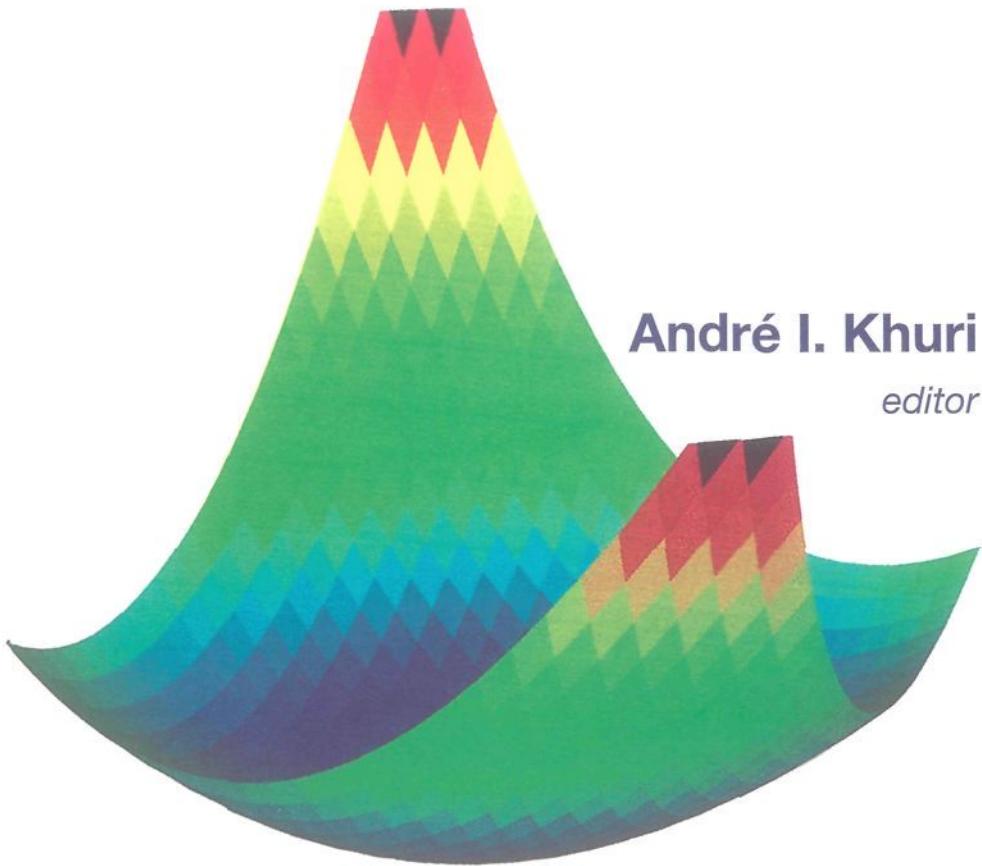
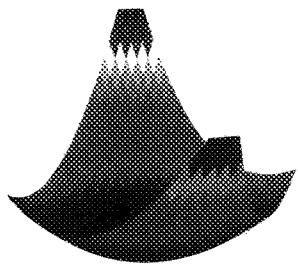


# RESPONSE SURFACE METHODOLOGY AND RELATED TOPICS

André I. Khuri

*editor*

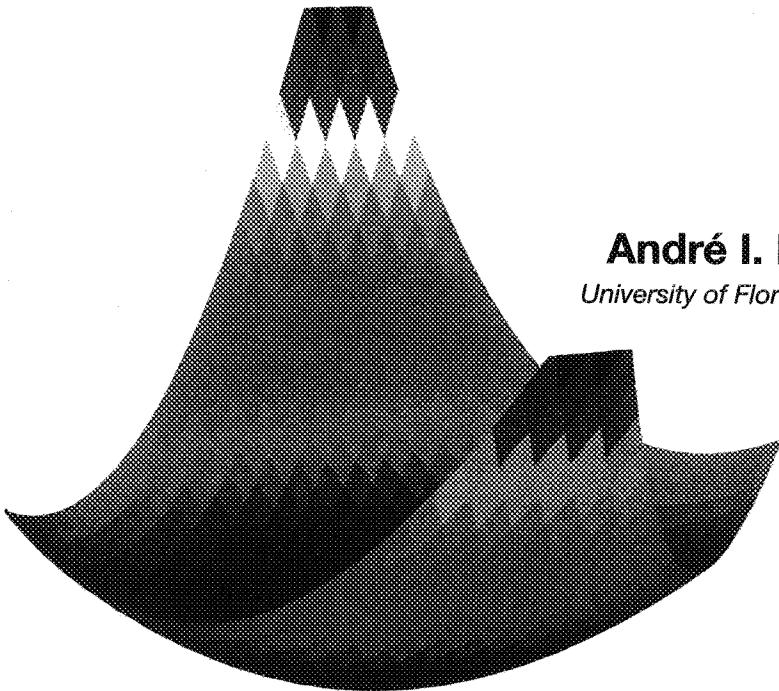




**R**ESENSE **S**URFACE  
**M**ETHODOLOGY **A**ND  
**R**ELATED **T**OPICS

This page is intentionally left blank

# RESPONSE SURFACE METHODOLOGY AND RELATED TOPICS



*editor*

**André I. Khuri**

*University of Florida, USA*

 **World Scientific**

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

*Published by*

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

**Library of Congress Cataloging-in-Publication Data**

Response surface methodology and related topics / [edited by] A.I. Khuri.

p. cm.

Includes bibliographical references.

ISBN 981-256-458-6 (alk. paper)

1. Response surfaces (Statistics) I. Khuri, André I., 1940-

QA279.R47 2005

519.5--dc22

2005049744

**British Library Cataloguing-in-Publication Data**

A catalogue record for this book is available from the British Library.

Copyright © 2006 by World Scientific Publishing Co. Pte. Ltd.

*All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.*

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. In this case permission to photocopy is not required from the publisher.

## Preface

*Response surface methodology* (RSM) began with the work of Box and Wilson (1951) in the *Journal of the Royal Statistical Society, Series B*. That work was motivated by the need to run experiments efficiently, by a proper choice of design, and to determine operating conditions on a set of controllable variables that give rise to an optimal response. The key ideas in the so-called *classical RSM* were developed using linear polynomial models, mainly first-degree and second-degree models, with continuous response variables assumed, for the most part, to be independently and normally distributed with constant error variances. Since the 1970's, RSM has gone through several stages of development that were aimed at making it more applicable to wider experimental situations under less stringent assumptions. The purpose of this volume is to provide an exposition of a variety of topics covering a wide range of methods and techniques used in RSM, including the more recent advances made in the field.

The present volume contains 17 chapters written by leading experts in RSM. The topics covered include factorial and fractional factorial designs, processes with high variation, split-plot experiments, response surface optima, robust parameter designs, applications in the treatment of cancer, designs for generalized linear models, non-standard designs, a review of mixture experiments, graphical techniques for comparing response surface designs, and designs for estimating the slope of a response surface. These chapters provide an overview of the key ideas that have shaped RSM.

This volume should be useful to researchers as well as practitioners interested in RSM's theory and potential applications. Those who have used RSM in the past, but have not kept up with its recent developments, will find the volume particularly helpful. All the volume's chapters were refereed.

I would like to thank the editors of World Scientific, and in particular, Ms. Kimberly Chua and Ms. Eng Huay Chionh, for making the publication of this volume possible. I am very much indebted to the authors of the

chapters for their valuable contributions and efforts. This project would not have been possible without their cooperation, particularly with regard to their refereeing of the chapters and meeting the Publisher's deadlines.

*André I. Khuri*

June 2005

Gainesville, Florida

## Contents

Preface	v
List of Contributors	xi
Chapter 1 Two-Level Factorial and Fractional Factorial Designs in Blocks of Size Two. Part 2 <i>Yuyun Jessie Yang and Norman R. Draper</i>	1
Chapter 2 Response Surface Experiments on Processes with High Variation <i>Steven G. Gilmour and Luzia A. Trinca</i>	19
Chapter 3 Random Run Order, Randomization and Inadvertent Split-Plots in Response Surface Experiments <i>Jitendra Ganju and James M. Lucas</i>	47
Chapter 4 Statistical Inference for Response Surface Optima <i>Dennis K. J. Lin and John J. Peterson</i>	65
Chapter 5 A Search Method for the Exploration of New Regions in Robust Parameter Design <i>Guillermo Miró-Quesada and Enrique del Castillo</i>	89
Chapter 6 Response Surface Approaches to Robust Parameter Design <i>Timothy J. Robinson and Shaun S. Wulff</i>	123

<b>Chapter 7 Response Surface Methods and Their Application in the Treatment of Cancer with Drug Combinations: Some Reflections</b>	159
<i>Kathryn S. Dawson, T. J. Eller, and Walter H. Carter, Jr.</i>	
<b>Chapter 8 Generalized Linear Models and Response Transformation</b>	173
<i>Anthony C. Atkinson</i>	
<b>Chapter 9 GLM Designs: The Dependence on Unknown Parameters Dilemma</b>	203
<i>André I. Khuri and Siuli Mukhopadhyay</i>	
<b>Chapter 10 Design for a Trinomial Response to Dose</b>	225
<i>Shenghua Kelly Fan and Kathryn Chaloner</i>	
<b>Chapter 11 Evaluating the Performance of Non-Standard Designs: The San Cristobal Design</b>	251
<i>Linda M. Haines</i>	
<b>Chapter 12 50 Years of Mixture Experiment Research: 1955–2004</b>	283
<i>Greg F. Piepel</i>	
<b>Chapter 13 Graphical Methods for Comparing Response Surface Designs for Experiments with Mixture Components</b>	329
<i>Heidi B. Goldfarb and Douglas C. Montgomery</i>	
<b>Chapter 14 Graphical Methods for Assessing the Prediction Capability of Response Surface Designs</b>	349
<i>John J. Borkowski</i>	
<b>Chapter 15 Using Fraction of Design Space Plots for Informative Comparisons between Designs</b>	379
<i>Christine M. Anderson-Cook and Ayca Ozol-Godfrey</i>	

