GRAPHICAL TOOLS, INCORPORATING COST AND OPTIMIZING CENTRAL COMPOSITE DESIGNS FOR SPLIT-PLOT RESPONSE SURFACE METHODOLOGY EXPERIMENTS

by

Li Liang

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

Doctor of Philosophy
in
Statistics

C.M. Anderson-Cook, Co-Chair
T.J. Robinson, Co-Chair
E.P. Smith
G.G. Vining
K. Ye

March 28, 2005

Blacksburg, Virginia

Keywords: 3-dimensional variance dispersion graph, fraction of design space, cost penalized evaluation, multiple criteria, optimal factorial levels, central composite structure

GRAPHICAL TOOLS, INCORPORATING COST AND OPTIMIZING CENTRAL COMPOSITE DESIGNS FOR SPLIT-PLOT RESPONSE SURFACE METHODOLOGY EXPERIMENTS

Li Liang

(Abstract)

In many industrial experiments, completely randomized designs (CRDs) are impractical due to restrictions on randomization, or the existence of one or more hard-to-change factors. Under these situations, split-plot experiments are more realistic. The two separate randomizations in split-plot experiments lead to different error structure from in CRDs, and hence this affects not only response modeling but also the choice of design. In this dissertation, two graphical tools, three-dimensional variance dispersion graphs (3-D VDGs) and fractions of design space (FDS) plots are adapted for split-plot designs (SPDs). They are used for examining and comparing different variations of central composite designs (CCDs) with standard, *V*- and *G*-optimal factorial levels. The graphical tools are shown to be informative for evaluating and developing strategies for improving the prediction performance of SPDs. The overall cost of a SPD involves two types of experiment units, and often each individual whole plot is more expensive than individual subplot and measurement. Therefore, considering only the total number of observations is likely not the best way to reflect the cost of split-plot experiments. In this dissertation, cost formulation involving the weighted sum of the number of whole plots and the total number of observations is discussed and the three cost adjusted optimality criteria are proposed. The effects of considering different cost scenarios on the choice of design are shown in two examples. Often in practice it is difficult for the experimenter to select only one aspect to find the optimal design. A realistic strategy is to select a design with good balance for multiple estimation and prediction criteria. Variations of the CCDs with the best cost-adjusted performance for estimation and prediction are studied for the combination of *D*-, *G*- and *V*-optimality criteria and each individual criterion.

Dedicated

To my husband, my son and my family

ACKNOWLEDGEMENTS

I would like to wholeheartedly thank my advisor Dr. Christine Anderson-Cook for her invaluable guidance and advice. Her encouragement, support, and thoughtfulness as a mentor and a friend are highly appreciated. I would also like to thank my co-advisor Dr. Timothy Robinson for his valuable advice and help throughout this research.

I would also like to thank my committee members: Dr. Keying Ye, Dr. Geoffrey Vining, and Dr. Eric Smith for their helpful comments and suggestions. I also like to thank the professors in Statistics department for enhancing my knowledge with their excellent teaching.

I would like to thank the friends at Blacksburg, with whom I spend great time in the past five years. Thanks go to Xin Zhong, Bo Jin, Ayca Ozol-Godfrey, J. D. Williams, and many others.

I would like to express my sincere gratitude and deep appreciation to my beloved parents for their love, encouragement, and believing in me over all these years. I also thank my brothers Yong and Long, and my sister Yan for their love and support. I would like to deeply thank my husband Xiaopeng for his love, support, and patience, without which this work would not have been done. Thanks also go to my son Eric for the great happiness he has brought to my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1 Introduction

Split-plot designs are often used in agriculture experiments due to restrictions of randomizations or existence of one or more hard-to-change factors. In practice, it is not unusual that the experiment is run in the way that the observations with same levels are not reset and leads to an actual split-plot design (SPD) rather than the designed completely randomized experiment. Recently, the importance of split-plot experiments in industrial applications is gaining more recognition. The research in this dissertation explores different aspects of split-plot designs for first and second order models. In Chapters 2 and 3, two graphical tools called three-dimensional variance dispersion graphs (VDG) and fractions of design space (FDS) plots are adapted for split-plot designs to summarize the prediction variance distribution throughout the entire design space, and in the whole plot and subplot spaces, respectively. A few examples belonging to two types of central composite designs (CCDs) are provided to show how to construct the plots and extract meaningful information for comparisons from them. With the aid of 3-D VDGs and FDS plots, strategies for improving the designs are developed. *V*- and *G*-efficiency can be improved by moving the factorial design points of the CCDs to different locations in the design spaces. The plots are also used to illustrate how the optimization strategies alter the prediction variance distribution. So far previous research has focused on the quality of estimation and prediction scaled by the total number of runs, which means that designs with more observations are penalized more. For SPDs, the cost of the experiment comes from both the number of whole plots and the number of subplots. In Chapter 4, the cost of split-plot designs combined with the estimation and prediction performances are considered for the evaluation of SPDs. Different scenarios for the whole plot and subplot cost ratios are discussed in this chapter and the cost adjusted estimation and prediction variance measures are presented. Then two examples are studied for first and second order models respectively to show the impact of different cost and variance scenarios in the choice of split-plot designs. Different variations of central composite designs are explored in this chapter and the desirable structures are found. Combining the strategies of optimizing CCDs from Chapters 2-4, that is, optimizing the factorial levels and optimizing the structure of whole plots and the settings of subplot levels within whole

plots, we study the best CCD for each individual criterion. Recommended *D*-, *G*- and *V*-efficient designs as well as designs which best balance a combination of the three criteria are presented. The robustness of the best designs selected from our study to the variance component ratio is also explored. The studies and results can provide helpful information for practitioners to choose split-plot designs.

1.1 Split-Plot Designs

Complete randomization in many industrial and agricultural experiments is frequently impractical due to constraints in time and/or cost or existence of one or more hard-to-change factors. In these situations, restrictions on randomization lead to split-plot designs (SPD), allowing certain factor levels to be randomly applied to the whole plot units and remaining factor levels randomly to the subplot units. Separate random errors in whole plot units and in subplot units are due to the two randomizations in the experiment. The resulting compound symmetric error structure affects not only estimation and inference, but also the choice of design.

1.1.1 Split-Plot Designs in Industry

In many industrial experiments we may prefer to not completely randomize the order of the runs. In these cases, the experiments often involve two categories of factors: those factors with hard or costly to change levels and those factors whose levels are relatively easy to change. When hard-to-change factors exist, it is in the practitioner's best interest to keep the number of times the levels of these factors are changed relatively small for the economy and easy management of the experiments. As a result, a reasonable experiment is to run multiple combinations of the easy-to-change factors at each setting of the hard-to-change (HTC) factors. Such a strategy results in a split-plot design (SPD) and is frequently more realistic than complete randomization of the combinations of the factors, which results in a completely randomized design (CRD).

Consider an example of a split-plot design in food industry. The goal of the experiment is to look for the best recipe of a cake mix. Several factors are thought to affect taste (y). Three factors are involved in the ingredients: flour, shortening and egg powder. Two other factors include oven temperature and cooking time. In experimentation, adjusting the temperature of the oven generally take considerable time, and cooking batches of the cake mixtures at one cooking time is more efficient. Complete randomization of each combination of the five factors is impractical. Since several distinct batches of cake batter (involving different levels of flour, shortening, and egg powder) can be quickly created, but changing oven temperature and baking time involve a great deal of time, the food scientist would like to fix the oven temperature and bake all compositions of the cakes at a particular temperature for a set period of time. In the experiment, the oven temperature and baking time (hard-to-change) are the whole plot factors, and the floor, shortening and the cake mix (easy-to-change) are the subplot factors. Other examples of split-plot designs are given in Hinkelmann and Kempthorne (1994), Montgomery (2001), Letsinger, Myers and Lentner (1996), Bisgaard (2000) and Kowalski and Vining (2001).

In industry, a great deal of attention is given to determining the optimal arrangement of treatment combinations (combinations of factor levels) in an experiment. In the determination of the best arrangement of treatment combinations, it is often assumed that the run order of treatment combinations will be completely randomized during experimentation. However, when hard to change factors are present, the run order is sometimes not completely randomized. For instance, a factorial experiment is intended to conduct to study the effect of the furnace temperature, the density of the nickel powder and the amount of binder for a sintering process of the nickel battery plates. Suppose each factor has three levels. The 27 runs are randomly ordered. However, in production, the data is often collected as follows: for any two consecutive observations where furnace temperature is at the same level, the furnace temperature is not reset. By not re-setting furnace temperature, the run order is not completely randomized. Consequently, this experiment is a split-plot experiment with furnace temperature serving as the hard-to-change factor. In this design, consecutive observations with the same temperature are actually in a same whole plot. Treating the experiment as a CRD would lead to incorrect

analysis of the response model. Ganju and Lucas (1999) discussed the detecting of the randomization restrictions and provide several recommendations for designing the experiment in a manner beneficial to both the running of the experiment and the analysis of the data.

In split-plot experiments, the experimental units (EUs) for the hard-to-change factors are often called “whole plots” and the experimental units for the easy-to-change factors are called “subplots”. The levels of hard-to-change factors are randomly applied to the whole plot units. Within each whole plot, the levels of easy-to-change factors are randomly assigned to subplot units. The two separate randomizations lead to an error structure in which observations within the same whole plot are not independent but are rather equally correlated due to being observed within the same whole plot. Observations from different whole plots, however, are assumed to be independent. The resulting error structure of observations within the same whole plot is a compound symmetric structure. This error structure must be accounted for not only when conducting inferences but also when determining the optimal design.

Because the subplot factors are replicated more often than the whole plot factors, split-plot experiments typically provide more information about the subplot factors. In addition there may be more natural variability associated with the whole plot factors. Hence the comparisons among the levels of the subplot factors have higher precision than those of the whole plot factors. Consequently, if the SPD is analyzed incorrectly as a CRD, whole plot effects might be incorrectly deemed significant. The subplot effects, on the other hand, might be incorrectly deemed insignificant. In the following section, the proper model and analysis for the SPDs are discussed.

1.1.2 Model and Analysis

The classic representation of split-plot experiments has its origin in agricultural experiments where the goals are generally to determine if significant differences exist among the levels of various treatments. For instance, consider a replicated factorial

experiment conducted with a split-plot scheme with one whole plot variable (A) and one subplot variable (B). Each of the *a* levels of the whole plot factor is replicated r_a times and each of the *b* levels of the subplot variable occurs once within each whole plot. Consequently, there are a total of r_a* *a* whole plots, each containing *b* subplots, for a total of *N*= r_a *a b* runs. In this split-plot experiment, the following model is appropriate

$$y_{ijkl} = \mu + r_{i(j)} + \alpha_j + e_{ij}^{W} + \beta_k + \alpha\beta_{jk} + \varepsilon_{ijk} : \; i = 1, \cdots, r_a, \; j = 1, \cdots, a \text{ ,and } \; k = 1, \cdots, b$$

where *y* is the response, μ is the global mean, $r_{i(j)}$ denotes the whole plot replicate effect, α_i is the whole plot effect, β_j is the subplot effect, $\alpha\beta_{ij}$ is the interaction of the whole plot factor and the subplot factor, e_{ij}^{W} denotes the experimental error from the whole plots, assumed to be i.i.d $N(0, \sigma_\delta^2)$ with σ_δ^2 denoting the variability among the whole plot units, and ε_{ijk} denoting the subplot error assumed to be i.i.d. $N(0, \sigma_e^2)$ with σ_e^2 representing the variability among subplot units. It is also assumed that e_{ij}^{W} and ε_{ijk} are independent. For more details see Kempthorne (1952) or Hinkelmann and Kempthorne (1994).

If the combinations of subplot factors selected are the same within every whole plot, the SPD is “Crossed”. Otherwise, it is “Non-Crossed”. Crossed SPDs are simple to implement and easy to analyze since the ordinary least square (OLS) estimates are equivalent to the generalized least square (GLS) estimates. The factorial design in the split-plot scheme is an example of this type SPD. In practice, there are often restrictions in the number of the subplots that can be run within a given whole plot. As a result, the crossed SPD might not be practical when the number of subplot factors is large. The SPD is “Balanced” if there is the same number of subplots within each whole plot. The example above is both crossed and balanced. The fractional factorial design in split-plot scheme is another example of balanced SPD. If designed properly, this type of design can be highly efficient and very appealing from a practical perspective. Vining, Kowalski, and Montgomery (2004) proposed general conditions for response surface designs such that the OLS=GLS and provided examples of balanced SPDs that satisfy the condition. Examples of these types of designs include the modified CCDs and modified BBDs. If

each whole plot factor level has only one whole plot in the design, it is known as a restricted SPD (RSPD). In the replicated factorial SPD with one whole plot variable and one subplot variable described above, the design would have been a RSPD had there been no replication. Another example of a RSPD is the standard CCD in the split-plot scheme where all the observations with same whole plot levels are put in one whole plot. For instance, a standard CCD with one whole plot variable is a restricted SPD if it has only five whole plots corresponding to the five levels $\pm\alpha$, ± 1 and 0. The standard and modified CCDs will be discussed in detail in this research. Goos and Vandebroek (2004) discussed the efficiency of RSPD and indicated that with high probability there are more efficient designs for parameter estimation between the CRD and RSPD.

The corresponding ANOVA analysis of the model above is shown in Table 1.1. When analyzing the split-plot experiment, the proper error term should be used for testing of parameter estimates. This ANOVA approach is straightforward, but for the non-balanced data or data with a large number of missing values, it has proven to be inefficient. In addition, the variance component estimates could be negative in the ANOVA analysis.

Table 1.1: ANOVA for split-plot designs.

Source	d.f.	E(MS)
Replicates	r_a-1	
Whole plot factor A	a-1	$\sigma_\delta^2 + b\sigma_\varepsilon^2 + r_a b\sum_j \alpha_j^2/(a-1)$
Whole plot error	(r_a-1)(a-1)	$\sigma_\delta^2 + b\sigma_\varepsilon^2$
Subplot factor B	b-1	$\sigma_\varepsilon^2 + r_a a\sum_k \beta_k^2/(b-1)$
Interaction A$\times$B	(a-1)(b-1)	$\sigma_\varepsilon^2 + r_a \sum_{j,k} \alpha\beta_{jk}^2 \Big/ [(a-1)(b-1)]$
Subplot error	(r_a -1)a(b-1)	σ_ε^2
Total	r_a ab-1	

In industrial experiments, the response is often a function of continuous factors and the goal is to develop a prediction model relating the response to the factors. Thus, unlike classical split-plot experiments where treatment comparisons were the focus,

interest is in developing a regression model. The resulting prediction model is known as a response surface model. In this research a mixed model is employed for modeling the split-plot experiment. When the experiment is run as a split-plot design with a whole plots, the following linear mixed model can be written to explain the variation in the $N\times 1$ response vector, **y,**

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Zu} + \boldsymbol{\varepsilon},$$

where **y** is the vector of N observations, $\boldsymbol{\beta}$ is the $p\times 1$ vector of fixed effect model parameters including the intercept, **X** is the design matrix containing both whole plot and subplot effects expanded to model form; $\mathbf{u}=(u_1,u_2,\ldots,u_a)'$ is the vector of random effects corresponding to the a whole plots with the u_i assumed to be i.i.d. $N(0,\sigma_\delta^2)$ with σ_δ^2 denoting the variability among whole plots. **Z** is the $N\times a$ incidence matrix of ones and zeroes where the ij^{th} entry is 1 if the j^{th} observation (j=1,...,N) belongs to the i^{th} whole plot (i=1,…,a). If observations are sorted by whole plots, we obtain $\mathbf{Z}=diagonal\{\mathbf{1}_{n_1},\cdots,\mathbf{1}_{n_a}\}$, where $\mathbf{1}_{n_i}$ is an $n_i\times 1$ column of one's and n_i is the number of subplots within the i^{th} whole plot and $\sum_{i=1}^{a} n_i = N$; $\boldsymbol{\varepsilon}=(\varepsilon_1,\varepsilon_2,...,\varepsilon_N)$ is the vector of residual errors where the ε_j are assumed to be i.i.d. $\mathrm{N}\left(0,\sigma_\varepsilon^2\right)$ where σ_ε^2 denotes the variability among subplot units. It is also assumed that **u** and $\boldsymbol{\varepsilon}$ are independent. Together, the forms of **X** and **Z** influence the quality of parameter estimation and model prediction.

To help illustrate notation, consider a replicated 3^2 factorial experiment conducted as a split-plot design with one whole plot variable and one subplot variable. All levels of the subplot variable are run within each level of the whole plot variable, and each level of the whole plot variable is replicated twice, yielding a total of six whole plots. Within each whole plot the three levels of subplot variable are randomly assigned to three subplots, giving a total of 18 subplots. The 18×6 incidence matrix **Z** and 6×1 vector of random effects **u** are defined as follows,

$$\mathbf{Z} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \underline{1}_3 & & & & & \\ & \underline{1}_3 & & & & \\ & & \underline{1}_3 & & & \\ & & & \underline{1}_3 & & \\ & & & & \underline{1}_3 & \\ & & & & & \underline{1}_3 \end{pmatrix} \text{ and } \mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \\ u_6 \end{pmatrix}.$$

From the model, the observations have mean $\mathbf{X\beta}$ and covariance matrix $\sum$, where $\sum = \sigma_\delta^2 \mathbf{Z}\mathbf{Z}^T + \sigma_\varepsilon^2 \mathbf{I}_N = \sigma_\varepsilon^2 \left[d\mathbf{Z}\mathbf{Z}^T + \mathbf{I}_N \right]$ where $d = \sigma_\delta^2 / \sigma_\varepsilon^2$ and denotes the ratio of whole plot variance to subplot variance. It is referred as the "variance component ratio" since it represents the relative strength of variations introduced by whole plot units in the experiment. Assuming that we have sorted observations by whole plots, i.e., $\mathbf{Z} = Diagonal\{\mathbf{1}_{n_1}, \cdots, \mathbf{1}_{n_a}\}$, we can conveniently write the covariance matrix in block-diagonal form as,

$$\sum = \begin{bmatrix} \sum_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sum_a \end{bmatrix},$$

where the $n_i \times n_i$ matrix $\sum_i$ is given by

$$\sum_i = \begin{bmatrix} \sigma_\varepsilon^2 + \sigma_\delta^2 & \cdots & \sigma_\delta^2 \\ \vdots & \ddots & \vdots \\ \sigma_\delta^2 & \cdots & \sigma_\varepsilon^2 + \sigma_\delta^2 \end{bmatrix}$$

and denotes the variances and covariances of observations within the i^{th} whole plot. In the example above, $\sum_i$ is a 3×3 matrix since there are 3 subplots within each whole plot. Note that the variance of an individual observation is the sum of the subplot and whole plot error variances, $\sigma_\varepsilon^2 + \sigma_\delta^2$. It is easy to show that the larger the variance component ratio is, the more strongly the observations within the whole plot are correlated to each other.

There are several approaches available to estimate the variance components and the coefficients of the fixed effect in the mixed model. These approaches include ANOVA, Henderson I, II, III methods (adaptation of ANOVA for unbalanced data), maximum likelihood (ML), restricted or residual ML (REML), Bayesian inference, etc. Searle (1992) provides detailed discussions of the advantages and disadvantages of these methods. Among them, REML is a modern approach and proven to work efficiently for unbalanced data cases. For balanced data, the REML estimates are identical to the standard ANOVA estimates if the latter estimates are positive. REML is the default method used in the SAS procedure PROC MIXED.

After the variance components are estimated, the linear mixed model equation (LMME) can be used for estimating the vector of fixed effects, $\boldsymbol{\beta}$, and for predicting the random term, $\mathbf{u}$. The estimated vector of model coefficients, $\hat{\boldsymbol{\beta}}$, obtained by this algorithm is the best linear unbiased estimates (BLUE) of $\boldsymbol{\beta}$, is identical to the generalized least square estimates (GLS) of $\boldsymbol{\beta}$, and is the maximum likelihood estimate (MLE) of $\boldsymbol{\beta}$ under normality. We will refer to this estimate as the generalized least squares (GLS) estimate given by the expression $\hat{\boldsymbol{\beta}} = [\mathbf{X}^T \Sigma^{-1} \mathbf{X}]^{-1} \mathbf{X}^T \Sigma^{-1} \mathbf{y}$. The covariance matrix of the estimated model coefficients is

$$Var(\hat{\boldsymbol{\beta}}) = [\mathbf{X}^T \Sigma^{-1} \mathbf{X}]^{-1}.$$

Even if the OLS estimates are equivalent to the GLS estimates, as they are for some SPDs, the covariance matrices of the estimates are not equivalent. Consequently, when conducting inferences, it is important to apply the correct variances. When the design is completely randomized, the variance of the parameter estimates is $Var\left(\hat{\boldsymbol{\beta}}\right) = \left(\mathbf{X}'\mathbf{X}\right)^{-1} \sigma^2$.

Comparing the expressions for the estimated model coefficients for split-plot designs and CRDs is important as it lends insight into the greater complexity associated with optimal design strategies for split-plot designs vs. for CRD. For example, if one wishes to obtain the optimal design in terms of ability to estimate model parameters, the optimal CRD depends only on the settings of the levels of the terms in **X**. The optimal split-plot design

in terms of parameter estimation will depend on the structure of **X**, the variance ratio, *d*, the number of whole plots, *a*, the dimensionality of each of the $\sum_i$'s (determined by the number of subplots within each of the whole plots), and the arrangement of subplot levels within whole plots.

Since the goal of designed experiments is often to gain a better understanding of the underlying relationship between the factors and the response, good prediction throughout the design space is important. Since the error structure in split-plot designs is different than that of completely randomized designs, the prediction variance expression differs from that of the CRD. The predicted value of the mean response at any location $\mathbf{x}_0$ is given by

$$\hat{\mathbf{y}}_0 = \mathbf{x}_0'\hat{\boldsymbol{\beta}} = \mathbf{x}_0'\left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y},$$

where $\mathbf{x}_0$ is the point of interest in the design space expanded to model form. The prediction variance for split-plot designs is then given by

$$Var\left(\hat{\mathbf{y}}_0\right) = \mathbf{x}_0'\left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{x}_0.$$

When the design is completely randomized, the prediction variance is $\sigma^2\mathbf{x}_0'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{x}_0$. For a CRD, $\sum = \sigma_\varepsilon^2\mathbf{I}$ and the optimal designs for prediction depend only on the settings of the factor levels in **X**. For split-plot designs, however, the variances of the model coefficient estimates as well as the variances of predicted values are not only a function of the settings of the factors in **X** but also the unknown variance components and the structure of $\sum$.

Often the variances of model coefficients and predictions are divided by the variance to get a unitless evaluation for the estimation and prediction qualities. For CRD, the scale is the variance of observations, σ^2. For split-plot design, however, there are two error terms because of the two independent randomizations and thus two variances. In the same principle, the variance of observation, $\sigma_\varepsilon^2+\sigma_\delta^2$, is used to scale the estimation or prediction variance for split-plot design and thus allows the evaluation of the quality of

the SPD is independent of the variations of the experiment units and only determined by the setting of factor levels in $\mathbf{X}$ and the whole plot structure in $\mathbf{Z}$, and the variance component ratio. The covariance matrix of model coefficient after scaled is

$$\frac{\left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}}{\sigma_\delta^2+\sigma_\varepsilon^2}=\left(\mathbf{X}'\left(\frac{\boldsymbol{\Sigma}}{\sigma_\delta^2+\sigma_\varepsilon^2}\right)^{-1}\mathbf{X}\right)^{-1}=(\mathbf{X}'\mathbf{R}^{-1}\mathbf{X})^{-1},$$

where $\mathbf{R}$ denotes the correlation matrix of the observations, and has block diagonal form for observations sorted by whole plots,

$$\mathbf{R}=\begin{bmatrix}\mathbf{R}_1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \mathbf{R}_a\end{bmatrix},$$

where each $n_i\times n_i$ matrix $\mathbf{R}_i$ is given by $\mathbf{R}_i=\begin{bmatrix}1 & \cdots & d/(1+d)\\ \vdots & \ddots & \vdots\\ d/(1+d) & \cdots & 1\end{bmatrix}$. The prediction variance scaled by variance is given by $\mathbf{x}_0'\left(\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\right)^{-1}\mathbf{x}_0$ for split-plot designs.

From the scale-free variance equations, we can learn that the quality of a split-plot design is strongly related to the variance component ratio, *d*, rather than the two individual variance components. A split-plot design may have different characteristics for quality of estimation and prediction for different *d* values. Therefore *d* is a critical factor in selecting split-plot designs. In practice, *d* value is often assumed known and the split-plot experiments are designed based on the particular *d* value. While we almost always do not know the exact value of variance component ratio at the design of experiments stage, a range of plausible values for *d* can be obtained from preliminary estimates from a pilot study or guessed based on previous knowledge. Bisgaard and Steinberg (1997) stated that for many applications the whole plot variance is usually larger than subplot variance in prototype experiments. Letsinger et al. (1996) studied a split-plot experiment in chemical industry with *d*=1.04. Vining, Kowalski and Montegomery (2004) estimated the variance terms using pure error and obtain the variance ratio 5.65. Webb, Lucas and Borkowski (2002) described an experiment with variance ratio 6.92 in a computer component manufacturing company. Kowalski, Cornell and Vining (2002) studied a mixture

experiment with process variables where the estimated variance ratio was 0.82. In this dissertation, specific values of d=0 (CRD), 1 and 10 are often considered in more details, representing the situations that the whole plot variance can't be distinguished from subplot variance from complete randomization, the whole plot variance is the same as and ten times than the subplot variance, respectively.

1.1.3 Optimal Split-Plot Designs

Many efficient designs have been developed for the CRDs, such as factorials and fractional factorials in first order models and the central composite designs (CCD) [Box &Wilson, 1951] and Box–Behnken designs (BBD) [Box & Behnken, 1960] for second order models. Due to the influence of the error structure, these highly efficient CRDs are not necessarily as efficient when the design has a split-plot randomization scheme.

Huang, Chen and Voelkel (1998) and Bingham and Sitter (1999) explore minimum-aberration (MA) two-level fractional factorial split-plot designs. Kowalski (2002) constructed 24-run SPDs using semifolding, which sacrifices some partial resolution of the whole plot factors to improve the partial resolution at the subplot levels, to get the desirable overall resolution. These designs are useful for screening experiments.

In SPDs, after the significant factors are selected from a set of variables, an experimental design for a second-order model becomes necessary for response modeling and the optimization. Some strategies for selecting optimal or high efficient SPDs in second-order model have been developed by Davison (1995), Ju & Lucas (2002) and Goos and Vandebroek (2001, 2004). These studies focused on the estimation properties of whole plot and subplot factors and the efficiency of the SPD. The criterion utilized to compare designs is the D-optimality criterion. They concluded that even if cost is not an issue, the split-plot design can be shown more efficient than the CRDs under some scenarios, especially for experiments with relatively large whole plot variability. Practitioners are often interested in using the final model for prediction as well. In this research, the prediction capability evaluation of SPDs for the second order model is the main focus.

Joiner & Campbell (1976) incorporate the different costs of changing factor levels and generate good run sequences considering a time effect. However, they consider the experiments as CRDs. In fact the experiments studied should have been considered under split-plot error structures. Similarly, if considering the different costs of whole plots and subplots, split-plot designs become superior to the CRDs. Bisgaard (2000) discusses the cost of split-plot experiment in the category of balanced SPD, but doesn't incorporate the cost into the quality evaluation of split-plot designs. In this research, more general cost scenarios are incorporated into the estimation and prediction evaluation of SPDs, and the influences of the different scenarios of error structures on the efficiency of the split-plot design are explored.

1.2 Design Optimality Criteria

Design optimality criteria characterized by some of the letters of the alphabet are often called alphabetical optimality criteria. Kiefer and Wolfowitz (1959) were among the first authors to develop these optimality criteria. These criteria are single number criteria where each measure intends to capture a different perspective of the 'goodness' of a design. Box and Hunter (1957), Box and Draper (1974, 1987), and Myers and Montgomery (page 304, 2002) discuss some of these properties for comparing designs.

1.2.1 *D*-optimality

Historically, the most popular design criterion has been the *D*-optimality criterion, which focuses on the estimation variance of model parameters through the moment matrix, defined as $\mathbf{M}(D) = \mathbf{X}'\mathbf{X}/N$ for a CRD. Here, $\mathbf{X}'\mathbf{X}$ is the information matrix of the parameter estimates and N denotes the total number of runs in the design and is used as a penalty for larger designs. The inverse of the moment matrix contains the variances and covariances of the estimated model parameters, scaled by the N/σ^2. For SPD, the moment matrix is $\mathbf{M}(D) = \mathbf{X}'\mathbf{R}^{-1}\mathbf{X}/N$. This criterion is important if the goal of the

experimental design is to choose the settings of the design points to achieve good estimation properties in the moment matrix.

D-optimality tries to maximize the determinant of the moment matrix, thus a *D*-optimal design is the design D^* with $\left|\mathbf{M}(D^*)\right| = Max_{D\in\Omega}\left|\mathbf{M}(D)\right|$, where Ω is the space of all possible designs. Under the assumptions of independent normal errors with constant variance for a completely randomized design, the determinant of $\mathbf{X}'\mathbf{X}$ is inversely proportional to the square of the volume of the confidence region for the regression coefficients, which reflects how well the coefficients are estimated. Hence the larger the determinant of $\mathbf{X}'\mathbf{X}$ is, the better the estimates of the model parameters are. For the first order or first order with interaction model, two-level orthogonal designs with levels at ± 1 extremes are *D*-optimal designs. Quite often, however, for second order models, there is no known *D*-optimal design.

The *D*-efficiency of a design D is defined as $D_{eff} = (\left|\mathbf{M}(D)\right| / \left|\mathbf{M}(D^*)\right|)^{1/p}$, where $\left|\mathbf{M}(D^*)\right|$ is the determinant of the moment matrix of the *D*-optimal design and *p* is the number of parameters in the model. The exponent, $1/p$, accounts for the *p* parameter estimates being assessed. Myers and Montgomery (2002) discuss some highly efficient response surface designs in the completely randomized setting such as the CCD and BBD. These designs are not necessarily highly efficient designs when implemented under split-plot randomization scheme. We explore in the following chapters how the efficiencies change when these popularly used response surfaces designs are run as SPDs.

1.2.2 *G*-optimality

In industrial situations, the aim of the practitioner is often to predict the response for a particular combination of the regressors or factors. Consequently, good prediction performance at various locations in the design space or in the entire design space can be important for assessing the choice of design. Box and Hunter (1957) defined the scaled prediction variance function (SPV) to examine how precise the estimated response is at

any particular location in the design space. A good design should have as small and stable SPV values throughout the entire region as possible.

The SPV for a completely randomized design is defined as $\nu(\mathbf{x}_0) = NVar(\hat{y}_0(\mathbf{x}_0))/\sigma^2$ $= N\mathbf{x_0}^T(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x_0}$ for a standard linear model, where $\mathbf{x}_0$ is the point of interest in the design space expanded to the model form, $\hat{y}_0(\mathbf{x}_0)$ is the estimated mean response at the location $\mathbf{x}_0$ used to predict the new observation y_0, and σ^2 is the variance of the observations. For a split-plot design, $\nu(\mathbf{x}_0)$ is calculated by $N\mathbf{x_0}^T(\mathbf{X}'\mathbf{R}^{-1}\mathbf{X})^{-1}\mathbf{x_0}$. The prediction variance is multiplied by the total number of runs of the design, *N*, to penalize larger designs and give a "per observation" evaluation of prediction performance. In addition, it is divided by the observation variability to be scale-free. We can see that the SPV values are determined by the design matrix $\mathbf{X}$, the correlation matrix $\mathbf{R}$, the location in the design space $\mathbf{x}_0$ and the design size *N*. Actually the SPV is merely a quadratic form function of the inverse of the moment matrix. For example, for a completely randomized 2^2 factorial design in first order model with levels ± 1, its moment matrix is $\mathbf{X}'\mathbf{X}/N = I$, thus $\nu(\mathbf{x}_0) = \begin{bmatrix}1 & x_{10} & x_{20}\end{bmatrix} I \begin{bmatrix}1 & x_{10} & x_{20}\end{bmatrix}^T = 1 + {x_{10}}^2 + {x_{20}}^2$.

G-optimality seeks to minimize the maximum SPV throughout the entire design region, i.e. a design D^* that satisfies $\min_{D\in\Omega}[max_{\mathbf{x}_0\in R}\nu(\mathbf{x}_0)]$, where *R* is the region of interest, is a *G*-optimal design. Intuitively, *G*-optimality tries to protect the experimenter against the worst-case scenario being too undesirable. An interesting and important result is that the lower bound for the maximum SPV for a completely randomized design is *p*, the number of parameters in the model, which represents the case with equal SPV values at any location in the design space (Myers and Montgomery, 2002). Consequently, *G*-efficiency can be defined as $G_{eff} = p/\max_{\mathbf{x}_0\in R}\nu(\mathbf{x}_0)$. For example, for the CRD factorial design mentioned above, *p*=3, if the design has a cuboidal region, $\max_D \nu(\mathbf{x}_0)$=3 thus the *G*-efficiency of this design is 1, hence the design is *G*-optimal. But for SPDs, the theoretical lower bound is not known, but it will no longer be *p* because of the complicated error structure and the role of variance component ratio.

1.2.3 *V*–optimality

Another design optimality criterion that addresses prediction variance is *V* (or *Q* or *IV*) - optimality. It studies the average SPV over the entire design region and thus requires integration over the whole region to determine its value. Although its computation is the most difficult, *V*-optimality is frequently used in terms of measuring an important characteristic, the overall performance of prediction, of the design. A *V*-optimal design is the one that minimizes the average SPV and $V(D^*) = \min_{D \in \Omega} \frac{1}{K} \int_R v(\mathbf{x}_0) d\mathbf{x}_0$, where *K* is the volume of the region of interest. For example, for a 2^2 factorial design with levels scaled between [-1,1] for each factor in a cuboidal region, $K=2^2=4$. The *V*-efficiency of design D is defined as $V_{eff} = V(D^*) \Big/ \frac{1}{K} \int_R v_D(\mathbf{x}_0) d\mathbf{x}_0$, where $V(D^*)$ is the average SPV for the *V*-optimal design. Myers and Montgomery (2002) indicate that the two-level first order orthogonal designs with levels set at the ± 1 extremes are also *V*-optimal design for completely randomized experiments.

Among the criteria discussed above, the *D*-optimality is most commonly used in computer-generated design since the calculations are simple. *G*-optimality is becoming somewhat more common in some statistical software packages. However, we should be aware that *V*-optimality is also an important and reasonable tool for choosing experimental design, especially in the situation that prediction performance is an important priority.

1.2.4 The Role of the Total Number of Runs, *N*, in Optimality Criteria

In order to fairly compare designs with different number of runs, for the three criteria we discussed so far, the scaled estimation and prediction variance are often evaluated. By dividing by the total number of design points, *N*, for the moment matrix or multiplying by *N* for the prediction variance, larger designs are penalized. This “per observation” evaluation is done based on the reasonable assumption that in most industrial settings

there is some cost involved in performing additional runs. For CRD, the cost of experiment can be thought to be proportional to the total number of design points used. This is not necessarily true for many split-plot experiments, however. Split-plot experiment involves two randomizations and thus two types of experiment units, and thus the cost of the SPD likely is a function of the cost related to the whole plots and subplots. The cost could also be related to the cost or time involved in changing the levels of the whole plot and subplot variables. In real life split-plot experiments, one might have to face different scenarios. For the situation that the cost of subplots is of main concern, the "per observation" evaluation is reasonable. However, one more common situation is that each individual whole plot is much more expensive than individual subplot. For instance, it may be very time-consuming to reset the levels of the whole plot variables like the speed of the product line or the temperature of the furnace. Therefore, a flexible cost based criteria should allow for different cost scenarios to be incorporated. In Chapter 4, details of various cost scenarios and their impacts on the choice of best split-plot design are discussed. In cases where little cost is associated with additional runs, the cost of experiment is not of primary concern and the practitioner may want to compare the absolute estimation or prediction variance among designs with different number of runs. In such cases, the un-scaled evaluation [scaling by $1/(\sigma_\delta^2 + \sigma_\varepsilon^2)$ is still done but when design size is not an issue, we may not wish to scale by N.] can be used.

1.2.5 Desirability Function for Combining Multiple Criteria

Each criterion discussed so far has focused on one aspect of the design's quality. Good design for one aspect might not yield a good design based on another criterion. Consequently, the practitioner might be in a dilemma of which aspect is the most important for the experiment. A natural and reasonable solution of that is to select the design having good (which might not be optimal) properties for each criterion and balanced performances for multiple criteria.

