. This implies that to uniformly minimize V^{*r} using a CCD, α should be as large as $\lfloor k$.

Note that values of $\alpha > \lfloor k$ were not investigated in this case, since equation (2.5.2) does not incorporate the scaling convention requiring

all design points to lie on or within a radius of $\sqrt{k}$.

Case 2:

To scale a design with radius R to a radius of $\sqrt{k}$, multiply the design matrix by $\sqrt{k}/R$.

The unscaled $X = [\underline{1}\ \underline{x}_1\ \underline{x}_2 \ldots \underline{x}_k\ \underline{x}_1^2 \ldots \underline{x}_k^2\ \underline{x}_1\underline{x}_2 \cdots \underline{x}_{k-1}\underline{x}_k]$

becomes $X = [\underline{1}\ (\sqrt{k}/R)\underline{x}_1 \ldots (\sqrt{k}/R)\underline{x}_k\ (k/R^2)\underline{x}_1^2 \ldots (k/R^2)\underline{x}_k\ (k/R^2)\underline{x}_1\underline{x}_2 \ldots (k/R^2)\underline{x}_{k-1}\underline{x}_k]$.

Hence, for the CCD without replicated axials,

$$(X'X)=\begin{bmatrix} N & \underline{0}' & (k/R^2)(F+2\alpha^2)\underline{1}_k' & 0 \\ \underline{0} & (k/R^2)(F+2\alpha^2)I_k & 0 & 0 \\ (k/R^2)(F+2\alpha^2)\underline{1}_k & 0 & (k^2/R^4)(2\alpha^4 I_k+F\underline{1}_k\underline{1}_k') & 0 \\ \underline{0} & 0 & 0 & (k^2/R^4)FI_1 \end{bmatrix}$$

$$(X'X)^{-1} = \begin{bmatrix} a & 0 & b\underline{1}_k' & 0 \\ 0 & cI_k & 0 & 0 \\ b\underline{1}_k & 0 & (d-e)I_k+e\underline{1}_k\underline{1}_k' & 0 \\ 0 & 0 & 0 & fI_1 \end{bmatrix},$$

where a, b, c, d, e, and f are redefined by setting $(X'X)(X'X)^{-1}=I_k$ and solving the equations simultaneously.

$$a = (2\alpha^4+kF)/D$$

$$b = -R^2(F+2\alpha^2)/kD$$

$$c = R^2/[k(F+2\alpha^2)]$$

$$d = (D+J)R^4/2\alpha^4Dk^2$$

$$e = JR^4/2\alpha^4k^2D$$

$$f = R^4/k^2F$$

$$J = (F+2\alpha^2)^2-NF$$

$$D = N(kF+2\alpha^4)-k(F+2\alpha^2)^2 \ .$$

When $\alpha > \sqrt{k}$ in the unscaled CCD, then the radius R of the design is $R=\alpha$.

$$V^*r(\hat{s_i}) = c^{ii} + \frac{r^2}{k} (4c^{iiii} + \sum_{j\neq i} c^{iijj})$$

For the CCD,

$$c^{ii} = \alpha^2N/k(F+2\alpha^2)$$

$$c^{iijj} = N\alpha^4/k^2F$$

$$c^{iiii} = dN.$$

Thus,

$$V^*r_{CCD} = \frac{N\alpha^2}{k(F+2\alpha^2)} + \frac{r^2\{2N^2[(k-1)F+2\alpha^4]-2N(k-1)(F+2\alpha^2)^2\}}{k^3\{N(kF+2\alpha^4)-k(F+2\alpha^2)^2\}} + \frac{r^2(k-1)N\alpha^4}{k^3F} \ .$$

$$\frac{\partial V^*r_{CCD}}{\partial \alpha} = \frac{k(F+2\alpha^2)2\alpha N-N\alpha^3 4k}{k^2(F+2\alpha^2)^2} + \frac{4r^2(k-1)N\alpha^3}{k^3F}$$

$$+ \frac{r^2}{k^3} \frac{\partial}{\partial \alpha} \frac{\{2N^2[(k-1)F+2\alpha^4]-2N(k-1)(F+2\alpha^2)^2\}}{N(kF+2\alpha^4)-k(F+2\alpha^2)^2} \ .$$

As long as the CCD contains at least one center run,

$$\frac{\partial}{\partial \alpha} V^*r_{CCD}(s_i) > 0 \text{ for all } r \text{ such that } 0\leq r\leq \sqrt{k} \text{ when } \alpha\geq \sqrt{k}.$$

Combining case 1 and case 2, the optimal α value, is $\alpha=\sqrt{k}$, and the theorem is proved.