<b>Chapter 16 Concepts of Slope-Rotatability for Second Order Response Surface Designs</b>	<b>409</b>
<i>Sung H. Park</i>	
<b>Chapter 17 Design of Experiments for Estimating Differences between Responses and Slopes of the Response</b>	<b>427</b>
<i>Shahriar Huda</i>	
<b>Subject Index</b>	<b>447</b>

This page is intentionally left blank

## List of Contributors

- Anderson-Cook, Christine:** 985 Los Pueblos Street, Los Alamos, NM 87544, USA, e-mail: c-and-cook@lanl.gov (Chap. 15).
- Atkinson, Anthony C.:** Department of Statistics, London School of Economics, London WC2A 2AE, UK, e-mail: a.c.atkinson@lse.ac.uk (Chap. 8).
- Borkowski, John J.:** Department of Mathematical Sciences, Montana State University, Bozeman, MT 59717, USA, e-mail: jobo@math.montana.edu (Chap. 14).
- Carter, Walter H., Jr.:** Department of Biostatistics, Virginia Commonwealth University, Richmond, VA 23298-0032, USA, e-mail: carter@hsc.vcu.edu (Chap. 7).
- del Castillo, Enrique:** Department of Industrial & Manufacturing Engineering & Department of Statistics, The Pennsylvania State University, 310 Leonhard Building, University Park, PA 16802, USA, e-mail: exd13@psu.edu (Chap. 5).
- Chaloner, Kathryn:** Department of Biostatistics, The University of Iowa, College of Public Health C22 GH, 200 Hawkins Drive, Iowa City, IA 52242, USA, e-mail: kathryn-chaloner@uiowa.edu (Chap. 10).
- Dawson, Kathryn:** Novartis Oncology, 180 Park Avenue, Building 105, Florham Park, NJ 07932-0675, USA, e-mail: kathryn.dawson@novartis.com (Chap. 7).
- Draper, Norman R.:** Statistics Department, University of Wisconsin, 1300 University Avenue, Madison, WI 53706-1532, USA, e-mail: draper@stat.wisc.edu (Chap. 1).
- Eller, T. J.:** SunTrust Bank, Mail Code CS-RVW-7900, 1001 Semmes Avenue, Richmond, VA 23224, USA, e-mail: tj.eller@Suntrust.com (Chap. 7).

- Fan, Shenghua K.**: Department of Statistics, California State University, East Bay, 25800 Carlos Bee Boulevard, Hayward, CA 94542-3087, USA, e-mail: kelly\_shfan@yahoo.com (Chap. 10).
- Ganju, Jitendra**: 4560 Horton Street, M/S X-600A, Chiron Corporation, Emeryville, CA 94608, USA, e-mail: jitendra\_ganju@chiron.com (Chap. 3).
- Gilmour, Steven G.**: School of Mathematical Sciences, Queen Mary, University of London, Mile End Road, London E1 4NS, UK, e-mail: s.g.gilmour@qmul.ac.uk (Chap. 2).
- Goldfarb, Heidi B.**: The Dial Corporation, 15101 North Scottsdale Road, Scottsdale, AZ 85254, USA, e-mail: goldfarb@dialcorp.com (Chap. 13).
- Haines, Linda**: Department of Statistical Sciences, University of Cape Town, Private Bag, Rondebosch 7700, South Africa, e-mail: lhaines@stats.uct.ac.za (Chap. 11).
- Huda, Shahriar**: Department of Statistics & O.R., Faculty of Science, Kuwait University, P. O. Box 5969, Safat-13060, Kuwait, e-mail: shuda@kuc01.kuniv.edu.kw (Chap. 17).
- Khuri, André I.**: Department of Statistics, 103 Griffin-Floyd Hall, P. O. Box 118545, University of Florida, Gainesville, FL 32611-8545, USA, e-mail: ufakhuri@stat.ufl.edu (Chap. 9).
- Lin, Dennis**: Department of Supply Chain & Information Systems, 483 Business Building, The Pennsylvania State University, University Park, PA 16802-3005, USA, e-mail: DKL5@psu.edu (Chap. 4).
- Lucas, James M.**: 5120 New Kent Road, J. M. Lucas and Associates, Wilmington, DE 19808, USA, e-mail: JamesM.Lucas@worldnet.att.net (Chap. 3).
- Miró-Quesada, Guillermo**: Bioprocess Research and Development, Lilly Technical Center-North, Indianapolis, IN 46221 (Chap. 5).
- Montgomery, Douglas C.**: Department of Industrial Engineering, Arizona State University, Tempe, AZ 85287-5906, USA, e-mail: doug.montgomery@asu.edu (Chap. 13).
- Mukhopadhyay, Siuli**: Department of Statistics, 103 Griffin-Floyd Hall, P. O. Box 118545, University of Florida, Gainesville, FL 32611-8545, USA, e-mail: siulimukherjee@gmail.com (Chap. 9).

- Ozol-Godfrey, Ayca:** 59 New Holland Village, Building #6, Nanuet, NY 10954, USA, e-mail: godfrea@wyeth.com (Chap. 15).
- Park, Sung H.:** Department of Statistics, College of Natural Sciences, Seoul National University, San 56-1, Shinrim-dong, Kwanak-ku, Seoul, 151-747, Korea, e-mail: parksh@plaza.snu.ac.kr (Chap. 16).
- Peterson, John J.:** Statistical Sciences Department, Mail Code UW281A, GlaxoSmithKline Pharmaceuticals, R&D, 709 Swedeland Road, King of Prussia, PA 19406-0939, USA, e-mail: john.peterson@gsk.com (Chap. 4).
- Piepel, Greg F.:** Statistical Sciences, K6-08, Pacific Northwest National Laboratory, P. O. Box 999, Richland, WA 99352, USA, e-mail: greg.piepel@pnl.gov (Chap. 12).
- Robinson, Timothy J.:** Statistics Department, University of Wyoming, Dept. 3332, 1000 E. University Avenue, Laramie, WY 82071, USA, e-mail: tjrobin@uwyo.edu (Chap. 6).
- Trinca, Luzia A.:** Departamento de Bioestatistica - IB, UNESP, Botucatu, CP 510, 18618-000, SP, Brazil, e-mail: ltrinca@ibb.unesp.br (Chap. 2).
- Wulff, Shaun S.:** Statistics Department, University of Wyoming, Dept. 3332, 1000 E. University Avenue, Laramie, WY 82071, USA, e-mail: wulff@uwyo.edu (Chap. 6).
- Yang, Yuyun (Jessie):** 6244 Via De Adrianna, San Jose, CA 95120, USA, e-mail: yuyun@sbcglobal.net (Chap. 1).

## CHAPTER 1

### TWO-LEVEL FACTORIAL AND FRACTIONAL FACTORIAL DESIGNS IN BLOCKS OF SIZE TWO. PART 2

Yuyun Jessie Yang<sup>1</sup> and Norman R. Draper<sup>2</sup>

<sup>1</sup>6244 Via De Adrianna, San Jose, CA 95120

<sup>2</sup>Department of Statistics, University of Wisconsin  
Madison, WI 53706

E-mail: <sup>1</sup>yuyun@sbcglobal.net, <sup>2</sup>draper@stat.wisc.edu

Two-level factorial and fractional factorial designs can be blocked in a variety of ways, depending on block size, on which factorial estimates are required to be estimated clear of blocks, and on the permissible number of runs to be performed. A previous (2003) paper by the same authors discussed blocks of size two for  $2^k$  designs when  $k = 2, 3, 4$  and 5. “Best” combination designs were given which provided the most pure (within block) estimates of main effects and two-factor interactions from choices that combined various confounding patterns. Extensions were also given for  $2^{k-p}$  fractional factorial designs with the same number of runs (that is, when the  $k-p$  value being considered is the same as the  $k$  value previously considered for the  $2^k$  designs). An example illustrated the use of follow-up inter-block analysis. The present article discusses larger designs of the same type, when  $k = 6, 7$  and 8. In addition, some “superior” sequential combinations of  $2^k$  designs with various choices of initial arrangements are given. The popular mirror-image pairing design is not the best initial arrangement in a sequence, but is a valuable one because it provides pure estimates of all the main effects.

#### 1. Introduction

In many experimental situations, it is desirable to group sets of experimental runs together in blocks. The block size is governed by many considerations and represents, in most experiments, the number of

runs that can be made without worrying (much) about variation caused by factors not being studied specifically in the experiment. Often a block is some natural interval of time (e.g., a week, a day, a work shift) or of space (an oven, a greenhouse, a work bench, a reactor) or of personnel (a research worker, a research team), and so on. An excellent discussion of some of the practical considerations that dictate the need to block experimental designs is given by Rosenbaum<sup>1</sup> (1999, p. 127), and a specific illustration is given by Yang and Draper<sup>2</sup> (2003, p. 294).