Derrinnger and Suich (1980) and Myers and Montgomery (2002, pp. 247-258) discuss the application of desirability function for exploring the optimum conditions for multiple

responses. Derrinnger and Suich (1980) developed a useful optimization procedure, using a desirability function to combine several objectives. The desirability function could be one-sided or two-sided depending on whether the response want to be maximized (or minimized) or be assigned a target value. If one of the responses is to be maximized, the desirability function is given by

$$desirability_i = \left\{ \frac{(\hat{y}_i - A_i)}{(B_i - A_i)} \right\}^s, \quad A_i \le \hat{y}_i \le B_i,$$

where the $\hat{y}_i$ is the one of the responses of interest, A_i and B_i define the acceptable range for the individual response. Hence when the response value increases, the condition is more desirable for this response. The power *s*, a positive value, represents how quickly the desirability will increase with respect to the changes in response value $\hat{y}_i$ and should be specified by the user according to the reality. If the desirability of the product would be greatly increased if the response value is considerably larger than the lower bound B_i, large value of *s* should be used. On the other hand, a small *s* value should be specified if the response value slightly above the lower bound is very desirable. Therefore, we can see that the desirability function has values ranging from 0 to 1. For the values outside the acceptable range, the desirability would be zero and the product is considered unacceptable. The two-sided desirability function is applied if we want the objective of optimization is as close as to the target value. It is defined as

$$desirability_i = \begin{cases} \left\{ \frac{(\hat{y}_i - A_i)}{(B_i - A_i)} \right\}^s, & A_i \le \hat{y}_i \le B_i \\ \left\{ \frac{(\hat{y}_i - C_i)}{(B_i - C_i)} \right\}^t, & B_i \le \hat{y}_i \le C_i \end{cases},$$

where A_i and C_i define the acceptable range for the response, and B_i is the target, i.e., the most desirable value. By specifying values of power *s* and *t*, the desirability function allows the practitioner to incorporate the subject knowledge about the role of the response in the total desirability of the product. If the goal of the product improvement involves multiple responses, the desirability functions of all responses are combined to get a single number. Usually the geometric mean of the desirability values for the *k* responses, $D = (d_1 \times d_2 \times \cdots \times d_k)^{1/k}$, is applied to obtain the overall assessment of the

desirability of the combined responses and D turns our to be the objective function of optimization. Still, D values ranges from 0 to 1. If one response value is unacceptable, the corresponding desirability will be zero and same zero value for D, then the overall assessment will fail to accept the condition. This is a merit of geometric mean and also the reason it is commonly used. Heredia-Langner, Montgomery, Carlyle and Borror (2004) apply desirability functions to find model-robust optimal designs. In their paper, the design has a desirability value based on an alphabetic optimality criterion for each model and the geometric mean of the values for all the models concerned is used as the overall evaluation of each design.

In the dissertation, we considered three efficiencies, i.e., D-efficiency for evaluating the quality of estimation, as well as V- and G-efficiencies for prediction. We found that the split-plot design may be desirable for having low average prediction variance, but it may have larger maximum prediction variance. Therefore, in practice, the situation may exist that the preferred design is one with the best balance between the three criteria. In Chapter 5, we will study the overall performance based on a combination of the three efficiencies for several variations of the central composite designs.

1.3 Graphical Tools

In the above discussions, design comparison was based on using single number optimality criteria to compare competing designs. Though easy to compute, we should be aware of their disadvantages. The decision of a best design is typically more complicated than can be summarized by a single number. For instance, when the practitioner is interested in predicting and wants to learn more about prediction variance distribution, i.e., if the prediction variance is stable in the entire design region, or where are the regions with best and worst prediction precision, graphical tools are very helpful for exploring the prediction properties of competing designs. We will discuss several commonly used graphical tools in the following section.

1.3.1 Variance Dispersion Graphs (VDG)

Giovannitti-Jensen and Myers (1989) proposed the variance dispersion graph (VDG) to evaluate prediction variance properties of response surface designs. It displays the distribution of scaled prediction variance (SPV) throughout a multidimensional region on two-dimensional graphs. When the practitioners' interest is in predicting the response over the experimental region of interest, VDGs have proven to be extremely useful and more informative for comparing designs than single-number optimality criteria such as *G*- or *V*-optimality.

The VDG was developed first for CRDs with a spherical design space. Assuming a design with a spherical region and a given standard linear models, three curves are displayed in the VDG, i.e., maximum, average and minimum scaled prediction variances. They are plotted against the spheres radius from zero up to the outer radius of the sphere covering the region of interest, to evaluate the prediction capacity. For example, the maximum SPV is defined as $\max_{\mathbf{x}\in U_r}[v(\mathbf{x})]$, where $U_r = \{\mathbf{x} : \sum_i x_i^2 = r^2\}$ denoting the spherical sub-region with radii *r* and maximum SPV values are found on this sub-region. The maximum curve provides where is the worst case thus the weakest area in the design space and how large is the maximum value. The average SPV is defined as $V_{avg}{}^r = \psi \int_{U_r} v(\mathbf{x}) d\mathbf{x}$, where $\psi^{-1} = \int_{U_r} d\mathbf{x}$ is the measure of area in the sphere with radii *r*, average values are found by integrating the SPVs over this sphere. The average curve helps us learn how stable the SPV distribution is at various distances from the center of the design space. The minimum SPV is $\min_{\mathbf{x}\in U_r}[v(\mathbf{x})]$. These graphs are informative and useful in terms of comparing designs on a fixed design space. Examples of the VDGs can be found in Myers et al. (1992), Vining (1993), Borkowski (1995), Trinca and Gilmour (1998), and Borror, Montgomery, and Myers (2002).

Some designs involve cuboidal design spaces rather than spherical ones. In the paper by Myers et al. (1992), the maximum, average and minimum prediction variances are still plotted over spherical radius, but only the sections of the spheres inside the cuboidal

region are assessed. VDGs in this approach show “bumps” over the sphere radius partially falling outside the cuboidal region of interest. Rozum and Myers (1991) rectified this by evaluating the maximum, average and minimum prediction variances over shrunken cubes with radius from zero up to one, which corresponds to the edge of the design space. Interpretation of the plots is the same as for spherical region. Moreover, Trinca and Gilmour (1998) used VDGs for comparing blocked response surface designs for both types of regions.

Piepel and Anderson (1992), and Piepel, Anderson, and Redgate (1993a, b) developed and demonstrated VDGs for response surface designs with polyhedral region and irregular-shaped regions, specifically mixture experiments with constraints. Using a shrunken region approach, they plot the variances along constant shrinkage values of the original design polyhedron or the constrained region. The shrunken regions begin at the boundaries of the original design space and shrink to the overall centroid of the region. Vining, Cornell, and Myers (1993) also develop the VDGs for mixture designs. In their approach they plot the prediction variances along the Cox directions, which are rays passing through the centroid of the constrained region and the vertices of the unconstrained simplex.

1.3.2 Three-Dimensional VDG (3-D VDG)

Goldfarb, Borror, Montgomery and Anderson-Cook (2004) developed three-dimensional VDGs to evaluate prediction variance of mixture-process experiments. The conventional two-dimensional VDGs consider only a summary of the combined spaces. For mixture experiments with process variables, the design spaces can be thought of as combinations of two sub spaces. In their work, distribution of prediction variance in the mixture and process spaces are evaluated separately to allow better understanding of the prediction performance of designs in different areas of the design space. In addition, by keeping one shrinkage factor (mixture or process shrinkage factor) at a constant value, one can examine the prediction variance distribution in the other space.

In Goldfarb et al (2004), the authors plot the shrinkage factor for the mixture space along the *x*-axis, the shrinkage factor for the process space along the *y*-axis, and the prediction variance surface along the *z*-axis. The shrinkage factor is equivalent to a radius when the process space is spherical, or represents shrunken cubes when the process space is cuboidal. For the mixture space, the shrunken regions have the same shape as the entire mixture space. The Shrinkage values of 0 to 1 represent the overall centroid and the edge of the corresponding sub design space, respectively. To calculate the scaled prediction variance, a large number of shrinkage points are uniformly sampled along the shrunken spheres/cubes, and maximum, average SPV are calculated over these points. The number of points sampled is chosen to be proportional to the area of each shrinkage region. The 3-D VDGs for mixture-process design can be viewed either as surface or contour plots. These plots are excellent tools to aid in selecting the designs, augmenting addition designs points, or selecting appropriate fractions of experiments when full mixture-process design has an unmanageably large number of runs.

1.3.3 Quantile Dispersion Graph (QDG)

The quantile dispersion graph (QDG) was introduced by Khuri, Kim and Um (1996). This method uses quantiles of the scaled prediction variance to describe the distribution of SPV in the entire design region. These plots display the quantiles of the SPV including the minimum and the maximum values within the shrunken regions. Khuri, Harrison and Cornell (1999) use the QDGs to examine the mixture experiments with constrained regions. QDG plots provide an overall picture and detail information of the SPV distribution throughout the constrained region, and the minimum and maximum values as well. However, there are frequently too many plots if we want to explore different shrinkage values. They are hard to construct and interpret, which lead to some practical difficulties.

1.3.4 Fractional Design Space (FDS)

Fractional design space plots were proposed by Zahran, Anderson-Cook and Myers (2003) to complement the variance dispersion graphs. The FDS plots display the fraction

of the design space where the SPV is less than or equal to any SPV value. These plots provide additional information on the distribution of SPV throughout the design space. It summarizes the SPV distribution without providing direct information about the specific location of particular SPV values, but manages to summarize the distribution with a single line.

FDS plot can be constructed by sampling a large number, say n, uniformly from locations throughout the design space. No matter what shapes of design space are considered, this uniformly sampling method could be employed. Then all the corresponding SPV values obtained are ordered and plotted against the quantiles (any value from 0 to 1). Hence in the FDS Plot, the maximum and minimum SPV over the entire region are conveniently displayed. If we plot a horizontal line with value p and $2p$, where p is the number of parameters in the model and represents the theoretical 100% and 50% G-efficiency for the completely randomized design, we can learn about the G-efficiency of this design. More of the fraction of the design space for SPV which is close to the minimum, the better the design will be. Moreover, the flatter the line, the more stable the SPV distribution for the design will be.

Goldfarb, Anderson-Cook, Borror and Montgomery (2004) modify FDS Plots for mixture and mixture-process experiments with irregular regions. They plotted global FDS curve and sliced FDS curves for SPV values in the mixture space at constant shrinkage factors for process variables. In this way, the experimenter is able to see which of the two spaces contributes more to changes in the SPV.

Chapter 2 Three-Dimensional Variance Dispersion Graphs (3-D VDGs)

Abstract

Complete randomization in many industrial and agricultural experiments is frequently impractical due to constraints in time, cost or existence of one or more hard-to-change factors. In these situations, restrictions on randomization lead to split-plot designs (SPD), allowing certain factor levels to be randomly applied to the whole plot units and remaining factor levels randomly to the subplot units. Separate random errors in whole plot units and in subplot units are due to the two randomizations in the experiment. The resulting compound symmetric error structure affects not only estimation and inference, but also the choice of design. In this paper, we first consider looking at the prediction properties of split-plot designs, expanding the comparison between designs beyond just looking at parameter estimation properties, and we present the 3-dimensional variance dispersion graphs (3-D VDGs) as a tool for evaluating the prediction capability of split-plot designs and for developing design strategies. A popular design for second-order models is the central composite design (CCD). By studying the distribution of the scaled prediction variance (SPV) in the 3-D VDGs, we demonstrate that the *G*- and *V*-efficiencies of standard CCDs can be improved upon by changing the factorial levels of the CCD.

2.1 Split-Plot Designs

Many industrial experiments involve two types of factors, some with levels hard or costly to change and others with levels that are relatively easy to change. Typical examples of hard-to-change factors include humidity, pressure and temperature. When hard-to-change factors exist, it is in the practitioner's best interest to minimize the number of times the

levels of these factors are changed. A common strategy is to run all combinations of the easy-to-change factors for a given setting of the hard-to-change factors. Such a strategy results in a split-plot design (SPD) and is more economically realistic than complete randomization of factor level combinations resulting in a completely randomized design (CRD).

In split-plot experiments, the levels of hard-to-change factors are randomly applied to the whole plot units. Within each whole plot, the levels of easy-to-change factors are randomly assigned to subplot units. The separate randomizations lead to a compound symmetric error structure of observations within the same whole plot, which must be accounted for not only when conducting inferences but also when determining an optimal design. Many efficient designs have been developed for the CRD, such as factorials and fractional factorials for first order models and the central composite designs (CCD) (Box and Wilson 1951) and Box–Behnken designs (BBD) (Box and Behnken 1960) for second order models. These highly efficient CRDs are not necessarily as efficient when the design has a split-plot structure. Huang, Chen and Voelkel (1998) and Bingham and Sitter (1999) explore minimum-aberration (MA) two-level fractional factorial split-plot designs, which are useful for screening experiments. Some strategies for selecting SPDs for second-order models have been developed by Davison (1995) and Goos and Vandebroek (2004). These studies focused on the estimation properties of SPDs. Practitioners are often interested in using the final model for prediction as well, especially when a second order model is being considered. Therefore, the prediction variance capability of split-plot designs is an important consideration and is the focus of this paper.

2.2 Model and Analysis

The two separate randomizations in split-plot designs result in correlated observations within each whole plot and independent observations from different whole plots. An

appropriate model for reflecting the error structure of a split-plot design is the linear mixed model

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Zu} + \boldsymbol{\varepsilon},$$

where $\mathbf{y}$ is the vector of N observations, $\boldsymbol{\beta}$ is the $p \times 1$ vector of fixed effect model parameters including the intercept, $\mathbf{X}$ is the design matrix expanded to model form; $\mathbf{u} = (u_1, u_2, \ldots, u_a)'$ is the vector of random effects corresponding to the a whole plots with the u_i assumed to be i.i.d. $N(0, \sigma_\delta^2)$ with σ_δ^2 denoting the variability among whole plots. Assuming that observations are sorted by whole plots, the $N \times a$ incidence matrix $\mathbf{Z}$ is given by $diagonal\{\mathbf{1}_{n_1}, \cdots, \mathbf{1}_{n_a}\}$, where $\mathbf{1}_{n_i}$ is an $n_i \times 1$ column of one's and n_i is the number of subplots within the i^{th} whole plot and $\sum_{i=1}^{a} n_i = N$; $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_N)'$ is the vector of residual errors where the ε_j are assumed to be i.i.d. $\mathrm{N}(0, \sigma_\varepsilon^2)$ where σ_ε^2 denotes the variability among subplot units. It is also assumed that $\mathbf{u}$ and $\boldsymbol{\varepsilon}$ are independent. Together, the forms of $\mathbf{X}$ and $\mathbf{Z}$ influence the quality of parameter estimation and the quality of model prediction.

The observations of a split-plot design have mean $\mathbf{X\beta}$ and covariance matrix Σ, where $\Sigma = \sigma_\delta^2 \mathbf{ZZ}' + \sigma_\varepsilon^2 \mathbf{I}_n = \sigma_\varepsilon^2 \left[d\mathbf{ZZ}' + \mathbf{I}_n \right]$ where $d = \sigma_\delta^2 / \sigma_\varepsilon^2$ denotes the ratio of whole plot error variance to subplot error variance. In this manuscript, we refer to d as the "variance component ratio" since it represents the relative strength of variations introduced by whole plot units in the experiment. Assuming observations sorted by whole plots allows one to conveniently write the covariance matrix in block diagonal form as

$$\Sigma = \begin{bmatrix} \Sigma_1 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \Sigma_a \end{bmatrix}$$

where each $n_i \times n_i$ matrix Σ_i is given by

$$\Sigma_i = \begin{bmatrix} \sigma_\varepsilon^2 + \sigma_\delta^2 & \cdots & \sigma_\delta^2 \\ \vdots & \ddots & \vdots \\ \sigma_\delta^2 & \cdots & \sigma_\varepsilon^2 + \sigma_\delta^2 \end{bmatrix}.$$

Note that the variance of an individual observation is the sum of the subplot and whole plot error variances, $\sigma_\varepsilon^2+\sigma_\delta^2$. A popular method for estimating the variance components is restricted maximum likelihood (REML). It is easy to show that the larger the variance component ratio is, the more strongly the observations within the whole plot are correlated to each other.

The vector of fixed effects parameters, $\boldsymbol{\beta}$, is estimated via generalized least squares, yielding

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y}.$$

The covariance matrix of the estimated model coefficients is given by

$$Var\left(\hat{\boldsymbol{\beta}}\right) = \left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}.$$

When the design is completely randomized, $Var\left(\hat{\boldsymbol{\beta}}\right)=\sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}$. The predicted value of the mean response at any location $\mathbf{x}_0$ is given by

$$\hat{\mathbf{y}}_0 = \mathbf{x}_0'\hat{\boldsymbol{\beta}} = \mathbf{x}_0'\left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y},$$

where $\mathbf{x}_0$ is the point of interest in the design space expanded to model form. In the CRD, the prediction variance is given by

$$Var\left(\hat{\mathbf{y}}\right) = \sigma^2\mathbf{x}_0'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{x}_0,$$

whereas for split-plot designs, it is given by

$$Var\left(\hat{\mathbf{y}}_0\right) = \mathbf{x}_0'\left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{x}_0.$$

Comparing the expressions for estimation and prediction variance for split-plot designs and CRDs is important as it lends insight into the greater complexity associated with optimal design strategies for split-plot designs vs. the CRDs. For example, if one wishes to obtain the optimal design in terms of ability to estimate the mean response, the optimal CRD depends only on the settings of the levels of the variables in **X** and the residual error variance. The optimal split-plot design in terms of prediction variance will depend on the structure of **X**, the variance ratio, *d*, the number of whole plots, *a*, and the dimensionality

of each of the $\sum_j$ (determined by the number of subplots within each whole plot), and subplot levels arrangements in whole plots.

2.3 Alphabetic Optimality Criteria for Comparing Competing Designs

Strategies for choosing an optimal design depend on the goal of the researcher. For instance, if the desire is to have quality model parameter estimates, a popular strategy is to find a design with high *D*-efficiency. The *D*-efficiency criterion is defined in terms of the scaled moment matrix of the parameter estimates. Goos and Vandebroek (2004) discuss strategies for obtaining *D*-optimal split-plot designs. If interest is in finding a design with precise prediction, *G*- and *V*- efficiencies of the design are popular choices for comparing competing designs. Although a great deal of attention in the literature has been given to studying the *G*- and *V*- efficiencies of CRDs, very little attention has been devoted to studying the *G*- and *V*- efficiencies of SPDs. It is the study of *G*- and *V*- efficiencies of SPDs that is the focus of this paper.

G- and *V*- criteria are based on the scaled prediction variance where the scaling is done to make the *G*- and *V*-efficiencies unitless as well as to penalize larger designs. As mentioned earlier, the expression for the prediction variance in CRDs is given by

$$Var\left(\hat{\mathbf{y}}\right) = \sigma^2 \mathbf{x}_0^{'} \left(\mathbf{X}^{'}\mathbf{X}\right)^{-1} \mathbf{x}_0$$

where the scaling is σ^2/N (observation error variance divided by the design size). The scaled prediction variance (SPV) for the CRD is then given by

$$SPV = \frac{NVar\left[\hat{\mathbf{y}}\left(\mathbf{x}_0\right)\right]}{\sigma^2} = N\mathbf{x}_0^{'} \left(\mathbf{X}^{'}\mathbf{X}\right)^{-1} \mathbf{x}_0 .$$

The scaling of the prediction variance for split-plot designs can be done in a similar fashion by scaling by $(\sigma_\delta^2 + \sigma_\varepsilon^2)/N$ (the observation error variance divided by the design size). The SPV for the split-plot design is then given by

$$SPV = \frac{NVar\left[\hat{\mathbf{y}}(\mathbf{x}_0)\right]}{\sigma_\delta^2 + \sigma_\varepsilon^2} = N\mathbf{x}_0{}' \left(\mathbf{X}' \left(\frac{\boldsymbol{\Sigma}}{\sigma_\delta^2 + \sigma_\varepsilon^2} \right)^{-1} \mathbf{X} \right)^{-1} \mathbf{x}_0$$

$$= N\mathbf{x}_0{}' \left(\mathbf{X}'\mathbf{R}^{-1}\mathbf{X} \right)^{-1} \mathbf{x}_0$$

where **R** denotes the correlation matrix of the $N \times 1$ vector of responses.

The scaling of the prediction variance by *N* provides a "per observation" value of the prediction variance. This is done because in most industrial settings, there is some cost involved in performing additional runs. In cases in which little cost is associated with additional runs, the practitioner may be interested in the relative absolute prediction variance among designs with different number of runs. In such cases, the un-scaled prediction variance [scaling by $1/(\sigma_\delta^2 + \sigma_\varepsilon^2)$ is still done but when design size is not an issue, we may not wish to scale by *N*]. The cost of split-plot experiments includes not only the cost of subplots/measurements, but also the cost of whole plots. Under some situations, whole plots may be more expensive than the subplot and measurement, which should be incorporated into the evaluation when taking the cost into account. Bisgaard (2000) discussed cost of the balanced split-plot experiments. Detailed work about cost penalized estimation and prediction evaluation for split-plot designs can be found in Liang, Anderson-Cook and Robinson (2005). In this paper, we assume that the experimental cost is only a function of the number of subplots, i.e., the number of observations, and use *N* to adjust for unequal sized split-plot designs.

Choosing the best design in terms of prediction variance involves selecting the design with the minimum average SPV (the *V*-optimal design) or the minimum maximum SPV (the *G*-optimal design) over the entire design space. Anbari and Lucas (1994) used the lower bound of the maximum scaled prediction variance for CRD, i.e., *p*, the number of model parameters, to evaluate *G*-efficiency of split-plot designs and claimed some super-efficient designs. Apparently *p* is not a reasonable lower bound for split-plot designs because the two errors and the different values of variance component ratio play role in the *G*-efficiency. It should be pointed out that the actual bounds for the *G*- and *V*-efficiencies for SPDs needs further investigation, and here we focus more on relative

efficiencies for comparisons between competing designs. It should also be pointed out that *G*- and *V*-criteria produce a single number with which to compare designs. In this manuscript we advocate the use of graphical techniques for comparing designs in which we consider the distribution of the average or maximum SPV over the entire design region. The single-number criteria values, in and of themselves, do not offer the practitioner a strategy for improving an existing design in terms of *G*- and *V*-efficiency. By providing a visual representation of the SPV (maximum or average) over the entire design region, it is possible to develop strategies for improving the existing designs.

2.4 Variance Dispersion Graphs (VDGs)

Giovannitti-Jensen and Myers (1989) proposed the Variance Dispersion Graph (VDG) to evaluate prediction properties of response surface designs. These two-dimensional plots show the patterns of scaled prediction variance (SPV) for a given model throughout the design space. When the practitioners' interest is in the quality of response predictions over the experimental region of interest, VDGs have proven to be extremely useful and more informative for comparing designs than single-number optimality criteria such as *D*-, *V*- or *G*-efficiency values. In standard VDGs for spherical design regions, maximum, average and minimum prediction variances curves are plotted versus radii from zero to the outer radius of the sphere. Previously, only completely randomized experiments have been considered. Examples of the VDGs can be found in Myers, Vining, Giovannitti-Jensen and Myers (1992), Vining (1993), Borkowski (1995), Trinca and Gilmour (1998), and Borror, Montgomery and Myers (2002). Goldfarb, Borror, Montgomery and Anderson-Cook (2003) developed 3-dimensional VDGs to evaluate prediction performance of mixture-process experiments. They evaluated mixture and process components of the design space separately to allow better understanding of the relative performance of designs in different regions of the design space.

2.5 Three-Dimensional VDGs (3-D VDGs) for Split-Plot Designs

Split-plot experiments essentially involve a design in the subplot factors overlaid on top of a design in the whole plot factors. Similar to mixture process experiments where the design space consists of two components (mixture and process), SPDs also involve two components (subplot and whole plot). Consequently, for evaluating SPDs, it is informative to examine the quality of the design in both the subplot space and whole plot space. In this section, we introduce the 3-dimensional VDG (3-D VDG) as a tool for examining the prediction variance properties over the entire split-plot design space as well as for any sub region.

A standard VDG applied to a split-plot design uses the *x*-axis corresponding to the single shrinkage factor of the entire design space and the prediction variance distribution is displayed along the *y*-axis. However, the influence of the variance components upon the prediction variance distribution for a split-plot design is likely to make it no longer symmetric in the whole plot and subplot spaces. For instance, for a standard CCD in the split-plot structure with one whole plot variable, *w*, and one subplot variable, *x*, the conventional VDG of maximum SPV and the proposed 3-dimensional VDG are provided in Figure 2.1 for a variance component ratio of d=10. In the 3-D VDG of maximum SPV, we observe that the maximum SPV differs widely at different locations in the whole plot space, *w*, but is somewhat uniform in the subplot space, *x*. In the traditional VDG, however, since only the maximum values in the circle $w^2 + x^2 = r^2$ are plotted at each value of radius *r*, one does not have the ability to decipher how the prediction variance changes over the whole plot and subplot spaces separately, and thus some interesting characteristics of the design are lost. Observing the difference between the whole plot space and the subplot space is helpful for understanding the quality of model prediction in the design space and any sub region. Moreover, it can be utilized to illustrate the important role of the variance component ratio on the performance of split-plot designs and is a helpful tool for devising strategies to improve the existing design. The conventional VDGs, however, are not capable of displaying such detailed information.

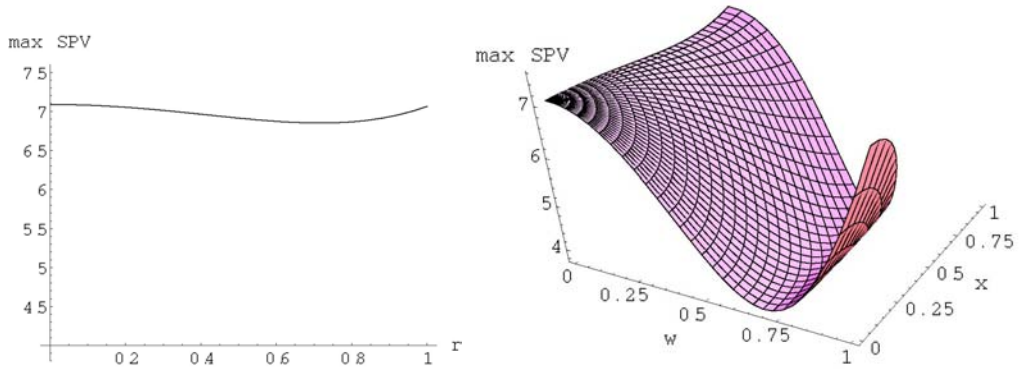


Figure 2.1 Conventional VDG (left) and 3-D VDG (right).

Although not the focus of this paper, the 3-D VDG also has application in robust parameter design (RPD). RPD experiments are often implemented with a split-plot randomization scheme due to the economy of experimentation or restrictions on randomization. In these situations, whole plot variables are often environmental factors that are hard or costly to change and subplot variables are control/design factors whose levels are relatively easy to change. In RPD, the desire is to find levels of the control factors that result in relatively flat profiles of the SPV as one moves along the whole plot space, which indicates that the process can be made robust to fluctuations in the levels of the environmental factors. 3-D VDGs allow the practitioner to study changes of the SPV values along the whole plot direction. The plots also enable the practitioner to visualize the distribution of prediction variance along the subplot space at specific locations in the whole plot space.

The 3-D VDGs plot the SPV as a function of ordered pairs of two shrinkage factors (w,x); one for the whole plot space (w) and the other for the subplot space (x). Shrinkage factors are used to indicate the shrunken design spaces and can be thought of as multipliers of the original design space. If the whole plot or subplot space is spherical, the shrinkage factor is the radius of the shrunken sphere; if the whole plot or subplot space is cuboidal, the shrinkage factor is the distance from overall center to the square face of shrunken cube. A shrinkage value of one would represent points on the edge of the whole plot or subplot space and a value of zero would indicate the center of the given design space. A shrinkage value of 0.5 would denote points that are half way between the center and the original outline of the given space. Figure 2.2 shows two examples of shrunken regions.

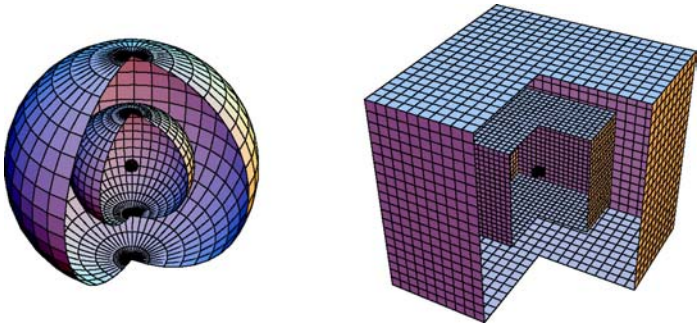


Figure 2.2 Cutaway view of shrunken spherical (left) and cuboidal (right) regions for shrinkage values of 0, 0.5 and 1.

In 3-D VDGs, the shrinkage factor for the whole plot space is plotted along the *w*-axis, the shrinkage factor for the subplot space along the *x*-axis, and the scaled prediction variance surface along the vertical axis. When several designs are compared, all designs are assumed to have the same design region of interest. Based on this assumption, all the designs are scaled to have shrinkage levels from 0 to 1. For instance in a spherical design region, the outermost design points are on the sphere with radius one. The coding of the variables will, however, be given in standard form.

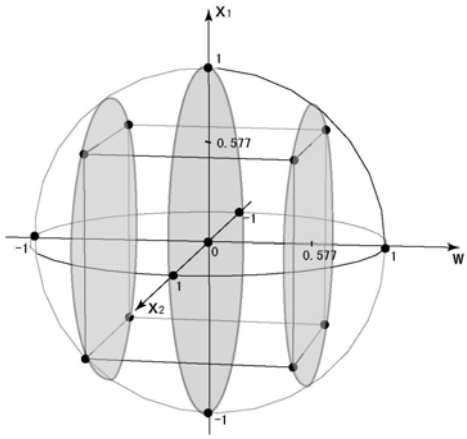


Figure 2.3 The spherical space for a central composite design with one whole plot variable and two subplot variables.

For cuboidal regions, the size of the subplot space is the same for any given whole plot level combination. For spherical regions, the size of the subplot space is different at different levels of the whole plot factors, and thus a restriction should be imposed on the combined whole plot and subplot shrinkage levels (*w, x*). Figure 2.3 illustrates a spherical design space for a central composite design. It has one whole plot variable, *w*, and two

subplot variables, x_1 and x_2. The 15 design points are indicated by dots in the plot. All the design points, with the exception of the center runs, are the same distance from the center, thus the design has a spherical design region. In the 3-D VDG for this example, the whole plot space is the one-dimensional space indicated by w, and the subplot space is the two-dimensional spherical region summarized by x, i.e., $x_1^2 + x_2^2 = x^2$. The subplot spaces are observed having different volumes at different whole plot levels. Due to the constraint $w^2 + x_1^2 + x_2^2 \le 1$ on the combinations of factor levels, a constraint $w^2 + x^2 \le 1$ determines the edge of the combined design space. For example, the three locations on the edge of design region in 3-D VDGs are indicated by the pairs (w=$1/\sqrt{3}$, x=$\sqrt{2/3}$), (w=0, x=1) and (w=1, x=0) corresponding to the design points at whole plot and subplot factorial levels, the points at whole plot center and subplot axial levels, and the points at whole plot axial levels and subplot center, respectively. This adaptation is different from the Goldfarb *et al*. (2003) plots where all of the mixture-process spaces consisted of the same mixture space crossed with same process space.

To characterize the variation over the shrunken regions we can look at the maximum or the average scaled prediction variance (SPV). The maximum would allow us to see the worst prediction variance for each shrinkages combination and could be used to examine the *G*-efficiency of the design where the goal is to minimize the maximum prediction variance. The average SPV provides information that can be used for the *V* (or *IV* or *Q*) type criterion which seeks to minimize the average prediction variance over the entire region. The shapes of the prediction variance surfaces show the behavior of the prediction variance as we move throughout the whole plot and subplot spaces. The minimum variance, the third summary traditionally included in VDGs, is not considered here for the sake of simplicity in the graphical presentation and because it is typically of least importance.

In the 3-D VDGs, the average and maximum SPV are calculated at each specific combination of w and x. The constraint, $w^2 + x^2 = 1$, for the overall spherical design region is incorporated into the surface and contour plots automatically. The outer edge of

the contour plots is the projection of the surface edge into the *w*-*x* plane and indicates the edge of the design space. By fixing a value on the *x*-axis and moving along the *w*-axis, one can observe the effect that whole plot level changes have on the scaled prediction variance. Similarly we can understand the effect of level changes in the subplot space on SPV by moving along the *x*-axis at constant *w*-values. In industrial split-plot experiments, the variation introduced by whole plots is usually larger than the variation in subplots. Hence *d* values larger than one are primarily considered in the paper. The 3-D VDGs provide a visualization of changes in the distribution of the SPV for various values of variance component ratio, *d*, and therefore provide a helpful means of choosing good split-plot designs under a variety of practical settings. In the following sections, two types of central composite designs (CCD) with split-plot error structures are studied for illustrating the utility of 3-D VDGs for evaluating competing designs and for developing strategies for design improvement.

2.6 Example 1

The first example illustrated is a standard CCD with split-plot structure with one whole plot factor and two subplot factors. The design listed in Table 2.1 is a restricted split-plot design (Letsinger, Myers and Lentner 1996; Goos and Vandebroek 2004), where each whole plot level is contained in only one whole plot. This is a popular implementation of the central composite design when split-plot randomization is present. We assume a full second order model with the fixed effects modeled as

$$f(w, x_1, x_2) = \beta_0 + \beta_1 w + \beta_2 x_1 + \beta_3 x_2 + \beta_{12} w x_1 + \beta_{13} w x_2 + \beta_{23} x_1 x_2 + \beta_{11} w^2 + \beta_{22} x_1^2 + \beta_{33} x_2^2 .$$

Figures 2.4 and 2.5 display the 3-D VDGs for the maximum and average SPV, respectively, when *d*=0, 1 and 10. The maximum and average SPV are very similar throughout the design space, indicating how little variability there is in the prediction at any given location in the combined whole plot and subplot spaces. When *d*=0, the design is a CRD. In this case, the subplot and whole plot errors are indistinguishable. The design is almost rotatable under this situation since the exact axial levels for a rotatable design

are at $\pm\sqrt[4]{F} = \pm\sqrt[4]{8} = \pm 1.682$, where F is the number of factorial runs (Myers and Montgomery 2002, pp. 309). The rotatability manifests itself in circular curves centered in the middle of the design space. The smallest prediction variance occurs near the factorial points with shrinkage value $1/\sqrt{3} \approx 0.577$.

Table 2.1 Example 1: Standard CCD with one whole plot variable, w, and two subplot factors, x_1 and x_2. It has three center runs, axial levels of whole plot and subplot variables are $\pm\alpha = \pm\sqrt{3} \approx \pm 1.732$.

Whole plot	w_1	x_1	x_2	No. of runs per whole plot	Location in w-x plane in 3-D VDGs	
					w	x
1	-1	±1	±1	4	$\sqrt{1/3}$	$\sqrt{2/3}$
2	1	±1	±1	4		
3	1.732	0	0	1	1	0
4	-1.732	0	0	1		
5	0	±1.732	0	2	0	1
	0	0	±1.732	2		
	0	0	0	3	0	0

The plots indicate that the relative size of whole plot error variance to subplot error variance (d) plays an important role in the distributions of the average and maximum SPV. When $d \neq 0$, the SPV distribution is no longer symmetric in the whole plot and subplot spaces. As d increases, the maximum and average SPV distributions become flat along the subplot direction (x), inferring that the location in the subplot space makes relatively little difference in the magnitude of the SPV. Consequently, predictions are considered to be stable along the subplot space. For large values of d, we also observe that the contours of the surface plots are nearly vertical, indicating that most changes of the maximum and average SPVs are along the whole plot space. The SPV values are smallest close to the factorial points in the whole plot space regardless of the position of the point in the subplot space. The SPV increases when moving towards the axial points or the center along the whole plot (w) direction. The maximum SPV occurs at the edge of whole plot space when $d \neq 0$.

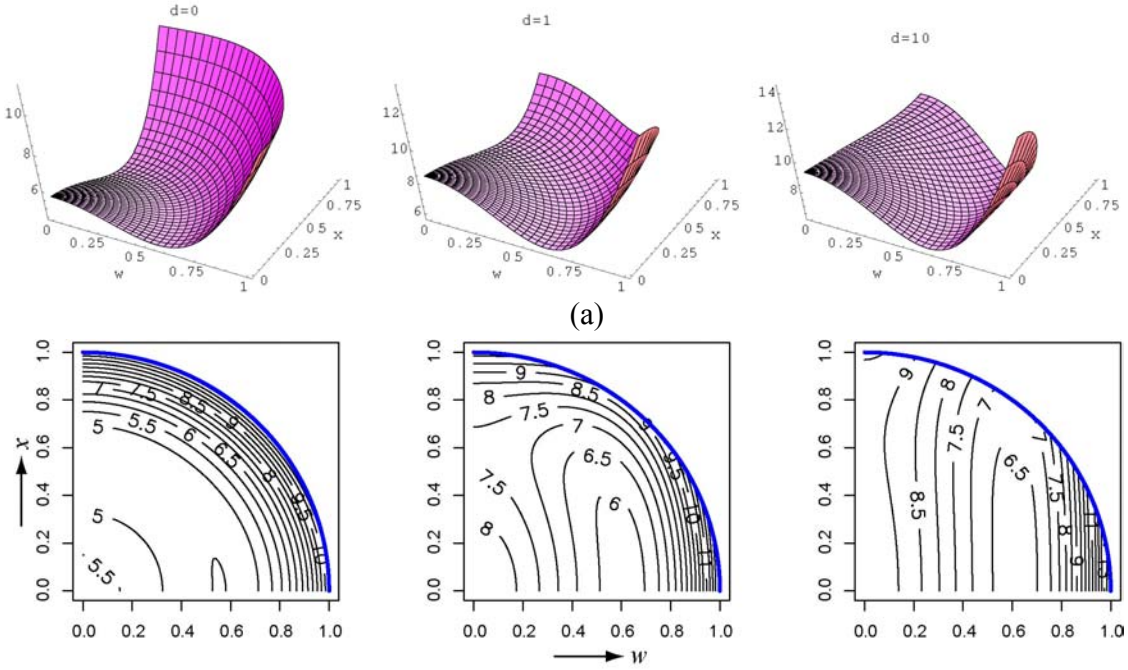


Figure 2.4 3-D VDGs of maximum SPV for the standard CCD in example 1, for d=0 (left), 1 (middle) and 10 (right). (a) Surface plots. (b) Contour plots.