APPENDIX A.3

REPLICATING CENTER RUNS AND AXIAL POINTS IN A CCD

First consider adding center runs to a CCD. Write

$$N = F + 2kn_a + n_0$$
$$= FA + n_0,$$

where n_a is the number of replicates of an axial point, F is the number of factorial points, FA is the sum of the number of factorials and the number of axials, and n_0 is the number of center runs. Let V_m denote V^{*r} for the CCD with m center runs. Similarly, let $c_m{}^{ii}$, $c_m{}^{iiii}$, and $c_m{}^{iijj}$ denote the appropriate elements of M^{-1} for the CCD with m center runs. To determine whether or not to add a center run, consider

$$V_m - V_{m+1} = c_m{}^{ii} - c_{m+1}{}^{ii} + \frac{r^2}{k}[4(c_m{}^{iiii} - c_{m+1}{}^{iiii}) + (k-1)(c_m{}^{iijj} - c_{m+1}{}^{iijj})].$$

If this quantity is negative, it is not beneficial to add the center run. If it is positive, an extra center run is helpful. It will be shown that adding a few center runs improves slope prediction precision in the outer portion of the design region more than it damages slope estimation variance near the center.

$$V_m - V_{m+1} = \frac{-1}{F+2n_a\alpha^2} + \frac{r^2}{k}\left[\frac{-(k-1)}{F} + 4(c_m^{iiii} - c_{m+1}^{iiii})\right].$$

$$c^{iiii} = \frac{N^2[b+(k-2)c]-N(k-1)a^2}{(b-c)\{N[b+(k-1)c]-ka^2\}},$$

where $a = F + 2n_a\alpha^2$

$b = F + 2n_a\alpha^4$

$c = F.$

To more easily manage the calculation notation, the following constants are defined.

$K1 = b+(k-2)c$

$K2 = (k-1)a^2$

$K3 = (b-c)[b+(k-1)c]$

$K4 = ka^2(b-c)$

$K5 = 2(FA)(K1)-(K2)$, where $FA=F+2kn_a$

$K6 = (FA)^2(K1)-(FA)(K2)$

$K7 = (K3)(FA)-(K4).$

Using these constants,

$$c_m^{iiii} - c_{m+1}^{iiii} = \frac{-(K1)(K3)m(m+1)-(K1)(K7)(2m+1)+(K3)(K6)-(K5)(K7)}{[(K3)m + (K7)][(K3)(m+1) + (K7)]}.$$

Thus,

$$V_m-V_{m+1} = \frac{-1}{F+2n_a\alpha^2} + r^2\left[\frac{-(k-1)}{kF} + 4\left(\frac{-K1K3m(m+1)-K1K7(2m+1)+K3K5-K5K7}{k(K3m + K7)[K3(m+1) + K7]}\right)\right]$$

$$= \quad -DIF1 \quad + r^2[\quad COEF \quad].$$

To determine for which radii this quantity is positive, that is, in what part of the region a center point will improve slope estimation, consider

$$V_m - V_{m+1} > 0 \quad \text{when} \quad r^2 > \text{DIF1/COEF}, \quad \text{if COEF} > 0$$
$$r^2 < \text{DIF1/COEF}, \quad \text{if COEF} < 0.$$

Note that DIF1>0, so if COEF<0, we should not add center runs. If COEF>0, adding center runs will improve prediction capability beyond a radius of $r=\sqrt{\text{DIF1/COEF}}$. Using a computer to calculate this critical radius and the differences V_m-V_{m+1}, it can be shown that for α near $\sqrt{k}$, the improvement obtained in the perimeter of the design region outweighs the damage done near the design center for m=0, 1, 2. This implies that up to three center runs are advantageous for these α values. When α is small, however, Center runs worsen slope estimation over the region. No more than three center runs should be used. Only $0<\alpha\leq\sqrt{k}$ was studied.

A similar strategy is used to determine the advantages of replicating axial points. Only $n_a=1$ and $n_a=2$ are compared.

$$V_1-V_2 = (c_1^{ii}-c_2^{ii}) + r^2\left[\frac{4}{k}(c_1^{iiii}-c_2^{iiii})+\frac{(k-1)}{k}(c_1^{iijj}-c_2^{iijj})\right]$$

$$= \quad \text{DIF1} \quad + r^2 \left[\quad \text{COEF} \quad \right],$$

where the subscripts denote the value of n_a. Replicating axials is beneficial when $V_1-V_2 > 0$.