When only a single factor is being examined, we have the case of a simple comparative experiment as described by Box, Hunter and Hunter<sup>3</sup> (1978, pp. 97-102). When more than two factors are involved, any allocation of the runs of a  $2^k$  or  $2^{k-p}$  design into pairs will involve confounding some effects of possible interest with the blocks. Thus more than one replicate of the design, blocked differently, will be necessary. Box, Hunter and Hunter<sup>3</sup> (1978, p. 341) discuss an example of this type in which a  $2^3$  design is blocked into pairs in four separate ways to give a total of  $4(2^3) = 32$  runs in 16 blocks. In each of the four portions of the design, different effects are confounded with blocks, so that an overall balance is achieved, and all main effects and interactions are estimable. Draper and Guttman<sup>4</sup> (1997) showed that for a  $2^k$  design,  $k2^k$  runs are needed to estimate all main effects and all interactions. In this article, however, we shall confine interest to estimating only main effects and two-factor interactions, while tentatively assuming that all interactions between three or more factors are zero. Using the notation  $x_{fi}$  for an interaction between  $x$  factors for  $x \geq 2$ , we can write this assumption as " $\geq 3fi = 0$ ." As might be anticipated, this can be done with more economical designs containing fewer blocks. We assume here that block effects are additive, representing changes in overall level only, and that there are no block-factor interactions. (In our 2003 paper<sup>2</sup>, we also discussed the inter-block analysis of such designs, in which it is assumed that block effects are random variables. Such an analysis would also apply here in similar circumstances.)

Because blocks of size two are very restrictive, any  $2^k$  design must be run several times over in various blocking configurations to estimate all main effects and 2fis. Consider, for example, the so-called "mirror image" or "foldover" pairs of runs, in which the levels of the factors are

completely reversed in the second run of the pair. Only main effects (and not 2fis) can be estimated from such pairings. Other pairs of blocked runs would thus be needed to estimate the 2fis, if we began with a foldover design. Moreover, in certain applications where  $2^k$  designs are already subject to some initial fractionation, it becomes impossible to form foldover pairs at all, as pointed out by Rosenbaum<sup>1</sup> (1999, p. 131). Thus we need to seek alternative blocking methods for blocks of size two.

In blocking entire  $2^k$  or  $2^{k-p}$  designs in blocks of size two in this article, we shall pair up runs using conventional ideas of blocking generators. Note that this is not the most general situation. A much wider problem would be to form all the possible pairs that could be chosen from  $2^k$  runs (there are  $2^k(2^k - 1)/2$  such pairs) and then to consider how to add pairs one at a time sequentially to form a useful design. (For  $k = 3$ , we would select from 28 pairs, for example.) We believe that a design chosen in this more general way would not be an improvement over the designs we shall choose by using blocking generators, simply because it is essential to build certain symmetries in order to estimate the effects. Moreover, such designs might not be resolvable, that is, might not permit division into sets of blocks, each set of which contains an entire  $2^3$  design within it. However, we have not investigated these wider issues.

For the designs we derive, estimates of the main effects and 2fis are made by least squares estimation. The model fitted will include a general mean, terms for all main effects and 2fis, and terms for  $B-1$  blocking variables, where  $B$  is the number of blocks of size two in the design. There are thus  $1 + k + k(k-1)/2 + B - 1 = B + k(k+1)/2$  terms in the model. The total number of runs is  $n = 2B$ . In order for the model to be estimable, all main effects and 2fis must be estimable internally *within* one or more blocks (pairs of runs) somewhere, and usually several times, in the design. (See Yang and Draper<sup>2</sup>, 2003, pp. 295-297 for a detailed discussion of this point.)

## 2. The Six Factor, 64 Runs, $2^6$ Design

To divide the 64 runs of a  $2^6$  design into 32 pairs of runs requires choice of five generators, which we select from the set of six main effects and

15 two-factor interactions. In all there are  $2^6 - 1 = 63$  possible divisions of this type, that is, there are 63 possible “blocking arrangements”, listed in Table 1. (It can be shown that Table 1 provides all the possible blocking arrangements (in pairs) for a  $2^6$  design, although they can also be re-described in other ways, using alternative generators. Choices of other generators will simply reproduce one of the 63 arrangements shown.) These 63 are divided into six “types” depending on the numbers of main effects and 2fis blocking generators used. Arrangement No 1, for example is defined by

$$\begin{aligned} I &= 1 = 2 = 3 = 4 = 5 \\ &= 12 = 13 = 14 = 15 = 23 = 24 = 25 = 34 = 35 = 45. \end{aligned} \quad (1)$$

The pairs of runs in this specific arrangement do not contain any pure comparisons for the five defining main effects and the ten 2fis shown in (1), but pure (within block) estimates can be made for the main effect 6 and the remaining 2fis, namely, 16, 26, 36, 46, and 56. Thus combinations of various of the 63 arrangements in Table 1 will permit estimation of particular effects in the overall design. We specifically seek combinations that will estimate *all* main effects and 2fis. These ideas are discussed more fully for smaller designs in Yang and Draper<sup>2</sup> (2003). Note that the foldover (mirror image) design is No. 63, type 6, in Table 1.

We consider the choice of  $d$  arrangements from the arrangements available. For  $k \leq 5$ , it was feasible to examine *all* combinations of arrangements, but limits on available space prevented this for larger  $d$ . For example, for  $k = 6$ , all combinations were examined for  $d \leq 5$  but, for larger  $d$ , we discuss only the “best” sequential combinations given a selected initial arrangement. Previously used selection criteria still apply, as described below.

### 3. Definitions and Notation

We need some additional notation in addition to that already mentioned. Specifically:



$t$  = blocking type, as in the first column of Table 1; when a vector  $t = (t_1, t_2, \dots, t_d)$  is used, it defines a combination of  $d$  blocking types used to form a design.

$x_{fi}$  already means an  $x$ -factor interaction, and we shall use  $1_{fi}$  to mean a main effect.

$u = 0, 1, \dots, d$  is the number of times an effect is estimated in a combination design.

$f_u$  is the non-negative integer frequency of  $u$ . For example if  $u = 0$  occurs twice in a combination design,  $f_0 = 2$ .

$f = (f_0, f_1, f_2, \dots, f_d)$  is a collection of the frequency values  $f_u$  for a combination design. This defines the *overall confounding pattern of a combination design*. It is sometimes abbreviated as *overall pattern* or simply OP.

$f_x = (f_0, f_1, f_2, \dots, f_d)^{(x_{fi})}$  is a collection of the frequency values  $f_u$  for the group of  $x$ -factor interactions. Here,  $x$  could be 1, 2, ..., or  $k$  and, for the combination designs within each overall pattern,  $f_1 + f_2 + \dots + f_k = f$ . This is called the *x-factor interaction group confounding pattern of a combination design*, or more simply, the *x<sub>fi</sub> group pattern*.

Each design has a set of  $f_i$  values,  $(f_1, f_2, \dots, f_k)$  which is called its *group pattern*, abbreviated in tables as GP. Because we focus only on main effects ( $1_{fi}$ ) and  $2_{fi}$  for this article, we mostly list only  $(f_1, f_2)$ . When this is done, statements of "different group patterns" refer to sets of  $(f_1, f_2)$  only.

The following notation is also used:

$(x_{fi}:f_u)$  = an  $f_u$  value which belongs to one of the  $f_x = (f_0, f_1, f_2, \dots, f_d)^{(x_{fi})}$ . Where restricted as above, this refers to the frequency values for  $1_{fi}$  and  $2_{fi}$  only.

$\text{Sum}_x(u)$ ,  $x = 1$  or  $2$ . If  $x = 1$ ,  $\text{Sum}_1(u)$ , is the sum of the  $u$  values in the main effects, and if  $x = 2$ , it is the sum of the  $u$  values in the two-factor interactions. Thus, for example,

$$\text{Sum}_1(u) = \sum_{u=1}^d u f_u, \text{ where } f_u \in (f_0, f_1, f_2, \dots, f_d)^{(1_{fi})}.$$

$\text{Std}_x(u)$ ,  $x = 1$  or  $2$ . If  $x = 1$ , it is the standard deviation of the  $u$  values in the main effects, and if  $x = 2$ , it is the standard deviation of the  $u$  values in the two-factor interactions.

Although our study examines  $u$  values, we do not save them all individually, in order to conserve computer storage space. The  $f$  and the  $f_i$ 's summarize the  $u$  patterns. The only information lost is the order of the  $u$  values, which is irrelevant because the order changes if the effects are renamed, but the  $f$  and the  $f_i$ 's do not. It is also unnecessary to store all the combination designs. Keeping track of the blocking arrangements (the row numbers in  $C_k$  that constitute each combination design) and the combinations of blocking types, namely the  $t$  values, is sufficient to identify the types of arrangements that constitute the various combination designs.

#### 4. Combination Design Selection Process

We compare the  $u$  value patterns of all possible combination designs. Cases that do not provide estimates of all  $1f_1$  and  $2f_1$  are discarded immediately if alternatives exist that *do* estimate these. The larger the  $u$  values, the better the design, in general. The best situation is to have a design whose  $u$  values are uniformly better than those for other competing designs. This can be examined directly if there are not many patterns of  $u$  values to examine. Otherwise, we instead examine the sums and standard deviations of the  $u$  values, that is,  $\text{Sum}_1(u)$ ,  $\text{Sum}_2(u)$ ,  $\text{Std}_1(u)$ , and  $\text{Std}_2(u)$ , by plotting them, leading to the selection of just a few designs for further detailed examination. Desirable are designs with high Sums and small Stds.