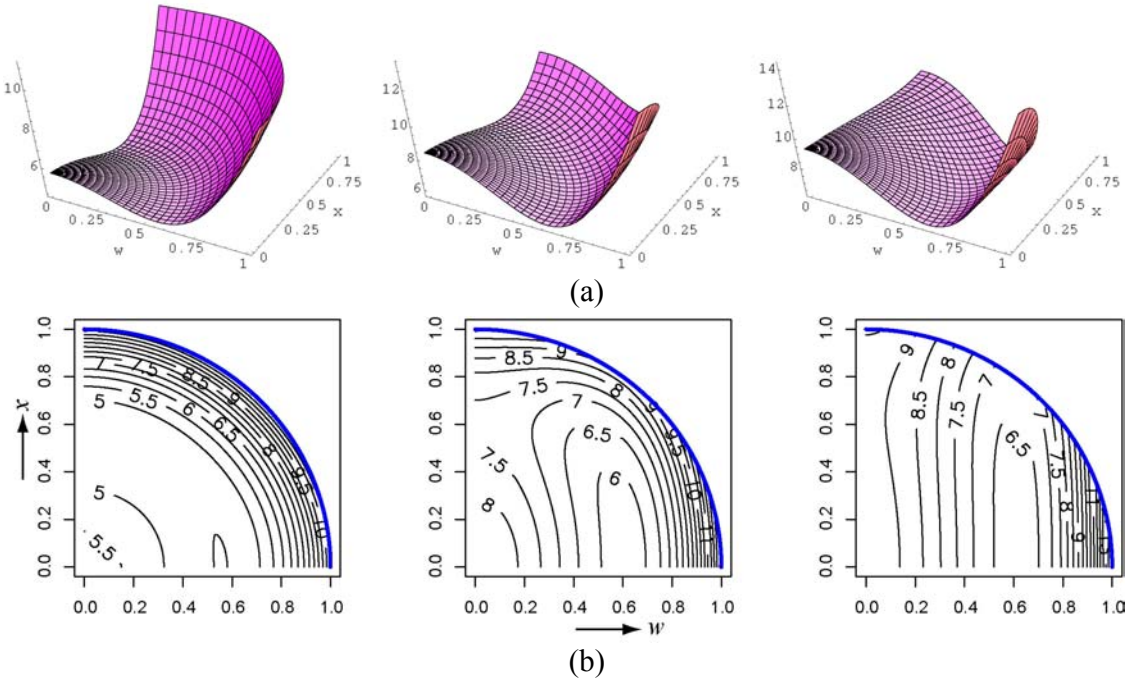


Figure 2.5 3-D VDGs of average SPV for the standard CCD in example 1, for d=0 (left), 1 (middle) and 10 (right). (a) Surface plots. (b) Contour plots.

For SPDs, the information for estimation of subplot terms (including subplot variables and interactions of whole plot and subplot variables) mainly comes from the contrasts of the subplot levels within each whole plot, and thus comparably precise estimates of subplot terms can be obtained as d increase. On the other hand, as d increases, whole plot terms are estimated less precisely. For large values of d, the subplot error variance contributes less to the SPV than the whole plot error variance. This is manifested in relatively smaller changes in the subplot contribution, which leads to stable SPV values along the subplot space but steep slopes along the whole plot space.

From the distributions of the maximum and average SPV, insight for improving the design can be obtained. For this design, the strategy for improving the prediction variance is to move design points such that the maximum or average prediction variance within the design region is lowered according to the practitioner's interest. The usual way of adjusting locations of design points is to adjust the axial levels for whole plot and subplot factors. If you fix the axial levels and adjust the factorial levels, this can be made equivalent (through recoding) to fixing the factorial levels and adjusting the axial levels. There are computational advantages by fixing the axial levels since the same sized region is obtained throughout. Symmetric factorial levels for each type of variables are assumed in the optimization.

Davison (1995) investigated the D- and V-efficiencies of the standard CCD with a split-plot error structure for the second order model. She showed that the standard CCD could be modified to better accommodate the bi-randomization scheme. In her work, V-optimal factorial values were obtained for different d values using the Nelder-Mead simplex algorithm. Since the 3-D VDGs provide information regarding the SPV distribution in the whole plot and subplot spaces separately, they are helpful for developing the optimization strategy of moving design points. It is also helpful for understanding how the design is improved and the price the optimal CCDs for one criterion has to pay in terms of another criterion.

2.6.1 *V*-optimal CCDs

When striving for a *V*-efficient design, one attempts to find a design with small average prediction variance over the entire design region. As seen from the 3-D VDGs in Figure 2.5 for the standard CCD, there are relatively high peaks' for the SPV at the center of the whole plot space. By moving design points toward the center of the whole plot space, it may be possible to improve prediction there. The locations of design points in the subplot space do not appear to affect the prediction variance as much as the locations in the whole plot space do. Table 2.2 lists optimal factorial levels, f1, for the whole plot variable and, f2, for the subplot variables for a range of possible values of *d*. In MATHEMATICA, the average values are calculated by integrating the scaled prediction variance over the entire spherical region. Then the optimal levels with the minimum average value can be obtained analytically. A constraint of $f_1^2 + 2f_2^2 \le 3$ is implemented for the whole plot and subplot factorial levels in order to keep all the design points inside the spherical region. Table 2.2 also provides relative efficiency values for evaluating the improvement obtained by using the optimal factorial levels instead of the standard levels ± 1. Note the ± 1 levels are common in practice but are not necessarily the most efficient levels for the factorial points. We define the relative efficiency as the ratio of the average scaled prediction variance of the CCD with optimal factorial levels to the average SPV of the standard CCD. Values less than one indicate improved *V*-efficiency. As *d* increases, the factorial points shift further away from the standard levels and yield increasingly efficient designs compared to the standard CCD. For instance, if the variance ratio is 10, the best factorial levels should be ± 0.51 for the whole plot variable and ± 1.17 for the subplot variables, which means the whole plot factorial points are moved much closer to the center of the whole plot space and the subplot factorial points are extended outwards. The average scaled prediction variance is reduced substantially by 15% by using the optimal levels (relative efficiency is 0.858 at d=10). The *V*-optimal levels here are virtually identical to those values proposed by Davison (1995).

Table 2.2: *V*-optimal factorial levels for example 1, f1 is the *V*-optimal whole plot factorial level, and f2 is the *V*-optimal subplot factorial level.

d	f1	f2	Relative Efficiency
0	1	1	1.0
1	0.78	1.09	0.965
10	0.51	1.17	0.858
50	0.34	1.20	0.816

The plots in Figure 2.6 reveal more information about how the optimization works. The contour plots in (b) indicate that the prediction variances at the center are reduced and the surfaces flatten since the whole plot factorial points are moved to the center. Figure 2.6(a) compares the average SPV for the *V*-optimal CCD to the average SPV of the standard CCD. The bottom surfaces at the center correspond to the *V*-optimal CCDs. The two surfaces cross at between 0.5 and 0.75 in the whole plot space (*w*), indicating that the trade-off in the improvement at center is an increase in the SPV values at the edge of the design space. Since the increase in SPV at the edge of the design space is offset by SPV decreasing at the broad center area, the average scaled prediction variance over the entire region is lowered. Also note that the maximum SPV for the *V*-optimal CCD is slightly larger than the maximum SPV for the design with standard levels.

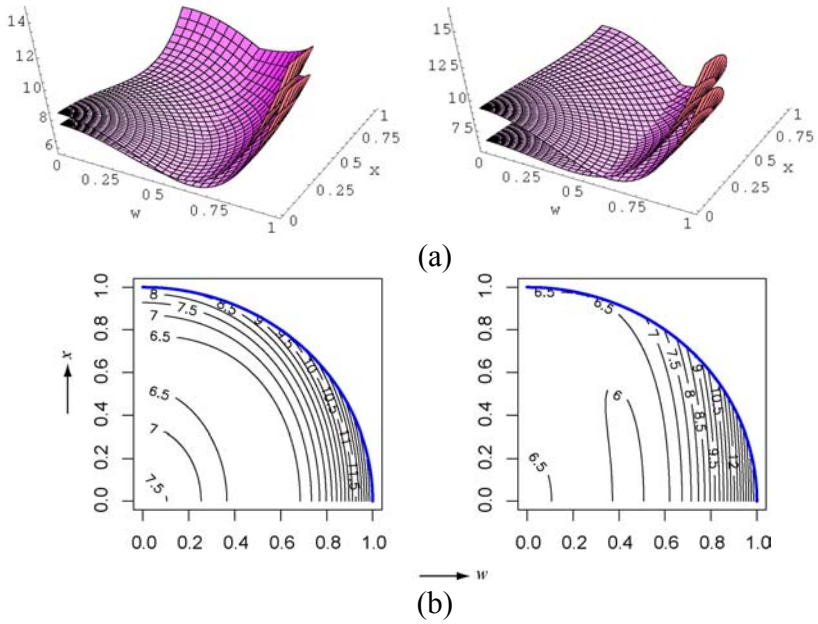


Figure 2.6 3-D VDGs of average SPV for the standard and *V*-optimal CCD in example 1, for *d*=1 (left) and 10 (right). (a) Surface plots of average SPV for the standard and *V*-optimal CCD. The two surfaces cross each other and the *V*-optimal CCD has smaller values at the center. (b) Contour plots of average SPV for the *V*-optimal CCD.

2.6.2 Robustness of the *V*-optimal CCDs to Changes of the Optimal Factorial Levels

Often the *V*-optimal CCDs may involve factor levels that are not practical to accommodate in production. As a result, one may be interested in the robustness of the design performance to near-optimal levels that might be more practical. This is especially true for the whole plot factors, whose levels are often hard to change. For instance, the designed optimal oven temperature may be 347.2 degrees. But due to practical constraints in setting the oven temperature, it may only be possible to adjust the temperature to 350 degrees. The desire is for the experiment to still be highly *V*-efficient even with slight alterations to the optimal factorial levels.

The average prediction variance for designs with factorial levels f1 and f2, different from the optimal levels, is studied for d=1 and 10 via Figure 2.7. The contour levels represent the relative *V*-efficiency – the ratio of the average SPV for designs with factorial levels f1 and f2 to the standard CCD with levels ± 1. The upper curve on the edge of the contour lines represents the constraint $f_1^2 + 2f_2^2 \leq 3$. Point (1,1) corresponds to the standard design and hence has a relative efficiency value equal to one. The dot on the edge of the shaded region indicates the optimal factorial levels. For instance, if d=1 (left plot), the optimal levels are f1=0.78, f2=1.09, and the corresponding relative *V*-efficiency is 0.965, the minimum relative efficiency for the optimization. The graph shows that designs with factorial levels within a wide range around the optimal levels are at least better than the standard CCD (i.e., relative efficiencies < 1).

For d=10, the optimal factorial levels are (f1=0.51, f2=1.17) and the minimum relative *V*-efficiency is 0.858 (Table 2.2). Suppose, for instance, the exact optimal factorial levels cannot be obtained and one can only operate at f1=0.5 and f2=1.1. In the right hand plot, this location is in the darker shaded region around the optimal levels, implying that the average scaled prediction variance of the design with (f1=0.5, f2=1.1) is at least 10% lower than the standard CCD. We can therefore conclude that the *V*-efficiency for the *V*-optimal CCD is robust to modest changes in the factorial levels about the optimal levels

and that highly *V*-efficient designs can be obtained by using the levels close to instead of exact optimal factorial levels.

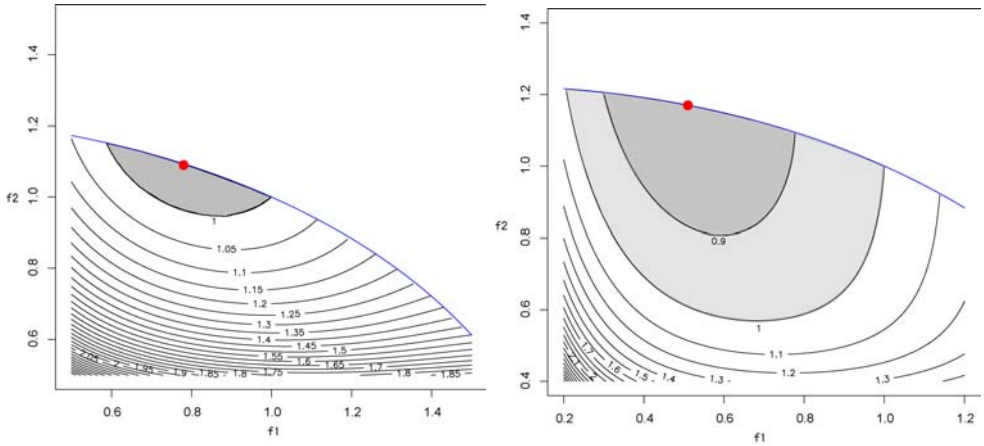


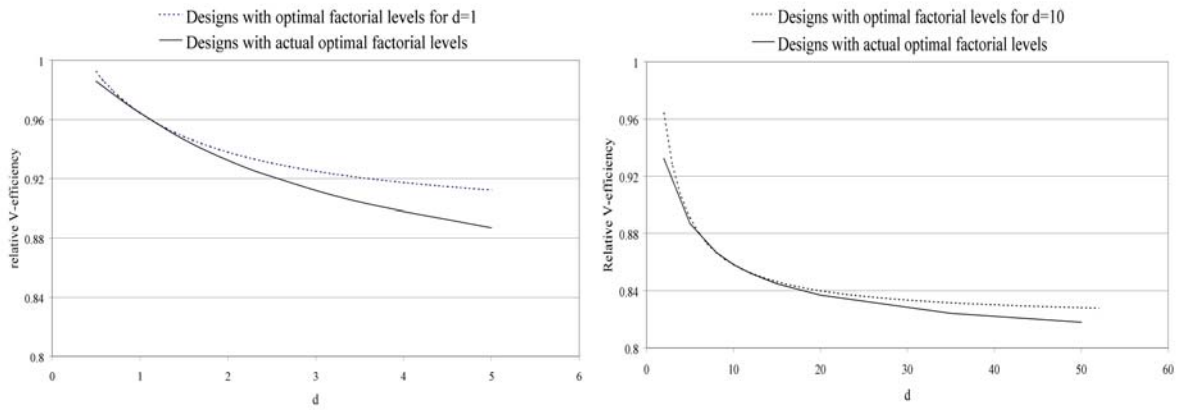
Figure 2.7 Contour plot of relative *V*-efficiency for the CCD with different whole plot (f1) and subplot factorial (f2) levels for d=1 (left) and 10 (right). "•": optimal factorial levels.

2.6.3 Robustness of the *V*-optimal CCDs to Changes in the Variance Component Ratio

Thus far, we have assumed known values of the variance component ratio, d. If it is possible to conduct pilot experiments, d may be estimated. Another option would be to guess the value of d based on the researcher's knowledge or based on field experiences. At times however, a pilot study is not feasible or the guess is far from the actual value of d. Consequently, it is important to study the robustness of the efficient designs to changes in d. One would expect that the optimal CCD for the guessed level of d would be less efficient than the optimal CCD for the actual value of d. In this section we examine the impact of misspecifying the value of d on the *V*-efficiency.

Figure 2.8 provides the framework for a study of robustness of the relative *V*-efficiency with respect to a misspecified variance component ratio. In Figure 2.8, the dashed curve denotes relative *V*-efficiencies for designs with specific factorial levels that are optimal only for an assumed d value (d=1 or 10), and the solid curve represents designs with actual optimal factorial levels for values of d ranging from 0.5 to 5 in (a) and 2 to 50 in (b). The relative efficiency denoted by the y-axis is defined as the ratio of average SPV of

the designs represented by the solid or dashed lines to the average SPV of the standard CCD. Lower values of relative efficiency indicate better designs. In Figure 2.8(a), the two curves meet at d=1, which implies that the whole plot factorial levels 0.78 and subplot factorial level 1.09 are optimal when the whole plot error variance is equal to the subplot error variance. At other values of d, if the design has factorial levels of f1=0.78 and f2=1.09, the design will be less efficient than the actual optimal CCDs which are represented by the solid curve. As d varies between 0.5 and 2, the efficiency reductions are negligible.



(a) (b)

Figure 2.8 Comparison of V-efficiencies for the design with optimal factorial levels based on the guessed d values to those for exactly V-optimal CCDs.

In Figure 2.8(b) similar results are observed. If the whole plot error variance is guessed to be roughly 10 times that of the subplot error variance, the optimal factorial levels are f1=0.51 and f2=1.17. It is possible that the actual variance ratio is between 5 and 30, values that are far from the guessed value. However, Figure 2.8(b) indicates that the efficiency of the design selected based on the guessed value of d is quite close to that of the optimal CCDs. Therefore, the improved V-efficient CCD has good V-efficiency even when there is an inaccurate guess or estimated value of d.

In summary, if one wishes to focus on optimizing the average SPV using V-efficiency, as d increases, the strategy for the CCD is to shift the whole plot factorial levels to the center of the design space. If he/she cannot run the exact optimal levels, then the V-efficiency of the improved CCD is quite robust to small changes in the optimal factorial levels. If the initial estimate of d is not accurate, the V-efficient CCD will still perform

close to optimal for a wide range of d values around the initial guess. It should be noted that the improvement in the overall average SPV comes with the price of exhibiting slightly higher maximum SPV (i.e. lower G-efficiency). Strategies for obtaining G-optimal CCDs are discussed in the next section.

2.6.4 *G*-optimal CCDs

Frequently, the practitioner is interested in keeping the worst prediction over the design region under control. In these situations, G-efficiency is a popular tool for comparing competing designs. Not only is the value of the maximum SPV important but it is also important to assess the location of the maximum SPV in the design space. In terms of protecting worst scenario, the 3-D VDGs can be quite informative tools for comparing and improving candidate designs. Since G-efficiency focuses on the area in the design region with the worst prediction (maximum SPV), while V-efficiency considers the overall performance of the SPV over the entire region, there are different strategies involved for obtaining G-efficient designs than there are for obtaining V-efficient designs.

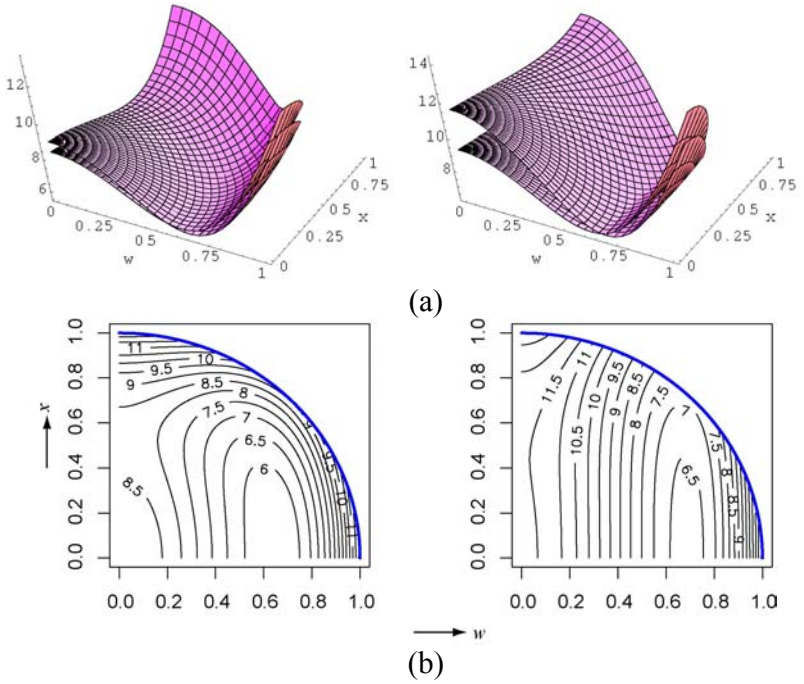


Figure 2.9 3-D VDGs of maximum SPV for the standard and G-optimal CCD in example 1, for d=1 (left) and 10 (right). (a) Surface plots of maximum SPV for the standard and G-optimal CCD. The two surfaces cross each other and the G-optimal CCD has larger values at the center. (b) Contour plots of maximum SPV for the G-optimal CCD.

We consider again the CCD. The *G*-optimal factorial levels for several values of *d* are listed in Table 2.3. The relative efficiency is the ratio of the maximum SPV of the *G*-optimal CCD to the maximum SPV of the standard CCD. Relative efficiency values less than one imply better designs than the standard CCD. The *G*-optimal factorial levels indicate a different strategy for design improvement than what was incorporated in finding *V*-optimal split-plot CCDs. The strategy then for obtaining a *G*-optimal CCD, is to move the whole plot factorial levels away from the center and then to move the subplot factorial levels towards the center.

Table 2.3 *G*-optimal factorial levels for example 1, h1 is the *G*-optimal whole plot factorial level, and h2 is the *G*-optimal subplot factorial level.

d	h1	h2	Relative Efficiency
0	1	1	1.0
1	1.13	0.92	0.927
10	1.22	0.80	0.897
50	1.25	0.75	0.884

Note that the *G*-optimal CCD improves the area with the maximum SPV by sacrificing the SPV at the center of the design space. The contours in Figure 2.9(b) show that the area with the maximum SPV for the *G*-optimal CCD is at the edge of the subplot space and center of the whole plot space. Comparing the location of the maximum SPV for the *G*-efficient CCD (seen in the contours of Figure 2.9) to the standard CCD (pictured in the contours of Figure 2.4) the locations of maximum SPV have shifted from (w=1, x=0) for the standard CCD to (w=0, x=1) for the *G*-optimal CCD. From Figure 2.9(a), we can observe that in the broad center area of the whole plot space the SPV values for the *G*-optimal CCDs are larger than those for the standard CCD. The implication is that the *G*-optimal CCDs have a larger average SPV and thus less *V*-efficient than the standard CCD. Utilizing plots of the distributions of the maximum and average SPV, the practitioner can discover the reasons why improving the design based on one criterion may lead to a sacrifice in the efficiency based on another criterion. The practitioner is now afforded the ability of developing a strategy of perhaps choosing a design that balances *G*- and *V*-efficiency values.

2.7 Example 2

In this example, we consider a standard CCD with two whole plot factors, w_1 and w_2, and two subplot factors, x_1 and x_2, for a full second order model (design points are listed in Table 2.4). The axial levels of all variables are $\alpha = \pm 2$, yielding a rotatable CCD when the run order is completely randomized. The design region is spherical.

Table 2.4 Example 2: standard CCD with two whole plot and two subplot factors. Its axial levels are ± 2 and 3 center runs.

Whole plot	w_1	w_2	x_1	x_2	No. of runs per whole plot
1	-1	-1	±1	±1	4
2	-1	1	±1	±1	4
3	1	-1	±1	±1	4
4	1	1	±1	±1	4
5	2	0	0	0	1
6	-2	0	0	0	1
7	0	2	0	0	1
8	0	-2	0	0	1
9	0	0	±2	0	2
	0	0	0	±2	2
	0	0	0	0	3

The maximum and average SPVs are displayed via the 3-D VDGs for d=0, 1 and 10 in Figure 2.10. Analytical calculations show large differences between the maximum and average SPVs at the edge of the whole plot space and thus large dispersion among the predictions variances in this area. Similar to Example 1, we observe that as d increases from 0, the distribution of the maximum and average SPV values is no longer symmetric in the whole plot and subplot spaces. For large values of d, the maximum and average SPV becomes quite stable in the subplot space (x), as evidenced by the nearly vertical contours in the right-most figures of 2.10(b) and (c). In the whole plot space, the maximum and average SPV increase rapidly at the edge, while from the center to the factorial levels of the whole plot space, the maximum and average SPV values are stable.

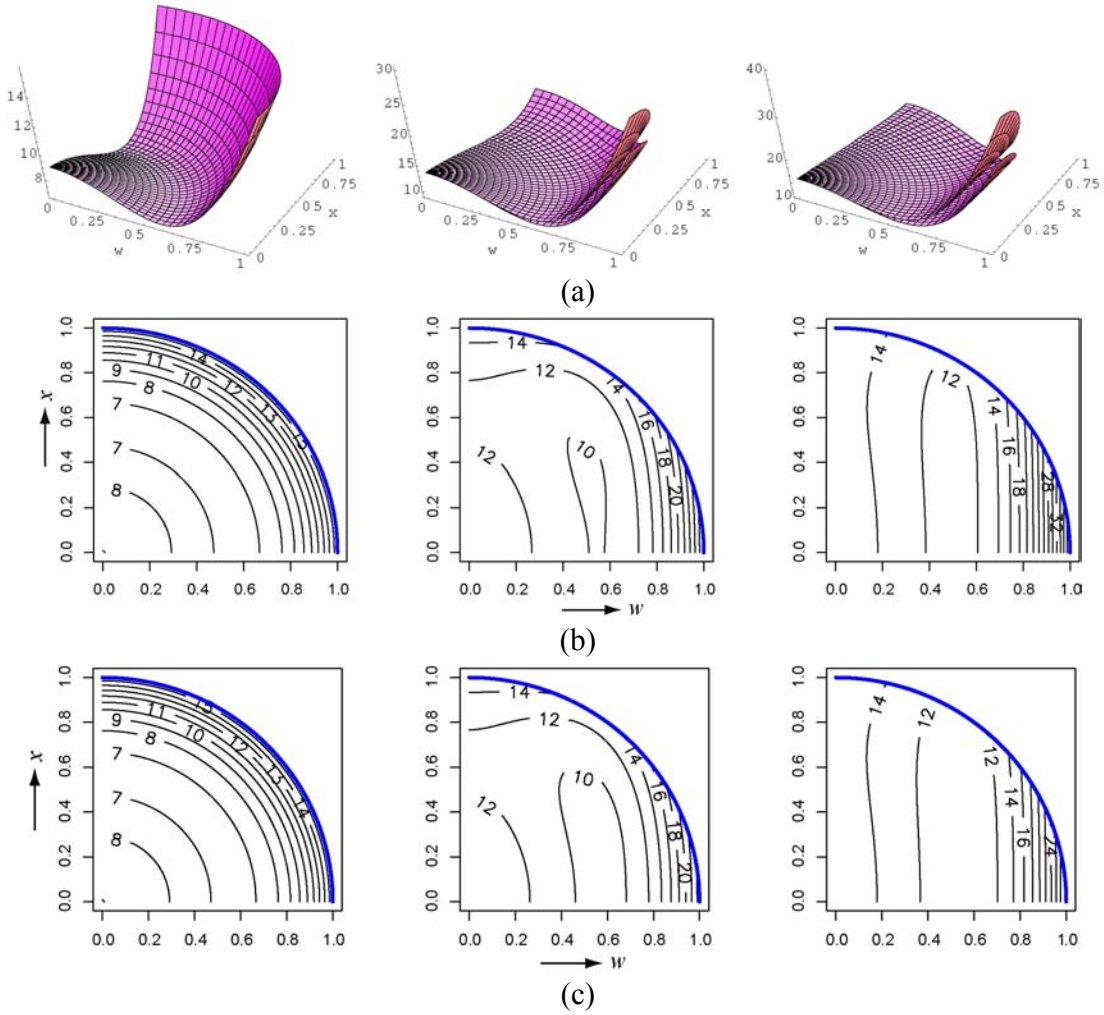


Figure 2.10 3-D VDGs for the standard CCD in example 2 for d=0 (left), 1 (middle) and 10 (right). (a) Surface plots of maximum and average SPV. (b) Contour plots of maximum SPV. (c) Contour plots of average SPV.

Compared to the design in Example 1 (one whole plot and two subplot variables), this design has a larger proportion of the region with smaller average and maximum SPV values. The V- and G-optimal factorial levels for this design are provided in Table 2.5. The strategy for improving the V- and G-efficiency of this design remains the same as in Example 1: for V-optimality, the whole plot factorial levels move in and subplot factorials move out, while the G-optimal whole plot factorial levels move outward and the G-optimal subplot levels move closer to the center. However, note that the V-optimal levels change little from the standard levels ± 1. Consequently, the V-efficiencies do not represent a significant improvement, inferring that the standard CCD for this case has

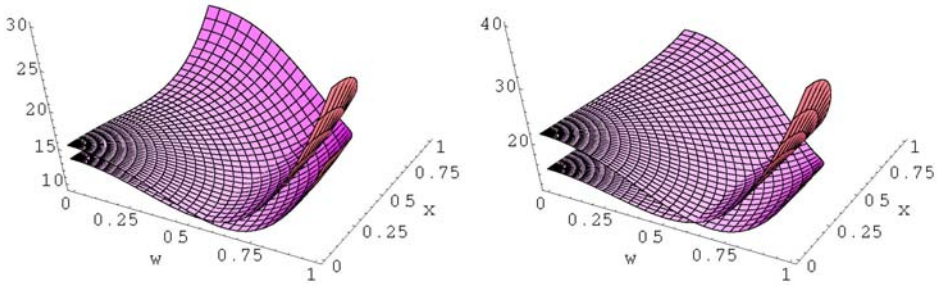
nearly optimal factorial levels. We do however note that the *G*-efficiency of this CCD with standard factorial levels can be improved substantially. Note also that the improvements in *G*-efficiency are much more substantial in this example than in the first example. For instance, when d=10, the *G*-optimal factorial levels are 1.21 for whole plot variables and 0.73 for subplot variables, yielding a relative efficiency of 0.602, which says that the maximum scaled prediction variance of the *G*-optimal CCD is almost 60% of that for the standard CCD, while in Example 1 the maximum SPV of the *G*-optimal CCD is 90% of that for the standard CCD and the optimal factorial levels for whole plot and subplot factors are 1.22 and 0.80, respectively.

Table 2.5 *V*- and *G*-optimal factorial levels for example 2, f1 and f2 are *V*-optimal whole plot and subplot factorial levels, respectively, and h1 and h2 for the *G*-optimal levels.

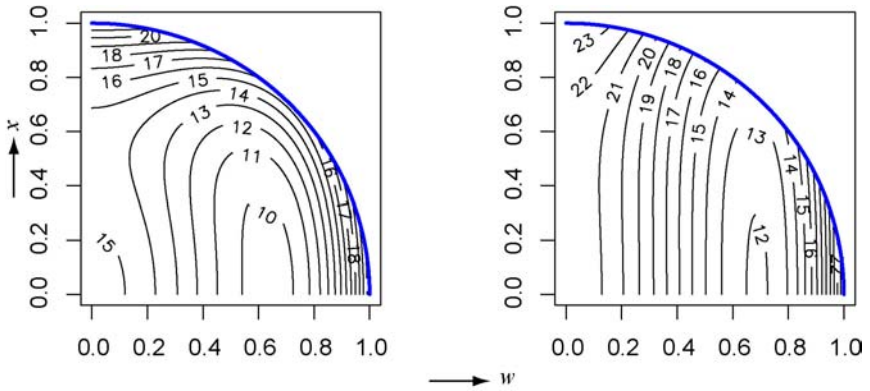
d	*V*-optimal			*G*-optimal		
	f1	f2	Relative efficiency	h1	h2	Relative efficiency
0	1	1	1	1	1	1
1	1	1	1	1.13	0.85	0.735
10	0.97	1.02	0.997	1.21	0.73	0.602
50	0.97	1.02	0.997	1.23	0.69	0.574

The plots in Figure 2.11 illustrate the differences in the distributions of the maximum SPV for the standard and *G*-optimal CCDs. The *G*-optimal CCDs have much lower SPV values at the edge of whole plot space. Also, the plots indicate a change in the location of the maximum SPV. The maximum values shown in Figure 2.11(b) occur at the edge of subplot space and the center of whole plot space (w=0, x=1) for the *G*-optimal CCDs. In the standard CCD, the maximum SPV occurs at the edge of whole plot space and the center of subplot space (w=1, x=0). Finally, we note that the *G*-optimal CCDs have much lower maximum SPVs than the standard CCD.

In summary, for the two whole plot variable and two subplot variables case, if the practitioner chooses to focus on overall prediction performance with *V*-optimality, the factorial levels for the standard CCD are nearly optimal. If however, one focus on protecting the worst prediction over the entire region (*G*-optimality), substantial improvements can be made by altering the design.



(a)



(b)

Figure 2.11 Surface plots of maximum SPV for the standard and *G*-optimal CCDs in example 2, for d=1 (left) and 10 (right). (a) Surface plots for the standard and *G*-optimal CCD. The two surfaces cross each other and the *G*-optimal CCD has larger values at the center. (b) Contour plots of maximum SPV for the *G*-optimal CCD.

2.8 Example 3

In split-plot experiments the major cost of a design is typically associated with changing the levels of whole plot variables. When a standard CCD is run within a split-plot randomization scheme as in the previous two examples, there is only one subplot within each of the axial level whole plots. Consequently, the standard CCD is not as efficient in terms of the additional cost of the whole plot units. In addition, the standard CCD has no replicates for the whole plot levels, making the whole plot pure error variance non-estimable. Vining, Kowalski and Montgomery (2004) [henceforth referred to as VKM] recommend that the practitioner consider a restriction on the minimum and maximum number of subplots within each whole plot in order to better utilize the available

experimental resources. They propose a modified CCD in which each whole plot contains the same number of subplots. When every whole plot contains an equal number of subplots, the design is said balanced. In addition, VKM only consider CCDs in which the ordinary least squares (OLS) estimates are equivalent to the generalized least squares (GLS) estimates. In this section we compare the modified CCD of VKM to the standard CCD when there are one whole plot variable and two subplot variables. The modified CCD is given in Table 2.6. We again assume a full second order model. Compared to the standard CCD, the modified CCD is a balanced SPD, where the balance is achieved by augmenting the whole plot axial levels with replicated center runs at the subplot levels as well as splitting the w=0 whole plot into separate whole plots for the subplot axial points and the subplot center runs. The standard CCD and modified CCD can be compared by observing Tables 2.1 and 2.6.

Table 2.6 Example 3: modified CCD with one whole plot variable and two subplot variables, axial level $\pm\alpha = \pm\sqrt{3} \approx \pm 1.732$.

Whole plot	w	x_1	x_2	No. of runs per whole plot
1	-1	±1	±1	4
2	1	±1	±1	4
3	1.732	0	0	4
4	-1.732	0	0	4
5	0	± 1.732	0	2
	0	0	± 1.732	2
6	0	0	0	4

The surface and contour plots of the maximum SPV for the modified CCD are provided in Figure 2.12. The plots of the average SPV are not provided due to the small differences between the maximum and the average SPVs across the design region. The 3-D VDGs allow the user to study the prediction performance across the design region as the variance ratio (d) changes from 0 (the CRD setting) to larger values. When d=0, the contours are not circular about the design center, indicating a loss of rotatability for the modified CCD. The worst SPV of 15.4 occurs at the center of whole plot space and the edge of subplot space (w=0, x=1).

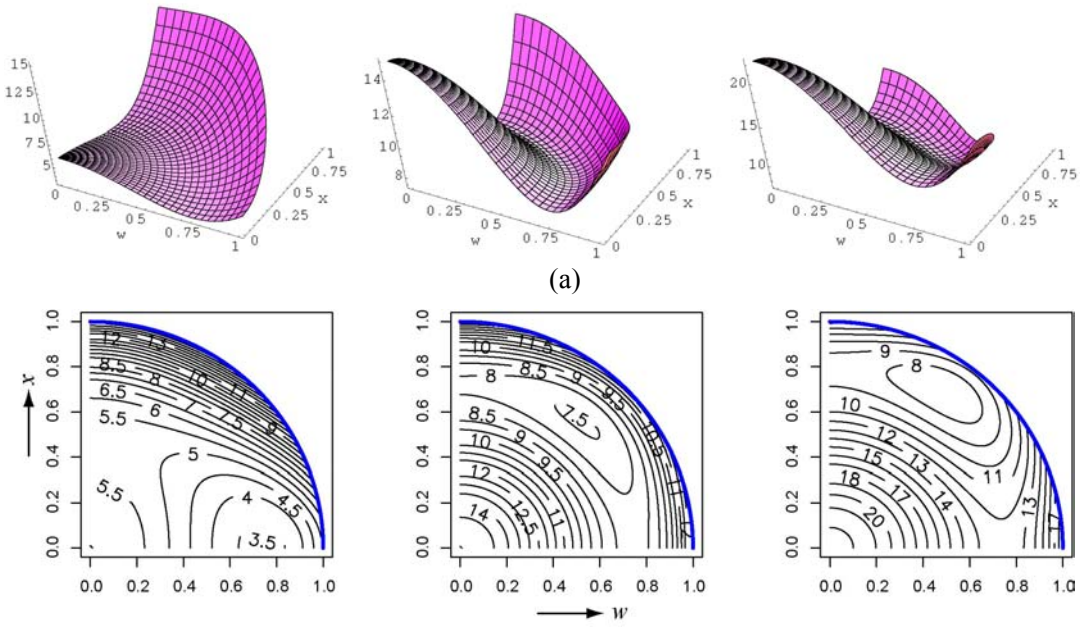


Figure 2.12 3-D VDGs of maximum SPV for the modified CCD in example 3, for d=0 (left), 1 (middle) and 10 (right). (a) Surface plots. (b) Contour plots.