$V_1 - V_2 > 0$ when $r^2 > -DIF1/COEF$, if $COEF > 0$

$r^2 < -DIF1/COEF$, if $COEF < 0$.

Table A.3.1 contains some results for several α values when k=3.

TABLE A.3.1

No. CENTER RUNS	α	$V_1 > V_2$ WHEN:
1	1.00	r > .4296
1	1.35	r < 0 (never)
1	1.682	r < .0953
1	1.732	r < .1357
3	1.00	r > .1245
3	1.35	r < 0 (never)
3	1.682	r < .1038
3	1.732	r < .1248

Note that DIF1 < 0 when $\alpha < \lfloor kF/(F+n_0) \rfloor$. So, when COEF>0, there is improvement in the region except near the center. The amount of improvement is greater than the damage done to the center. On the other hand, as α becomes close to $\lfloor kF/(F+n_0) \rfloor$, there is more damage done to slope estimation over most of the region than there is improvement near the center.

APPENDIX A.4

Theorem 2.7.1: If V^{*r} as defined in (2.3.2) and (2.3.3) is used to measure and compare the slope estimation precision capabilities of two standard RS designs, the maximum difference in V^{*r} from one design to the other is located at the design perimeter, that is, at $r=\lfloor k$, provided that

$$\left| \frac{2(c_1^{ii}-c_2^{ii})}{4(c_1^{iiii}-c_2^{iiii})+(\sum_{j\neq i} c_1^{iijj}-\sum_{j\neq i} c_2^{iijj})} \right| \leq 1.$$

Proof: Denote the calculation of V^{*r} in the slope parallel to the x_i axis using design m, by

$$V_m = V_m^{*r}(\hat{s}_i) = c_m^{ii} + r^2[(4/k)c_m^{iiii} + (1/k)\sum_{j\neq i} c_m^{iijj}].$$

Then,

$$V_1-V_2 = c_1^{ii}-c_2^{ii}+r^2[(4/k)(c_1^{iiii}-c_2^{iiii})+(1/k)(\sum_{j\neq i} c_1^{iijj}-\sum_{j\neq i} c_2^{iijj})]$$
$$= A + r^2B.$$

It needs to be shown that V_1-V_2 is largest at the radius $r=\lfloor k$, which is the largest design radius possible using V^{*r}. Consider the following four cases:

1. $A \geq 0$ and $B \geq 0$
2. $A < 0$ and $B < 0$
3. $A \geq 0$ and $B < 0$

and

4. $A < 0$ and $B \geq 0$.

case 1: If A and B are both nonnegative, then $\max_r |V_1-V_2|$ occurs when r is as large as possible, ie when $r=\lfloor k$.

case 2: If both A and B are negative, then $\max_r |V_1-V_2|$ is equivalent to case 1, since $|A+r^2B| = -1(A+r^2B)$.

case 3: If A is nonnegative and B is negative, then $A+r^2B\leq 0$ when $r^2\geq -A/B$, and $A+r^2B\geq 0$ when $r^2\leq -A/B$. The biggest negative value occurs for fixed A and B when r^2 is as large as possible, since $-A/B\geq 0$. r^2 is largest at $r=\lfloor k$. The biggest positive value is A and occurs at r=0. Comparing the size of the largest negative difference to the size of the largest positive difference,

$$A + r^2B \leq -A \quad \text{when } r^2 \geq -2A/B.$$

So, if $|2A/B|\leq k$, then the largest difference is at $r=\lfloor k$.

case 4: This case is parallel to case 3. If $|2A/B|\leq k$ then the largest difference is at $r=\lfloor k$.

Now, to complete cases 3 and 4,

$$|2A/B| = \left| \frac{2k(c_i^{ii}-c_2^{ii})}{4(c_1^{iiii}-c_2^{iiii})+(\sum_{j\neq i} c_1^{iijj}-\sum_{j\neq i} c_2^{iijj})} \right| \leq k$$

iff

$$\left| \frac{2(c_1^{ii}-c_2^{ii})}{4(c_1^{iiii}-c_2^{iiii})+(\sum_{j\neq i} c_1^{iijj}-\sum_{j\neq i} c_2^{iijj})} \right| \leq 1. \qquad (A.2.1)$$

All standard response surface designs compared using V^{*r} in this dissertation satisfy (A.2.1), thus the theorem is proved.

The two page vita has been removed from the scanned document. Page 1 of 2

The two page vita has been removed from the scanned document. Page 2 of 2