#### 5. Case $k = 6, d = 2$

There are, in total,  $\binom{6^3}{2} = 1953$  combination designs to examine. The overall pattern  $f = (15, 32, 16)$  applies to all of them. This means that, in each possible combination design, 15 effects have no internal (that is, within block or intra-block) estimates, 32 have one internal estimate, and 16 effects have two internal estimates. The numbers total to 63, which

accounts for all possible main effects and interactions of all orders. In what follows, however, we consider only 1fi and 2fi estimation in evaluating these designs in more detail. Within this (15, 32, 16) pattern lie 37 different group patterns, in none of which can all 1fis and 2fis be estimated. Table 2 shows all 14 group patterns that have (1fi:  $f_0 = 0$ ), that is, the 14 group patterns that provide estimates of all main effects. The best of these is the group pattern GP6 with (2fi:  $f_0 = 3$ ) which estimates all but three 2fis. It has the following u values for 1fis and 2fis:

GP6: (1, 1, 1, 1, 2, 2) and (0, 0, 0, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2).

This means it provides one internal estimate of four main effects, two internal estimates of two main effects, does not provide estimates of three 2fi, provides one internal estimate of eight 2fi and two internal estimates of four 2fi. This is the best result possible for combining two ( $d = 2$ ) of the designs in Table 1. The 45 arrangements that form the optimal group pattern GP6 are given in Yang<sup>5</sup> (2002, Appendix, List 1).

Table 2.  $k = 6, d = 2$ ; the 14 group patterns that have (1fi:  $f_0 = 0$ ).

	1fi ( $f_0, f_1, f_2$ )			2fi ( $f_0, f_1, f_2$ )				Number of combination designs
GP1	0	1	5	10	5	0	(5, 6)	6
GP2	0	2	4	6	8	1	(5, 5)	15
GP3	0	2	4	7	8	0	(4, 6)	15
GP4	0	3	3	4	9	2	(4, 5)	60
GP5	0	3	3	6	9	0	(3, 6)	20
GP6	0	4	2	3	8	4	(4, 4)	45
GP7	0	4	2	4	8	3	(3, 5)	60
GP8	0	4	2	7	8	0	(2, 6)	15
GP9	0	5	1	4	5	6	(3, 4)	60
GP10	0	5	1	6	5	4	(2, 5)	30
GP11	0	5	1	10	5	0	(1, 6)	6
GP12	0	6	0	6	0	9	(3, 3)	10
GP13	0	6	0	7	0	8	(2, 4)	15
GP14	0	6	0	10	0	5	(1, 5)	6

The basic rule for this design type is to select two type 4 arrangements, i.e.,  $(t_1, t_2) = (4, 4)$ , that do *not* have common 1fi block generators. For example,  $(B_1, B_2, B_3, B_4, B_5) = (1, 2, 36, 46, 56)$  and  $(3, 4, 16, 26, 56)$  together will give such a design.

Table 3 shows an example of an optimal design of type GP6, consisting of 128 runs in 64 blocks of size 2. Blocks 1-32 are defined by the blocking generators  $(B_1, B_2, B_3, B_4, B_5) = (5, 6, 14, 24, 34)$ , and blocks 33-64 are defined by  $(B_1, B_2, B_3, B_4, B_5) = (3, 4, 16, 26, 56)$ .

Table 3. An example of a best combination design for  $k = 6, d = 2$ .

Block	Factor No.						Block	Factor No.					
	1	2	3	4	5	6		1	2	3	4	5	6
1	+	+	+	-	-	-	33	+	+	-	-	+	-
	-	-	-	+	-	-		-	-	-	-	-	+
2	+	+	+	-	+	-	34	+	+	+	-	+	-
	-	-	-	+	+	-		-	-	+	-	-	+
3	+	+	+	-	-	+	35	+	+	-	+	+	-
	-	-	-	+	-	+		-	-	-	+	-	+
4	+	+	+	-	+	+	36	+	+	+	+	+	-
	-	-	-	+	+	+		-	-	+	+	-	+
5	-	+	+	-	-	-	37	-	+	-	-	+	-
	+	-	-	+	-	-		+	-	-	-	-	+
6	-	+	+	-	+	-	38	-	+	+	-	+	-
	+	-	-	+	+	-		+	-	+	-	-	+
7	-	+	+	-	-	+	39	-	+	-	+	+	-
	+	-	-	+	-	+		+	-	-	+	-	+
8	-	+	+	-	+	+	40	-	+	+	+	+	-
	+	-	-	+	+	+		+	-	+	+	-	+
9	+	-	+	-	-	-	41	+	-	-	-	+	-
	-	+	-	+	-	-		-	+	-	-	-	+
10	+	-	+	-	+	-	42	+	-	+	-	+	-
	-	+	-	+	+	-		-	+	+	-	-	+
11	+	-	+	-	-	+	43	+	-	-	+	+	-
	-	+	-	+	-	+		-	+	-	+	-	+
12	+	-	+	-	+	+	44	+	-	+	+	+	-
	-	+	-	+	+	+		-	+	+	+	-	+
13	-	-	+	-	-	-	45	-	-	-	-	+	-
	+	+	-	+	-	-		+	+	-	-	-	+

Table 3. (Continued)

Block	Factor No.						Block	Factor No.					
	1	2	3	4	5	6		1	2	3	4	5	6
14	-	-	+	-	+	-	46	-	-	+	-	+	-
	+	+	-	+	+	-		+	+	+	-	-	+
15	-	-	+	-	-	+	47	-	-	-	+	+	-
	+	+	-	+	-	+		+	+	-	+	-	+
16	-	-	+	-	+	+	48	-	-	+	+	+	-
	+	+	-	+	+	+		+	+	+	+	-	+
17	+	+	-	-	-	-	49	+	+	-	-	-	-
	-	-	+	+	-	-		-	-	-	+	+	
18	+	+	-	-	+	-	50	+	+	+	-	-	-
	-	-	+	+	+	-		-	-	+	-	+	+
19	+	+	-	-	-	+	51	+	+	-	+	-	-
	-	-	+	+	-	+		-	-	-	+	+	+
20	+	+	-	-	+	+	52	+	+	+	+	-	-
	-	-	+	+	+	+		-	-	+	+	+	+
21	-	+	-	-	-	-	53	-	+	-	-	-	-
	+	-	+	+	-	-		+	-	-	+	+	+
22	+	+	-	-	+	-	54	-	+	+	-	-	-
	+	-	+	+	+	-		+	-	+	-	+	+
23	-	+	-	-	-	+	55	-	+	-	+	-	-
	+	-	+	+	-	+		+	-	-	+	+	+
24	-	+	-	-	+	+	56	-	+	+	+	-	-
	+	-	+	+	+	+		+	-	+	+	+	+
25	+	-	-	-	-	-	57	+	-	-	-	-	-
	-	+	+	+	-	-		-	+	-	-	+	+
26	+	-	-	-	+	-	58	+	-	+	-	-	-
	-	-	+	+	+	-		-	+	+	-	+	+
27	+	+	-	-	-	+	59	+	-	-	+	-	-
	-	+	+	+	-	+		-	+	-	+	+	+
28	+	-	-	-	+	+	60	+	-	+	+	-	-
	-	+	+	+	+	+		-	+	+	+	+	+
29	-	-	-	-	-	-	61	-	-	-	-	-	-
	+	+	+	+	-	-		+	+	-	-	+	+
30	-	-	-	-	+	-	62	-	-	+	-	-	-
	+	+	+	+	+	-		+	+	+	-	+	+
31	-	-	-	-	-	+	63	-	-	-	+	-	-
	+	+	+	+	-	+		+	+	-	+	+	+
32	-	-	-	-	+	+	64	-	-	+	+	-	-
	+	+	+	+	+	+		+	+	+	+	+	+

## 6. Case $k = 6, d=3$

The  $\binom{6^3}{3} = 39711$  combination designs lead to two overall confounding patterns, namely, OP1:  $f = (7, 24, 24, 8)$  and OP2:  $f = (15, 0, 48, 0)$ . None of the designs in the latter estimate all the 1fi and 2fi, and so they are dropped from further consideration. OP1 contains 39060 combination designs and 228 group patterns, but only three of these 228 provide a full set of 1fi and 2fi estimates. These three group patterns are displayed in Table 4.