Similar to the standard CCD with split-plot structure, changes in d play an important role in the distribution of the SPV for the modified CCD. For split-plot structures (i.e., d >0), the location with the worst prediction variance is at the centroid of entire spherical region (w=x=0). The distribution of the maximum SPVs along the subplot and whole plot spaces is nearly symmetric. As d increases, the maximum SPV values at the design center become greater than the values on the other region of the design space. Consequently, as d increases, the broad center area represents the locations with less precise prediction. This is generally not a desirable characteristic since the practitioner often expects to have better prediction performance at the center of the design space. The surface plots for the comparison of the standard CCD and the modified CCD are provided in Figure 2.13. The modified CCD, in terms of the maximum and average SPV, is less desirable than the standard CCD throughout most of the design region. As pointed out, this is especially true near the center of the design space and when the whole plot units account for a large proportion of the variability of observations (which results in larger values of d). This is due in part to the subplot quadratic terms being estimated exclusively with the whole plot error variance for modified CCD. Although the modified CCD of VKM has one more

whole plot than the standard CCD, its structure is less efficient than the restricted split-plot structure for CCD, thus the prediction performance of the modified CCD is less desirable than the standard CCD. More than the single number criteria, like *G*- and *V*-efficiencies, the 3-D VDGs provide the practitioner with a more complete comparison of the two adaptations of CCD in split-plot structure in terms of prediction variance.

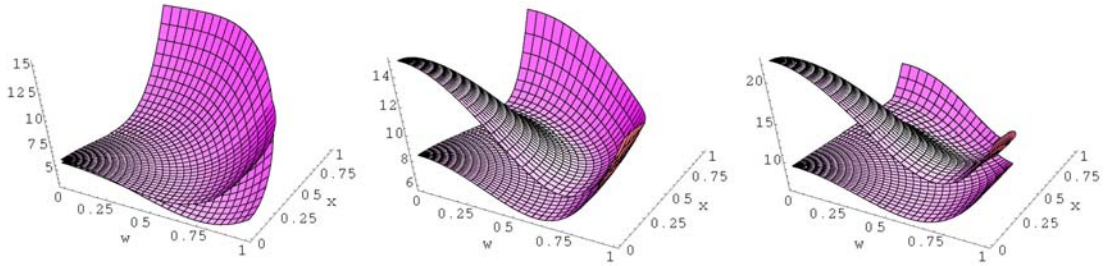


Figure 2.13 Surface plots of maximum SPV for the modified and the standard CCD with one whole plot variable and two subplot variables, for d=0 (left), 1 (middle) and 10 (right). The modified CCD has larger values in almost entire region.

2.9 Conclusions

Although a great deal of attention in the literature has been given to studying the prediction variance properties of CRDs, very little attention has been devoted to studying the prediction variance properties of SPDs. In terms of alphabetic optimality criteria, previous research involving strategies for determining optimal split-plot designs has focused mainly on parameter estimation (*D*-efficiency). Little work has been done regarding developing strategies for choosing split-plot designs on the basis of prediction variance properties (*G*- and *V*-efficiency). Although single number criteria such as the maximum SPV or average SPV are useful in comparing designs, graphical displays of the distribution of the maximum and average SPV over the entire design region provide a more complete comparison of designs. In this paper, we present the three-dimensional variance dispersion graphs (3-D VDGs) as a method of comparing the distributions of the maximum and average SPV for competing designs. Combining the contour and surface plots, we can examine the distribution of prediction variance throughout the entire design space as well as for the whole plot and subplot spaces individually. 3-D VDGs provide a

simple visual way for better understanding the prediction performance of split-plot designs in design space, especially when one is interested in the strengths and weaknesses for a candidate design in a specific area of the design space.

Using 3-D VDGs, we have demonstrated that it is possible to improve existing designs in terms of their prediction variance properties. In the examples provided, we have demonstrated that the 3-D VDGs offer the development of intuitive strategies for improving the prediction variance properties of the popularly used CCD. We have also demonstrated that our *G*- and *V*-optimal CCDs are quite robust to changes in the optimal factorial levels as well as to misspecifications of the variance component ratio.

Although all the examples illustrated in this paper involve the spherical design space, the same methodology of construction and interpretation can easily be applied to cuboidal regions and constrained design spaces. When the design region is cuboidal, shrunken cubes are considered and no constraint is imposed on the combinations of the whole plot and subplot shrinkage factors. When the design space is a constrained region, the shrunken regions are located on a sub-region of the full simplex, and suitable constraints might be considered for combinations of the shrinkage factors. In addition, other functions of the prediction variance can easily be plotted using the 3-D VDGs. Other examples of cost-based optimality criteria based on *G*- and *V*-efficiencies are given in Liang et al (2005). This flexibility can make the use of 3-D VDGs adaptable to a broad range of optimization problems.

Chapter 3 Fraction of Design Space (FDS) Plots

Abstract

In industrial experiments, restrictions on the execution of the experimental runs or existence of one or more hard-to-change factors often lead to split-plot experiments, where there are two types of experimental units and two independent randomizations. The resulting compound symmetric error structure and the setting of whole plots play important role in the performance of split-plot experiments. When the practitioner is interested in predicting the response, a response surface design for a second order model such as a central composite design (CCD) is often applied. The prediction variance of these designs under split-plot structure can be of interest. In this paper, fraction of design space (FDS) plots are adapted to split-plot designs. In addition to the global curve exploring the entire design space, sliced curves at various whole plot levels are presented to study prediction performance for sub regions in the design space. The different sizes of the constrained sub regions are accounted for by the proportional size of the sliced curves. The construction and use of the FDS plots are demonstrated through two examples of the standard CCD in split-plot schemes. The roles of the two types of variables in the prediction variance under different values of variance component ratio are also discussed.

3.1 Introduction

Many industrial experiments involve two types of factors, some with levels that are hard or costly to change and others with levels that are relatively easy to change. When hard-to-change factors exist, the experimenter may want to reduce the number of times the levels of these factors are changed. A common strategy is to run all combinations of the easy-to-change factors for a given setting of the hard-to-change factors. Such a strategy results in a split-plot design (SPD) and is frequently more realistic than complete randomization of the desired factor level combinations, which results in a completely

randomized design (CRD). In split-plot experiments, there are two types of experimental units – whole plots and subplots, and the levels of hard-to-change factors are randomly applied to the whole plot units. Within each whole plot, the levels of easy-to-change factors are randomly assigned to subplot units. The separate randomizations lead to the correlated observations within the same whole plot and thus a compound symmetric error structure, which must be accounted for not only when conducting inferences but also when determining an optimal design. Examples of split-plot designs are given in Letsinger, Myers and Lentner (1996) and Kowalski and Vining (2001).

When the experiment is run as a split-plot design with a whole plots, the following linear mixed model can be written to explain the variation in the $N\times 1$ response vector, **y,**

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}.$$

Regarding notation, $\boldsymbol{\beta}$ is the $p\times 1$ vector of fixed effect model parameters including the intercept, **X** is the $N\times p$ fixed effects design matrix; **Z** is an $N\times a$ incidence matrix of ones and zeroes where the ji^{th} entry is 1 if the j^{th} observation (j=1,...,N) belongs to the i^{th} whole plot (i=1,…,a); $\mathbf{u}$ is an $a\times 1$ vector of random effects for whole plot units where the elements are assumed *i.i.d* $\mathrm{N}\left(0,\sigma_\delta^2\right)$ with σ_δ^2 denoting the variability among whole plots; $\boldsymbol{\varepsilon}$ is the $N\times 1$ vector of residual errors where the elements are assumed *i.i.d* $\mathrm{N}\left(0,\sigma_\varepsilon^2\right)$ and σ_ε^2 denotes the variation among subplot units. It is also assumed that $\boldsymbol{\delta}$ and $\boldsymbol{\varepsilon}$ are independent.

The covariance matrix of the responses in a split-plot design is

$$\mathrm{Var}\left(\mathbf{y}\right) = \boldsymbol{\Sigma} = \sigma_\delta^2\mathbf{Z}\mathbf{Z}' + \sigma_\varepsilon^2\mathbf{I}_N = \sigma_\varepsilon^2\left[d\mathbf{Z}\mathbf{Z}' + \mathbf{I}_N\right]$$

where $\mathbf{I}_N$ is an $N\times N$ identity matrix and $d=\sigma_\delta^2/\sigma_\varepsilon^2$ represents the variance component ratio. For simplicity of presentation, we will assume that the observations are sorted by the whole plots, implying $\mathbf{Z} = diagonal\{\mathbf{1}_{n_1},\cdots,\mathbf{1}_{n_a}\}$, where $\mathbf{1}_{n_i}$ is an $n_i\times 1$ vector of one's and n_i is the size of the i^{th} whole plot. Assuming the diagonal form of $\mathbf{Z}$ allows one to conveniently write the covariance matrix as

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \boldsymbol{\Sigma}_a \end{bmatrix}$$

where each $n_i \times n_i$ matrix $\boldsymbol{\Sigma}_i$ is given by

$$\boldsymbol{\Sigma}_i = \begin{bmatrix} \sigma_\varepsilon^2 + \sigma_\delta^2 & \cdots & \sigma_\delta^2 \\ \vdots & \ddots & \vdots \\ \sigma_\delta^2 & \cdots & \sigma_\varepsilon^2 + \sigma_\delta^2 \end{bmatrix}$$

and $\boldsymbol{\Sigma}_i$ denotes the covariance matrix of responses for the i^{th} whole plot. Note that the variance of an individual observation is the sum of the subplot and whole plot error variances, $\sigma_\varepsilon^2 + \sigma_\delta^2$. A popular method for estimating the variance components is restricted maximum likelihood (REML).

The vector of fixed effects parameters, $\boldsymbol{\beta}$, is estimated via generalized least squares, yielding

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X} \right)^{-1} \mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{y} .$$

The covariance matrix of the estimated model coefficients is given by

$$Var\left(\hat{\boldsymbol{\beta}} \right) = \left(\mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X} \right)^{-1} .$$

The predicted mean response at any location $\mathbf{x}_0$ is given by

$$\hat{y}\left(\mathbf{x}_0 \right) = \mathbf{x}_0' \hat{\boldsymbol{\beta}},$$

where $\mathbf{x}_0$ is the point of interest in the design space expanded to model form. The prediction variance at $\mathbf{x}_0$ is given by

$$Var\left[\hat{y}\left(\mathbf{x}_0 \right) \right] = \mathbf{x}_0' \left[\mathbf{X}' \boldsymbol{\Sigma}^{-1} \mathbf{X} \right]^{-1} \mathbf{x}_0 .$$

When the design is completely randomized, $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}$ and the optimal design depends only on the settings of the factor levels in **X**. In considering optimal split-plot designs, the situation becomes more complex than it is for CRDs due to the complexities introduced by the more general form of $\boldsymbol{\Sigma}$. For instance, if one is interested in the best split-plot design in terms of model prediction, the prediction is not only a function of the settings of

the factors in **X** but is also a function of the number of whole plots, the variance component ratio, *d*, the dimensionality of each of the $\sum_j$ (determined by the number of subplots within each whole plot), and the arrangement of subplot levels within whole plots.

Often when considering prediction variance as an objective function for determining an optimal design, the prediction variance is scaled by the variance of observational error to make the quantity scale free and by the design size in order to penalize larger designs. Similar to the scaled prediction variance (SPV) calculation equation for CRDs (Myers and Montgomery, 2002), the scaled prediction variance for split-plot designs is obtained by multiplying the prediction variance by the total number of runs, *N*, and then dividing by the observational error variance ($\sigma_\delta^2 + \sigma_\varepsilon^2$). Thus for SPDs we have

$$\text{SPV} = N\mathbf{x}_0'\{\mathbf{X}'\sum^{-1}\mathbf{X}\}^{-1}\mathbf{x}_0 \big/ (\sigma_\varepsilon^2 + \sigma_\delta^2) = N\mathbf{x}_0'[\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}]^{-1}\mathbf{x}_0,$$

where $\mathbf{R} = diagonal\{\mathbf{R}_1, \cdots, \mathbf{R}_a\}$ with $\mathbf{R}_i$ denoting the correlation matrix of observations within whole plot *i*.

The correlation matrix, **R**, is a function of the unknown variance component ratio, *d*. The performance of a SPD in terms of prediction variance is strongly related to the value of the variance component ratio, *d*. This range of *d* values is based primarily on *d* values that have been reported in the literature for industrial split-plot experiments. Bisgaard and Steinberg (1997) stated that the whole plot variance is usually larger than subplot variance in prototype experiments. Letsinger et al. (1996) studied a split-plot experiment in chemical industry with $\sigma_\delta^2 / \sigma_\varepsilon^2 = 1.04$. Vining, Kowalski and Montgomery (2004) estimated the variance terms using pure error and reported a variance ratio 5.65. Webb, Lucas and Borkowski (2002) described an experiment with a variance ratio of 6.92 in a computer component manufacturing experiment. Kowalski, Cornell and Vining (2002) studied a mixture experiment with process variables where the estimated variance ratio was 0.82. In this chapter, we consider *d* values of 0 (CRD), 1 and 10, representing situations in which the whole plot is not distinguished from subplot, the whole plot error variance is the same and ten times that of the subplot error variance, respectively.

Graphical tools

Alphabetical criteria, like *D*-, *G*-, or *V*-optimality criteria, are often employed when evaluating designs, however, the decision of a best design is typically more complicated than can be completely summarized by a single number. For instance, if the practitioner is interested in prediction, the *V*-criterion, for example, yields information concerning the average prediction variance over the design region but single value is incapable of providing information regarding whether or not the prediction variance is stable in the entire design region or which regions exhibit best and worst prediction precision. In these situations, graphical tools are more informative for the exploration of properties of competing designs and allow for more detailed understanding of prediction performance throughout the design region. Variance dispersion graphs (VDGs) and fraction of design space (FDS) plots are two popularly used graphical tools for comparing designs. VDGs were proposed by Giovannitti-Jensen and Myers (1989) for CRDs with spherical design region. In the plots, minimum, maximum and average SPV values are plotted versus the radius of the shrunken spheres. VDGs have been extended to designs with cuboidal and irregularly constrained design regions via the use of shrinkage factors, i.e., multipliers of the original edges of the design space. The shrinkage factors range from 0 to 1, with 0 indicating the design center and 1 representing the outer perimeter of the given space (see Khuri, Harrison and Cornell 1999 regarding the use of shrinkage factors). Liang, Anderson-Cook, Robinson and Myers (2004) utilize the shrinkage concept for the creation of three-dimensional variance dispersion graphs (3-D VDGs) for studying the distribution of prediction variance in a split-plot design space. In 3-D VDGs, the average or maximum SPV values are plotted versus combined shrinkage values for whole plot and subplot spaces. This allows the practitioner to examine the prediction variance in the whole plot and subplot spaces separately. 3-D VDGs have been shown to be useful for evaluating and comparing prediction performance of split-plot designs.

Another tool for studying the distribution of prediction variance across the design space is the fraction of design space (FDS) plot. The fraction of design space (FDS) plot was introduced by Zahran, Anderson-Cook and Myers (2003) as a complement to VDGs.

While VDGs are informative in terms of the distribution of the SPV in the design space, FDS plots present the cumulative distribution function of all of the SPV values from the entire space, showing clearly the minimum, maximum and quantiles of the SPV distribution. In the FDS plot, SPV values are plotted versus the fraction of the design space that has SPV values at or below a given value. Goldfarb, Anderson-Cook, Borror and Montgomery (2004) developed FDS plots for mixture and mixture-process experiments with irregular regions. Since mixture-process experiments typically involve different characteristics of the SPV over the mixture and process spaces, Goldfarb et al. demonstrated that sliced FDS plots can be constructed to allow for more detailed views of the different sub-spaces.

3.2 FDS Plots for Split-Plot Designs

In considering FDS plots for split-plot designs, one can conceptualize a global plot for the entire design region or sliced plots for the whole plot or subplot spaces individually. Depending on the interest of the practitioner (global or regional), the plot is constructed by sampling the design space. To construct an FDS plot which characterizes the global behavior of the SPV over the entire design space, a large number of points, say n, are uniformly randomly sampled from the design space. SPV values are then calculated at each point in the sample via the equation $N\mathbf{x}_0'[\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}]^{-1}\mathbf{x}_0$ where $\mathbf{x}_0$ is a given location in the design space. The obtained SPV values are then ordered and plotted against the quantiles $(1/n, 2/n,\ldots,1)$. In the resulting curve, the minimum and maximum SPV values are displayed at FDS=0 and 1, respectively. The slope of the curve indicates how quickly the SPV reaches its maximum value in the design space. A desirable characteristic of a design is that its SPV values are minimized. In terms of the FDS plot, this characteristic manifests itself in an FDS curve that has small SPV values with a relatively flat slope, inferring a stable SPV distribution. Using global FDS plots to compare designs, curves with lower values and flatter slopes denote better designs.

Since split-plot designs involve two sub spaces, namely the whole plot space and subplot space, it is not unreasonable to expect the SPV values to change according to different patterns as we move around in the whole plot or subplot spaces. One reason for the existence of different patterns in the whole plot and subplot spaces is the impact of the variance component ratio, *d*. As *d* increases (i.e. whole plot error variance increases in relative importance in terms of observational error variance), whole plot factor terms are estimated less precisely. This manifests itself in different slopes of prediction variance profiles along the whole plot and subplot spaces. Therefore, when comparing designs, it may be informative to examine the difference in SPV profiles within the whole plot and subplot spaces. Sliced FDS curves can be constructed to examine the SPV distribution within the subplot space at any specified whole plot shrinkage level. To construct these plots, points are uniformly sampled throughout the subplot space with the whole plot factor fixed at a given level. Alternately, if one were interested in profiles in the whole plot space for a given subplot level, sampling could take place with points selected uniformly throughout the whole plot space for the given subplot level.

In producing sliced FDS plots it is important to keep in mind the restriction of the spherical nature of the overall design space. For example, consider an experiment with one whole plot factor, *w*, and two subplot factors, x_1 and x_2. For a cuboidal region, regardless of the shrinkage value of *w* selected for the whole plot factor, the valid subplot space remains a square region with $0 \le x_1 \le 1$ and $0 \le x_2 \le 1$. If a spherical region with radius one is of interest and the whole plot shrinkage factor is specified at a value *w*, the valid subplot space is the circle with $0 \le x_1^2 + x_2^2 \le 1 - w^2$ due to the constraint $w^2 + x_1^2 + x_2^2 \le 1$ in the spherical design region. Therefore, at different whole plot levels, the subplot spaces have different sized sub regions in the design space. In order to reflect the proper area explored by the slices, the range of FDS values on the *x*-axis are shown to be proportional to the volume of the studied subplot space for the sliced curve at any specified whole plot shrinkage level. To do this we divide it by a standard volume to display how large the subspaces are relative to each other. The maximum volume of the subplot space among all the whole plot levels is used as the standardizing volume for

sliced subplot spaces. Using this principle for a cuboidal region, the proportional size of all of the subplot spaces and the length of corresponding slice curves are one. However, for spherical regions, the sizes of the subplot spaces at different whole plot levels may not be the same. The maximum subplot space occurs at w=0 with the corresponding subplot space as the circle $x_1^2 + x_2^2 \le 1$ and thus the standardizing subplot space volume is π. After standardizing (dividing by π), the lengths of all of the sliced FDS curves are different for the various whole plot levels. To aid in understanding the principle of different sliced curves at different whole plot levels, consider the information in Table 3.1 and the corresponding design space plotted in Figure 3.1, which shows the design space of a central composite design for one whole plot factor, w, and two subplot factors, x_1 and x_2 (settings provided in Table 3.2).

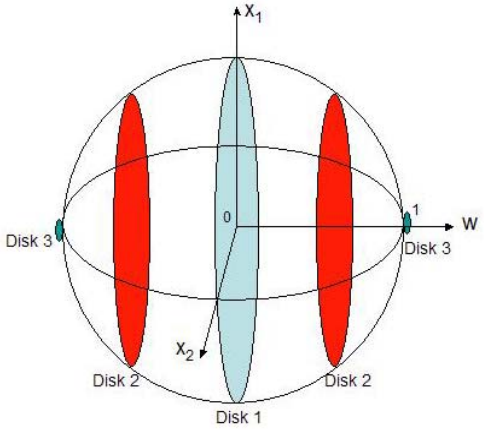


Figure 3.1 Spherical space for a design with one whole plot and two subplot variables.

The axial levels for the three factors are $\sqrt{3}$ and thus the design has a spherical design region. All shrinkage levels displayed in Table 3.1 are scaled such than the outermost design points are on the sphere with radius 1. The subplot space is largest at the center of whole plot space, hence the resulting sliced FDS curve at w=0 will extend to FDS=1. The smallest slice is at w=$\pm$ 1 representing the axial points at the edge of whole plot space. The corresponding subplot space is only a single point and hence the resulting FDS curves at w=$\pm$ 1 will be displayed as a dot. The shrinkage values $w = \pm 1/\sqrt{3}$ correspond

to locations with the factorial levels in the whole plot space. The subplot spaces at these locations have proportional volume 2/3, thus the resulting FDS curve will extend to 2/3.

Table 3.1 The proportional sizes of the sliced FDS curves at different whole plot shrinkage values for a design with spherical region as in Figure 3.1.

Whole plot shrinkage value (w)	Location in the whole plot space	Possible values subplot shrinkage values (x_1, x_2)	Region in Figure 3.1	Volume of the disk	Proportional size of the curve
0	center	$x_1^2 + x_2^2 \le 1$	disk 1	π	1
$\pm 1/\sqrt{3}$	factorial	$x_1^2 + x_2^2 \le 2/3$	disk 2	$2\pi/3$	2/3
± 1	axial	$x_1 = 0, x_2 = 0$	disk 3	0	0

Sliced FDS curves can be used to examine the SPV distribution in the subplot or whole plot spaces, depending on the interests of the practitioner. If interest lies in the behavior of the SPV throughout the subplot space, a flat FDS slice would indicate that SPV values are relatively stable in the subplot space at the given whole plot shrinkage level. From slices at different whole plot levels, one can observe the regions with good or poor prediction in both the whole plot and subplot spaces. The top slice (associated with largest SPV values) corresponds to the whole plot region with maximum SPV. This information can aid the practitioner who wishes to improve the design by augmenting with additional design points since the slices can help the practitioner focus on regions where attention should be focused. The slice plots can also be used to judge which of the two spaces, whole plot space or subplot space, contributes more to changes in the SPV values. If the slices at different whole plot levels are flat and far apart, the whole plot space has a larger effect on the SPV distribution. If the slices have rapidly increasing slopes and are close to each other and to the global FDS curve, the location in the subplot space has a larger effect on the SPV values, which also means that the SPV is relatively stable in the whole plot spaces and robust to the changes of the whole plot levels. If the slices are somewhere between horizontal and steeply sloped, then both whole plot and subplot spaces contribute to changes in the SPV values. Therefore, from the relative positions of the slices, information about the SPV distribution in the whole plot space can be provided.

It is important to note that FDS plots are not meant to compete with 3-D VDGs when evaluating split-plot designs but are to be viewed as an additional tool for making informative comparisons. In the following sections, two examples involving a split-plot CCD are studied to illustrate the utility of global and sliced FDS plots. We also seek to illustrate how the FDS plots and 3-D VDGs can be used together for obtaining a better understanding of prediction variance properties of SPDs.

3.3 Example 1

This example considers a central composite design (CCD) in a split-plot setting with one whole plot factor, w, and two subplot factors, x_1 and x_2. A spherical design region is assumed and the following full second order model in the fixed effects is studied

$$f(w, x_1, x_2) = \beta_0 + \beta_1 w + \beta_2 x_1 + \beta_3 x_2 + \beta_{12} w x_1 + \beta_{13} w x_2 + \beta_{23} x_1 x_2 + \beta_{11} w^2 + \beta_{22} x_1^2 + \beta_{33} x_2^2 .$$

The factor combinations are listed in Table 3.2. Note that each whole plot level is contained in only one whole plot in this design. It is a popular set-up for the central composite design run in a split-plot setting.

Table 3.2 Standard CCD with one whole plot variable and two subplot variables. It has axial level $= \pm\sqrt{3} \approx \pm 1.732$. The levels in the table are given in the standard form.

Whole plot	w	x_1	x_2	No. of runs in the whole plot
1	1	±1	±1	4
2	-1	±1	±1	4
3	1.732	0	0	1
4	-1.732	0	0	1
5	0	± 1.732	0	2
	0	0	± 1.732	2
	0	0	0	3

The global FDS curves for the standard CCD and four sliced curves for the subplot space at specified whole plot shrinkage values are presented in Figure 3.2 for different values of the variance ratio, d=0, 1 and 10. The SPV distribution across the subplot space is plotted for each of the four whole plot shrinkage values, w=0, ± 0.29, ± 0.58 and ± 1. These slices display the SPV distribution across the subplot space at the center, between the

center and the factorial levels, the factorial levels, and the axial levels in the whole plot space, respectively. The four corresponding lengths of the curves are 1, 0.92, 0.67 and 0 to reflect the volume of each sub region being examined.

The minimum and maximum scaled prediction variance in the design space and the distribution profile can be read from the global FDS curves. For instance, when d=10, the minimum SPV value of 6.2 is observed at FDS=0. The maximum value of 14.3 is observed at FDS=1. Note that the maximum SPV is related to the design's G-efficiency. For the standard CCD with d=0, corresponding to a CRD, the global FDS curve has a diagonal slope, implying that the SPV increases gradually across the design region. As d increases, the global curves have relatively flat slopes in more than 95% of the design space, and then increase rapidly in a tiny portion of the design space. This implies that the proportion of the design region with small SPV values increases as d increase.

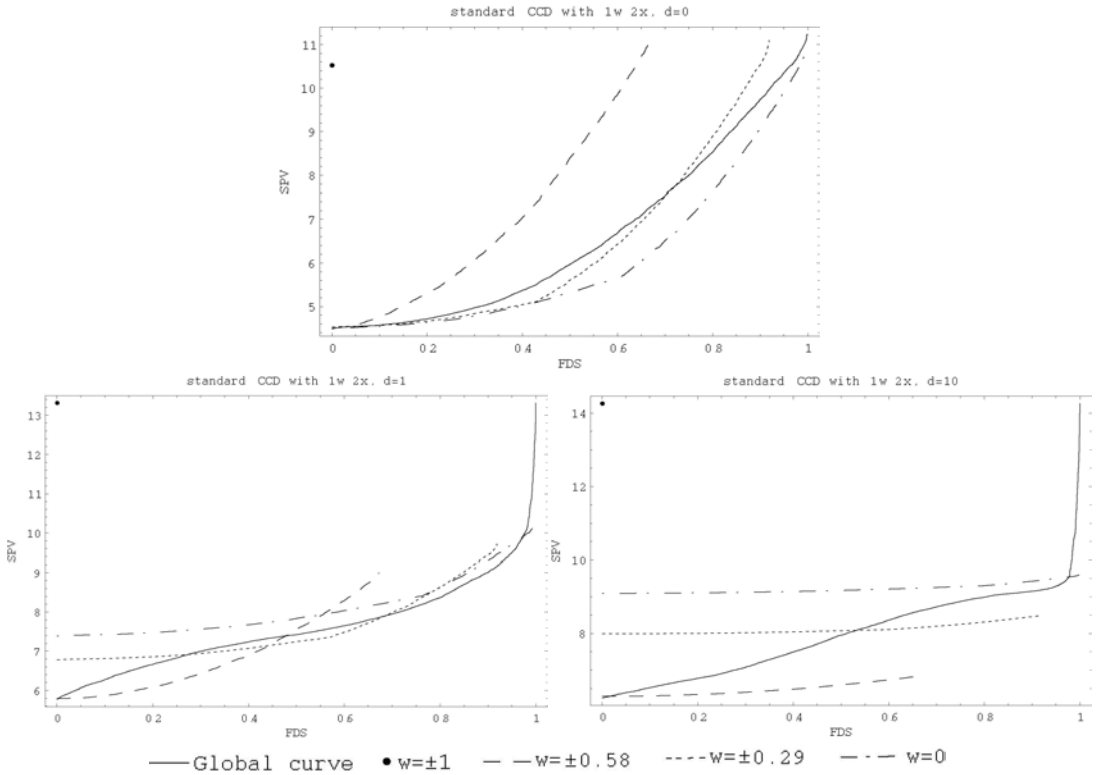


Figure 3.2 Global and sliced FDS plots for the standard CCD with one whole plot variable and two subplot variables.

For d=0, the slices are dispersed on both sides of the global FDS curve. All of the slices except for the dot at $w=\pm 1$ originate from the minimum SPV and all slices have similar slopes, implying the balanced contributions to the SPV values from the whole plot and subplot spaces, which matches the characteristics of a rotatable design where all sub spaces contribute equally to the SPV. Note that for a completely randomized design, the exact axial levels for a rotatable CCD are $\pm\sqrt[4]{F} = \pm\sqrt[4]{8} = \pm 1.682$, where F is the number of factorial runs (see Myers and Montgomery 2002, pp. 309). As d increases, i.e., the proportion of the whole plot error variance in the observational error variance increases, the SPV distribution loses its symmetry, which is visualized by the flatter slices and larger differences displayed between the slices at different whole plot levels. When value of d is quite large, for instance, d=10, the slices are flat and parallel to each other, implying that the whole plot space contributes most to the changes in the SPV values whereas the SPV values are relatively stable across the subplot space.

In the plot for d=10, the smallest SPV values are associated with the slice corresponding to $w=\pm 0.58$, revealing that the most precise prediction in the subplot space occur at the whole plot factorial levels. As one moves away from the whole plot factorial levels towards the center of the whole plot space from $w=\pm 0.58$ to w=0 or towards the edge of the whole plot space at $w=\pm 1$, the SPV values increase. The SPV value for the dot at $w=\pm 1$ is same as the maximum SPV from the global curve, implying the worst case in the entire space is the edge of whole plot space. In summary, the SPV distribution across the whole plot space is a W-shaped surface with the lowest values near w=0.58 and w=-0.58. The SPV values at the center of the whole plot space are lower than they are at the edge of the whole plot space.

Liang et al. (2004) studied this design using 3-D VDGs. Figure 3.3 displays the surface plots for maximum SPV for a variety of d values. The w-axis represents the whole plot shrinkage level, the x-axis represents the summarized subplot shrinkage level with $x = \sqrt{x_1^2 + x_2^2}$, the vertical axis represents the maximum SPV value for the sub region with the given whole plot and subplot shrinkage levels w and x. Note that this plot only

focuses on maximum SPV for each combination of w and x, while the FDS plot provides information regarding the distribution of actual SPV values. From the 3-D VDGs we can see the near-rotatable SPV distribution for d=0, since along quarter circles of fixed radii on the 3-D VDG, the maximum SPV values are nearly constant. For large values of d, the maximum SPV experiences large changes along the whole plot space but for fixed levels of w, the surface is relatively flat across the x-axis.

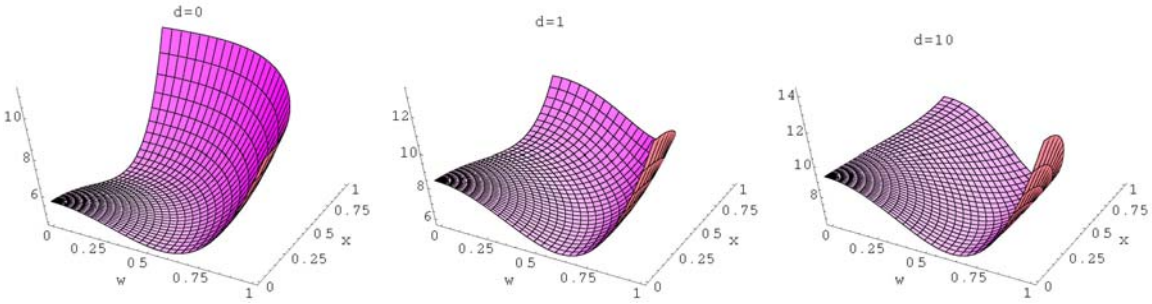


Figure 3.3 3-D VDG surface plots for maximum SPV of standard CCD with one whole plot variable and two subplot variables.

Liang et al (2004) use the 3-D VDGs to show that the design can be improved in terms of both V- (minimize the average SPV value) and G-efficiency (minimize the maximum SPV value) by altering the factorial levels. This approach is different to the usual method of changing the axial levels α. In their approach, the axial levels are fixed to keep the design within the same region. Since the distribution of the SPV values in the whole plot and subplot spaces is different, the factorial levels for the whole plot and subplot factors are moved in opposite directions when making the design more efficient. The V-optimal values of the whole plot and subplot factorial levels, f1 and f2, and G-optimal factorial levels, h1 and h2, for a variety of variance component ratios are listed in Table 3.3. More details are available in Liang et al. (2004).

Table 3.3 Optimal factorial levels in terms of G- and V-efficiency.

d	V-optimal			G-optimal		
	f1	f2	Relative efficiency	h1	h2	Relative efficiency
0	1	1	1	1	1	1
1	0.78	1.09	0.965	1.13	0.92	0.927
10	0.51	1.17	0.858	1.22	0.8	0.897
50	0.34	1.20	0.816	1.25	0.75	0.884

Utilizing the FDS plots, we can make informative comparisons between the CCDs with the standard, *G*-optimal and *V*-optimal factorial levels. Comparisons among the designs are first made utilizing global FDS plots. The global FDS plots for d=1 and 10 are presented in Figure 3.4 and display similar pattern of SPV distribution. As we can see from Figure 3.2, for most of the design space, the standard CCD has much smaller SPV values than its maximum SPV. The maximum SPV for the standard CCD (indicated by the symbol "." at FDS=1) is less than that of *V*-optimal CCD and is larger than the maximum SPV for the *G*-optimal CCD. The plots for the *V*-optimal CCD (dashed curve) have flatter slopes (implying a more stable SPV distribution) than the other two designs for approximately 90% of the design region. However, the tradeoff in using the *V*-optimal CCD is that it has a higher maximum value and if one is interested in predictions at the edge of the design space, the *V*-optimal CCD may not be the most desirable design. Alternatively, the *G*-optimal CCD (dashed-dot curve) has smaller maximum values (indicated by the symbol "+") than the standard and *V*-optimal CCDs but larger SPV values than the other two designs in almost the entire region. Thus, in achieving the smallest maximum SPV, the *G*-optimal CCD sacrifices precision in more than 95% of the design region. Note that traditionally SPDs have been compared using single number alphabetic optimality criteria. This example illustrates the fact that a single number criterion such as *G*- or *V*- criteria, while informative, does not reveal the behavior of the SPVs throughout the design region. Although the 3-D VDGs reveal the behavior of functions of the SPVs throughout the design region, they are not designed to allow one to quantify the proportion of the design region exhibiting certain behaviors.

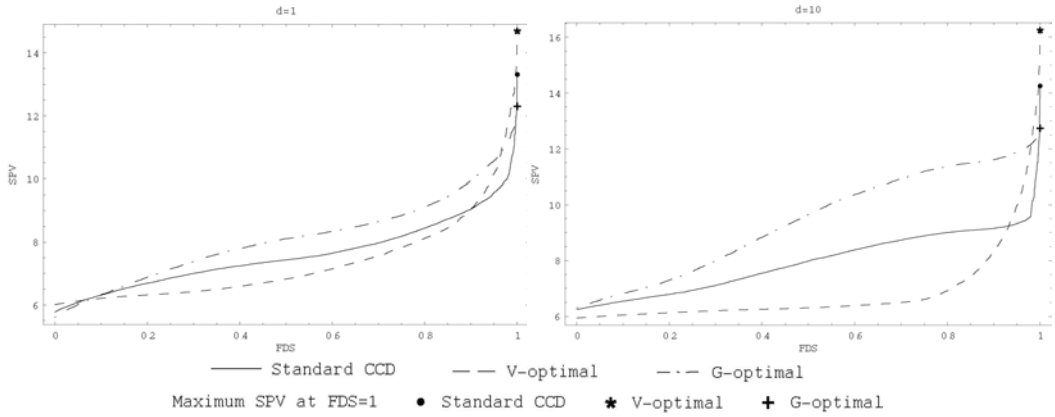


Figure 3.4 Global FDS plots for the CCDs with the standard, *G*- and *V*-optimal factorial levels.

By studying the sliced FDS curves for the *G*- and *V*-optimal CCDs (shown in Figure 3.5), detailed differences in the SPV distributions between the three designs can be observed. For the *V*-optimal CCD (represented in the top two plots of Figure 3.5), the three curves at w=0, ± 0.29 and ± 0.58 have flat slopes and the values are close to the minimum SPV in the global curve. Moreover, the dot ($w=\pm 1$) corresponds to the maximum SPV for the entire design region, since this dot has the same SPV as indicated by the symbol "*" at FDS=1 in Figure 3.4. The FDS slices indicate that when optimizing in terms of *V*-efficiency, the SPV is quite stable at the center and in the middle of whole plot space but goes up quickly when moving towards the edge, which implies a U-shaped SPV distribution in whole plot space. The SPV distribution in the whole plot space for the *G*-optimal CCD has a similar W-shape to that observed in the standard CCD in Figure 3.2. However, the slices at w=0 (dash-dot line) and $w=\pm 1$ have comparably large SPV values, showing that *G*-optimization forces a large sacrifice in prediction performance at the center of whole plot space. Therefore, for the *V*-optimal CCD, the SPV distribution in the whole plot space has two nearly equal peak values at the center and the edge.