Table 4.  $k = 6, d = 3$ ; the group patterns that can estimate 1fi's and 2fi's (A indicates the number of combination designs available).

		u values for 1fi's and 2fi's	t	A
OP1	GP47	(1,1,2,2,2,3) & (1,1,1,1,1,1,2,2,2,2,2,3,3)	(3,4,4)	360
OP1	GP76	(1,1,1,2,2,3) & (1,1,1,1,1,1,2,2,2,2,2,2,3,3)	(3,3,4)	360
OP1	GP86	(1,1,1,2,2,2) & (1,1,1,1,1,1,2,2,2,2,2,2,3,3,3)	(3,3,3)	120

We see that GP47 (just) provides the best choice, provided we rate 1fi estimation more important than 2fi estimation. Such a design consists of two type 4 arrangements and a type 3 arrangement, i.e.,  $t = (3, 4, 4)$  in Table 4. Thus an example of this best design type would be to choose a suitable type 3 arrangement to add to Table 3. The basic rule for such a choice is as follows. Choose a type 3 arrangement that has a 1fi blocking variable taken from each of the type 4 blocking variable sets and then add a third 1fi blocking variable that does not appear in either type 4 arrangement. Then, complete the blocking generators with two 2fi chosen from Table 1 to obtain a type 3 arrangement. To the arrangements **(5, 6, 14, 24, 34)** and **(3, 4, 16, 26, 56)** of Table 3, we can add, for example, the blocking generators **(5 or 6, 3 or 4, 1 or 2)**. We choose here **(2, 4, 6, \*, \*)** and can then see from an inspection of type 3 arrangements of Table 1, that blocking arrangement No. 36, namely, **(2, 4, 6, 15, 35)** will be suitable. The runs of this arrangement are given in Table 5. The 96 run design obtained from the combination of Tables 3 and 5 turns out to be not only a best combination design for  $k = 6, d = 3$ , but also a best sequential combination (for the addition of 64 runs).

to follow up an initial type 4 arrangement consisting of blocks 1-32 of Table 3. (This would be true whatever the choice of the type 3 arrangement, according to the indicated rule, was made.)

Table 5. An example of a best arrangement to add for  $k = 6, d = 3$ .

Block	Factor No.						Block	Factor No.					
	1	2	3	4	5	6		1	2	3	4	5	6
1	+	-	+	-	-	-	17	+	-	-	-	-	-
	-	-	-	-	+	-		-	-	+	-	+	-
2	+	+	+	-	-	-	18	+	+	-	-	-	-
	-	+	-	-	+	-		-	+	+	-	+	-
3	+	-	+	+	-	-	19	+	-	-	+	-	-
	-	-	-	+	+	-		-	-	+	+	+	-
4	+	+	+	+	-	-	20	+	+	-	+	-	-
	-	+	-	+	+	-		-	+	+	+	+	-
5	+	-	+	-	-	+	21	+	-	-	-	-	+
	-	-	-	-	+	+		-	-	+	-	+	+
6	+	+	+	-	-	+	22	+	+	-	-	-	+
	-	+	-	-	+	+		-	+	+	-	+	+
7	+	-	+	+	-	+	23	+	-	-	+	-	+
	-	-	-	+	+	+		-	-	+	+	+	+
8	+	+	+	+	-	+	24	+	+	-	+	-	+
	-	+	-	+	+	+		-	+	+	+	+	+
9	-	-	+	-	-	-	25	-	-	-	-	-	-
	+	-	-	-	+	-		+	-	+	-	+	-
10	-	+	+	-	-	-	26	-	+	-	-	-	-
	+	+	-	-	+	-		+	+	+	-	+	-
11	-	-	+	+	-	-	27	-	-	-	+	-	-
	+	-	-	+	+	-		+	-	+	+	+	-
12	-	+	+	+	-	-	28	-	+	-	+	-	-
	+	+	-	+	+	-		+	+	+	+	+	-
13	-	-	+	-	-	+	29	-	-	-	-	-	+
	+	-	-	-	+	+		+	-	+	-	+	+
14	-	+	+	-	-	+	30	-	+	-	-	-	+
	+	+	-	-	+	+		+	+	+	-	+	+
15	-	-	+	+	-	+	31	-	-	-	+	-	+
	+	-	-	+	+	+		+	-	+	+	+	+
16	-	+	+	+	-	+	32	-	+	-	+	-	+
	+	+	-	+	+	+		+	+	+	+	+	+

## 7. Case $k = 6, d = 4$

There are  $\binom{6^3}{4} = 595665$  combination designs, which split into the overall patterns of Table 6. The examination of the 178 group patterns that estimate all 1fis and 2fis is tedious, but a plot of  $\text{Sum}_2(u)$  versus  $\text{Sum}_1(u)$ , defined in Section 3, reduces the number of group patterns to 45. A subsequent plot of  $\text{Std}_2(u)$  versus  $\text{Std}_1(u)$  quickly isolates, as superior choices, the three group pattern types shown in Table 7.

Table 6. Overall patterns for case  $k = 6, d = 4$ .

	<b>f</b>	<b>A</b>	<b>B</b>	<b>C</b>
OP1	(3, 16, 24, 16, 4)	546840	1254	173
OP2	(7, 0, 48, 0, 8)	9765	74	2
OP3	(7, 8, 24, 24, 0)	39060	228	3
Total		595665		178

A = No. of combination designs

B = No. of group patterns

C = No. of group patterns that estimate all 1fi and 2fi

Table 7. Superior group patterns for case  $k = 6, d = 4$ .

		u values for 1fis and 2fis [ $\text{Sum}_1(u)$ , $\text{Sum}_2(u)$ ] & [ $\text{Std}_1(u)$ , $\text{Std}_2(u)$ ]	<b>t</b>	<b>A</b>
OP1	GP172	(2,2,2,2,2,3) & (1,1,2,2,2,2,2,2,2,3,3,4,4) [13, 35] & [0.408, 0.900]	(3, 3, 3, 4)	360
OP2	GP 24	(2,2,2,2,2,2) & (2,2,2,2,2,2,2,2,2,2,4,4,4) [12, 36] & [0.000, 0.828]	(3, 3, 3, 3)	30
OP3	GP 23	(2,2,2,3,3,3) & (1,1,1,2,2,2,2,2,3,3,3,3,3,3)	(3, 4, 4, 4)	120

A = No. of combination designs

There is clearly some preference leeway here for saying which choice is “best”, depending on how one weights the criteria that have been applied to choose these group patterns. (OP3, GP23) provides the most (15) 1fi internal estimates for example. However the estimation pattern for (OP2, GP24) shows less variability. We note that (OP3, GP23) has a t-pattern of (3, 4, 4, 4); this means that it provides a sequential possibility

to our previous example, as follows. We can add a type 4 arrangement to any of the (3, 4, 4) patterns that arise in the  $k = 6, d = 3$  case dealt with earlier. The added arrangement must be one that uses 1fi block generators that have *not* been used in type 4 arrangements already chosen. For example, suppose we had already used these arrangements of Table 1:

Arrangement 27	<b>1 3 5 26 46</b>	(Type 3)
Arrangement 42	<b>1 2 36 46 56</b>	(Type 4)
Arrangement 51	<b>3 4 16 26 56</b>	(Type 4).

The only possible addition is thus a Type 4 using (5, 6, \*, \*, \*), and we see from Table 1 that the only possibility is

Arrangement 56    **5 6 14 24 34** (Type 4).

Another example is derived from the specific design represented in Tables 3 and 5. This is made up of

Arrangement 56	<b>5 6 14 24 34</b>	(Type 4)
Arrangement 51	<b>3 4 16 26 56</b>	(Type 4)
Arrangement 36	<b>2 4 6 15 35</b>	(Type 3).

The only possible addition is thus a Type 4 using (1, 2, \*, \*, \*), and we see from Table 1 that the only possibility is

Arrangement 42    **1 2 36 46 56** (Type 4).

We recall that we can actually estimate all 1fi and 2fi with only three arrangements, as described in the  $k = 6, d = 3$  case. However, if we begin sequentially with a foldover type design with mirror image pairs of runs, four arrangements are needed.

## 8. Case $k = 6, d = 5$

We omit the parallel detailed discussion for this case and simply present Table 8, similar in format to Table 7, showing some superior choices. An example of the design types shown in the first and second items in Table 8 is further explained in the  $k = 6, d = 6$  sequential case which follows.

Table 8. Selected superior group patterns for case  $k = 6, d = 5$ .

		$u$ values for 1fis and 2fis			
		[ $\text{Sum}_1(u)$ , $\text{Sum}_2(u)$ ] & [ $\text{Std}_1(u)$ , $\text{Std}_2(u)$ ]	$t$	A	
OP1	GP179	(3,3,3,3,3,3) & (2,2,2,2,2,2,2,2,4,4,4,4,4) [18, 42] & [0.000, 1.014]	(3,3,4,4,4)	360	
OP4	GP105	(3,3,3,3,3,3) & (2,2,2,2,2,2,2,2,4,4,4,4,4) [18, 42] & [0.000, 1.014]	(3,3,4,4,4)	60	
OP4	GP285	(2,3,3,3,3,3) & (2,2,2,2,2,2,3,3,3,3,3,4,4,4,4) [17, 43] & [0.408, 0.834]	(3,3,3,4,4)	360	
OP5	GP 63	(2,2,3,3,3,3) & (2,2,2,2,3,3,3,3,3,3,3,4,4,4) [16, 44] & [0.516, 0.704]	(3,3,3,3,4)	90	

A = No. of combination designs

## 9. Case $k = 6, d = 6$

We do not give a full enumeration of the  $\binom{6^3}{6} = 67945521$  possible combinations for this case. Instead we discuss the way in which it would be sensible to build up to a superior pattern sequentially, proceeding through the values  $d = 1, 2, 3, 4, 5$  and 6, and making use of sensible choices as the design is enlarged stage by stage. We make use of the information already established in foregoing sections. A review of superior choices for successive  $d$ -values shows that it makes sense to consider the following pattern of  $t$ -values for  $d = 2$  through 5: (4, 4), (3, 4, 4), (3, 4, 4, 4), and (3, 3, 4, 4, 4). When  $d = 3$ , we already can estimate all 1fi's and 2fi's. We would need to begin the pattern with an initial type 4 arrangement to achieve this sequence. The following example illustrates the types of choices to achieve such a pattern.

Suppose we make the initial choice of a type 4 arrangement as No. 56 in Table 1 with blocking generators (5, 6, 14, 24, 34). The second arrangement should be another type 4 that does not use the same 1fi blocking generators; for example No. 51, (3, 4, 16, 26, 56). The third, of type 3, should share one 1fi with each of the previous choices and have a third, different 1fi. There are eight possible choices, from which we select No. 36, (2, 4, 6, 15, 35). The fourth arrangement selected from Table 1 should be of type 4 and such that the 1fi generators were not used by the previously chosen type 4 estimators. The only possible

choice is No. 42, (1, 2, 36, 46, 56). The 5<sup>th</sup> arrangement should be a type 3 arrangement that has the three 1fi block variables *not* used in the previous type 3 arrangement. This is (1, 3, 5, 26, 46), No. 27. (This is a superior design type, as indicated in Section 8.) The 6<sup>th</sup> arrangement can be the mirror-image pairing, No. 63, (16, 26, 36, 46, 56), which will enhance the estimation of main effects and provide a well balanced design overall. Note that  $d = 3$  sets are sufficient to estimate all 1fi's and 2fi's. Table 9 summarizes the internal estimation patterns that result from all possible choices of the 6<sup>th</sup> arrangement, including No. 63, while holding the first five choices fixed as explained above.

## 10. Sequential Designs for $k = 7$ Factors

We comment only briefly on this case because, with  $2^7 = 128$  runs in 64 blocks of size two, designs are very large. Moreover, three ( $d = 3$ ) such sets are required to obtain estimates of all 1fi's and 2fi's, and such a design is *not* an extension of the best choices of two-set designs. Only one type of group pattern is suitable. This is obtained as follows: Pick a type 4 arrangement that has three 1fi generators, for example,  $(B_1, B_2, B_3, B_4, B_5, B_6) = (1, 2, 3, 47, 57, 67)$ . Add a second type 4 arrangement with one 1fi generator in common and two not, for example,  $(1, 4, 5, 27, 37, 67)$ . The third arrangement should have: (a) a 1fi generator that has appeared only once from the first arrangement; (b) a (different) 1fi generator that has appeared only once from the second arrangement; and (c) a 1fi generator not used before. An example is  $(2 \text{ or } 3, 4 \text{ or } 5, 6 \text{ or } 7, *, *)$  which provides eight choices, with the asterisked generators completely defined by the specific choices made earlier. A table similar to Table 1, but for  $k = 7$ , shows that if  $(2, 4, 6, 7, *, *)$  is selected, the appropriate design is generated by  $(2, 4, 6, 7, 15, 35)$ . The internal estimation pattern for this  $k = 7$ ,  $d = 3$  design is

$$(1,1,1,2,2,2,3) \& (1,1,1,1,1,1,1,2,2,2,2,2,2,2,3,3,3).$$

Table 9. Summary information for sequential choice of a sixth arrangement added to numbers (56, 51, 36, 42, 27) for the case  $k = 6$ ,  $d=6$ .

Add	Type	u values for 1fis and 2fis [Sum <sub>1</sub> (u), Sum <sub>2</sub> (u)] & [Std <sub>1</sub> (u), Std <sub>2</sub> (u)]
1–6	1	(3,3,3,3,3,4) & (2,2,2,2,2,3,3,3,4,4,4,5,5) [19, 47] & [0.408, 1.125]
7–8, 11–13, 15, 18, 20, 21	2	(3,3,3,3,4,4) & (2,2,2,2,2,3,3,3,3,4,4,5,5,5) [20, 50] & [0.516, 1.234]
9–10, 14, 16–17, 19	2	(3,3,3,3,4,4) & (2,2,2,3,3,3,3,3,3,4,4,4,5,5) [20, 50] & [0.516, 0.976]
22–26, 31–32, 37–41	3	(3,3,3,4,4,4) & (2,2,2,2,3,3,3,3,3,4,4,5,5,5) [21, 51] & [0.548, 1.183]
28–30, 33–35	3	(3,3,3,4,4,4) & (2,2,3,3,3,3,3,3,3,4,4,4,5,5) [21, 51] & [0.548, 0.910]
43, 45, 48, 50, 52, 55	4	(3,3,4,4,4,4) & (2,2,2,2,3,3,3,3,4,4,5,5,5) [22, 50] & [0.516, 1.234]
44, 46–47, 49, 53–54	4	(3,3,4,4,4,4) & (2,2,2,3,3,3,3,3,4,4,4,4,5,5) [22, 50] & [0.516, 0.976]
57–62	5	(3,4,4,4,4,4) & (2,2,2,2,2,2,3,3,4,4,4,4,5,5) [23, 47] & [0.408, 1.125]
63	6	(4,4,4,4,4,4) & (2,2,2,2,2,2,2,2,4,4,4,4,4,4) [24, 42] & [0.000, 1.014]

## 11. Sequential Designs for $k = 8$ Factors

Four sets of 256 runs in 128 blocks of two are needed to estimate all 1fi's and 2fi's, although all 1fi and all but one 2fi can be estimated with three sets of 256 runs in 128 blocks of two. An example is the following. We first combine two type 5 combinations with one common 1fi blocking generator, for example,  $(B_1, B_2, B_3, B_4, B_5, B_6, B_7) = (1, 2, 3, 48, 58, 68, 78)$  and  $(1, 4, 5, 28, 38, 68, 78)$ . The third choice is a type 5 arrangement

selected from possibilities (2 or 3, 4 or 5, 6 or 7 or 8, \*, \*, \*, \*) using, in other words, one of the non-common 1fi's from each of the first two selections, one 1fi not used before, and the asterisked values determined from the first three choices. A suitable choice is, for example, (2, 4, 6, 18, 38, 58, 78). This triple of arrangements leaves one 2fi unestimated. There are many possibilities for the choice of a fourth arrangement. The ones that provide the most (22) internal 1fi estimates are those of type 7, for example, (8, 17, 27, 37, 47, 57, 67). The pattern of internal estimates for this quadruple combination is then:

$$(2,2,2,3,3,3,3,4) \text{ & } (1,1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,2,2,2,2,3,3,3,3,3)$$

and  $\text{Sum}_1(u) = 22$ ,  $\text{Sum}_2(u) = 52$ ,  $\text{Std}_1(u) = 0.707$  and  $\text{Std}_2(u) = 0.756$ . Further examples of, and additional details about, designs of the type discussed in this article are given in Yang<sup>5</sup> (2002).

## Acknowledgements

We gratefully acknowledge the help of Professor Gregory Reinsel in providing additional hard disc capacity for our computing at the University of Wisconsin, and we mourn his death at age 56 on May 5, 2004. We also thank a referee for a number of helpful suggestions.

## References

1. P. Rosenbaum, *Technometrics*, 125 (1999).
2. Y. J. Yang and N. R. Draper, *J. Quality Technology*, 294 (2003).
3. G. E. P. Box, W. G. Hunter and J. S. Hunter, *Statistics for Experimenters*, Wiley (1978); 2nd edn. (2005).
4. N. R. Draper and I. Guttman, *J. Quality Technology*, 71 (1997).
5. Y. J. Yang (2002). Ph.D. Thesis, University of Wisconsin Statistics Department, Madison.

## CHAPTER 2

# RESPONSE SURFACE EXPERIMENTS ON PROCESSES WITH HIGH VARIATION

Steven G. Gilmour and Luzia A. Trinca

*School of Mathematical Sciences, Queen Mary, University of London  
Mile End Road, London E1 4NS, UK*

*E-mail: s.g.gilmour@qmul.ac.uk*

*Departamento de Bioestatística - IB, UNESP  
Botucatu, CP 510, 18618-000, SP, Brazil*

*E-mail: ltrinca@ibb.unesp.br*

In industries such as food engineering, pharmaceuticals, bioprocessing and agrochemicals the use of biological materials means that, when experiments are being performed, run to run variation is typically higher than in many other engineering applications. This means that some of the smallest response surface designs are inappropriate and the traditional design principles of randomization, replication and blocking are more important. Recent work for response surface experiments of this type are reviewed in this chapter. Topics covered include the choice of run size, three-level designs, including some qualitative factors, arranging designs in various block structures, restricted randomization leading to multi-stratum designs and issues in the analysis of blocked and multi-stratum designs.

### 1. Introduction

In the first paragraph of their seminal paper, Box and Wilson<sup>5</sup> explained that they had developed response surface methodology (RSM) for chemical processes, but wrote “we believe that the methods will be of value in other fields where experimentation is sequential and the error fairly small”. In the discussion of their paper, D. R. Read pointed out that in biochemical processes, the standard deviation of observations is typically 10 or more times that in Box and Wilson’s examples and wrote “A considerable amount of replication, either absolute or hidden, is required in such cases, and it will be desirable to make efficient use of this replication”. Much research

on response surface designs has concentrated on producing designs that are smaller than Box and Wilson's central composite design. See the texts by Box and Draper<sup>4</sup>, Khuri and Cornell<sup>19</sup> and Myers and Montgomery<sup>23</sup> and the review by Myers *et al.*<sup>24</sup> for this and other research in RSM. Here, we summarize recent work on larger designs for processes with high variation.

Our own interest in these designs arose out of regular contacts we had with researchers in the School of Food Biosciences at the University of Reading, where we helped many experimenters to design their experiments and analyze their data. However similar problems arise in experiments in the pharmaceutical, biotechnology and agrochemicals industries. The common feature of these industries is that they all work with biological materials and so have to accept the natural variation that exists in living organisms.