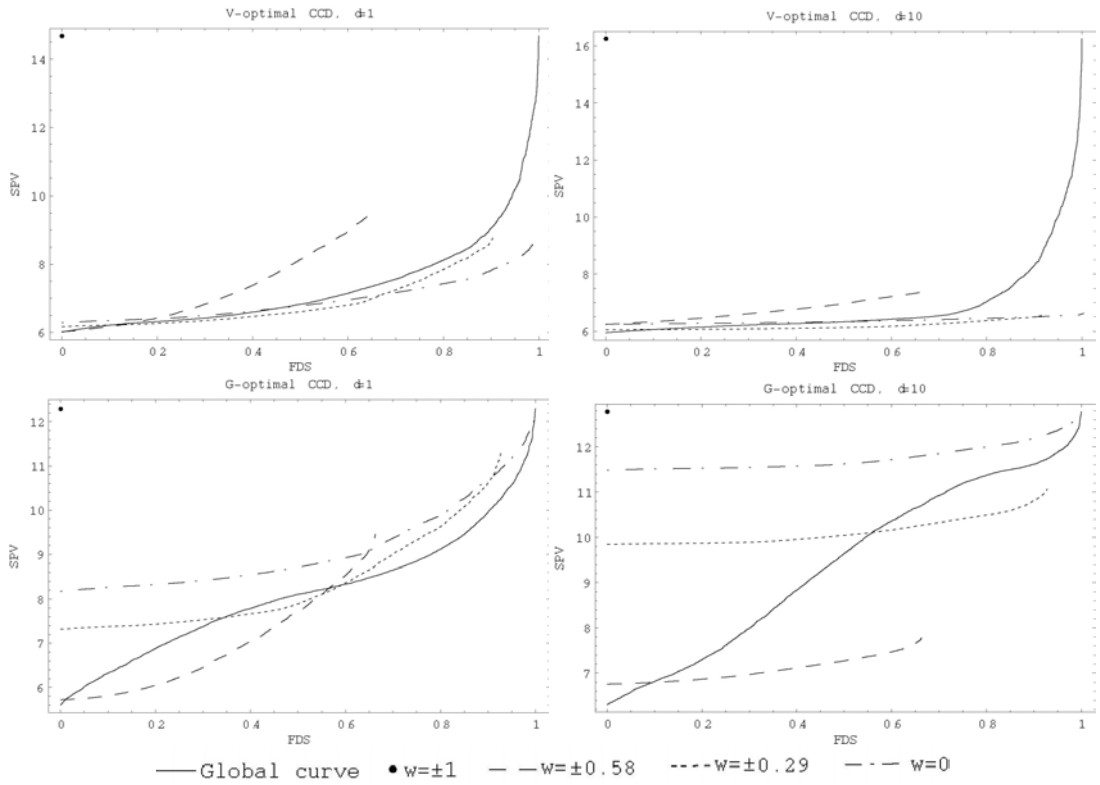


Figure 3.5 Global and slice FDS plots for *G*- and *V*-optimal CCDs.

3.4 Example 2

The second example considers a standard CCD with two whole plot factors, w_1 and w_2, and two subplot factors, x_1 and x_2, for a full second order model. The set up of the design is provided in Table 3.4. In this section, the global and slice FDS plots combined with 3-D VDGs are employed to study the design's prediction variance properties.

Table 3.4 Standard CCD with two whole plot and two subplot factors.

Whole plot	w_1	w_2	x_1	x_2	No. of runs per whole plot
1	-1	-1	±1	±1	4
2	-1	1	±1	±1	4
3	1	-1	±1	±1	4
4	1	1	±1	±1	4
5	2	0	0	0	1
6	-2	0	0	0	1
7	0	2	0	0	1
8	0	-2	0	0	1
9	0	0	±2	0	2
	0	0	0	±2	2
	0	0	0	0	3

Figure 3.6 displays the global and sliced FDS curves for various variance component ratios (d=0, 1 and 10). Note that since this design has two whole plot variables, the slices of the subplot space are constructed at constant whole plot level combinations. Since the design region is now a 4-dimensional hypersphere, the restriction on the subplot space for each combination of w_1 and w_2 corresponds to $x_1^2 + x_2^2 \le 1 - (w_1^2 + w_2^2)$. In the plots, four whole plot shrinkage value combinations $w_1 = w_2 = 0$, $w_1 = w_2 = \pm 0.5$, $w_1 = \pm 1$ and $w_2 = 0$, and $w_1^2 + w_2^2 = 1$ are specified. The shrinkage value at $w_1 = w_2 = 0$ is taken at the center of the whole plot space and corresponds to the largest subplot space slice. The FDS profile here extends to FDS=1. The slice at $w_1 = w_2 = \pm 0.5$ corresponds to the slice of the subplot space taken at the factorial levels of the whole plot space, i.e., a circle of $x_1^2 + x_2^2 = 1 - (0.5^2 + 0.5^2) = 0.5$. Standardizing it by the maximum volume of the subplot spaces, i.e., π, the FDS line extends to FDS=0.5. The whole plot level combination $w_1^2 + w_2^2 = 1$

corresponds to the locations (the circle) at the edge of the whole plot space. Note that the whole plot combinations of $w_1 = \pm 1$ and $w_2 = 0$ represent special points on the circle denoting the axial points in the w_1 direction in the whole plot space. For these two combinations of the whole plot factor levels the corresponding subplot levels are zero, i.e., $x_1 = 0, x_2 = 0$. Hence the slice for $w_1 = \pm 1$ and $w_2 = 0$ is displayed as a dot at FDS=0. Note that the longer profile associated with $w_1^2 + w_2^2 = 1$ indicates that a larger sub region than what is associated with the sub region corresponding to $w_1 = \pm 1$ and $w_2 = 0$.

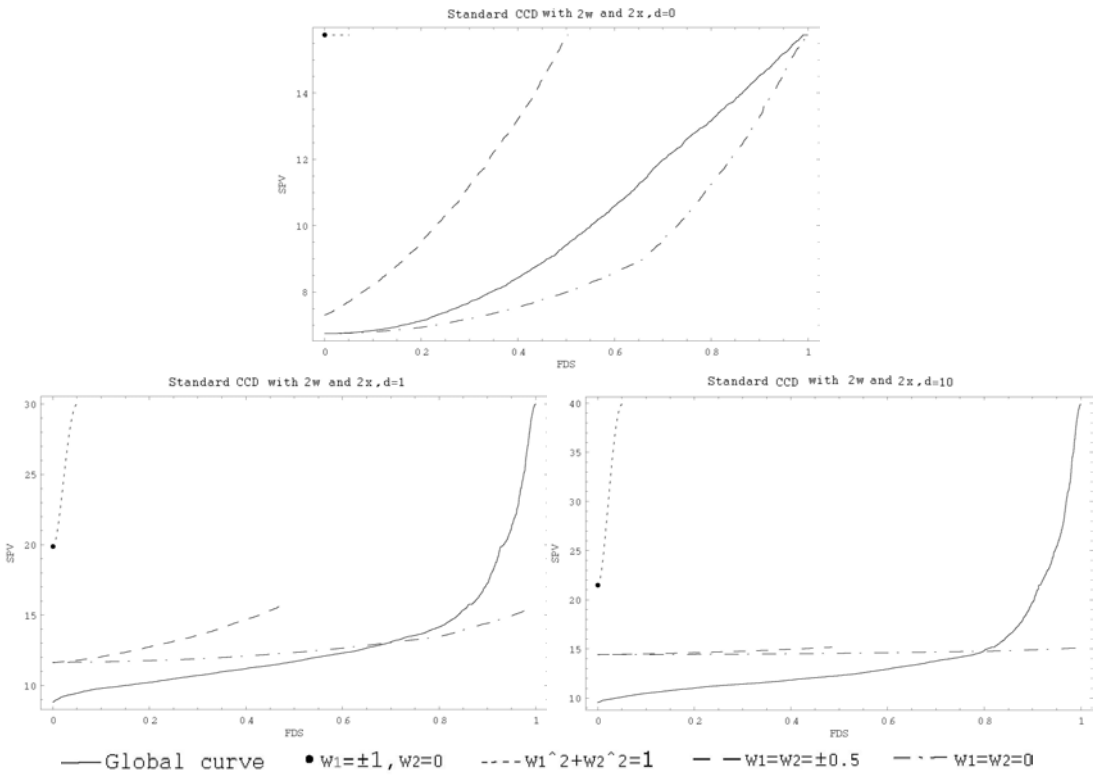


Figure 3.6 Global and sliced FDS plots for standard CCD with two whole plot variables.

For *d*=0 (CRD), the global curve has a relatively diagonal slope, implying that the SPVs are not stable across the entire design region. As *d* increases, the global FDS curves display a relatively stable SPV distribution (global curves are flat) for more than 85% of the design space and then the SPV values increase rapidly to the maximum. Note that this increase only occurs in a small portion of the design region. The 3-D VDGs associated with the distributions of the average and maximum SPV are displayed in Figure 3.7.

Depending on the location in the design space, each shrinkage factor combination (*w,x*) corresponds to a different volume of the sub region. For example, for any point (*w,x*) in the 3-D VDGs, the corresponding subplot space is determined by $x_1^2 + x_2^2 = x^2$ and the whole plot space is $w_1^2 + w_2^2 = w^2$. As the values of (*w,x*) become larger, the associated sub region also becomes larger. However, since 3-D VDGs provide the same emphasis to each combination of (*w,x*), the differences in volumes are difficult to visualize using the 3-D VDGs alone. Hence for estimation of *V*-efficiency, the practitioner must mentally rescale the 3-D VDG plots to focus more on larger values of the (*w,x*), which are the outer regions of the design space. It is commonly these outer regions where poorer prediction can occur and which comprise a large proportion of the total design space. The FDS plots eliminate the need for this mental re-scaling by automatically weighting the volumes and displaying the different volumes of the sub space via differing lengths of the FDS profiles.

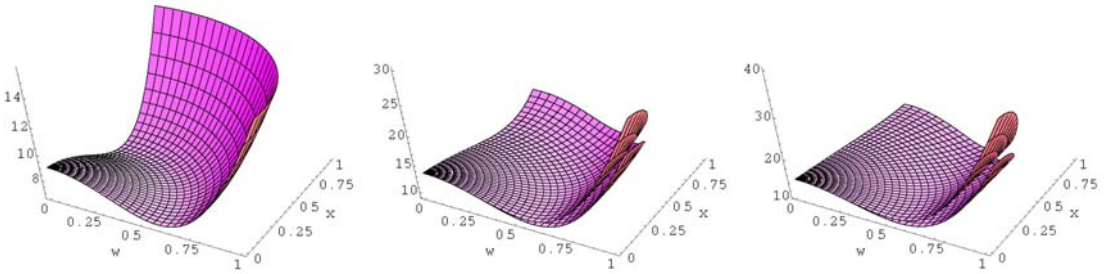


Figure 3.7 3-D VDG surface plots for maximum and average SPV of standard CCD with two whole plot variables for *d*=0 (left), 1 (middle) and 10 (right).

From the slopes and actual values of the FDS slices in Figure 3.6, the distribution of the SPV in the subplot and whole plot spaces can be examined. For *d*=0 (CRD), the slices are spread on both sides of the global curve and have relatively steep slopes, implying that both the whole plot and subplot variable locations contribute to changes in SPV values. As *d* increases, the SPV distribution is stable from the center out to the factorial levels in the whole plot space. This is indicated by the relatively flat slices at $w_1 = w_2 = 0$ and $w_1 = w_2 = \pm 0.5$. The SPV values increase as we move towards the edge of the whole plot space. Large changes in the SPV values at the edge of the whole plot space are observed with the steepest slope occurring at the slice taken at $w_1^2 + w_2^2 = 1$. This implies that the

SPV values are different at locations at the edge of the design space, and thus rotatability is lost for the split-plot scheme. In these plots, at FDS=0, the global curve is under the slice curves, showing that the subplot spaces at the chosen whole plot level combinations do not contain the global minimum.

This example illustrates that the FDS plots provide prediction variance evaluation by providing the user with an idea of the proportion of the design space experiencing a particular behavior. Another graphical technique for comparing split-plot designs is the 3-D VDGs. Although the 3-D VDGs are informative in their own right, they are not designed to reflect the size of the design space. As a result, FDS plots provide a useful complement to the 3-D VDGs, especially as the dimensions of the split-plot design space increases.

3.5 Conclusions

In this chapter, we have proposed the use of FDS plots for evaluating split-plot designs. We have demonstrated that FDS plots can be a useful tool for evaluating split-plot designs and they are especially useful when combined with the information from 3-D VDGs. FDS plots help visualize information about the entire range of SPV values across the design space and are easy to interpret, allowing a more complete evaluation and comparison of competing split-plot designs than simply comparisons of designs in terms of single-number *G*- or *V*-efficiencies. Since the FDS plots treat each location in the design space with same weight, the outer portion of the design space have more influence on the FDS curves because they represent a large proportion of the total region. This is extremely helpful to better understand the design's average prediction variance. The 3-D VDGs visually tend to underestimate the influence of SPVs at the edge of design space on the *V*-efficiency. Therefore, the FDS plots are more sensible for assessing *V*-efficiency than 3-D VDGs do. When FDS plots are used in conjunction with the 3-D VDGs, the plots provide the ability to make an overall evaluation of prediction performance.

For split-plot designs, the sliced FDS curves are extremely helpful for examining the distribution of the SPV in the whole plot and subplot spaces separately. The FDS slices also allow for better understanding of the influence of the split-plot error structure through the variance component ratio, d, on the choice of design. Since the characteristics of SPV distribution in the whole plot and subplot spaces can be read from the slices, the slice curves provide information about which of the two spaces contributes more to the changes of the SPV values. This can help the user who may wish to optimize or augment the design in the most efficient manner.

In this chapter, when scaling the split-plot design by the design size, the inherent assumption is that the cost of the experiment is only a function of the number of subplots. This assumption may or may not be valid, depending on the particular experiment. Often the whole plots are more expensive than the subplots and measurements in industrial split-plot experiments. Therefore, for split-plot designs, it may be more appropriate to scale the split-plot design according to the relative costs associated with whole plots and subplots. Liang, Anderson-Cook and Robinson (2005) present cost penalized prediction variance for considering different cost scenarios for split-plot designs. Incorporating a different cost scenario into the graphical tools is straightforward and can be done by replacing the SPV by the cost adjusted prediction variance on the vertical axis in the FDS plots.

For split-plot designs with cuboidal regions, the subplot spaces at different whole plot levels all have the same volume. Hence all the FDS profiles would have the same length. Otherwise the mechanism to generate the FDS curves and the interpretations of the plots remains the same as those discussed in this chapter for spherical regions.

Chapter 4 Cost Penalized Evaluations for SPDs

Abstract

Comparisons between different designs have traditionally focused on balancing the quality of estimation and prediction with the overall size of the design. For split-plot designs with two levels of randomization, the total number of observations likely does not accurately summarize the true cost of the experiment, since different costs are likely associated with setting up the whole plot and subplot levels. In this paper, we present several flexible measures for design assessment based on *D*-, *G*- and *V*-optimality criteria that take into account potentially different cost scenarios for the split-plot designs. The new measures are illustrated with two examples: a 2^3 factorial experiment for first-order model, where all possible designs are considered, and selective designs for a three-factor second order model.

4.1 Introduction

The use of response surface methods generally begins with a process or system involving a response y that depends on a set of k controllable input variables (factors) $x_1, x_2, \ldots, x_k$. To assess the effects of these factors on the response, an experiment is conducted in which the levels of the factors are varied and changes in the response are noted. The size of the experimental design (number of distinct level combinations of the factors as well as number of runs) depends on the complexity of the model the user wishes to fit. Limited resources due to time and/or cost constraints are inherent to most experiments, and hence, the user typically approaches experimentation with a desire to minimize the number of experimental trials while still being able to adequately estimate the underlying model.

There are many different ways to assess a design's capability to estimate the underlying model. For instance, one can focus on the quality of parameter estimates (often quantified by the *D*-criterion) or the precision of model predictions such as with the minimization of average prediction variance (*V*, *Q* or *IV*-criteria) or minimization of the worst prediction variance (*G*-criterion). The use of these 'alphabetic optimality' criteria for comparing competing designs is well documented (see Myers and Montgomery 2002, pp. 390-402).

Embedded in the calculation of each of the alphabetic optimality criteria is the information matrix of the estimated model parameters. The information matrix is based on the error structure and the error structure is a function of how one randomizes the run order of the experimental trials. Frequently, experiments are designed assuming a completely random run order. However, if the levels of one or more factors' are difficult and/or costly to change, the practitioner is not as inclined to run the experiment using a completely randomized run order. Instead, the practitioner may select a run order involving fewer changes of the hard-to-change factor. The resulting experiment generally involves two separate randomizations: one for the run order of the levels of the hard-to-change factors and one within each level combination of the hard-to-change factors where the run order of all possible combinations of the easy-to-change variables is randomized. When there are separate randomizations for hard-to-change factors and easy-to-change factors, the error structure is that of a split-plot design. Letsinger, Myers, and Lentner (1996), Ganju and Lucas (1999), and Ju and Lucas (2002) describe split-plot designs resulting from factor levels on consecutive runs of an experiment not being reset.

Although the selected experimental design may have nice statistical properties for complete randomization, if the design is run as a split-plot, the design's statistical properties may not be well understood and hence are unlikely to be optimal. Ganju and Lucas (1999) point out that split-plot designs chosen by random run order are not as appealing as those chosen by a good design strategy. Design strategies for split-plot randomization have received considerable attention in the literature of late. Huang, Chen and Voelkel (1998) and Bingham and Sitter (1999) derive minimum aberration two-level fractional factorial designs for screening experiments. Anbari and Lucas (1994)

considered the *G*-optimality criterion for several competing split-plot designs. Goos and Vandebroek (2001 and 2004) proposed exchange algorithms for constructing *D*-optimal split-plot designs. Liang, Anderson-Cook, Robinson, and Myers (2004, 2005) considered graphical techniques for assessing competing split-plot designs over the design region in terms of *G*- and *V*-efficiency. Bisgaard (2000), noting that the benefits of running a split-plot design are the savings obtained by reducing the number of whole plot setups, formulated cost functions indicating the relative costs of performing each of the subplot tests to the cost of setting up the individual whole plot tests.

With the exception of Bisgaard (2000), optimal strategies for split-plot designs, in large part, have been focused on objective criteria that reflect the statistical properties of the design (*D*-criterion involves quality of parameter estimates, while *G*- and *V*-optimality criteria involve quality of prediction). The fact that the split-plot error structure is a result of hard-to-change factors often implies that there is greater cost/time involved in changing the levels of the whole plot factors than there is in changing the levels of the subplot factors. As a result, the practitioner may desire to use design selection criteria that not only reflect the statistical properties of the design but also the cost of the experiment. For instance, suppose we have three factors, one hard-to-change (w) and two easy-to-change (x_1 and x_2) variables, each at two levels run as a full factorial design. Consider the three competing designs in Table 4.1 with 2, 4 and 8 whole plots, respectively. The numbers in the columns for each design specify which whole plot will contain that combination of factor levels. For any design that has more than one observation per whole plot, the level of w remains unchanged within a whole plot, making it possible to collect these observations without changing whole plot levels. Goos and Vandebroek (2004) state that Design 2 is the best possible eight run split-plot design for estimating the pure linear model in terms of *D*-efficiency. The authors go on to state that when there is more variation among whole plots than there is among subplots, Design 1 is always more *D*-efficient than Design 3, the completely randomized design.

Although Design 2 may be the best possible 8 run split-plot design in terms of *D*-efficiency, a consideration of 'cost' in terms of ease of experimentation, time, etc., may

suggest other alternatives. For instance, suppose the only appreciable cost in experimentation is due to changes in the hard-to-change factor. If this is the case, Design 1 is twice as appealing as Design 2 (in terms of cost since Design 1 requires two set-ups for the levels of w and Design 2 requires four) and four times more appealing than Design 3 (in terms of cost since Design 3 involves eight set-ups). Clearly there is a trade-off between "good statistical properties" and "cost of experimentation". Although alphabetic optimality criteria are useful in determining split-plot designs, these criteria do not reflect the different costs that are likely associated with hard-to-change and easy-to-change factors.

Table 4.1 Eight runs factorial design in split-plot structure. The columns for Design 1, 2 and 3 indicate the index of the whole plot, and corresponding w level represents the whole plot level.

w	x_1	x_2	Design 1	Design 2	Design 3
-1	-1	-1	1	1	1
-1	-1	+1	1	2	2
-1	+1	-1	1	2	3
-1	+1	+1	1	1	4
+1	-1	-1	2	3	5
+1	-1	+1	2	4	6
+1	+1	-1	2	4	7
+1	+1	+1	2	3	8

Dompere (2004) states "there are two important sides to any decision…the two sides are simply the *costs* that may be incurred in order to receive the *benefits* that may be associated with a particular decision." Relating this to split-plot experimentation, a design that may have nice statistical properties in terms of the estimated model may not be appealing from a cost perspective. In the split-plot setting, we often need to find the right balance of designs with good statistical properties and are within the experimenter's budget.

In this manuscript, we propose cost adjusted D-, G-, and V-optimality criteria for split-plot designs. We adjust the D-, G-, and V-optimality criteria for cost, where the expressions for cost are similar to that of Bisgaard (2000) with adjustments made to allow for unequal whole plot size. With the cost adjusted optimality criteria, the user is

presented with single objective functions that simultaneously account for the desired statistical property (efficient parameter estimation or model prediction) and cost of experimentation. Utilizing the new objective functions, we demonstrate strategies for choosing optimal split-plot designs and then illustrate these ideas with two examples. In the next section we discuss the cost formulations and then we develop the appropriate cost adjusted *D*-, *G*-, and *V*-expressions for split-plot designs. Finally, two examples are provided which demonstrate the trade-off between 'good statistical properties' and 'cost reduction'.

4.2 Cost Formulations

In practice, if a completely randomized experiment is run, it is generally the case that changing the levels of a factor is uniformly difficult across all factors. As a result, the cost or time associated with the experiment is related only to the number of experimental units (EUs). In split-plot experiments, there are generally two types of EUs – whole plots and subplots. In industrial experimentation, whole plot factors are those factors whose levels are hard/costly to change and subplot factors are those factors whose levels are relatively easy to change. Thus, in considering the cost of running an industrial split-plot experiment, the total cost is a function of both the cost associated with whole plot units as well as the cost associated with subplot units. Similar to Bisgaard (2000), we write the cost of a split-plot experiment as

$$C = C_{WP}a + C_{SP}N \tag{4.1}$$

where C denotes the total cost of the experiment, a denotes the number of whole plot units, N is the total number of subplot units, and C_{WP} and C_{SP} are the costs associated with the individual whole plot and subplot, respectively. Note that the cost incurred by measurement is considered a part of cost of subplot.

In practice, it may be difficult for the practitioner to ascertain the exact costs associated with whole plots or subplots, i.e. precise values for C_{WP} and C_{SP}, but it may be more

feasible to specify the relative cost of these quantities, i.e. $r = C_{SP}/C_{WP}$. Hence, the cost of the experiment is proportional to $a + rN$, i.e.

$$C \propto a + rN . \tag{4.2}$$

Writing the cost in this manner allows flexibility for specifying the relative costs of the two cost components without having to specify their exact values. Generally speaking, C_{WP} is greater than C_{SP} due to the more time/effort involved with changing levels of the whole plot factors. As C_{WP} increases relative to C_{SP} , r approaches zero. On the other hand, if obtaining the measurement of the response for each observation is expensive, then C_{SP} may be larger compared to C_{WP} and r will increase. When $C_{WP} = C_{SP}$, $r = 1$.

It is noteworthy that the completely randomized design (CRD) can be thought as a special case of a split-plot design where each observation can be treated as a separate whole plot. In CRDs the number of whole plots and subplots are equal and the total cost of the experiment is given by

$$C = \left(C_{WP} + C_{SP}\right) N = C_{WP}\left(1 + r\right) N . \tag{4.3}$$

The expression in (4.3) is proportional then to the standard penalty of N commonly used for the *D*-, *G*-, and *V*-optimality criteria in completely randomized experiments. In the next section, we review the general model for split-plot designs and present some cost adjusted alphabetic optimality criteria (*D*-, *G*-, and *V*-) for split-plot designs that utilize the expressions for cost discussed above.

4.3 The Split-Plot Model and Cost Adjusted *D*-, *G*-, and *V*-optimality Criteria

When the experiment is run as a split-plot design with a whole plots, the following linear mixed model can be written to explain the variation in the $N \times 1$ response vector, **y,**

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\delta} + \boldsymbol{\varepsilon} .$$

Regarding notation, $\mathbf{X}$ is the $N \times p$ design matrix expanded to model form for p parameters including the intercept; $\mathbf{Z}$ is an $N \times a$ classification matrix of ones and zeroes where the ji^{th} entry is 1 if the j^{th} observation (j=1,...,N) belongs to the i^{th} whole plot

(i=1,…,a); $\boldsymbol{\delta}$ is an $a\times 1$ vector of whole plot random effects where the elements are assumed *i.i.d* $\mathrm{N}\left(0,\sigma_{\delta}^{2}\right)$ with σ_{δ}^{2} denoting the variability among whole plots; $\boldsymbol{\varepsilon}$ is the $N\times 1$ vector of residual errors where the elements are assumed *i.i.d* $\mathrm{N}\left(0,\sigma_{\varepsilon}^{2}\right)$ and σ_{ε}^{2} denotes the variation among subplots. It is also assumed that $\boldsymbol{\delta}$ and $\boldsymbol{\varepsilon}$ are independent.

The covariance matrix of the responses in a split-plot design is

$$\mathrm{Var}\left(\mathbf{y}\right) = \boldsymbol{\Sigma} = \sigma_{\delta}^{2}\mathbf{Z}\mathbf{Z}' + \sigma_{\varepsilon}^{2}\mathbf{I}_{N} = \sigma_{\varepsilon}^{2}\left[d\mathbf{Z}\mathbf{Z}' + \mathbf{I}_{N}\right]$$

where $\mathbf{I}_{N}$ is an $N\times N$ identity matrix and $d=\sigma_{\delta}^{2}/\sigma_{\varepsilon}^{2}$ represents the variance component ratio. For simplicity of presentation, we will assume that observations are sorted by the whole plots, implying $\mathbf{Z} = diagonal\{\mathbf{1}_{n_1},\cdots,\mathbf{1}_{n_a}\}$, where $\mathbf{1}_{n_i}$ is an $n_i\times 1$ vector of one's and n_i is the size of the i^{th} whole plot. Assuming the diagonal form of $\mathbf{Z}$ allows one to conveniently write the covariance matrix as

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \boldsymbol{\Sigma}_a \end{bmatrix}$$

where each $n_i\times n_i$ matrix $\boldsymbol{\Sigma}_i$ is given by

$$\boldsymbol{\Sigma}_i = \begin{bmatrix} \sigma_{\varepsilon}^{2}+\sigma_{\delta}^{2} & \cdots & \sigma_{\delta}^{2} \\ \vdots & \ddots & \vdots \\ \sigma_{\delta}^{2} & \cdots & \sigma_{\varepsilon}^{2}+\sigma_{\delta}^{2} \end{bmatrix}$$

and $\boldsymbol{\Sigma}_i$ denotes the covariance matrix of responses for the i^{th} whole plot. Note that the variance of an individual observation is the sum of the subplot and whole plot error variances, $\sigma_{\varepsilon}^{2}+\sigma_{\delta}^{2}$. A popular method for estimating the variance components is restricted maximum likelihood (REML).

The vector of fixed effects parameters, $\boldsymbol{\beta}$, is estimated via generalized least squares, yielding

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y}\,.$$

The covariance matrix of the estimated model coefficients is given by

$$Var\left(\hat{\boldsymbol{\beta}}\right) = \left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}.$$

When the design is completely randomized, $Var\left(\hat{\boldsymbol{\beta}}\right) = \sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}$. Comparing the expressions for the estimated model coefficients for split-plot designs and CRDs is important as it lends insight into the greater complexity associated with optimal design strategies for split-plot designs vs. for CRDs. For example, if one wishes to obtain the optimal design in terms of ability to estimate model parameters, the optimal CRD depends only on the settings of the levels of the terms in **X**. The optimal split-plot design, in terms of parameter estimation, will depend on the structure of **X**, the variance ratio, *d*, the number of whole plots, *a*, the dimensionality of each of the $\boldsymbol{\Sigma}_j$ (determined by the number of subplots within each whole plot), and the arrangement of subplot levels within whole plots.

Cost Adjusted *D*-optimality Criterion

Strategies for choosing an optimal design depend on the goal of the researcher. If the desire is to have quality model parameter estimates, one strategy is to find a design with high *D*-efficiency. The *D*-criterion is defined in terms of the scaled moment matrix. For CRDs, the scaled moment matrix is $\mathbf{M} = \mathbf{X}'\mathbf{X}/N$, i.e., the estimation variance scaled by σ^2/N. Note that scaling by σ^2, the variance of the observation, causes $\mathbf{M}$ to be unitless and the scaling by $1/N$ causes **M** to be reflective of design size. Since the cost of a CRD is determined by the design size, the scaling by $1/N$ is essentially a scaling for cost. The *D*-optimal design is then the design that maximizes the determinant of M, that is, $|\mathbf{M}| = |\mathbf{X}'\mathbf{X}|/N^p$, where *p* denotes the number of model parameters.

The moment matrix for split-plot designs is $\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}$. Scaling the information matrix in a similar fashion to the scaling in the CRDs, we can define the scaled moment matrix for split-plot designs as

$$\mathbf{M} = \frac{\left(\sigma_\delta^2 + \sigma_\varepsilon^2\right)\left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)}{cost}.$$

Note that $\left(\sigma_\delta^2 + \sigma_\varepsilon^2\right)\left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right) = \left(\mathbf{X}'\left(\frac{\boldsymbol{\Sigma}}{\sigma_\delta^2 + \sigma_\varepsilon^2}\right)^{-1}\mathbf{X}\right) = (\mathbf{X}'\mathbf{R}^{-1}\mathbf{X})$, where $\mathbf{R}$ denotes the observational correlation matrix. Rewriting $\mathbf{M}$ we have

$$\mathbf{M} = \frac{\left(\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\right)}{cost}$$

Since $\mathbf{R}$ is the correlation matrix, $\mathbf{M}$ is unitless as desired. Since the cost of a split-plot experiment is not as simple as the design size, N, we must adjust by an expression for cost that allows for potentially different costs associated with whole plots and subplots. A natural divisor is the expression for cost provided in (4.2), yielding $\mathbf{M}=\left(\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\right)\Big/(a+rN)$. The cost adjusted D-efficiency is then defined as

$$D_{eff} = \frac{\left|\mathbf{M}(D)\right|^{1/p}}{\underset{D\in\boldsymbol{\Omega}}{Max}\left|\mathbf{M}(D)\right|^{1/p}},$$

where $\mathbf{M}(D)$ is the scaled moment matrix for split-plot design D, $\boldsymbol{\Omega}$ denotes the space of all possible split-plot designs for the given model, and p denotes the number of fixed effect model parameters. Since the upper bound for the determinant is generally unknown, we can consider relative performance for two or more designs by looking at the cost adjusted D-criteria,

$$D = \left|\mathbf{M}\right|^{1/\mathrm{p}} = \frac{\left|\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\right|^{1/\mathrm{p}}}{a+rN}. \qquad (4.4)$$

The cost adjusted D-value for a given split-plot design will be calculated using equation (4.4). Note that if the expression in (4.4) is not divided by $a+rN$, the D-value is only indicative of the quality of the design in terms of its precision in estimating the fixed effect model parameters.

Cost Adjusted *G*- and *V*-optimality Criteria

The predicted value of the mean response at any location $\mathbf{x}_0$ for a SPD is given by

$$\hat{\mathbf{y}}_0 = \mathbf{x}_0'\hat{\boldsymbol{\beta}} = \mathbf{x}_0'\left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y},$$

where $\mathbf{x}_0$ is the point of interest in the design space expanded to model form. The prediction variance is then given by

$$Var\left(\hat{\mathbf{y}}_0\right) = \mathbf{x}_0'\left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{x}_0.$$

If interest is in finding a design with precise estimates of the mean, the *G*- and *V*-efficiencies of the design are popular choices for evaluating competing designs. As with *D*-efficiency, the desire is to work with a scale free quantity that provides a penalty for design cost. In the CRD, the prediction variance is given by

$$Var\left(\hat{\mathbf{y}}_0\right) = \sigma^2\mathbf{x}_0'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{x}_0$$

and the usual scaling is σ^2/N (variance of observations divided by the design size or cost). The scaled prediction variance for the CRD is then given by

$$SPV = \frac{NVar\left[\hat{\mathbf{y}}\left(\mathbf{x}_0\right)\right]}{\sigma^2} = N\mathbf{x}_0'\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{x}_0.$$

The scaling of the prediction variance for split-plot designs can be done in a similar fashion by scaling by $(\sigma_\delta^2+\sigma_\varepsilon^2)/(a+rN)$ (observational error variance divided by the design cost). The cost penalized prediction variance (CPPV) for the split-plot design is then given by

$$\begin{aligned} CPPV &= \frac{(a+rN)Var\left[\hat{\mathbf{y}}\left(\mathbf{x}_0\right)\right]}{\sigma_\delta^2+\sigma_\varepsilon^2} = (a+rN)\mathbf{x}_0'\left(\mathbf{X}'\left(\frac{\boldsymbol{\Sigma}}{\sigma_\delta^2+\sigma_\varepsilon^2}\right)^{-1}\mathbf{X}\right)^{-1}\mathbf{x}_0 \\ &= (a+rN)\mathbf{x}_0'\left(\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\right)^{-1}\mathbf{x}_0 \end{aligned}. \qquad (4.5)$$

By minimizing the average or maximum CPPV over the entire design region, one can obtain the best balance between quality and cost in terms of *V*- or *G*-efficiency. Anbari and Lucas (1994) used the lower bound of the maximum scaled prediction variance for the CRD, i.e., *p*, the number of model parameters, to evaluate the *G*-efficiency of split-plot designs and claimed some super-efficient designs. The number of parameters, *p*, is not a reasonable lower bound for split-plot designs due to the more complicated error structure of the model and the role that the variance component ratio plays in the

computation of the *G*-efficiency. It should be pointed out that the actual bounds for the *D*-, *G*-, and *V*-efficiencies for SPDs needs further investigation, and here we focus more on relative efficiencies for comparisons between competing designs.

In the following sections we examine first and second order model SPDs utilizing the cost adjusted *D*-, *G*-, and *V*-criteria developed above. For simplicity of presentation, we discuss experiments with three factors – one whole plot variable, *w*, and two subplot variables, x_1 and x_2. The next section involves a study of a first order design in which the candidate set of design points is the 8 runs of a 2^3 factorial design. All possible sequences of run orders are permuted and the corresponding split-plot designs are constructed. By evaluating the estimation and prediction properties of these designs with and without cost penalization, we demonstrate that the selected design is influenced by not only the split-plot error structure but also the relative costs between whole plots and subplots. The best design in terms of a joint consideration of cost and quality is often different from the optimal design when only quality is considered. As might be expected, designs with a smaller number of whole plots are preferred as the whole plots become more expensive. In the following section, five variations of the central composite design (CCD) are studied for the second order model. The study shows that under different scenarios, the cost adjusted evaluations penalize the designs with larger numbers of whole plots proportional to the relative cost of the whole plots to subplots. We also provide some design strategies for second order split-plot designs.

4.4 Factorial 2^3 Split-Plot Designs for a First Order Model

In this section, we consider an example with 8 design points for a first-order model with fixed effects modeled as $f(w, x_1, x_2) = \beta_0 + \beta_1 w + \beta_2 x_1 + \beta_3 x_2$. We will assume that *w* is a hard-to-change factor and that x_1 and x_2 are easy to change. In this section we will evaluate all possible split-plot designs involving the 8 runs of a 2^3 factorial experiment in terms of their cost adjusted *D*-, *G*- and *V*-efficiencies for different combinations of the cost ratio, *r*, and the variance component ratio, *d*. To study all possible split-plot designs

in this setting, we will first consider all of the 8! = 40,320 run orders. To construct the SPD for a given run order we assume that consecutive runs with the same whole plot level are not reset and are consequently part of the same whole plot. Due to the easy-to-change nature of the subplot levels, the subplot levels within each whole plot are always assumed being reset. The *D*-, *G*- and *V*-criteria (both cost adjusted and unadjusted) will then be computed for each design under different combinations of the cost ratio (*r*) and variance component ratio (*d*). The unadjusted *D*-, *G*- and *V*-criteria provide an indication of the design's ability to provide precise parameter estimates and predictions, respectively. The cost penalized *D*-, *G*- and *V*-criteria provide an indication of these properties when its expense is taken into account. It should be noted that Joiner and Campbell (1976) considered the importance of run order in 2^3 designs when levels of one factor are not reset but did not consider a split-plot error structure in their comparisons of designs.