Although experimentation in these applications is often sequential, some of the small designs, such as a  $2^2$  factorial plus 3 centre points, often recommended will not give reliable results. Similarly, the value of experimenting along the path of steepest ascent is doubtful, as any change in response observed could be due to error. Thus, although experimentation is sequential, experiments should be designed in fairly large stages.

Traditionally the first stage in experimentation in a response surface study involves using a small design to fit the first order model. In our applications several replicates of such a design would be required to obtain reliable estimates of the parameters of this model. However, because of the hidden replication properties of the factorial structure, we can make use of the larger number of runs to allow higher order terms to be fitted without seriously damaging the estimation of linear main effects. It is often sensible to use a design which allows the second order polynomial model to be fitted. After using the fitted model to decide which combination of factor levels is best, some follow up runs may be made to check the predictions. In the applications considered here, experimentation is not so expensive as to make this impossible, as it may be in some industrial applications.

Box and Draper<sup>4</sup> discussed the competing effects of bias and variance and showed how large bias should lead to experiments being conducted in limited regions of experimentation. Since error is larger in biological processes, variance is more important and the region of experimentation should be larger.

Blocking, to separate the larger sources of variation from the estimation of the response surface model, is also much more important than it is usually considered to be in RSM. In many applications, large day to day variation can be expected because of changes in environmental conditions,

degradation of materials and other uncontrollable causes. The experimental runs should then be grouped into blocks with the units within a block being run on the same day. Since it may be possible to do only a few runs each day, the blocks will often be fairly small. We always randomise the run order within blocks.

In summary, in experiments using biological materials, we have to emphasise the traditional principles of experimental design (which were developed for biological applications), namely randomization, replication and blocking. Thus the designs required for these experiments are quite different from the sequences of small designs recommended in standard RSM. Full or regular fractional factorial designs are not always appropriate either, as they are not specifically intended for use with quantitative factors.

Designs are required which allow the second order polynomial to be fitted and are large enough to allow for blocking and some testing of lack of fit of the model, while still leaving enough degrees of freedom to get a reliable estimate of pure error. Three or four level designs are appropriate to achieve all of this, although sometimes it is necessary to include some two level factors as well. In bioprocessing, unit costs per run are high relative to fixed costs and follow up runs can be made to check the conclusions from the model fitting, so three level designs give an economical and effective way of experimenting.

## 2. Design Strategy

The choice of a good design involves:

- (1) a clear statement of the objectives, which should lead to the choice of a set of treatments (combinations of levels of the factors);
- (2) consideration of the experimental units (i.e. runs) which can be used and any expected patterns of variability among them, which will lead to a blocking structure;
- (3) consideration of any restrictions on which treatments can be applied to which units, e.g. hard to change factors, which might lead to multi-stratum structures;
- (4) given 1-3, finding a combinatorial, algorithmic or *ad hoc* method of constructing a design which meets these requirements.

See Bailey<sup>2,3</sup> and Mead<sup>21</sup> for similar descriptions of the process of designing experiments in more general contexts and Coleman and Montgomery<sup>6</sup> for a description in the context of industrial experiments. This is also dis-

cussed in many textbooks on the design of experiments, e.g. Mead<sup>20</sup> and Montgomery<sup>22</sup>. Box and Draper<sup>4</sup> and Myers and Montgomery<sup>23</sup> discussed considerations in selecting a response surface design, mainly from the viewpoint of processes with low variation.

Because of the several requirements a design has to meet there are advantages in separating the overall design building process into three stages:

- (1) first, a decision is made regarding how many runs should be made;
- (2) next, a treatment set with this number of runs is chosen to have a number of desirable properties;
- (3) finally, this treatment set is arranged in blocks in such a way that these properties are maintained.

There is some disagreement in the literature about the benefits of separately choosing a treatment design and then optimally blocking it, as opposed to choosing a globally optimum design. Cook and Nachtsheim<sup>7</sup> mentioned, as benefits of blocking a given treatment design, the fact that well known designs (such as regular fractional replicates and central composite designs) could be used, the ability to check for lack of fit and the fact that blocks might turn out to have been ineffective. Trinca and Gilmour<sup>29</sup> also argued that a design has to be chosen with several useful properties and that designs such as central composite designs do this. Such designs can then be arranged in blocks in a way that preserves these properties as well as possible. On the other hand, Atkinson and Donev<sup>1</sup> and Goos and Vandebroek<sup>15</sup> argued that, since a  $D$ -optimal treatment design, arranged in blocks  $D$ -optimally, is not necessarily a globally  $D$ -optimal blocked design, a one-stage choice of  $D$ -optimal design is better.

The argument for separating the construction of the treatment design and its arrangement into blocks hinges on a proper understanding of what blocking is. We take the traditional, fisherian, view that blocking is defined as a restriction to the randomization of the design. The particular type of randomization used defines the initial analysis of variance and, implicitly, the initial model that will be fitted. Therefore a particular blocking structure is chosen so that the derived analysis will be as informative as possible. If the randomization is conducted correctly the derived model will be a linear mixed model, with random effects corresponding to each level of randomization, e.g. blocks and runs, or rows, columns and runs.

The treatment design is chosen to meet the objectives of the experiment. In practice, the vast majority of experimenters prefer response surface designs, such as the central composite design, to  $D$ -optimal designs, because

they have many desirable properties, even though they are less efficient for estimating the parameters of the second order polynomial model. It does not seem reasonable that this logical and carefully considered choice should be drastically reversed just because the experiment is randomized in a different way. To put it another way, if the only objective of the experiment is to estimate the parameters as well as possible, in the sense of minimizing the size of their joint confidence region, then one should use a  $D$ -optimal design. If the objectives are more complex, one should use a design with many desirable properties. This choice is independent of how the experiment will be randomized.

The initial form of data analysis in a RS experiment of this type with  $q$  factors involves fitting, and checking for lack of fit of, the second order polynomial model,

$$\mu = \beta_0 + \sum_{i=1}^q \beta_i x_i + \sum_{i=1}^q \beta_{ii} x_i^2 + \sum_{i=1}^{q-1} \sum_{j=i+1}^q \beta_{ij} x_i x_j, \quad (1)$$

where  $\mu$  is the expected response, given  $\mathbf{x}$ ,  $\mathbf{x}' = [x_1, \dots, x_q]$ ,  $x_1, \dots, x_q$  are the levels of the factors, coded to be between  $-1$  and  $1$ , and responses from different runs are assumed to be independent with constant variance  $\sigma^2$ . If a three level design has been used the lack of fit parameters are associated with some of the other components of the two factor interactions and with higher order interactions. Others of these higher order effects will be aliased with terms already in the second order model. If serious lack of fit is found, one or two higher order terms will be added to the model to see if they can explain the lack of fit.

If no serious lack of fit of the second order model is found, it is often useful to try to simplify the model by dropping some of the interaction and quadratic parameters which seem to be small and possibly some of the linear terms if some factors do not seem to have any effect on the responses.

This analysis is a form of stepwise variable selection. Starting with the second order model, one or two higher order terms are added if necessary and then terms are dropped one by one if they do not seem to be important. A term is not dropped if any term to which it is marginal (Nelder<sup>26</sup>) is still in the model, e.g. the linear effect of a factor is not dropped if any higher order effects involving that factor are still in the model. One of the objectives of this model simplification is to aid the interpretation of the results. The experiments are not run simply to allow predictions to be made and the optimal combination of factor levels to be identified. Understanding of how changing the levels of the factors leads to changes in the response is

essential.

The experiments must be designed with a view to carrying out this analysis as efficiently as possible. Thus a design should allow:

- (1) the second order model to be fitted efficiently;
- (2) some check for lack of fit of this model;
- (3) variable selection to be as nearly as possible unambiguous; and
- (4) the final model to be fitted efficiently.

If only the first objective were important, then a *D*-optimal or weighted-*A*-optimal design for the second order model would be appropriate. However, the second requirement implies that the design should allow the fitting of some higher order terms and should allow sufficient degrees of freedom for estimating pure error. The third requirement implies that orthogonality, or near orthogonality, of the parameter estimates is important and it is very useful to achieve *factorwise balance*, i.e. equal variance for each effect of the same type. The final requirement implies that the design should give good estimates of the parameters of any submodel of the second order polynomial.

### 3. Choice of Size of Experiment

A useful concept for the first stage in building a design is the *resource equation* (Mead<sup>20</sup>). We will require at least  $n$  runs in blocks of size  $n_b$ , where

$$n = \frac{n}{n_b} + \frac{q(q+3)}{2} + n_{lof} + n_{pe}, \quad (2)$$

$\frac{n}{n_b}$  degrees of freedom are needed to allow for block effects,  $\frac{q(q+3)}{2}$  are needed to estimate the response surface model,  $n_{lof}$  is a small number (typically 5-10) of degrees of freedom for estimating higher order terms to check for lack of fit and  $n_{pe}$  is a small number (typically 5-15) of degrees of freedom for estimating pure error.

This should be regarded as giving a minimum sensible size for the experiment. It is also necessary to consider the efficiency of the design chosen. This can be assessed either through formal power studies or, more often in practice, by simply looking at the standard errors of parameter estimates that will arise from a proposed design. Of course, economic and practical considerations will also affect the size of the experiment chosen.