As will be demonstrated later, design performance is strongly related to the value of the variance component ratio, *d* In this manuscript, we consider *d* values of 0.5, 1 and 10, representing situations in which the whole plot error variance is half, the same and ten times that of the subplot error variance, respectively. This range of *d* values is based primarily on values of *d* which have been reported in the literature for industrial split-plot experiments. Bisgaard and Steinberg (1997) stated that the whole plot variance is usually larger than subplot variance in prototype experiments. Letsinger et al. (1996) studied a split-plot experiment in chemical industry with $\sigma_\delta^2 / \sigma_\varepsilon^2 = 1.04$. Vining, Kowalski and Montgomery (2004) estimated the variance terms using pure error and reported a variance ratio 5.65. Webb, Lucas and Borkowski (2002) described an experiment with a variance ratio of 6.92 in a computer component manufacturing experiment. Kowalski, Cornell and Vining (2002) studied a mixture experiment with process variables where the estimated variance ratio was 0.82. Very little, work has been done in terms of considering costs in industrial split-plot experiments. In this manuscript, we consider cost ratios of *r*=0 (experimental costs depend only on the whole plots), 0.5 (one whole plot is twice as expensive as one subplot), and 1 (each subplot cost equally to that of each whole plot).

Of the 40,320 possible run orders involving the 8 runs of a 2^3 factorial design, there are 31 distinct split-plot designs in terms of their information matrices. Examples of the 31 distinct split-plot designs are provided in Table 4.2. Run orders resulting in equivalent split-plot designs (in terms of isomorphism) are those satisfying one of the following rules:

1. Run orders only differ in the fact that all whole plot levels are switched (i.e. changed from +1 to -1 or vice versa)
2. Run orders whose subplot levels remain the same but the order of the whole plots switches.
3. Run orders whose whole plot levels are the same but the subplot levels are switched from +1 to -1 and -1 to +1.
4. Run orders whose whole plot levels are the same but the subplot combinations within each whole plot are permuted.

To illustrate the equivalence rules above, consider design 2 in Table 4.2 with three whole plots composed of two, four and two subplots respectively. Assume that the run order sequence (1,2,3,4,5,6,7,8) corresponds to the following eight runs (w, x_1, x_2)= (-1,-1,-1), (-1,-1,1), (-1,1,-1), (-1,1,1), (1,-1,-1), (1,-1,1), (1,1,-1) and (1,1,1). Design 2 in Table 4.2 is then conducted with the following sequencing: (1,4), (5,6,7,8), and (2,3). Rule 1 above implies for example that the design with sequencing (5,8), (1,2,3,4), and (6,7) is equivalent to Design 2. Rule 2 implies that the design with sequencing (2,3), (5,6,7,8) and (1,4) is equivalent to Design 2. Rule 3 implies that a design with sequencing (4,1), (8,7,6,5), and (3,2) is equivalent to Design 2. Rule 4 implies that designs that differ only in permutations of the subplot combinations within the whole plots are equivalent. As an example, a design with sequencing (4,1), (6,5,8,7) and (3,2) is equivalent to Design 2.

Table 4.2 31 distinct designs. Each combination (w, x_1, x_2) represents a design point and indicates the levels of the three factors at this point, a is the number of whole plots, ID is the identification of the design in the list. The units in the same cell of the table are within the same whole plot.

ID	a	Whole plots	
1	2	(-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	3	(-1, -1, -1) (-1, 1, 1)	(1, -1, -1) (1, -1, 1) (1, 1, -1) (1, 1, 1)	(-1, -1, 1) (-1, 1, -1)					
3	3	(-1, -1, -1) (-1, -1, 1)	(1, -1, -1) (1, -1, 1) (1, 1, -1) (1, 1, 1)	(-1, 1, -1) (-1, 1, 1)					
4	3	(-1, -1, -1)	(1, -1, -1) (1, -1, 1) (1, 1, -1) (1, 1, 1)	(-1, -1, 1) (-1, 1, -1) (-1, 1, 1)					
5	4	(-1, -1, -1) (-1, -1, 1)	(1, -1, -1) (1, -1, 1)	(-1, 1, -1) (-1, 1, 1)	(1, 1, -1) (1, 1, 1)				
6	4	(-1, -1, -1) (-1, -1, 1)	(1, -1, -1) (1, 1, -1)	(-1, 1, -1) (-1, 1, 1)	(1, -1, 1) (1, 1, 1)				
7	4	(-1, -1, -1)	(1, -1, -1) (1, -1, 1)	(-1, -1, 1) (-1, 1, -1) (-1, 1, 1)	(1, 1, -1) (1, 1, 1)				
8	4	(-1, -1, -1)	(1, 1, 1)	(-1, -1, 1) (-1, 1, -1) (-1, 1, 1)	(1, -1, -1) (1, -1, 1) (1, 1, -1)				
9	4	(-1, -1, -1)	(1, -1, -1)	(-1, -1, 1) (-1, 1, -1) (-1, 1, 1)	(1, -1, 1) (1, 1, -1) (1, 1, 1)				
10	4	(-1, -1, -1)	(1, -1, 1)	(-1, -1, 1) (-1, 1, -1) (-1, 1, 1)	(1, -1, -1) (1, 1, -1) (1, 1, 1)				
11	4	(-1, -1, -1) (-1, 1, -1)	(1, -1, -1) (1, 1, 1)	(-1, -1, 1) (-1, 1, 1)	(1, -1, 1) (1, 1, -1)				
12	4	(-1, -1, -1)	(1, -1, -1) (1, 1, 1)	(-1, -1, 1) (-1, 1, -1) (-1, 1, 1)	(1, -1, 1) (1, 1, -1)				
13	4	(-1, -1, -1) (-1, 1, 1)	(1, -1, -1) (1, 1, 1)	(-1, -1, 1) (-1, 1, -1)	(1, -1, 1) (1, 1, -1)				
14	5	(-1, -1, -1)	(1, -1, -1) (1, -1, 1)	(-1, -1, 1)	(1, 1, -1) (1, 1, 1)	(-1, 1, -1) (-1, 1, 1)			
15	5	(-1, -1, -1)	(1, -1, -1) (1, -1, 1)	(-1, 1, -1)	(1, 1, -1) (1, 1, 1)	(-1, -1, 1) (-1, 1, 1)			
16	5	(-1, -1, -1)	(1, -1, 1)	(-1, 1, -1)	(1, -1, -1) (1, 1, -1) (1, 1, 1)	(-1, -1, 1) (-1, 1, 1)			
17	5	(-1, -1, -1)	(1, -1, -1)	(-1, -1, 1)	(1, -1, 1) (1, 1, -1) (1, 1, 1)	(-1, 1, -1) (-1, 1, 1)			
18	5	(-1, -1, -1)	(1, -1, -1) (1, -1, 1)	(-1, 1, 1)	(1, 1, -1) (1, 1, 1)	(-1, -1, 1) (-1, 1, -1)			
19	5	(-1, -1, -1)	(1, -1, -1)	(-1, 1, 1)	(1, -1, 1) (1, 1, -1) (1, 1, 1)	(-1, -1, 1) (-1, 1, -1)			
20	5	(-1, -1, -1)	(1, -1, 1)	(-1, 1, 1)	(1, -1, -1) (1, 1, -1) (1, 1, 1)	(-1, -1, 1) (-1, 1, -1)			
21	5	(-1, -1, -1)	(1, -1, -1) (1, 1, 1)	(-1, -1, 1)	(1, -1, 1) (1, 1, -1)	(-1, 1, -1) (-1, 1, 1)			
22	5	(-1, -1, -1)	(1, -1, -1) (1, 1, 1)	(-1, 1, 1)	(1, -1, 1) (1, 1, -1)	(-1, -1, 1) (-1, 1, -1)			
23	6	(-1, -1, -1)	(1, -1, 1)	(-1, 1, -1)	(1, 1, 1)	(-1, -1, 1) (-1, 1, 1)	(1, -1, -1) (1, 1, -1)		
24	6	(-1, -1, -1)	(1, -1, -1)	(-1, -1, 1)	(1, -1, 1)	(-1, 1, -1) (-1, 1, 1)	(1, 1, -1) (1, 1, 1)		
25	6	(-1, -1, -1)	(1, -1, -1)	(-1, -1, 1)	(1, 1, -1)	(-1, 1, -1) (-1, 1, 1)	(1, -1, 1) (1, 1, 1)		
26	6	(-1, -1, -1)	(1, -1, -1)	(-1, -1, 1)	(1, 1, 1)	(-1, 1, -1) (-1, 1, 1)	(1, -1, 1) (1, 1, -1)		
27	6	(-1, -1, -1)	(1, -1, -1)	(-1, 1, 1)	(1, 1, 1)	(-1, -1, 1) (-1, 1, -1)	(1, -1, 1) (1, 1, -1)		
28	6	(-1, -1, -1)	(1, -1, 1)	(-1, 1, 1)	(1, 1, -1)	(-1, -1, 1) (-1, 1, -1)	(1, -1, -1) (1, 1, 1)		
29	7	(-1, -1, -1)	(1, -1, -1)	(-1, -1, 1)	(1, -1, 1)	(-1, 1, -1)	(1, 1, -1) (1, 1, 1)	(-1, 1, 1)	
30	7	(-1, -1, -1)	(1, -1, -1)	(-1, -1, 1)	(1, 1, 1)	(-1, 1, -1)	(1, -1, 1) (1, 1, -1)	(-1, 1, 1)	
31	8	(-1, -1, -1)	(1, -1, -1)	(-1, -1, 1)	(1, -1, 1)	(-1, 1, -1)	(1, 1, -1)	(-1, 1, 1)	(1, 1, 1)

As previously mentioned, two SPDs are distinct if their information matrices are different. Clearly the information matrices of two SPDs involving different numbers of whole plots will be distinct since the structure of the covariance matrix, $\boldsymbol{\Sigma}$, depends on the number of whole plots. To illustrate two distinct SPDs which have the same number of whole plots, we will focus on designs 2 and 3 in Table 4.2. Although both designs are equally appealing from a cost perspective ($C = 8r+3$ in both cases), they are not equally appealing in terms of their information matrices. Note for design 2 that both linear main effects in x_1 and x_2 can be estimated within each of the three whole plots (i.e. contrasts exist for both x_1 and x_2 within each whole plot). In design 3, however, contrasts for x_1 only exists within whole plot 2 whereas contrasts for x_2 exist within all three whole plots. There is thus less information regarding the linear main effect for x_1 in design 3 than there is in design 2.

For purposes of our discussion, we will refer to the class of designs which are equivalent to design 2 as 'pattern A' and to the class of designs which are equivalent to design 3 as 'pattern B'. Pattern A designs are more efficient for estimation and prediction than pattern B designs. To further illustrate the comparison between pattern A and pattern B designs, their respective information matrices are provided below when the variance ratio is d=1 (i.e. equal whole plot and subplot error variances):

$$I(A) = \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} = \begin{bmatrix} 2.13 & 0.53 & 0 & 0 \\ 0.53 & 2.13 & 0 & 0 \\ 0 & 0 & 8 & 0 \\ 0 & 0 & 0 & 8 \end{bmatrix}$$

and

$$I(B) = \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} = \begin{bmatrix} 2.13 & 0.53 & 0 & 0 \\ 0.53 & 2.13 & 0 & 0 \\ 0 & 0 & 5.33 & 0 \\ 0 & 0 & 0 & 8 \end{bmatrix}$$

where $\mathbf{X}$ is the 8×4 design matrix given by $\mathbf{X}=[\mathbf{1},\mathbf{w},\mathbf{x}_1,\mathbf{x}_2]$ and $\mathbf{1}$ is the column of ones, $\mathbf{w}$ is the vector of whole plot levels, and $\mathbf{x}_1$ and $\mathbf{x}_2$ denote the vector of subplot levels. The determinant of *I(A)* is 273.07 and the determinant of *I(B)*=182.04. Note from the information matrices above, the designs are equivalent in terms of their information for

the whole plot main effect and the effect of x_2 but pattern B SPDs have less information for estimating x_1 than pattern A SPDs.

Comparing 2^3 SPDs in terms of *D*-efficiencies

We begin our study and comparisons of 2^3 SPDs in terms of their ability to precisely estimate model parameters as indicated by their *D*-criteria. Throughout our discussion, a first order model is assumed. When cost is not an issue, designs can be evaluated by the unadjusted *D*-criteria $\left(i.e.\ \left|\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\right|^{1/p}\right)$. The best 5 designs in terms of their unadjusted *D*-efficiencies are provided in Table 4.3 for various values of the variance component ratio, *d*. It is interesting to note that the best two designs in terms of their *D*-efficiencies remain the same as *d* changes from 0.5 to 10.

Interestingly, the *D*-optimal 2^3 SPD (design 13 in Table 4.2) involves 4 whole plots, each containing 2 subplots. Note that this is not the completely randomized design (CRD) (design 31 in Table 4.2), implying that often split-plot designs are more efficient than completely randomized designs when hard-to-change factors exist. Goos and Vandebroek (2004) provided similar conclusions regarding the superiority of the split-plot scheme over the CRDs when cost is not a consideration. Also note that the *D*-optimal 2^3 SPD is similar to the 'pattern A' designs discussed earlier in that contrasts for the subplot main effects occur within each whole plot.

Table 4.3 The best 5 designs with the best quality for *D*-efficiency, i.e., the 5 highest values of $\left|\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\right|^{1/4}$. Higher value indicates more information for the parameters and thus more precise estimates. *a* is the number of whole plots and ID represents the index of the design in Table 4.2. The sequence of whole plot sizes are listed in $(n_1,n_2,\ldots,n_a)$, where n_i is the number of subplot within the i^{th} whole plot of the design, for $i=1,\ldots,a$.

	a	ID	Whole plot sizes		*a*	ID	Whole plot sizes		*a*	ID	Whole plot sizes
d=0.5	4	13	2,2,2,2	*d*=1	4	13	2,2,2,2	*d*=10	4	13	2,2,2,2
	5	22	2,2,1,2,1		5	22	2,2,1,2,1		5	22	2,2,1,2,1
	6	28	1,1,2,2,1,1		6	28	1,1,2,2,1,1		4	12	1,2,3,2
	6	27	1,1,2,2,1,1		6	27	1,1,1,1,2,2		6	28	1,1,2,2,1,1
	7	30	1,1,1,1,1,2,1		5	21	2,2,1,2,1		5	21	2,2,1,2,1

It is important to note that the comparisons thus far in terms of the *D*-criterion have been made assuming that cost is not an issue. Using the cost adjusted *D*-criterion, the best 5 designs are reported in Table 4.4. Note that using the cost adjusted *D*-criterion results in a different set of five optimal designs. As with the set of five optimal designs in terms of the unadjusted *D*-criterion, this set remains similar for different values of the variance component ratio. This implies that even if the guess of the variance component ratio, *d,* is not precise, the best design for the guessed value of *d* may still be optimal or highly efficient. For instance, if the guess is d=1 and the actual value is d=3, the selected design based on d=1 may still be optimal or at least highly *D*-efficient. This is good news for the practitioner who often must select a design with little idea of the true value of *d*.

Table 4.4 The best 5 designs with the best cost adjusted estimation precision (calculated by equation (4.4)). The larger value indicates higher cost penalized *D*-efficiency and thus more desirable.

	a	ID	Whole plot sizes		*a*	ID	Whole plot sizes		*a*	ID	Whole plot sizes
r=0 d=0.5	2	1	4,4	r=0 d=1	2	1	4,4	r=0 d=10	2	1	4,4
	3	2	2,4,2		3	2	2,4,2		3	2	2,4,2
	3	4	1,4,3		3	4	1,4,3		3	4	1,4,3
	3	3	2,4,2		3	3	2,4,2		4	13	2,2,2,2
	4	13	2,2,2,2		4	13	2,2,2,2		3	3	2,4,2
r=0.5 d=0.5	2	1	4,4	r=0.5 d=1	2	1	4,4	r=0.5 d=10	4	13	2,2,2,2
	3	2	2,4,2		3	2	2,4,2		3	2	2,4,2
	4	13	2,2,2,2		4	13	2,2,2,2		2	1	4,4
	3	4	1,4,3		3	4	1,4,3		4	12	1,2,3,2
	3	3	2,4,2		4	12	1,2,3,2		3	4	1,4,3
r=1 d=0.5	4	13	2,2,2,2	r=1 d=1	4	13	2,2,2,2	r=1 d=10	4	13	2,2,2,2
	3	2	2,4,2		3	2	2,4,2		3	2	2,4,2
	2	1	4,4		4	12	1,2,3,2		4	12	1,2,3,2
	4	12	1,2,3,2		2	1	4,4		5	22	1,2,2,2,1
	3	4	1,4,3		5	22	1,2,2,2,1		2	1	4,4

Note that when the only appreciable experimental cost involves the cost of whole plots (r=0), the design with 2 whole plots is optimal and designs with small number of whole plots are preferred in terms of the cost adjusted *D*-criterion. The design with only two whole plots, however, has the unappealing property of non-estimability of whole plot error. This presents problems in the analysis in terms of variance component estimation

and is thus not a desirable design in practice. If each individual whole plot is equally expensive as each subplot (r=1), designs with moderate numbers of whole plots are preferred. The balanced design with 4 whole plots (optimal in terms of the unadjusted *D*-criterion) is best and design 2 from Table 4.2 consisting of 3 whole plots turn out to be second best. In summary, when cost is of concern, the most desirable SPDs are those with smaller numbers of whole plots than when cost is not a consideration.

Comparing 2^3 SPDs in terms of *V*- and *G*-efficiencies

The study of 2^3 SPDs in terms of parameter estimation when cost is not a consideration is thoroughly discussed by Goos and Vandebroek (2001, 2004). Often of interest to practitioners when choosing optimal designs is the ability of the design to precisely estimate the mean response. Very little work has been done in conjunction with choosing optimal SPDs in terms of their predictive capabilities. We now consider *V*- and *G*-efficiencies for the 31 possible 2^3 SPDs. Assuming a cuboidal region, the average prediction variance is calculated by

$$\frac{1}{8}\int_{-1}^{1}\int_{-1}^{1}\int_{-1}^{1}(PV \text{ or } CPPV)\, dw dx_1 dx_2$$

where *PV* denotes the prediction variance unadjusted for cost and *CPPV* denotes the cost adjusted prediction variance. The maximum prediction variance is obtained via a search of prediction variances over the design region $\{(w, x_1, x_2)$: $-1 \le w \le 1$, $-1 \le x_1 \le 1$ and $-1 \le x_2 \le 1\}$. Similar to our discussions involving the *D*-criterion, we first present optimal designs when cost is not a consideration. The 5 designs with the smallest average and maximum prediction variance are listed in Table 4.5 for various values of the variance component ratio.

SPDs with large number of whole plots (more than 6 in this example) are highly *V*-efficient, while highly *G*-efficient SPDs often include designs with smaller or moderate numbers of whole plots. When the variability among whole plots accounts for a small or medium proportion of the observational variance (d=0.5 or 1), the CRD is the most *V*-efficient design and exhibits the second highest *G*-efficiency. When the whole plot error

variance dominates the observational variance (d=10), the CRD is still highly V- and G-efficient. This observation is different to what was observed for the CRD in terms of D-efficiency where SPDs were often superior to the CRD.

Table 4.5 The best 5 designs in terms of average and maximum prediction variance. Smaller value indicates better performance in terms of V- and G-efficiency.

	Average prediction variance (V)			Maximum prediction variance (G)		
	a	ID	Whole plot sizes	a	ID	Whole plot sizes
d=0.5	8	31	1,1,1,1,1,1,1,1	6	28	1,1,1,1,2,2
	7	30	1,1,1,1,1,2,1	8	31	1,1,1,1,1,1,1,1
	6	28	1,1,1,1,2,2	4	13	2,2,2,2
	6	27	1,1,1,1,2,2	4	11	2,2,2,2
	7	29	1,1,1,1,1,1,2	5	21	1,2,1,2,2
d=1	8	31	1,1,1,1,1,1,1,1	6	28	1,1,1,1,2,2
	7	30	1,1,1,1,1,2,1	8	31	1,1,1,1,1,1,1,1
	6	28	1,1,1,1,2,2	4	13	2,2,2,2
	6	27	1,1,1,1,2,2	4	11	2,2,2,2
	7	29	1,1,1,1,1,1,2	5	21	1,2,1,2,2
d=10	6	28	1,1,1,1,2,2	6	28	1,1,1,1,2,2
	7	30	1,1,1,1,1,2,1	6	25	1,1,1,1,2,2
	6	25	1,1,1,1,2,2	8	31	1,1,1,1,1,1,1,1
	6	26	1,1,1,1,2,2	4	13	2,2,2,2
	8	31	1,1,1,1,1,1,1,1	4	11	2,2,2,2

Tables 4.6 and 4.7 present the five best designs for average and maximum CPPV. Recall that the CPPV adjusts the prediction variance according to the cost of experimentation. For the most part, the set of optimal designs remains consistent across the values of the variance component ratio, d. This is true for all of the comparisons discussed thus far. Comparing the results in Table 4.5 with those in Tables 4.6 and 4.7, we note that when cost is a consideration, the best designs tend to involve fewer numbers of whole plots than when cost is not an issue. As r approaches 0 (whole plots are increasingly more expensive than subplots), the best designs tend to involve fewer whole plots than in situations when $r > 0$.

Comparing Table 4.4 with Tables 4.6 and 4.7, we notice that the optimal cost adjusted D-efficient designs tend to have fewer whole plots than the optimal cost adjusted V- and G-efficient designs. It is interesting to note that some designs consistently exhibited good D-

, *V*-, and *G*-efficiencies. Design 13 from Table 4.2 exhibits good *D*-, *V*-, and *G*-efficiencies regardless of whether one adjusts for costs. Designs 1 and 2 from Table 4.2 possess good *D*-, *V*-, and *G*-efficiencies when cost is an issue but their *V*- and *G*-efficiencies were less impressive as the cost of whole plot approached that of the subplot.

Table 4.6 The best 5 designs in terms of average CPPV. Smaller value indicates better design and higher cost adjusted *V*-efficiency.

	a	ID	Whole plot sizes		*a*	ID	Whole plot sizes		*a*	ID	Whole plot sizes
r=0 *d*=0.5	2	1	4,4	*r*=0 *d*=1	2	1	4,4	*r*=0 *d*=10	2	1	4,4
	3	2	2,4,2		3	2	2,4,2		4	13	2,2,2,2
	3	3	2,4,2		3	3	2,4,2		4	11	2,2,2,2
	3	4	1,4,3		4	13	2,2,2,2		4	12	1,2,3,2
	4	13	2,2,2,2		3	4	1,4,3		4	6	2,2,2,2
r=0.5 *d*=0.5	4	13	2,2,2,2	*r*=0.5 *d*=1	4	13	2,2,2,2	*r*=0.5 *d*=10	6	28	1,1,1,2,2,1
	4	11	2,2,2,2		4	11	2,2,2,2		6	25	1,1,1,1,2,2
	2	1	4,4		4	12	1,2,3,2		6	26	1,1,1,1,2,2
	3	2	2,4,2		5	22	1,2,1,2,2		5	22	1,2,1,2,2
	4	12	1,2,3,2		4	6	2,2,2,2		5	21	1,2,1,2,2
r=1 *d*=0.5	4	13	2,2,2,2	*r*=1 *d*=1	4	13	2,2,2,2	*r*=1 *d*=10	6	28	1,1,1,2,2,1
	4	11	2,2,2,2		5	22	1,2,1,2,2		6	25	1,2,1,2,2
	5	22	1,2,1,2,2		6	28	1,1,1,1,2,2		6	26	1,2,1,2,2
	4	12	1,2,3,2		4	11	2,2,2,2		7	30	1,1,1,1,1,2
	5	21	1,2,1,2,2		5	21	1,2,1,2,2		5	22	1,2,1,2,2

In comparing Tables 4.3 and 4.5 (comparisons unadjusted for cost) to the results in Tables 4.4, 4.6, and 4.7 (cost adjusted comparisons), it is clear that taking the relative cost of the whole plots and subplots into consideration makes an important difference when evaluating SPDs. Frequently in industrial settings, the cost of the whole plots dominates the experimental cost. In these cases, split-plot designs with fewer numbers of whole plots can have the best overall cost penalized quality. However, other necessary properties, like the ability to estimate variance components, should also be considered. When cost and quality are both important, split-plot designs with a moderate number of whole plots are usually preferred. The best designs perform well for a wide range of *d* values, indicating that these designs are somewhat robust to initial guesses of the variance component ratio.

Table 4.7 The best 5 designs in terms of maximum CPPV. Smaller value indicates better design and higher cost adjusted *G*-efficiency.

	a	ID	Whole plot sizes		a	ID	Whole plot sizes		a	ID	Whole plot sizes
r=0 *d*=0.5	2	1	4,4	*r*=0 *d*=1	2	1	4,4	*r*=0 *d*=10	2	1	4,4
	4	13	2,2,2,2		4	13	2,2,2,2		4	13	2,2,2,2
	3	2	2,4,2		4	11	2,2,2,2		4	11	2,2,2,2
	3	3	2,4,2		4	6	2,2,2,2		4	6	2,2,2,2
	3	4	1,4,3		3	2	2,4,2		4	10	1,1,3,3
r=0.5 *d*=0.5	2	1	4,4	*r*=0.5 *d*=1	4	13	2,2,2,2	*r*=0.5 *d*=10	6	28	1,1,1,1,2,2
	4	13	2,2,2,2		4	11	2,2,2,2		4	13	2,2,2,2
	4	11	2,2,2,2		4	6	2,2,2,2		4	11	2,2,2,2
	4	6	2,2,2,2		2	1	4,4		4	6	2,2,2,2
	3	2	2,4,2		6	28	1,1,1,1,2,2		6	25	1,1,1,1,2,2
r=1 *d*=0.5	4	13	2,2,2,2	*r*=1 *d*=1	4	13	2,2,2,2	*r*=1 *d*=10	6	28	1,1,1,1,2,2
	4	11	2,2,2,2		4	11	2,2,2,2		4	13	2,2,2,2
	4	6	2,2,2,2		6	28	1,1,1,1,2,2		6	25	1,1,1,1,2,2
	2	1	4,4		4	6	2,2,2,2		4	11	2,2,2,2
	6	28	1,1,1,1,2,2		5	21	1,2,1,2,2		4	6	2,2,2,2

Discussions

If the design to be run were selected at random from the 40,320 possible experiments as sometimes occurs in industry (this was noted by Gunju and Lucas, 1999), it is helpful to have some indication of how likely we are to get a design with a reasonable efficiency. For this simple example with only a manageable number of run combinations, it is possible to obtain the relative frequency of the 31 distinct designs by sampling. To obtain a better design, a large number of designs can be sampled from the design space by randomly generating sequences of run orders. The final design selected should be the best from among those sampled.

A sampling approach outlined by Joiner and Campbell (1976) involves altering the probability of changing the whole plot and subplot levels in consecutive runs. The sampling strategy proceeds as follows: the level of each factor is generated at random and independently of other factors with the only restriction being that the sampling of the eight runs is performed without replacement. Between runs, the decision of whether to change the factor level or not is determined by draw from a Bernoulli distribution with a fixed probability of changing level, Pw for whole plot factors and Px for subplot factors.

Different factors may have different probabilities of changing levels depending on the difficulty associated with changing levels in practice.

We explore the chance to get the best or highly efficient designs by examining the efficiency and frequency profile of the distinct designs in the design space. Simulation shows approximately equal chances of getting efficient designs from permutation of the full set of possible designs or from sampling with Pw=Px=0.5 for all variables and the results are displayed in Figure 4.1. In the plots, relative efficiency (RE) is plotted on the x-axis, indicating how close the design is to the most efficient design among the 31 designs in terms of the *D*-, *V*- and *G*-efficiency with and without the cost adjustment. The *y*-axis represents the cumulative frequency for getting the designs that are at least as efficient as the design with RE=*x*. For instance, if a point (*x*=0.9, *y*=0.5) is on the curve, then the chance of obtaining a design that is at least 90% efficient is 50%. If the sampling technique is applied and there are *m* generated designs in the sample, then 50% of the *m* designs are among the top 10% in terms of efficiency, implying that sampling works well for searching good designs. In Figure 4.1, the top plots represent the evaluation not taking the cost of experiment into consideration, the middle plots correspond to the situation that the cost of whole plot dominates the experiment cost (*r*=0), and the bottom plots for the situations that whole plot is even expensive as subplot (*r*=1). The two variance ratios, *d*=1 and 10 are on the left and right sides, respectively.

The plots reveal that sampling is extremely effective in finding a design with high *V*-efficiency, for instance, we have 20%-80% chance of obtaining a design that is at least 90% *V*-efficient, except for the situation in which *d*=1 and *r*=0. In terms of *D*- and *G*-efficiency, it is less likely that sampling will yield a highly efficient design. The chance of finding the highly efficient designs by sampling is substantially lower when whole plots are extremely expensive (*r*=0), which indicate that the inadvertent split-plot designs are not desirable under some conditions. However, if the desirable properties of superior split-plot designs are known (such as a smaller number of whole plots for expensive cost ratio situations), we may improve the chance to get highly efficient design by adjusting the probability of level changes for each factor. For instance, from the quality and cost summary of the designs in Table 4.7 we know that a design with a small number of whole

plots is desirable when r=0. Therefore, by reducing the probability of changing whole plot levels appropriately, i.e., Pw=0.1 or 0.2, the chance of obtaining more efficient designs would be enhanced. However, this technique is based on the knowledge of the desirable characteristics of optimal designs The more knowledge we have about the characteristics of good designs, the more efficient the sampling can be. When the number of factors increases or the model considered becomes more complex, doing an exhaustive search of all possible designs rapidly becomes impractical. Hence, starting with several sensible designs based on good design of experiment principles may be a more realistic starting point.

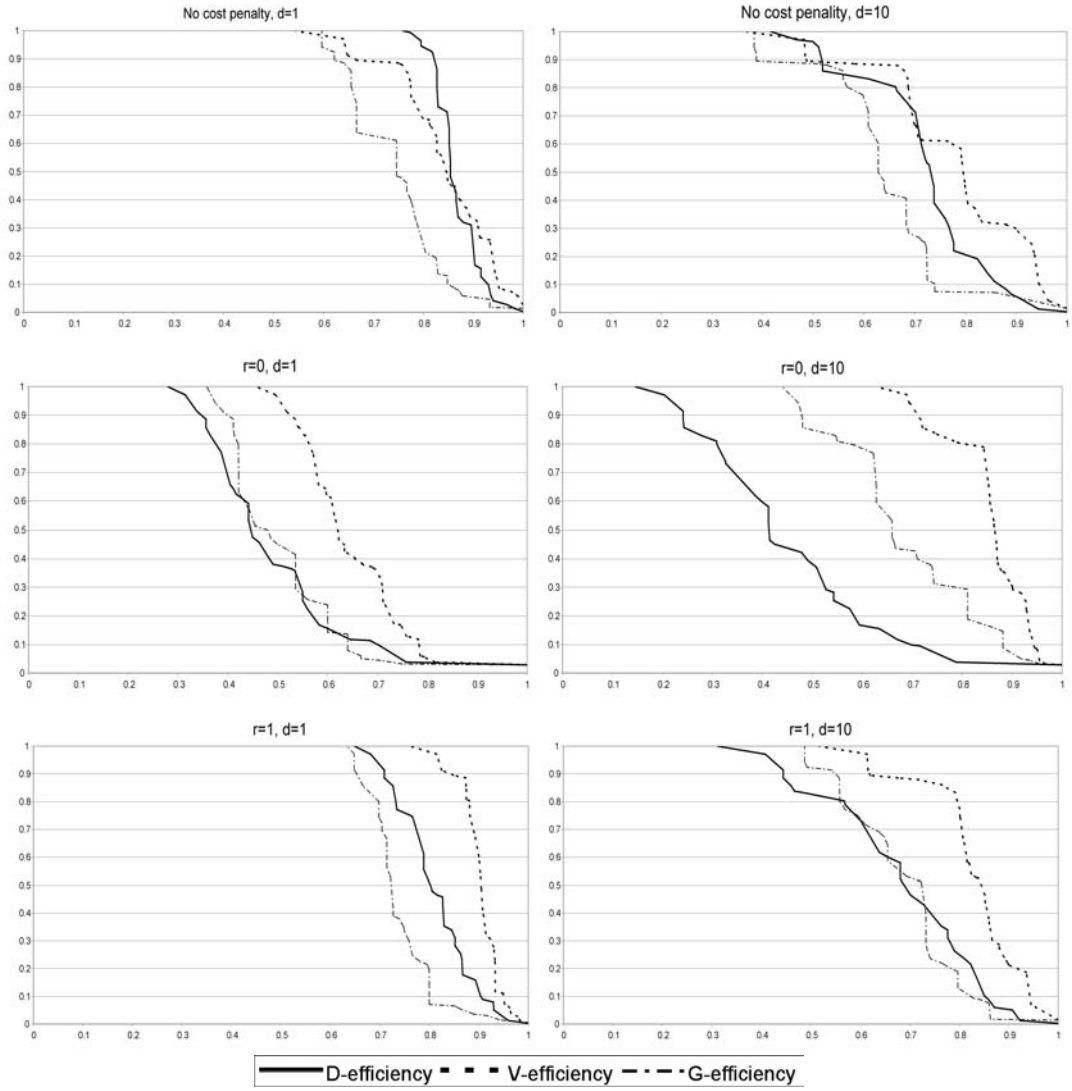


Figure 4.1 Accumulated frequency of getting highly efficient design for *D*-, *G*- and *V*-optimality under different combinations of cost ratio and variance component ratio.

4.5 Cost Adjusted Evaluations of Central Composite Designs with a Second Order Model

Second order models are important models in response surface methodology. Designs such as the central composite design (CCD) and the Box-Behnken design (BBD) are popular in practice and are known to be highly efficient in the CRD setting. The statistical properties of these designs in the split-plot randomizations scheme are not well understood. As we have observed for the 2^3 designs, the variance component ratio and the structure of the covariance matrix $\boldsymbol{\Sigma}$ play important roles in design performance. Also, as mentioned earlier, in split-plot experimentation, cost is frequently an issue and should be considered when determining an appropriate design. Letsinger, Myers, and Lentner (1996) compared popularly used response surface designs, such as the CCD, the BBD, the small composite design and full 3-level factorial designs in terms of *D*- and *V*-efficiencies (unadjusted for cost) for split-plot randomizations scheme. The CCD consistently performs better than the other designs for various values of *d* in terms of both estimation and average prediction performance in the design region. In this section, we compare variations of the central composite designs in terms of both unadjusted and cost adjusted *D*-, *G*-, and *V*-criteria and demonstrate strategies for selecting split-plot designs. We consider an experiment with one whole plot variable and two subplot variables and assume the following full second order model in the fixed effects

$$f(w, x_1, x_2) = \beta_0 + \beta_1 w + \beta_2 x_1 + \beta_3 x_2 + \beta_{12} w x_1 + \beta_{13} w x_2 + \beta_{23} x_1 x_2 + \beta_{11} w^2 + \beta_{22} {x_1}^2 + \beta_{33} {x_2}^2 .$$

The standard CCD run with a split-plot randomization scheme is provided in Table 4.8 and in our discussion, this version of the CCD will be referred to as 'standard' as well as 'D1'. The standard CCD has been commonly referred to as a restricted split-plot design (RSPD), meaning that there is only one whole plot for each whole plot combination in the design [see Letsinger et al (1996) and Goos and Vandebroek (2001)]. Since the statistical properties of the CCD within a split-plot structure have not been well understood, the formulation in Table 4.8 has been a common choice for the split-plot CCD in practice. Note that there is no replication for the whole plot levels, implying that the whole plot pure error is non-estimable for this design. In addition, the design is quite unbalanced,

where a "balanced" SPD is one in which all whole plots contain the same number of subplots. For the standard CCD, there is only one subplot within each axial level whole plot while there are 4 plus the number of center runs number of subplots within the whole plot corresponding to the center level of the whole plot variable (w=0). Assuming that the predominant cost in split-plot experimentation is due to changes in the whole plot levels, the standard CCD is inefficient in using the whole plots from a cost perspective.

Table 4.8 D1 - standard CCD with one whole plot variable, w, and two subplot variables, x_1 and x_2. The axial levels for all variables are $\pm\alpha = \pm\sqrt{3} \approx \pm 1.732$, total number of runs is 16 ($N=16$), it has 5 whole plots (a=5) and the whole plot sizes are (4,4,1,1,6).

Whole plot	w	x_1	x_2	No. of Runs per whole plot
1	-1	±1	±1	4
2	1	±1	±1	4
3	1.732	0	0	1
4	-1.732	0	0	1
5	0	± 1.732	0	2
	0	0	± 1.732	2
	0	0	0	2

Vining, Kowalski and Montgomery (2004) [henceforth referred to as VKM] recommend imposing minimum and maximum whole plot size restrictions. They propose a modified CCD (given in Table 4.9 and referred to as D2 in this manuscript), which satisfies the analysis condition that the generalized least square (GLS) estimates of the fixed effect parameters are equal to the ordinary least square (OLS) estimates. VKM assume that the only appreciable cost in SPDs is due to the changes in the whole plot levels (r=0). Under this assumption, they postulate that D2 is a desirable design. VKM point out that D2 is appealing in that, if adding more whole plot center runs with subplot centers, it allows for the estimation of pure whole plot error variance due to the replications of the subplot center runs within the whole plot centers (replicated WPs 6 here). VKM also point out that D2 is balanced and balance is attractive in the eyes of the practitioner due to ease of experimentation. VKM do not however, report on the quality of estimation and prediction of D2. Liang, Anderson-Cook, Robinson and Myers (2004) used graphical tools to study and compare D2 with D1 in terms of scaled prediction variance (SPV), and showed that D2 performs poorly compared to D1.

We also consider some other variations of the CCDs that may also be of interest. By adjusting the number of whole plots, the subplot levels for the runs within whole plots and the total number of observations, some alternate designs are obtained which may represent sensible variations of the CCDs. Design D3 in Table 4.10 has the same number of runs as D1 but with one more whole plot. In D3, the six whole plot center runs are split into two whole plots in such a way as to improve estimation for the whole plot and subplot terms. This is expected to bring benefit when the subplot cost account for a medium or a huge part of the cost of the experiment and the total number of runs is important, since this design has an additional whole plot but is still close to the minimal number of observations. Note that this design, like D1, is not a balanced SPD and could be viewed as less practical from a cost perspective when the whole plots are expensive due to only having one subplot within whole plots 3 and 4.

Table 4.9 D2 - Modified CCD by VKM. Its total number of runs is 24, 6 whole plots and the whole plot sizes are (4,4,4,4,4,4) – balanced design.

Whole plot	w	x_1	x_2	No. of runs per whole plot
1	-1	±1	±1	4
2	1	±1	±1	4
3	1.732	0	0	4
4	-1.732	0	0	4
5	0	± 1.732	0	2
	0	0	± 1.732	2
6	0	0	0	4

Since center runs improve the precision of estimating quadratic effects, design D4 in Table 4.11 modifies D3 by augmenting each whole plot with an additional subplot center run. When experimental cost is primarily a function of the number of whole plots, this design is expected to obtain better performance since it only involves five whole plots versus the six whole plots in D2 and D3.

D5 in Table 4.12 is obtained by adding one subplot center run to each whole plot of D2 and some of the extra observations are removed to control the total size of the design. It has the same size of observations and same number of whole plots as D2, but as we will observe, this design performs much better than the modified CCD (D2).

Table 4.10 D3 – same number of observations as D1, but there is one more whole plot; total number of runs is 16, 6 whole plots and the whole plot sizes are (4,4,1,1,3,3).

Whole plot	w	x_1	x_2	No. Runs per whole plot
1	-1	±1	±1	4
2	1	±1	±1	4
3	1.732	0	0	1
4	-1.732	0	0	1
5	0	± 1.732	0	2
	0	0	0	1
6	0	0	± 1.732	2
	0	0	0	1

Table 4.11 D4 - N=22, 5 whole plots and the whole plot sizes are (5,5,3,3,6).

Whole plot	w	x_1	x_2	No. Runs per whole plot
1	-1	±1	±1	4
		0	0	1
2	1	±1	±1	4
		0	0	1
3	1.732	0	0	3
4	-1.732	0	0	3
5	0	± 1.732	0	2
		0	± 1.732	2
		0	0	2

Table 4.12 D5 - N=24, 6 whole plots and the whole plot sizes are (5,5,3,3,5,3).

Whole plot	w	x_1	x_2	No. Runs per whole plot
1	-1	±1	±1	4
		0	0	1
2	1	±1	±1	4
		0	0	1
3	1.732	0	0	3
4	-1.732	0	0	3
5	0	± 1.732	0	2
		0	± 1.732	2
		0	0	1
6	0	0	0	3

Comparisons among the five designs will involve different combinations of the variance component ratio (d=0.5, 1 and 10) and the cost ratio (r =0, 0.5 and 1). Since the five CCDs considered involve different numbers of total observations, we will compare the designs based on cost adjusted D-, G-, and V-criteria where the adjustment is based different r values and solely on the design size, N. The cost penalization by N is the same adjustment as that popularly used for CRDs where design cost is based solely on the

number of runs. It is an indicative of situations in which the subplot/measurement cost is the only appreciable experimental cost in SPDs. The comparisons of the five designs in terms of their cost adjusted *D*-, *G*-, and *V*-criteria are presented in Table 4.13.

The best and second best designs are given in "bold" fonts and identified by "*****" and "+", respectively. When the cost is only a function of the number of whole plots (r=0), D4 is consistently the best or second best design in terms of both estimation and prediction for all values of the variance ratio considered. This result is intuitive since this design has the least number of whole plots. It appears that augmenting subplot center runs in the whole plots of the CCDs is a useful strategy when the cost of experimentation is due solely to whole plot costs. Note that this is intuitive since subplots are relatively cheap to run. When d has a large value, implying large whole plot variability, more whole plots are required to obtain acceptable estimation precision for whole plot terms, and therefore D5 has best average prediction performance. Interestingly, the standard CCD (D1) performs very well in terms of *D*-efficiency for all variance ratios and for all cost ratios. D3 and D1 are the two best designs in terms of *G*- and *V*-efficiencies when N is used to penalize the design. The modified CCD (D2) is consistently worse than the other designs in terms of *D*-, *G*-, and *V*-criteria for all variance component ratios and cost ratios considered. Finally we note that for these two cost scenarios highly *V*-efficient SPDs have more whole plots than the highly *D*- and *G*-efficient SPDs, a conclusion which is consistent with those from example 1 where a first order model was considered.

Table 4.13 The cost penalized *D*-, *G*- and *V*-efficiency of the five variations of CCDs, which are cost adjusted *D*-efficiency, maximum and average cost penalized prediction variance (CPPV), calculated from equation (4.4) and (4.5), for different combinations of d and r value, including the extreme case of penalization by the total number of runs, N. The best design and second best designs are in "Bold" fonts.

	Design	Cost penalization				
		$r = 0$	$r = 0.1$	$r = 0.5$	$r = 1$	N
		d=0.5				
CPD	D1 (4,4,1,1,6)	**0.51+**	**0.384+**	**0.195***	**0.121***	**0.158***
	D2 (4,4,4,4,4,4)	0.47	0.334	0.156	0.093	0.117
	D3 (4,4,1,1,3,3)	0.42	0.327	0.178	**0.113+**	**0.156+**
	D4 (5,5,3,3,6)	**0.582***	**0.404***	**0.182+**	0.108	0.132
	D5 (5,5,3,3,5,3)	0.502	0.358	0.167	0.100	0.125

Average CPPV	D1 (4,4,1,1,6)	2.3	3.04	**5.98+**	**9.661+**	**7.36+**
	D2 (4,4,4,4,4,4)	2.218	3.11	6.65	11.089	8.872
	D3 (4,4,1,1,3,3)	2.531	3.21	**5.91***	**9.28***	**6.749***
	D4 (5,5,3,3,6)	**1.966***	**2.83+**	6.29	10.617	8.65
	D5 (5,5,3,3,5,3)	**2.004+**	**2.805***	6.012	10.020	8.016
Maximum CPPV	D1 (4,4,1,1,6)	3.737	4.933	**9.716***	**15.695***	**11.958+**
	D2 (4,4,4,4,4,4)	3.649	5.108	10.946	18.244	14.595
	D3 (4,4,1,1,3,3)	4.433	5.615	10.343	**16.253+**	**11.820***
	D4 (5,5,3,3,6)	**3.047***	**4.388***	**9.751+**	16.454	13.407
	D5 (5,5,3,3,5,3)	**3.458+**	**4.841+**	10.373	17.289	13.831
d=1						
CPD	D1 (4,4,1,1,6)	**0.598+**	**0.453+**	**0.23***	**0.142***	**0.187***
	D2 (4,4,4,4,4,4)	0.507	0.362	0.169	0.102	0.127
	D3 (4,4,1,1,3,3)	0.482	0.381	0.207	**0.132+**	**0.181+**
	D4 (5,5,3,3,6)	**0.666***	**0.463***	**0.208+**	0.123	0.151
	D5 (5,5,3,3,5,3)	0.571	0.408	0.190	0.114	0.143
Average CPPV	D1 (4,4,1,1,6)	2.331	3.077	**6.062+**	**9.792+**	**7.459+**
	D2 (4,4,4,4,4,4)	2.351	3.291	7.053	11.755	9.404
	D3 (4,4,1,1,3,3)	2.455	3.11	**5.728***	**9.002***	**6.547***
	D4 (5,5,3,3,6)	**2.065+**	**2.973+**	6.607	11.148	9.086
	D5 (5,5,3,3,5,3)	**2.028***	**2.840***	6.085	10.141	8.113
Maximum CPPV	D1 (4,4,1,1,6)	3.915	5.168	**10.180+**	**16.445+**	**12.529+**
	D2 (4,4,4,4,4,4)	3.75	5.25	11.25	18.75	15
	D3 (4,4,1,1,3,3)	4.645	5.883	10.838	17.031	**12.386***
	D4 (5,5,3,3,6)	**2.980***	**4.291***	**9.535***	**16.091***	13.111
	D5 (5,5,3,3,5,3)	**3.310+**	**4.634+**	*9.930*	16.550	13.240
d=10						
CPD	D1 (4,4,1,1,6)	**1.854+**	**1.405***	**0.713***	**0.442***	**0.579***
	D2 (4,4,4,4,4,4)	1.203	0.859	0.401	0.241	0.301
	D3 (4,4,1,1,3,3)	1.455	1.149	**0.623+**	**0.397+**	**0.546+**
	D4 (5,5,3,3,6)	**1.956***	**1.358+**	0.611	0.362	0.445
	D5 (5,5,3,3,5,3)	1.656	1.183	0.552	0.331	0.414
Average CPPV	D1 (4,4,1,1,6)	2.356	3.109	**6.124+**	**9.893+**	**7.539+**
	D2 (4,4,4,4,4,4)	2.677	3.748	8.072	13.387	10.708
	D3 (4,4,1,1,3,3)	**2.149+**	**2.722***	**5.013***	**7.878***	**5.731***
	D4 (5,5,3,3,6)	2.303	3.316	7.368	12.434	10.133
	D5 (5,5,3,3,5,3)	**2.058***	**2.881+**	6.173	10.289	8.231
Maximum CPPV	D1 (4,4,1,1,6)	**4.193+**	**5.535***	**10.901***	**17.610***	**13.417+**
	D2 (4,4,4,4,4,4)	5.591	7.827	16.773	27.955	22.364
	D3 (4,4,1,1,3,3)	4.982	6.311	**11.626+**	**18.269+**	**13.286***
	D4 (5,5,3,3,6)	**3.972***	**5.720+**	12.710	21.448	17.476
	D5 (5,5,3,3,5,3)	4.723	6.612	14.168	23.614	18.891

*: best designs; +: second best designs.

In practice, more realistic situations are frequently between the above two extreme scenarios of r=0 and scaled prediction variance (N is used), that is, both the cost of whole

plots and cost of subplot have to be considered. For these situations, we focus on design comparisons involving the columns associated with r = 0.1, 0.5 and 1 in Table 4.13. When r=1 or 0.5, the standard CCD is the best in terms of D- and G-efficiency, and it has the second best average prediction (V-efficiency). D3 is the most V-efficient design and is second best in terms of D- and G-efficiency. Note that both designs (D1 and D3) involve the least total number of runs (N=16), making them more competitive when the total size of design is important. D3 has one more whole plot than D1. For the comparisons done here, V-efficient CCDs have more whole plots than the D- and G-efficient CCDs. This is consistent with the conclusions of example 1 for the first-order model.

In D3, the additional center runs in the axial levels whole plots help with the estimation of quadratic terms. Splitting the whole plot at w=0 for D1 into two whole plots for D3 does not substantially compromise the estimation of the subplot quadratic terms, however, it does improve the estimation of the whole plot quadratics. On the other hand, although the modified CCD (D2) has good properties from an analysis standpoint, the quality of estimation and prediction is poor, because the subplot axial points are assigned separately from subplot center runs and hence the estimation of subplot quadratics is with whole plot error.

More detailed comparisons of the CCDa can be made by considering the distributions of prediction variance across the design region for each design. Liang *et al.* (2004, 2005) proposed the use of three-dimensional variance dispersion graphs (3-D VDGs) and fraction of design space (FDS) plots for comparing SPDs. These graphical tools show the locations with best or worst prediction over the entire design region as well as for any particular sub region (where sub region implies whole plot region only or subplot region only). In Figure 4.2, the global FDS plots for the five designs are displayed. See Zahran, Anderson-Cook and Myers (2003) for more details on FDS plots. The horizontal "FDS" axis represents the fraction of the design space with cost adjusted prediction variance less or equal to the given CPPV value from the vertical axis. The maximum prediction variance is displayed by the CPPV-value at FDS=1 and the average of the values on the curve shows the average prediction variance over the entire region, and thus the values of

G- and *V*-efficiency summarized in Table 4.13 can be read from the plots. When d=1 and r=0, D5 is the best design and this is manifested in the FDS curve having the smallest values and a flatter slope. As the whole plot variance increases relative to the subplot variance, D5 is still the most desirable design, thus demonstrating that the choice of the best design is robust to changes in the values of the variance component ratio. Note that D3 with the 6 whole plots and 16 observations performs much better when d increases from 1 to 10. This is due to the fact that larger numbers of whole plots improves the performance of SPDs when the proportion of whole plot error variance in the systematic variation increases. This design is superior to the other designs for situations when subplot is equally expensive as whole plot, r=1. Note also that the superiority of D3 over the others is robust to changes in the value of d.

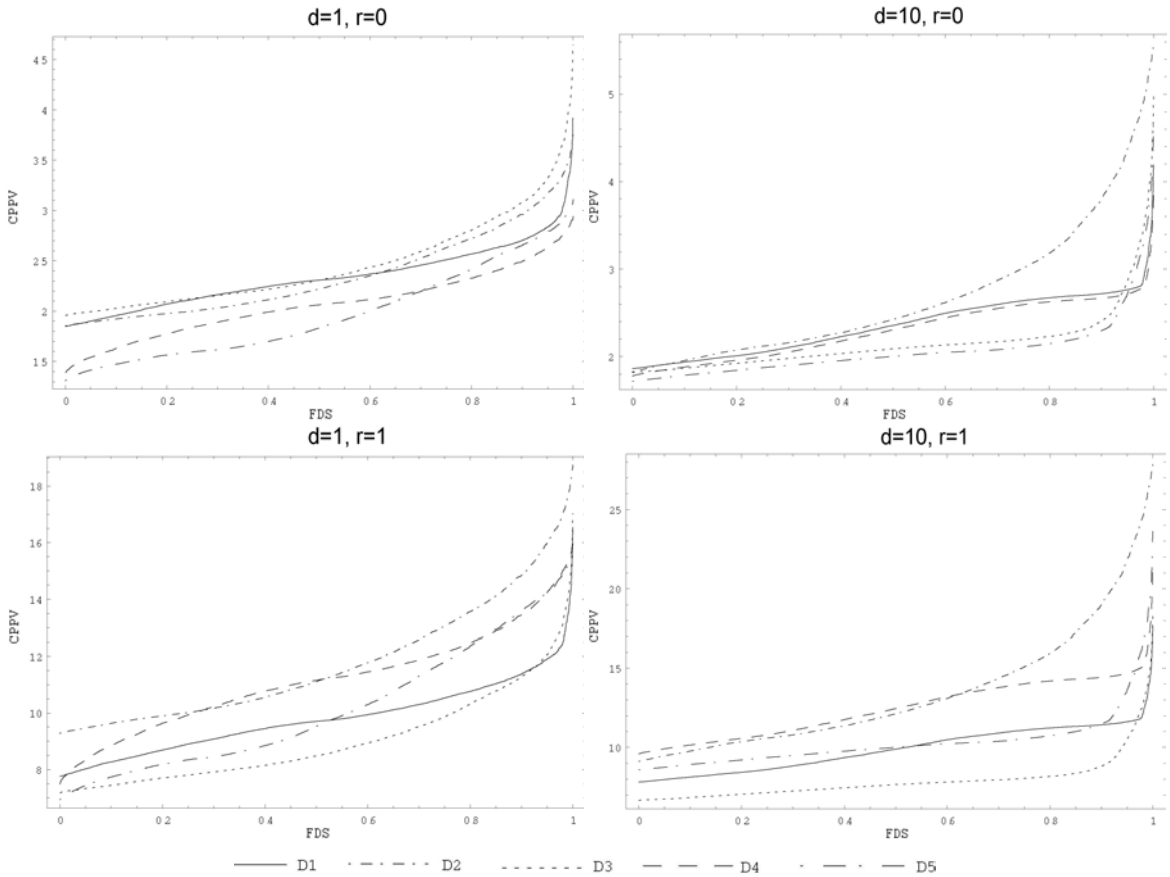


Figure 4.2 FDS plots for the five candidate designs under different scenarios of cost ratio and variance component ratio.

In Figure 4.3, 3-D VDGs are plotted for the average CPPV for the two best designs (D3 and D5) under different cost and variance ratio values. See Liang, *et al*, (2004, 2005) for more details on these plots. In the plots, "*w*" indicates the distance of the location from the center in the whole plot space, and "*x*" represents the location in the subplot space. The vertical axis denotes the average cost adjusted prediction variance at a location in the combined space (*w,x*). The plots show a common characteristic of the prediction variance distribution for desirable designs in terms of *V*-efficiency. Namely, the prediction variance is stable and relatively small for the broad center area of the design space and prediction deteriorates for a small portion at the edge of the whole plot space. These plots can be helpful in understanding the advantages and weaknesses of the designs in terms of their prediction capabilities.

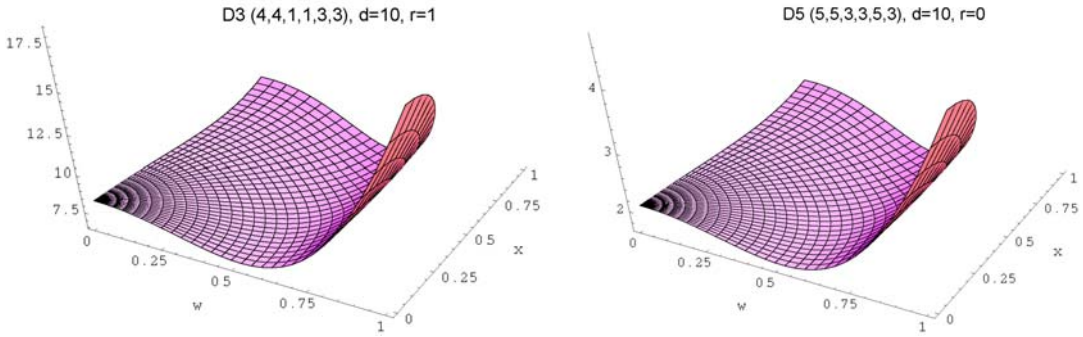


Figure 4.3 Surface plots of average CPPV for D3 and D5 at different scenarios.

From comparisons of the five designs one can learn that under different weighting conditions of cost and quality, the best design differs significantly. When the whole plots are extremely expensive, one may minimize the number of whole plots and then seek to achieve the best performance by assigning as many subplots as possible within each whole plot. When the subplot/measurement are comparably expensive to the whole plot, the design with fewer runs is desirable. When the subplot/measurement cost dominates the experimental cost, the prediction variance scaled by the design size (*N*) is appropriate for evaluating SPDs. In this case, smaller sized designs are desirable.

In addition, the example provides helpful information for the practitioner when choosing a split-plot design for a second order model. The standard CCD has good performance

when cost is incorporated, providing support for extending this type of response surface design from the CRD setting to the SPD setting. Moreover, if whole plot size is limited, for instance, the maximal number of subplots accommodated in the whole plot can not exceed six in practice, then running all combinations of subplot factors within each whole plot is not feasible. Thus we might have incomplete subplot levels within whole plots. The intuition would lead to the most balanced setting in the whole plots, such as the modified CCD does. However, this example shows that exact balanced or symmetric setting of the subplot levels may be less efficient. For instance, it may be better to assign the subplot center runs with axial or factorial levels in the same whole plot rather than separating the axial points and center runs.

4.6 Conclusions

For different problems or under different conditions of the experiment in real life, the experimenter may want to focus on different combinations of performance and cost for assessing split-plot experiments. Incorporating the split-plot structure and cost scenarios into the evaluation of split-plot design is helpful to better understand their effect on the overall desirability of the design. The proposed cost penalized *D*-efficiency, average and maximum cost penalized prediction variance (cost adjusted *V*- and *G*-efficiencies), including a special case of the scaled prediction variance, provide strategies for the practitioners rather than choosing the designs arbitrarily based on the available resources. Different r ratios require the experimenter to evaluate the relative costs of the whole plot and subplot units, as measured in time, effort or money to change the levels of two types of experimental variables and for measuring the observations. This should be based on the understanding of the practical conditions for running the experiments. From our comparisons of different designs, we learn that the standard scaled prediction variance evaluation commonly used for completely randomized design only makes sense under some special cases for split-plot design, and thus the generalization of SPV from CRD to SPD should be done carefully.

Although constraints on time or cost may suggest to the practitioner to use as small size of design as possible, other desirable properties, such as the ability to estimate the whole plot and subplot error terms, and adequate precision for estimating the whole plot factor effects should also be taken into account.

In industrial experiments, an important problem for the practitioner is to select a response surface design with a desirable structure when there are restrictions on randomization. This study shows that adapting central composite designs in a variety of ways can help improve performance for different cost and variance ratio scenarios. Some desirable strategies for assigning subplot levels within whole plots are also provided, which argue against the intuition that a balanced design is always preferable. However, there may be some benefits during analysis.

While the estimation of cost ratio can be obtained from understanding of the conditions for a given split-plot experiment, the variance component ratio is probably not as easy to estimate, if a pilot study or previous data are not available. However, the study in the two examples implies that the choice of a highly efficient design is frequently quite robust to the change of variance component ratio value. If the guessed or estimated d value is slightly different from the obtained value once the experiment has been run, the chosen design frequently remains highly efficient. This robustness means that good performance is likely even when the split-plot design is selected based on limited information about the variance component ratio.

Finally, this new approach to flexibly balancing quality and cost will hopefully provide practitioners with more realistic mechanisms for selecting a design that is most appropriate to their requirements.

Chapter 5 Evaluation for Multiple Criteria

5.1 Introduction

In Chapter 4, we have studied several variations of central composite designs (CCDs) with one whole plot and two subplot variables when taking the cost of the experiment into consideration. We learned that when the whole plots are expensive, a design with fewer whole plots is more desirable. When the subplots and measurements are costly, the restricted SPD CCD (D1) is the best in terms of estimation and minimizing the maximum prediction variance. This is due to the fact that D1 involves a small total number of runs. To lower average prediction variance over the entire design region, more whole plots are required. Therefore, D5 and D3 are the best designs with the lowest average CPPV under the situations that whole plots are expensive (r values close to 0) and the cost of subplot and measurement is comparable to cost of whole plot (r=1), respectively.

In Chapter 2, 3-D VDGs are used to formulate optimization strategies involving the factorial levels of a restricted SPD CCD. The G- and V-optimal levels are obtained analytically and we demonstrated that significant improvements can be realized by optimizing the factorial levels. The directions of the factorial level changes for G- and V-optimization are opposite, inferring that the optimal factorial levels for one criterion result in worse performance in terms of another criterion. It also means that the efficiency of one aspect is sacrificed to improve the performance in another aspect and a design cannot be improved for all optimality criteria simultaneously by changing the factorial levels. When good performance across several criteria is most important, the practitioner might have to consider the balance between the different criteria to avoid making the difficult decision of choosing just one particular criterion. The goal of high efficiency for all criteria of interest may be a more desirable choice. D-efficiency for estimation and V- and G-efficiency for prediction are the optimality criteria most commonly used in the literature, and in this chapter we evaluate split-plot designs using these criteria, which are treated with equal importance. In the following sections, we study the comprehensive

performance of the restricted SPD CCD (RSPD CCD), D3 and D4 with standard, *D*-, *G*- and *V*-optimal factorial levels.

When the practitioner seeks to find a design that exhibits good estimability of model parameters as well as quality of prediction, it may be of interest to evaluate a design in terms of a combined measure of the *D*-, *G*- and *V*-efficiencies. In general, the overall performance should be a function of the combined efficiencies for the multiple criteria. Derrinnger and Suich (1980) and Myers and Montgomery (2002, pp. 247-258) use desirability function for searching the optimum conditions for multiple responses. The desirability is a power function of the response and thus it ranges from 0 to 1 with different changing rate with respect to the response value. By specifying the power parameter in the desirability function for each response, the subject knowledge about the role of each response in the total desirability of the product is incorporated into the assessment of the multiple responses. Heredia-Langner, Montgomery, Carlyle and Borror (2004) apply desirability function approach to study model-robust designs. A desirability value is assigned to the design for each model and use the geometric mean of the values for all the models considered as the overall evaluation of design's model robustness. In this chapter, we want to assess the design based on the three criteria. Each criterion has already been in the range of (0,1), so the function combining the three efficiencies can be used as the objective function to assess the overall performance of a SPD and to compare between different designs. The geometric mean is one of them. It is desirable for that, if the value is unacceptable for any of the multiple criteria, the overall performance is unacceptable. Another option is to use arithmetic mean, which put the same weight to each criterion and penalize the poor design less than the geometric means does. The weighted average also could be used to allocate different weights to the multiple criteria to reflect the situation that one of the criteria could be more important than another. In this chapter, we consider the case that the three efficiencies are equally important. If one of the efficiencies is more important than the others for a particular experiment, the weighted arithmetic mean or weighted geometric mean can be employed to incorporate the importance of each criterion in the whole desirability of the designs. The arithmetic mean and geometric means are compared in the following section and we demonstrate

that the two assessing approaches provide similar comparison results. Therefore, the product is chosen as the overall assessment for remaining comparisons.

Another related question concerns the performance of CCDs with standard factorial levels to CCDs with optimal factorial levels. Recall from Chapter 4 that we determined the best CCDs (namely D3 and D4) based on particular scenarios for the cost and variance component ratios. In this section, we determine the optimal factorial levels for these designs in terms of *D*-, *G*- and *V*-efficiency, and discuss the degree of improvement that can be obtained for these designs using the optimal factorial levels. As mentioned previously, the practitioner may wish to know how these designs perform across all criteria combined. As a result, we study the designs from Chapter 4 in terms of their combined performances.

This study also compares between the different designs, say the restricted SPD CCD (D1), D3 and D4 with standard and optimal levels, for balance between multiple criteria and for each individual criterion. The cost scenarios are taken into consideration for comparing variations of the CCDs for individual and multiple criteria. All the designs have three variables: one whole plot variable, w, and two subplot variables, x_1 and x_2, and the second order model is considered.

5.2 Optimizing Factorial Levels for the Restricted SPD CCD (D1) and Overall Performance Based on Multiple Criteria

The restricted SPD CCD is a common way to run the central composite design for split-plot experiments (refer Table 4.1 for details of the set up of the design). The 3-D VDG and FDS plots (in Figures 2.4, 2.5 and 3.2) display the characteristics of its prediction performance. We note the following observations: the prediction variance is stable in the subplot space but varies more along the whole plot space; close to the center of the whole plot space there is a relatively large prediction variance, and the worst prediction occurs at the edge of the whole plot space. In Chapter 2, *V*- and *G*-optimal levels are listed.

Here, an exhaustive search for the optimal whole plot and subplot factorial levels, f1 and f2, determines the *D*-optimal levels, which are summarized in Table 5.1 along with the previously noted *V*- and *G*-optimal levels. Note that here the restricted SPD CCD with the standard factorial levels f1=f2=± 1 are used as a basis for comparison to evaluate the optimization. The relative *D*-efficiency is defined as the ratio of the *D*-efficiency of the optimal design to that of the RSPD CCD, and the relative *V*- and *G*-efficiencies are the ratios of the *V*- and *G*-efficiencies for the restricted SPD CCD with standard levels to those of the CCDs with the optimal levels. As a result, values larger than 1 indicates an improvement in efficiency upon using the optimal factorial levels. Note the resulting gains in efficiency hold for all cost scenarios, because the total number of runs and the whole plots sizes do not change for different factorial levels. For *D*-optimization, we can see that the *D*-optimal levels are close to the standard factorial levels. The *D*-efficiencies increase less than 5%, which shows that estimations of the parameters do not improve significantly.

Table 5.1 *D*-, *G*- and *V*-optimal factorial levels for the RSPD CCD with one whole plot variable and two subplot variables. Under the columns of *D*-, *G*- and *V*-optimal, f1 and f2 are the optimal factorial levels in terms of *D*-, *G*- and *V*-efficiency respectively, and the relative efficiency is compared to the RSPD CCD with standard levels, and larger values indicate more improvements by using the optimal factorial levels.

d	*D*-optimal			*V*-optimal			*G*-optimal		
	f1	f2	RE	f1	f2	RE	f1	f2	RE
0	0.9998	1.001	1.001	1	1	1	1	1	1
1	0.924	1.036	1.004	0.78	1.09	1.036	1.13	0.92	1.079
10	0.906	1.044	1.007	0.51	1.17	1.166	1.22	0.80	1.115

When we wish to consider balancing multiple criteria, for instance, *D*-, *G*- and *V*-optimality simultaneously, the arithmetic mean and geometric mean are two of popularly used functions to find the design with the best balance between different criteria. Here we consider the three criteria are equally important, and thus the average (arithmetic mean) and product are utilized (the product is equivalent to geometric mean because of the fixed number of criteria considered and for simplicity of calculation). We found that for these and all remaining comparisons in this chapter, whether the average or product are utilized, the design ranking remains unchanged, so the product of efficiencies is used as

the overall efficiency for the three criteria in the remaining work. Table 5.2 provides the overall efficiencies for variance component ratios of d=1 and d=10, where the efficiencies are all relative efficiency to the RSPD CCD (D1) with standard levels. Larger values indicate better performance.

From Table 5.2, we note that there are negligible differences between the average and product. The *D*-optimal D1 is not only better than the standard D1 in terms of *D*-efficiency but it is also better in terms of *V*-efficiency. The *V*- or *G*-optimal D1, however, has lower efficiencies than the standard D1 for all the optimality criteria other than itself. When all three efficiencies are combined, the standard RSPD CCD (D1) is the best for a small variance ratio, d=1, and second best for a large variance ratio, d=10. The *D*-optimal CCD is also a good choice. However, as noted above, the *D*-optimal levels are close to the standard levels ± 1.

Table 5.2 Performance of the standard and optimal RSPD CCDs in terms of *D*-, *G*- and *V*-efficiency and their overall performance for multiple criteria.

		Standard	*D*-optimal	*V*-optimal	*G*-optimal
d =1	*D*-eff	1	1.004	0.992	0.964
	V-eff	1	1.021	1.036	0.949
	G-eff	1	0.963	0.909	1.076
	Average	1	0.996	0.979	0.996
	Product	**1**	0.986	0.934	0.983
d =10	*D*-eff	1	1.007	0.903	0.864
	V-eff	1	1.053	1.164	0.849
	G-eff	1	0.966	0.876	1.113
	Average	1	1.009	0.981	0.942
	Product	1	**1.024**	0.920	0.817

Based on the information shown in Table 5.2, the standard and *D*-optimal designs are the two designs with the best overall performance for the three optimality criteria. In practice, the exact value of the variance component ratio, d, is often unknown and in these cases, the RSPD CCD with the standard factorial levels may be a better choice for practitioners since the *D*-optimal design requires that the user finds the optimal f1 and f2 values and these values are based on the anticipated variance ratio.

5.3 Optimizing Factorial Levels for D3 and Overall Performance for Multiple Criteria

Recall from Chapter 4, the design D3 is obtained by adapting the RSPD CCD by splitting the whole plot at w=0 into two equal-sized whole plots (see Table 4.10 for a complete description of the design). The total number of runs of the design remains the same but the number of whole plots increases. When the subplot and measurement cost is comparable to the whole plot cost, the RSPD CCD and D3 are desirable choices. Because the good performance of average prediction variance requires more whole plots, D3 is better than the RSPD CCD in terms of V-efficiency, while the RSPD CCD is better than D3 for estimation and maximum prediction variance.

The cost penalized prediction variance distribution of D3 is displayed in Figure 4.3(a) for the case of r=1 and d=10. The pattern of the CPPV distribution is consistent for different scenarios of cost ratio, since the number of whole plots and total number of observations remains constant for different cost ratios. We observe that the CPPV is stable in the broad center region and increases rapidly in a small portion of region at the edge of the whole plot space. To lower the average prediction variance, small values in most of the region is desirable, and thus D3 might have a prediction variance distribution close to the optimal setting. On the other hand, to improve the G-efficiency, only the location with the worst prediction, i.e., the edge of the whole plot space, is of interest. By moving the factorial points, the prediction variance at this location can be easily lowered. Meanwhile, the prediction variance in the center might be somewhat sacrificed. The D-, G- and V-optimal factorial levels for D3 are listed in Table 5.3. The relative efficiencies are defined as the ratio of the efficiency values for D3 with standard levels ± 1 to the efficiency values associated with the optimized D3.

Table 5.3 D-, G- and V-optimal factorial levels for D3, which has sixteen design points and six whole plots, the setting of whole plot sizes are (4,4,1,1,3,3).

d	D-optimal			V-optimal			G-optimal		
	f1	f2	RE	f1	f2	RE	f1	f2	RE
0	1	1.001	1.001	1	1	1	1	1	1
1	0.944	1.027	1.002	0.91	1.04	1.005	1.15	0.85	1.107
10	0.928	1.035	1.004	0.64	1.14	1.046	1.45	0.66	1.334

Although the *V*-optimal factorial levels change considerably for different variance ratios, the relative *V*-efficiency does not improve significantly, implying that the design's average prediction variance distribution does not change much. As a result, one can conclude that the *V*-efficiency of D3 is robust to the factorial level changes, and the design D3 with standard levels has efficiency close to the *V*-optimal D3. The shift in the factorial levels locations for *G*-optimization is opposite to that for *V*-optimization, and substantial improvement is possible for this criterion. For instance, when d=10, the maximum prediction variance drops more than 30% by moving the whole plot factorial levels outwards and moving the subplot factorial points into the center. We also note that the standard levels are close to the *D*-optimal factorial levels and only a small increase in *D*-efficiency is possible. Therefore, when the *G*-efficiency is of primary interest, the *G*-optimal D3 should be chosen. But if *D*- and *V*-efficiencies are of interest, D3 with standard levels is quite efficient, which should provide more confidence for the practitioner to choose the standard factorial levels for the design D3 when information about the variance ratio is not available.

Table 5.4 Performance of the standard and optimal D3's in terms of *D*-, *G*- and *V*-efficiency and their overall performance for multiple criteria.

		Standard	*D*-optimal	*V*-optimal	*G*-optimal
d =1	*D*-eff	1	1.002	1.001	0.904
	V-eff	1	1.004	1.005	0.942
	G-eff	1	0.970	0.952	1.107
	Product	**1**	0.976	0.957	0.943
d =10	*D*-eff	1	1.004	0.954	0.792
	V-eff	1	1.016	1.046	0.850
	G-eff	1	0.971	0.892	1.334
	Product	**1**	0.991	0.890	0.898

The performances of D3 with different factorial levels are studied in Table 5.4. The relative efficiencies are compared to the D3 with the standard factorial levels, f1=f2=$\pm$ 1 and a larger value indicates better performance. From the table, it is observed that the *V*- and *G*-optimal design have much worse performances for the criteria other than itself, especially for large variance ratios. This implies that when focusing on a single criterion, one may obtain a design that performs poorly in terms of other criterion. A natural

consequence of this is that the *V*- and *G*-optimal D3s perform poorly in terms of the combination of all three criteria. D3 with the standard levels ± 1, however, is uniformly the best design when considering multiple criteria simultaneously regardless of the relative size of whole plot variability to the subplot variance.

5.4 Optimizing Factorial Levels for D4 and Overall Performance for Multiple Criteria

When whole plots are extremely expensive, the cost ratio, *r*, can be close to 0. Under this cost scenario, the practitioner would like not only to limit the number of whole plots, but also to have good quality of estimation and prediction by augmenting more designs points in the available whole plots. D4 is appealing for this case, because it has the least number of whole plots for CCDs with one whole plot variable and one subplot center run is augmented in each factorial level whole plot compared to D1 (D4 has 5 whole plots and 22 design points, the setting of whole plot sizes is (5,5,3,3,6), please refer to Table 4.11 for details of the factor levels setting).

In this section, we first use the 3-D VDG to study the prediction variance distribution over the spherical design space when variance structure changes. Since the cost penalized prediction variance is independent of the cost scenarios, we consider the prediction variance without any scaling for the design size, *N* or cost of the design. Note that the distribution pattern of the prediction variance is the same for any cost scenarios. Figure 5.1 display the maximum prediction variance when d=0 (CRD), 1 and 10. For a large variance ratio (d=10), the pattern for design D4 is similar to RSPD CCD (D1) with large changes in prediction variance across the whole plot space but small changes across the subplot space. As a result, we will observe that strategies for improving the maximum or average prediction variance for D4 will involve moves in the whole plot factorial levels of greater magnitude than moves in the subplot factorial levels.

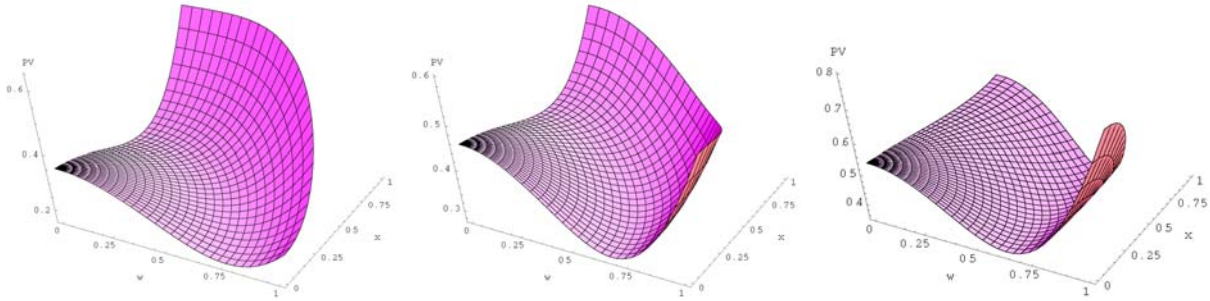


Figure 5.1 Surface plot of maximum prediction variance for D4 with standard factorial levels when d=0 (CRD) (left), d=1 (middle) and d=10 (right).

The *D*-, *G*- and *V*-optimal levels for this design are provided in Table 5.5. *V*-optimization moves the whole plot factorial points towards the center and subplot factorial points outwards. The changes of factorial levels are substantial and *V*-efficiency is improved by more than 10% for large d values (d=10). The *G*-optimal designs reduce the maximum prediction variance to some extent and the movement of the factorial levels for optimizing the *G*-criterion is observed to be in the opposite direction to the movement for *V*-optimization only for d=10. For small d values, the whole plot factorial points are moved towards the center. Figure 5.1 illustrates the reason of different patterns of *G*-optimization from the 3-D VDGs. For d=0 and 1, the location with the worst prediction is the edge of subplot space at the center of whole plot space, while for d=10 the maximum prediction variance is at the edge of whole plot space and center of subplot space. Due to adding the subplot center runs in some of the whole plots, the location being optimized changes for different d values, and therefore optimization strategies in terms of the *G*-criterion show no consistent pattern across the various d values. The *D*-optimal factorial levels are similar (around 0.86 for whole plot variable and 1.06 for subplot variables) for different variance component ratios, which implies the *D*-optimal D4 is quite robust to the variance ratios. Note, however, the *D*-efficiency is not substantially improved using the *D*-optimal factorial levels.

Table 5.5 *D*-, *G*- and *V*-optimal factorial levels for D4, which has 22 design points and five whole plots, the setting of whole plot sizes are (5,5,3,3,6).

d	*D*-optimal			*V*-optimal			*G*-optimal		
	f1	f2	RE	f1	f2	RE	f1	f2	RE
0	0.875	1.057	1.010	0.865	1.06	1.021	0.93	1.03	1.026
1	0.856	1.065	1.014	0.716	1.115	1.080	0.95	1.02	1.052
10	0.852	1.067	1.015	0.49	1.17	1.182	1.21	0.87	1.100

The overall qualities for D4 with standard and optimal levels are summarized in Table 5.6. The *D*-optimal D4 is uniformly best for different variance component ratios. For small values of variance ratio (d=1), D4 with standard factorial levels is worse and its overall efficiency is almost 10% lower than the *D*-optimal D4. For d=10, the standard D4 is the second best. In a summary, D4 with the *D*-optimal factorial levels has the best quality for the three criteria. In addition, the *D*-optimal factorial levels are robust to changes in variance ratio. Therefore, practitioners can choose *D*-optimal levels at (f1=0.86, f2=1.06) for various values of variance component ratio.

Table 5.6 Performance of the standard and optimal D4's in terms of *D*-, *G*- and *V*-efficiency and their overall performance for multiple criteria.

		Standard	*D*-optimal	*V*-optimal	*G*-optimal
d =1	*D*-eff	1	1.015	1.001	1.003
	V-eff	1	1.059	1.080	1.022
	G-eff	1	1.017	0.978	1.051
	Product	1	1.093	1.057	1.078
d =10	*D*-eff	1	1.016	0.923	0.921
	V-eff	1	1.085	1.182	0.853
	G-eff	1	0.952	0.880	1.100
	Product	1	1.018	0.995	0.958

From the study of the three variations of CCDs, that is, the restricted SPD CCD (D1), D3 and D4, the *D*-optimal designs are desirable choices when the three optimality criteria are concerned. Hence, we may conjecture that the *D*-optimal SPD usually has the optimal or highly efficient performance. Goos and Vandebroek (2004) study *D*-optimal SPDs but did not consider their performances for other optimality criteria, such as *V*- and *G*-efficiency. Because of the historical focus on estimation and the intensive computational demands, the *V*- and *G*-optimal SPDs have been explored less frequently in the literature. So these results may allow for the expansion of Goos and Vandebroek's work of the *D*-optimal SPDs to more general situations, and is good news for practitioners to exploit available optimal designs in the literature.

5.5 Comparison of Variations of CCDs for Individual and Multiple Criteria When Considering Different Cost Scenarios

In Chapter 4, the relative performances of the variations of the CCDs were evaluated and we obtained the best structure for the CCDs by selecting the number of whole plots, the sizes of whole plots and the subplot factor levels settings within whole plots, under different scenarios of variance component ratio and cost ratio. In the above sections, the optimal factorial levels within each type of the split-plot structure of CCD are studied for individual optimality criterion. Then for the various ways of setting the factorial levels, their overall performances for the three criteria are evaluated. In practice, we would like to consider the structure of the CCD and the optimal factorial levels simultaneously in order to obtain highly efficient designs. In this section, the overall performances are first studied for the three types of CCD structures. Next, we discuss their performance for each individual optimality criterion.

Since D1, D3 and D4 with the standard and *D*-optimal factorial levels are the best designs obtained from Section 5.2-5.4 with best overall performances (refer to Table 5.2, 5.4 and 5.6 or details), they are compared in Table 5.7 to explore the effect of structures of the CCD and optimizing the factorial levels. All of the relative efficiencies are calculated compared to RSPD CCD (D1) with the standard factorial settings. For these comparisons, because of the different numbers of whole plots and observations, the particular cost scenarios do influence the relative performance of the designs. Three cost scenarios are considered:

- When no cost is taken into consideration, only the quality of estimation and prediction are of interest;
- r=0 represents that the whole plots are extremely expensive;
- r=1 represents the situation that adding one subplot has equal cost to adding a whole plot in the split-plot experiment.

The r values between 0 and 1 may be close to the scenarios commonly encountered in practice, so we use the two extreme values of this range to show the influence of the cost scenario on choice of design for multiple criteria. The designs D1, D3 and D4 with the

standard and *D*-optimal levels are compared under the different scenarios of cost ratios and variance component ratios (*d*=1 and 10).

Table 5.7 Performance of the designs D1, D3 and D4 with the standard and *D*-optimal factorial levels are compared in terms of individual and multiple criteria.

			D1 Std.	D1 *D*-opt	D3 Std.	D3 *D*-opt	D4 Std.	D4 *D*-opt
No cost	*d* =1	*D*-eff	1	1.004	0.968	0.970	1.115	1.131
		V-eff	1	1.021	1.140	1.145	1.129	1.196
		G-eff	1	0.963	1.011	0.981	1.315	1.338
		Product	1	0.986	1.116	1.090	1.656	**1.810**
	d =10	*D*-eff	1	1.007	0.942	0.945	1.055	1.071
		V-eff	1	1.053	1.316	1.337	1.023	1.110
		G-eff	1	0.966	1.010	0.981	1.056	1.005
		Product	1	1.024	**1.251**	1.239	1.139	1.195
r=0	*d* =1	*D*-eff	1	1.003	0.806	0.808	1.114	1.131
		V-eff	1	1.021	0.950	0.954	1.130	1.196
		G-eff	1	0.963	0.843	0.817	1.315	1.338
		Product	1	0.986	0.646	0.630	1.655	**1.809**
	d =10	*D*-eff	1	1.007	0.785	0.788	1.055	1.071
		V-eff	1	1.053	1.097	1.114	1.023	1.111
		G-eff	1	0.966	0.842	0.818	1.056	1.005
		Product	1	1.025	0.724	0.717	1.139	**1.196**
r=1	*d* =1	*D*-eff	1	1.004	0.924	0.926	0.867	0.880
		V-eff	1	1.021	1.088	1.092	0.878	0.930
		G-eff	1	0.963	0.965	0.936	1.022	1.040
		Product	**1**	0.986	0.971	0.947	0.779	0.851
	d =10	*D*-eff	1	1.006	0.899	0.902	0.820	0.833
		V-eff	1	1.053	1.256	1.276	0.796	0.864
		G-eff	1	0.966	0.964	0.937	0.821	0.782
		Product	1	**1.024**	1.088	1.078	0.536	0.562

If the quality of the split-plot design is the most important and the cost of the experiment is not an issue, the strategy of selecting the best design depends on the variance ratio as well. When the variability of the subplots accounts for a half of the observation variance (*d*=1), the *D*-optimal D4 is the best design, showing that the additional total observations have paid off; if the whole plot variance is relatively large (*d*=10), more whole plots are required to precisely estimate whole plot terms, and D3 is best, even it has much less design points. Note that adding one more whole plot is more helpful than adding the six extra observations under this situation. If the cost of whole plots dominates the cost of

experiment (r=0), the D-optimal D4 is best regardless of the variance ratio values, inferring that in D4 augmenting more subplots while keeping the least number of whole plots is an effective strategy. Since the whole plots are so expensive, adding one whole plot to D1 (the RSPD CCD with 5 whole plots) would penalize the design by 20%, so D3 is not appealing when r=0. If the cost of an additional whole plot and a new observation are equal (r=1), designs with less total number of runs are more desirable, but the choice of the best design also depends on the variance ratio. If d=1, D1 with the standard levels is best; if d=10, D3 with the standard levels is the best. We can interpret this as the following: when the whole plots account for large proportion of the variations in the observations, more whole plots are required to have better estimation for whole plot terms, hence resulting in greater efficiency. Under the scenario that adding one whole plot cost as much as adding one observation, we would like to increase the number of whole plots to obtain better performance.

In a summary, when the multiple criteria are our objective, the structure of the CCD has the greatest influence on performance. In other words, the number of whole plots and subplots within the whole plots should be carefully chosen based on the cost scenario of the split-plot experiment and the value of variance component ratio. Optimizing the factorial levels is not as important as the structure of CCD if balance between multiple criteria is of interest.

Often the practitioner knows his/her main objective of the design, for instance, one desires a design to insure that the worst maximum prediction variance is minimized, or desires most of the design region have good prediction. In situations where a specific interest is stated, individual criterion should be considered. We have considered several types of CCD structures along with the corresponding optimal factorial levels for each structure. We have compared these versions of the CCDs and have examined whether the standard factorial levels result in efficiencies that are close to the optimal for the given individual criterion. The summaries of the comparisons are presented in Table 5.8. The cost of the design plays an important role in selecting the structure of the CCDs. Hence choosing the best CCD structure has the first priority for selecting the best CCD with efficient factorial levels. Moreover, optimizing factorial levels can be very beneficial

when individual criterion is concerned. For *D*-criterion, the discussions in Section 5.2-5.4 show that designs with standard levels have *D*-efficiencies that are close to *D*-optimal for a wide range of variance component ratio values. When the value of variance ratio cannot be obtained precisely, the standard D1 or D4 structures are good choices. Regarding the *V*-criterion, when there is relatively large whole plot error variance or comparative subplot cost to whole plot, the *V*-optimal D3 structure is best. In Section 5.2 it was pointed out that D3 structure with standard factorial levels is very close to the *V*-optimal D3, implying that for this situation, the standard D3 would be a good choice, since it is robust to the variance component ratio. *G*-efficiency considers only the location with the maximum prediction variance rather than average evaluation of all terms or all locations as *D*- or *V*-efficiency does. As a result, when choosing a design on the basis of its *G*-criterion, it is important to realize that optimizing the factorial levels helps to control the worst prediction, but doing so would compromise the prediction variances in the rest of the design region.

Table 5.8 Best design and its factorial levels for individual criterion. The *D*-, *G*- or *V*-optimal design represents the design of the given structure with the *D*-, *G*- or *V*-optimal factorial levels.

	D-criterion	*V*-criterion	*G*-criterion
No cost	*D*-optimal (standard) D4	*V*-optimal D4 for d=1 *V*-optimal (standard) D3 for d=10	*G*-optimal D4
r=0	*D*-optimal (standard) D4	*V*-optimal D4	*G*-optimal D4
r=1	*D*-optimal (standard) D1	*V*-optimal (standard) D3	*G*-optimal D1

5.6 Conclusions

Thus far, most work in the literature studying the split-plot designs is based on one single criterion, such as, one of the *D*-, *G*- and *V*-optimality criteria. One disadvantage of a single criterion is that the best design found based on one aspect of design properties often is not highly efficient for another aspect. This forces the practitioner to choose the most important property of the experiment of interest, which is often a difficult task. In practice, the practitioner would like to select a design with balance between different criteria, that is, the one highly efficient for all aspects.

In this chapter the comprehensive performance of each variation of central composite designs with different optimized factorial levels are examined. It is found that the CCDs with standard or *D*-optimal factorial levels perform best for the combined three criteria. The *D*-optimal CCDs adjust the standard levels moderately and thus keep good balance between *D*-, *G*- and *V*-efficiency. The *V*-/*G*-optimal CCDs makes a huge adjustment for optimizing the average/maximum prediction variance, but have lowered *G*-/*V*-efficiency on the other hand. The CCD with standard factorial levels is uniformly highly efficient for each CCD structure, and its good performance is robust to the change of variance component ratio. When the precise estimate or previous knowledge of variance ratio cannot be obtained, the CCDs with standard factorial levels are best and safest choice.

When taking the cost of experiment into consideration, we found that the most important aspect in choice of design is to select the correct structures of CCD according to the cost scenario, such as the structure of whole plots and the settings of subplot levels within whole plots. This is the principle regardless of whether single criterion or multiple criteria are interested. If a single criterion is the objective of design, based on the correct CCD structure, optimizing the factorial levels can make the design better, especially for the *G*-optimality criterion. If the overall performance of *D*-, *G*- and *V*-efficiency is of interest, first the CCDs with correct whole plot structure should be selected. For the best CCD structure in the specific cost scenario, the standard factorial levels are shown with the best overall performance and its superiority is robust to the variance ratio. Optimizing the factorial levels does not help much.

Therefore, the purpose of the split-plot experiment is very important to the choice of design. A conservative but useful strategy in practice is to select the design based on the balance for different criteria. Based on the overall performance of multiple criteria, the selected SPD is shown often highly efficient for many aspects of design properties. However, if the practitioner knows which particular aspect is the most important, chances are the design can be improved more in terms of the given criterion. For instance, for *G*-efficiency, by optimizing the factorial levels the design selected is doing much better.

Chapter 6 Summary and Future Work

In this dissertation, split-plot designs (SPDs) for first and second order models are explored. First, the graphical tools of three-dimensional variance dispersion graphs (3-D VDGs) and fraction of design space (FDS) plots are adapted for split-plot designs. They are shown to be informative tools for examining the prediction variance distribution for split-plot designs. With the aid of the 3-D VDGs and FDS plots, strategies are developed to improve the standard central composite designs (CCDs) in a spherical region for *D*-, *V*- and *G*-optimality by optimizing factorial levels in the whole plot and subplot spaces, respectively.

Second, the cost of split-plot experiment is taken into consideration for evaluating the performance of split-plot designs. A flexible approach for incorporating cost structure for different scenarios that might be encountered in industrial experimentation is discussed. Then the cost penalized estimation and prediction optimality criteria are presented. The impact of different cost and variance ratio scenarios on the choice of preferred design is illustrated with the two examples, a factorial experiment with 8 runs for first-order model and several variations of central composite designs for second order model. Strategies to obtain highly efficient designs are discussed and some recommendations that might be applicable in practice are made. It is also found that a number of different designs are selected based on different optimality criteria. Therefore, it is not uncommon in practice that one selects a best design based on the overall performance of multiple criteria. The variations of CCD structures and optimizing the factorial levels are combined together to study the performance for each criterion and the comprehensive performance for *D*-, *G*- and *V*-efficiencies in Chapter 5.

In many industrial experiments, there are restrictions of randomization, or exist the factors with levels hard or costly to change and the other factors with levels relatively easy to change. In these cases the split-plot experiments are realistic choices from efficiency and economy perspectives. It sometimes happens in industrial experiments that

completely randomized designs are actually conducted as split-plot experiments. The proper analysis and choice of design are important. In addition, the three important design optimality criteria, *D*-, *G*- and *V*-efficiency, are reviewed. Among them, the *D*-optimality is the alphabetical criterion primarily considered in the literature for optimal split-plot designs. Typically, the evaluation of the designs' prediction capability is more complicated than a single number criterion can accommodate. Graphical tools for design assessment are presented to complement the optimality criteria. The work on optimal and high efficient SPDs for first-order model and second-order model are also discussed briefly.

In Chapters 2 and 3, new adaptations of three-dimensional variance dispersion graphs (3-D VDGs) and fractions of design space (FDS) plots are proposed for split-plot designs. The 3-D VDGs display the distribution of scaled prediction variance (SPV) for the entire design space by separating the region into whole plot and subplot spaces, respectively. The minimum and maximum SPV values and the locations where they occur are also displayed in the 3-D VDGs. For *V*-efficiency, the 3-D VDGs often underestimate the influence of the edge area, which represents a substantial proportion of the total design space and also often has large prediction variance. The global FDS plots provide the overall summary of prediction variances throughout the entire design space. It is easy to read the maximum SPV (associated with the *G*-efficiency) and minimum SPV value at FDS=1 and 0 from the curve, respectively, as well as estimate an approximate average SPV (*V*-efficiency) from the slope and values of the curve. The FDS plots treat each location in the design space equally, which is helpful for evaluating the *V*-efficiency. Hence the FDS plots provide complementary information to the 3-D VDGs. Moreover, the sliced FDS curves are obtained by examining the SPV distribution in the subplot space while holding the whole plot shrinkage factors at constant values. From the slope of each slice, the SPV distribution in the subplot space at a specified whole plot level can be described. From the relative positions of the slices at different whole plot levels, we can examine the effect of each type of variables on SPV values and learn the SPV distribution in the whole plot space. The slices provide useful supplementary information

to the single global FDS curve. By combining the 3-D VDGs and FDS plots, the prediction performance of the design can be visualized.

The new graphical tools are used to examine the scaled prediction variance for two types of central composite designs, the restricted SPD CCDs and the modified CCD proposed by Vining, Kowalski and Montgomery (2004). Beyond single number efficiencies, the different distributions of SPV for the two designs are illustrated and the comparisons of their strength and weakness in prediction quality are aided by the graphical tools. Using 3-D VDGs, we demonstrated that it is possible to improve existing designs in terms of their prediction variance properties. From the 3-D VDGs, we learn which areas of the design space are most important for each of the criterion concerned. Based on the *G*- and *V*-criteria, different optimal designs are obtained, which implies that a single strategy does not work universally. We have also demonstrated that the obtained *G*- and *V*-optimal CCDs are quite robust to small changes in the optimal factorial levels as well as to the misspecifications of the variance component ratio. In addition to the SPV, other functions of the prediction variance, for instance the cost adjusted prediction variance, can easily be plotted using the 3D VDGs and FDS plots. This flexibility makes the use of graphical tools adaptable to a broad range of optimization problems

In industrial settings, the experiments often involve hard-to-change (HTC) factors, and often the relative costs of changing level for this type of factors or of the experimental units can vary considerably based on the nature of the application. A practitioner would likely to take the actual total cost of the split-plot experiment into consideration for evaluating the design. Different cost scenarios are discussed in Chapter 4, and the cost adjusted *D*-, *G*- and *V*-efficiencies are proposed to allow practitioners to assess the split-plot experiments combining the performance and the cost based on available resources. The cost is proportional to a linear function of the number of whole plots and the total number of design points, with different cost scenarios being made possible by altering the cost ratio. We also include two special cases where cost is not an issue and whole plot costs nothing additional.

A three factor factorial experiment with 8 runs for the assumed first-order model, as well as five variations of the three factors CCDs for a second order model are studied in Chapter 4 to explore the effect of cost scenarios on the selection of split-plot designs. When the cost of individual whole plot is much more than pooled cost of subplot and measurement, designs with a smaller number of whole plots are preferred. These designs should exploit the resource of whole plots as much as possible by augmenting more observations within each whole plot. If subplot and measurement are expensive, the total number of design points should be kept moderate in exchange for increasing the number of whole plots. The arrangement of subplot levels within whole plots can be very influential to the overall performance of the designs. The restricted SPD CCD is shown to be a good choice under this scenario. Moreover, some desirable strategies for assigning subplot levels within whole plots are provided, which argue against the intuition that a balanced design is always preferable.

To utilize the cost adjusted criteria in practice, an estimate of the cost ratio can be obtained based on understanding of the conditions for running a given split-plot experiment. However, if a pilot study or previous data are not available and the experimenter has limited information about the whole plot and subplot error variances, the variance component ratio is probably not as easy to estimate as the cost ratio. The study in the two examples shows that the choice of a highly efficient design is often quite robust to changes in the variance ratio values. If the guessed or estimated d value is slightly different from the actual value, the selected split-plot design based on the guessed value remains highly efficient. Therefore, the cost adjusted performances flexibly consider a balance between quality and cost, and will hopefully provide practitioners with more realistic mechanisms for selecting a SPD that is most appropriate to their situations.

Thus far, most work in the literature has studied split-plot design based on one single criterion, such as, one of the D-, G- and V-optimality criteria. One disadvantage of a single criterion is that the best design found based on one criterion could have poor performance for another criterion. For instance, in Chapter 2 and 3, we found that the V-optimal RSPD CCD improves average prediction variance by sacrificing its G-efficiency.

Therefore, determining the purpose of a particular experiment is very important before deciding one particular design. A simple alternative that offers a reasonable solution in practice is that the practitioner can select a design that is highly efficient for all aspects and with the best balance between multiple criteria. In Chapter 5, the geometric means of *D*-, *G*- and *V*-efficiency are employed to examine the overall performance of the different variations of CCDs with different factorial levels. The CCD with standard factorial levels is shown optimal or having high efficiency in terms of the comprehensive performance. The superior performances of standard factorial levels over the *V*- and *G*-optimal factorial levels for combined multiple criteria is robust to the changes of variance component ratio, which assure that the CCD with standard factorial levels is a safe and good choice, when the practitioner is not able to obtain the precise estimate of variance component ratio or does not have much knowledge about its range.

If the cost of experiment is also an issue the experimenter wants to consider, the most important aspect in choice of design is to select the appropriate structures of CCDs based on the cost scenario of the experiment. The structure of CCDs includes the number and sizes of the whole plots, as well as the settings of subplot levels within whole plots. Regardless of whether a single criterion or multiple criteria are of interest, designing the structure of CCD according to the cost and variance ratio scenario should be the first step. Next, for a single criterion, optimizing the factorial levels does help improve the efficiency of the design with best structure, especially for the *G*-optimality. If the overall performance for a combination of *D*-, *G*- and *V*-efficiency is of interest, the standard factorial levels are shown to have the best balance between the three efficiencies and optimizing the factorial levels does not yield substantial improvement. It is shown that the CCDs selected based on the multiple criteria are often highly efficient for each of the individual criterion. However, if a particular aspect is known most important, a design selected based on this criterion can be better because that the factorial levels can be adjusted to be more efficient.

Future Work

As discussed in Chapter 1, assessing split-plot designs is much more complicated than assessing completely randomized designs. Many factors, such as the factor combinations in the design matrix $\mathbf{X}$ and the structure of whole plots, and the way assigning the subplot levels within whole plots, all affect the performance of SPDs. Moreover, the value of variance component ratio plays an important role in the design properties and choice of optimal SPD. Incorporating a realistic summary of the relative costs for the whole plots and subplots can alter the choice of the most desirable design.

In this dissertation, some highly efficient designs have been identified, however an exhaustive search of optimal split-plot designs for the vast array of scenarios has not been performed. In particular, split-plot design optimization for the second order model case is considerably more complicated than for the first order model. Occasionally there may be constraints on the shape of the design region. Therefore, designing split-plot experiments can be an extremely large-scale combinatorial problem. A computer-generated design is often the solution for this type of problems. Heuristic optimization methods have been used as the alternatives to classical methods in computer-generated designs because they do not need a closed form mathematical formulation, they can deal with non-continuous space, and they approach a best solution quickly by avoiding being trapped at locally optimal solutions.

Among the various heuristic optimization methods, genetic algorithms (GA) are one of the emerging methods used for designing of experiments. See related work in Park, Robinson, Montgomery and Borror (2005) and Heredia-Langner et al. (2004). The GA is based on the principle of natural evolution. It simulates the natural selection mechanism in the biologic system to guide search of the solutions. A population of candidates should be provided first, and then the individuals in this population strive for survival. The GAs seek to breed the solutions via processes analog to the natural biological procedures, such as cloning and crossover (mating). In the GAs, a fitness measure or evaluation function on which to judge the goodness of the designs is flexible and user-specified. It is often an

optimality criterion, such as *D*-, *G*- or *V*-efficiency, or a desirability function that combines one or more of these measures. For each individual in the population, the designs are evaluated and selected for next generation based on the fitness measure. This procedure is repeated iteratively until no improvement in the solution for several consecutive generations. The GA is a computationally intensive optimization approach and provides an efficient algorithm for the exchange of information between the designs. Combined with other heuristic optimization methods, they can provide more exploration ability and more efficiency in finding a best solution. Drain, Carlyle, Montgomery, Borror and Anderson-Cook (2004) use genetic algorithm simulated annealing (GASA) method for evaluating prediction error variance and slope estimation variance for designs with noise factors. It is shown that the GASA adopts the advantages of GA and simulated annealing and thus provides a significant improvement in computational efficiency and more exploitation ability.

Using GA to find optimal split-plot designs for the second order model case under various cost and variance ratio scenarios, will allow better understanding of the tradeoffs between quality and cost. Goos and Vandebroek (2001, 2004) use exchange algorithm to search *D*-optimal SPDs for cuboidal region. However, optimal SPDs for prediction have not been studied in literatures. In the future, GAs will be used to find the optimal SPDs for *G*- and *V*-efficiencies first. Differences would be expected between the *G*-/*V*- and the *D*-optimal SPDs. We also want to find the optimal split-plot design for spherical design regions. Second, the cost scenarios will be incorporated into the optimal SPDs searching. In this case, there may be several natural mechanisms for specifying the cost based constraint on the total size of the design. We may wish to restrict the total number of design points, the total number of whole plots, or the total cost of the split-plot experiments. Each of these constraints will require different methods for generating offspring from the parents within the GA. Finally, the SPD with best balance for *D*-, *G*- and *V*-efficiency can be explored using GA.

Some other desirable properties could also be incorporated into the selection of an optimal design. For instance, we may want the design to be able to estimate the whole

plot error while keeping the number of whole plots as small as possible. Another consideration may be to limit the whole plot sizes because of some physical constraints on maximum number of subplot units that can be accommodated within the whole plot units. After the GA generates highly efficient designs, the above properties can be considered to help selecting the best one from the candidate designs obtained.

Bibliography

1. Anbari, F. T. and Lucas, J. M. (1994). "Super-Efficient Designs: How to Run Your Experiment for Higher Efficiency and Lower Cost". *ASQC Technical Conference Transactions*, pp. 852-863.
2. Bingham, D. and Sitter, R. S. (1999). "Minimum-Aberration Two-Level Fractional Factorial Split-Plot designs". *Technometrics*, 41, pp. 62-70.
3. Bisgaard, S. (2000). "The design and analysis of $2^{k-p} \times 2^{q-r}$ split-plot experiments". *Journal of Quality Technology*, 32, pp. 39-56.
4. Bisgaard, S. and Steinberg, D. M. (1997). "The design and Analysis of $2^{k-p} \times s$ Prototype Experiments". *Technometrics*, 39, pp. 52-62.
5. Borkowski, J. J. (1995). "Spherical prediction-variance properties of central composite and Box-Behnken designs". *Technometrics*, 37, pp. 399-410.
6. Borror, C. M., Montgomery, D. C., and Myers, R. H. (2002). "Evaluation of statistical designs for experiments involving noise variables", *Journal of Quality Technology*, 34 (1), pp. 54-70.
7. Box, G. E. P. and Behnken, D. W. (1960). "Some New Three-Level Designs for the Study of Quantitative Variables". *Technometrics*, 2, pp. 455-475.
8. Box, G. E. P. and Draper, N. R. (1974). "On Minimum-Point Second-Order Designs". *Technometrics*, 16, pp. 613-616.
9. Box, G. E. P. and Draper, N. R. (1987). Empirical Model Building and Response Surfaces. John Wiley & Sons, New York, NY.
10. Box, G. E. P. and Hunter, J. S. (1957). "Multi-factor Experimental Designs for Exploring Response Surfaces". *Annals of Mathematical Statistics*, 28, pp. 195-241.
11. Box, G. E. P. and Jones, S. (1992). "Split-Plot Designs for Robust Product Experimentation". *Journal of Applied Statistics*, 19, pp. 3-26.
12. Box, G. E. P. and Wilson, K. B. (1951). "On the Experimental Attainment of Optimum Conditions". *Journal of Royal Statistical Society*, Series B, 13, pp. 1-45.

13. Davison, J. J. (1995). “Response Surface Designs and Analysis for Bi-Randomization Error Structures”. Ph.D. dissertation, Virginia Polytechnic Institute and State University, Blacksburg, VA.
14. Derringer, G. and Suich, R. (1980). “Simultaneous Optimization of Several Response Variables”, *Journal of Quality Technology*, 12, pp. 214-219.
15. Dompere, K. K. (2004). Cost-Benefit Analysis and the Theory of Fuzzy Decisions: Identification and Measurement Theory. Springer-Verlag: New York.
16. Drain, D., Carlyle, W. M., Montgomery, D. C., Borror, C. M. and Anderson-Cook, C. M. (2004). “A Genetic Algorithm Hybrid for Constructing Optimal Response Surface Designs”, *Quality and Reliability Engineering International* (In press).
17. Ganju, J. and Lucas, J. M. (1999). “Detecting randomization restrictions caused by factors”. *Journal of Statistical Planning and Inference*, 81, pp. 129-140.
18. Giovannitti-Jensen, A., and Myers, R. H. (1989). “Graphical Assessment of the Prediction Capacity of Response Surface Designs”. *Technometrics*, 31, pp. 159-171.
19. Goldfarb, H.B., Borror, C. M., Montgomery, D.C., Anderson-Cook, C.M. (2004). "Three-Dimensional Variance Dispersion Graphs for Mixture-Process Experiments". *Journal of Quality Technology*, 36, pp. 109-124.
20. Goldfarb, H.B., Anderson-Cook, C.M., Borror, C.M., Montgomery, D.C. (2004). "Fraction of Design Space to Assess the Prediction Capability of Mixture and Mixture-Process Designs". *Journal of Quality Technology*, 36, pp. 169-179.
21. Goos, P. and Vandebroek, M. (2001). “Optimal Split-Plot Designs”. *Journal of Quality Technology*, 33, No. 4, pp. 436-450.
22. Goos, P. and Vandebroek, M. (2004). “Outperforming Completely Randomized Designs”. *Journal of Quality Technology*, 36, No. 1, pp. 12-26.
23. Heredia-Langner, A., Montgomery, D. C., Carlyle, W. M. and Borror, C. M. (2004). “Model Robust Optimal Designs: A Genetic Algorithm Approach”. *Journal of Quality Technology*, 36, pp. 263-279.
24. Hinkelmann, K. and Kempthorne, O. (1994). Design and Analysis of Experiments, Volume I: Introduction to Experimental Design. John Wiley & Sons, New York, NY.
25. Huang, P., Chen, D., and Voelkel, J. O. (1998). “Minimum-Aberration Two-Level Split-Plot Designs”. *Technometrics*, 40, pp. 314-326.

26. Joiner, B. L. and Campbell, C. (1976). “Designing Experiments When Run Order is Important”. *Technometrics*, Vol 18, No. 3, pp. 249-259.
27. Ju, H. L. and Lucas, J. M. (2002). “ Lk Factorial Experiments With Hard-To-Change and Easy-To-Change Factors”. *Journal of Quality Technology*, 34, No. 4, pp. 411-421.
28. Kempthorne, O. (1952). The Design and Analysis of Experiments. John Wiley & Sons, New York, NY.
29. Khuri, A. I., Harrison, J. M. and Cornell, J. A. (1999). “Using Quantile Plots of the Prediction Variance for Comparing Designs for a Constrained Mixture Region: An Application Involving a Fertilizer Experiment”. *Applied Statistics*, 48, pp. 521-532.
30. Khuri, A. I., Kim, H. J. and Um, Y. (1996). “Quantile Plots of the Prediction Variance for Response Surface Designs”. *Comput. Statist. Data Anal.*, 22, pp. 395-407
31. Kiefer, J. and Wolfowitz, J. (1959). “Optimum Designs in Regression Problems”. *Canadian Journal of Mathematics*, 12, pp. 363-366.
32. Kowalski, S. M. (2002). “24 Run Split-Plot Experiments For Robust Parameter Design”. *Journal of Quality Technology*, 34, No. 4, pp. 399-410.
33. Kowalski, S. M., Cornell, J. A. and Vining, G. G. (2002). “Split-Plot Designs and Estimation Methods for Mixture Experiments with Process Variables”. *Technometrics*, 44, pp. 72-79.
34. Kowalski, S. and Vining, G. G. (2001). “Split-Plot Experimentation for Process and Quality Improvement”. *Frontiers in Statistical Quality Control* 6, pp.335-350.
35. Letsinger, J. D., Myers, R. H., and Lentner, M. (1996). “Response Surface Methods for Bi-Randomization Structure”. *Journal of Quality Technology*, 28, pp. 381-397.
36. Liang, L., Anderson-Cook, C. M., Robinson, T., and Myers, R. H. (2004). “Three-Dimensional Variance Dispersion Graphs For Split-Plot Designs”. Technical report 04-4, Dept. of Statistics, Virginia Tech, Blacksburg, VA.
37. Liang, L., Anderson-Cook, C. M. and Robinson, T. (2005), “Cost Penalized Estimation and Prediction Evaluation for Split-Plot Designs,” Technical report 05-1, Department of Statistics, Virginia Tech, Blacksburg, VA.

38. Liang, L., Anderson-Cook, C. M., Robinson, T., and Myers, R. H. (2005), "Fraction of Design Space Plots For Split-Plot Designs," Technical report 05-2, Dept. of Statistics, Virginia Tech, Blacksburg, VA.
39. Montgomery, D. C. (2001). Design and Analysis of Experiments. John Wiley & Sons, New York, NY.
40. Myers, R. H. and Montgomery, D. C. (2002). Response Surface Methodology. John Wiley & Sons, Inc.: New York.
41. Myers, R. H., Vining, G., Giovannitti-Jensen, A., and Myers, S. L. (1992). "Variance Dispersion Properties of Second-Order Response Surface Designs". *Journal of Quality Technology*, 24, pp. 1-11.
42. Park, Y., Robinson, T., Montgomery, D. C. and Borror, C. M. (2005)."A Genetic Algorithm Approach to Construct G-optimal Design When Noise Variables Exist". *Journal of Quality Technology* (In press).
43. Piepel, G. and Anderson, C. M. (1992). "Variance Dispersion Graphs for Designs on Polyhedral Regions". *Proceedings of the Section on Physical and Engineering Sciences*, American Statistical Association, pp. 111-117.
44. Piepel, G., Anderson, C. M, and Redgate, P. E. (1993a). "Variance Dispersion Graphs for Designs on Polyhedral Regions - Revisited". *Proceedings of the Section on Physical and Engineering Sciences*, American Statistical Association, Alexandria, Virginia, pp. 102-107.
45. Piepel, G., Anderson, C. M, and Redgate, P. E. (1993b). "Response Surface Designs for Irregularly-Shaped Regions" (Parts 1, 2, and 3). *Proceedings of the Section on Physical and Engineering Sciences*, American Statistical Association, Alexandria, Virginia, pp. 205-227.
46. Rozum, M. A. and Myers, R. H. (1991). "Variance Dispersion Graphs for Cuboidal Regions". Paper Presented at *Joint Statistical Meetings*, American Statistical Association, Atlanta, GA.
47. Searle, S. R. (1992). Variance Components. John Wiley & Sons, New York, NY.
48. Trinca, L. A. and Gilmour, S. G. (1998). "Variance dispersion graphs for comparing blocked response surface designs". *Journal of Quality Technology*, 30, pp. 314-327.

49. Vining, G. G. (1993). “A computer program for generating variance dispersion graphs (Corr: V25 p333)”. *Journal of Quality Technology*, 25, pp. 45-58.
50. Vining, G. G., Cornell, J. A., and Myers, R. H. (1993). “A Graphical Approach for Evaluating Mixture Designs”. *Journal of the Royal Statistical Association Series C*, 42, pp.127-138.
51. Vining, G. G., Kowalski, S. and Montgomery, D. C. (2004). “Response Surface Designs within a Split-plot Structure”. *Journal of Quality Technology* (to appear).
52. Webb, D., Lucas, J. M. and Borkowski, J. J. (2004). “Factorial Experiments when Factor Levels Are Not Necessarily Reset”. *Journal of Quality Technology*, 36, pp. 1-11.
53. Zahran, A. R., Anderson-Cook, C. M., and Myers, R. H. (2003). “Fraction of Design Space to Assess Prediction Capability of Response Surface Designs”. *Journal of Quality Technology*, 35, pp. 377-386.

Vita

Li Liang, daughter of Wenyuan Liang and Yulan Shi, was born on September 11, 1975 in Shanxi, China. In 1992, she graduated from Kangjie High School, Yuncheng, China. In 1997, she graduated from Tsinghua University and received a Bachelor of Science degree in Electrical Engineering. She received a Master of Engineering degree from the same department in 2000. She started her graduate study in Statistics in August 2000 at Virginia Polytechnic Institute and State University. She worked as a graduate assistant in the department and a consultant at the Statistical Consulting Center. She completed the requirements for a M.S degree in Statistics in December 2000, and was awarded the Boyd Harshbarger Award for the academic achievement among the first year graduate students. She received her Ph.D. in Statistics in May 2005 at Virginia Polytechnic Institute and State University.