More unusual is the situation where the proposed experiment is much bigger than the minimum size given by the resource equation. In this case

large numbers of degrees of freedom will be used in estimating either high order terms or pure error. It is worth considering first if a smaller experiment could be run, although this might be undesirable for reasons of efficiency. If the proposed size is still considered appropriate, it is worth considering whether there are additional factors that could be studied, as no increase in the size of the experiment will be needed to accommodate them. Alternatively, it is worth considering whether the run to run variance could be better controlled by using smaller blocks or an additional block factor.

### Example 1

Dehydration is one of the major methods of food preservation and demand from the food industry for dehydrated potatoes is increasing. High temperature puffing is a process that leads to better quality dehydrated potatoes. Varnalis *et al.*<sup>31</sup> reported the results of an experiment to optimise the pre-treatment and drying conditions for the production of high quality potato cubes. Four factors were to be studied, as in Table 1, so that  $q = 4$ . Only six runs could be made per day and day to day differences in the response were considered likely. Therefore, it was decided to use blocks of size  $n_b = 6$ .

Table 1. Factors and levels for Example 1.

Factor	Units	Levels		
		-1	0	1
Blanching time	min	2	4	6
Sulfiting time	min	2	6	10
Initial drying time	min	40	60	80
Puffing time	s	40	50	60

Allowing 15 degrees of freedom for lack of fit and pure error, the resource equation is

$$\begin{aligned} n &= \frac{n}{6} + 14 + 15 \\ \Rightarrow n &= 34.8. \end{aligned}$$

Therefore 6 blocks, each of 6 runs, seems like a reasonable size for this experiment and was, in fact, the size chosen.

### Example 2

Industrial processes for the extraction of edible oil from oilseeds usually use solvents. Safety considerations have prompted attempts to use water

for extraction, but these were usually unsuccessful due to the low oil yields obtained. The use of enzymes in aqueous extraction to try to improve the yields had been suggested and Rosenthal *et al.*<sup>28</sup> reported an experiment to study the effects of several process variables on the yields from aqueous enzymatic extraction. In addition to four process factors, two types of enzyme, protease and cellulase, were used, which defines a qualitative two-level factor, so that  $q = 5$ . The process factors studied are shown in Table 2. Because the experiment was carried out on a small laboratory scale, heterogeneity of the runs was not expected to be a problem and so no blocking was used, i.e.  $n_b = n$ .

Table 2. Factors and levels for Example 2.

Factor	Units	Levels		
		-1	0	1
Particle size	μm	212.5	449.5	855
Liquid-to-solid ratio		0.05	0.125	0.2
Enzyme concentration	%	0.1	0.45	2
Time of hydrolysis	min	30	60	120

Since one of the factors is qualitative and therefore has no quadratic parameter, we have to reduce the degrees of freedom for the treatment model by 1 in the resource equation. Allowing 15 degrees of freedom for lack of fit and pure error, this becomes

$$n = 1 + 19 + 15 = 35.$$

In fact, 40 runs were made, just slightly greater than this minimum reasonable size for the experiment.

### Example 3

Freeze drying has become an economically important process and freeze dried coffee is a commercially successful product. However, its quality could be improved if the retention of volatile aroma compounds was increased. An experiment was carried out to study the effect on aroma retention of the five factors given in Table 3. There was no obvious blocking factor, so initially we might take  $n_b = n$ .

Allowing 15 degrees of freedom for lack of fit and pure error, the resource equation is

$$n = 1 + 20 + 15 = 36.$$

Table 3. Factors and levels for Example 3.

Factor	Units	Levels		
		-1	0	1
Pressure	Pa	30	50	70
Solids content	%	10	20	30
Slab thickness	cm	1	1.5	2
Temperature	°C	25	35	45
Freezing		slow	medium	fast

In this experiment, there was severe pressure on resources, so it was decided to make only 30 runs. Other complications arose from the fact that the pressure was more difficult to change than the other factors, which made it desirable to change the pressure setting as few times as possible. This will be discussed further in Section 6. Some aspects of this experiment were described by Gilmour *et al.*<sup>14</sup>

#### 4. Choice of Treatments

Having decided that the experiment should have  $n$  runs, the next task is to find an appropriate treatment set. The most commonly used response surface designs are the central composite design and the Box-Behnken design. However, these are rather smaller than the experiments we recommend for the applications discussed here. In Example 1, we have four factors. The four-factor central composite design has, typically, around 30 runs, depending on the number of centre points used, and the four-factor Box-Behnken design has 27 runs. It was decided that 36 runs was an appropriate size for this experiment. It is possible to replicate some of the points in one of the standard designs, but it is not clear which points should be replicated and whether better designs are available.

##### 4.1. Subset Designs

Gilmour<sup>11</sup> introduced a rich class of designs for factors with three equally spaced levels, namely *subset designs*. Let the factor levels be coded  $-1$ ,  $0$  and  $1$ . The designs are made up of subsets of the  $3^q$  factorial design defined by Hoke<sup>17</sup> as follows. Let  $S_r$ ,  $r = 1, \dots, q$ , be the subset of points of the  $3^q$  factorial design which lie on the hypersphere of radius  $\sqrt{r}$  about the centre point,  $S_0$ . Thus  $S_r$  contains all points which have  $r$  factors at  $\pm 1$  and the remaining  $q - r$  factors at  $0$ . Hoke went on to further divide the subsets in order to obtain designs with as few runs as possible. Gilmour<sup>11</sup> studied designs made up of combinations of the  $S_r$ . A design made up of

a combination of  $S_r$  subsets will be denoted by  $S_{r_1} + S_{r_2} + \dots$  and  $c_r S_r$  means that the points in the subset  $S_r$  are replicated  $c_r$  times. Gilmour developed the properties of these designs in more detail and presented some extensions.

This class of designs includes many that allow the second order model to be fitted. In order to do so, we require:

- $c_r > 0$  for at least two  $r$  and  $c_r > 0$  for at least one  $r$  with  $1 \leq r \leq q-1$  so that all quadratic parameters can be estimated;
- $c_r > 0$  for at least one  $r \geq 2$  so that all interactions can be estimated;

Note that the subset  $S_r$  contains  $\binom{q}{r} 2^r$  points, which consist of a  $2^r$  factorial design at levels  $-1$  and  $1$  for each combination of  $r$  factors, with the other  $q-r$  factors held at  $0$ .

Gilmour<sup>11</sup> showed that subset designs allow all parameters to be estimated orthogonally to each other, except for the quadratic parameters which are correlated with each other and with the intercept, and that subset designs are factorwise balanced. Thus, designs in this class should be useful for variable selection and for fitting submodels of the full second order model. The choice of design from within this class will be made on the basis of efficiency of estimating the parameters, ensuring sufficient degrees of freedom for pure error and so on.

### Example 1 (cont'd)

We require a design for four factors in 36 runs. The subset  $S_4$  has 16 runs,  $S_3$  has 32 runs,  $S_2$  has 24 runs,  $S_1$  has 8 runs and  $S_0$  has 1 run, so the subset designs available are those shown in Table 4. Along with the designs are shown their variances (appropriately scaled) for estimating each parameter, from which a weighted  $A$  criterion could be evaluated, and the value of the  $D$  criterion,  $D = |\mathbf{X}'\mathbf{X}|^{\frac{1}{p}}$ , where  $\mathbf{X}$  is the design matrix for the second order polynomial model and  $p = 15$  is the number of parameters in this model.  $S_3 + 4S_0$  has the highest  $D$ -efficiency, gives relatively good estimation for all effects and is, unusually, orthogonal for estimating all effects including the quadratic effects. This design was recommended by Edmondson<sup>10</sup> and studied further by Davis and Draper<sup>8</sup>. It does, however, have only 3 degrees of freedom for estimating pure error and so for this experiment  $S_4 + 2S_1 + 4S_0$ , which is almost as good in terms of efficiencies and allows 11 degrees of freedom for pure error, was preferred.

Table 4. Designs and their properties for Example 1.

Design	Variances $\times \frac{n}{\sigma^2}$			$D$
	$\beta_i$	$\frac{1}{2}\beta_{ii}$	$\beta_{ij}$	
$S_4 + 2S_1 + 4S_0$	1.8	1.668	2.25	13.94
$S_4 + S_1 + 12S_0$	2	3.45	2.25	11.99
$S_3 + 4S_0$	1.5	1.125	2.25	15.64
$S_2 + S_1 + 4S_0$	2.571	1.109	9	7.80
$S_2 + 12S_0$	3	1.25	9	7.61

#### 4.2. Fractional and Incomplete Subset Designs

Since subset designs are made up from two-level factorial designs in subsets of the factors, more designs can be obtained by replacing some of these factorial designs by fractional replicates. If the fractions are of Resolution V or higher, the properties of the subset designs for the second order model are not affected. However, if the fractions are of Resolution IV, then the orthogonality between interactions is lost and, depending on the fraction used, factorwise balance might be lost. If the fractions are of Resolution III, then the orthogonality between linear and interaction effects is lost and if they are of Resolution II, then the orthogonality between linear effects is lost. We use the notation  $\frac{1}{2^m} S_r^R$  to denote a fraction of the subset  $S_r$ , made up of  $\frac{1}{2^m}$  fractional replicates of Resolution- $R$ .

#### Example 1 (cont'd)

Using half replicates of the  $2^4$  fractions of either Resolution III or IV in  $S_4$  and using half replicates of Resolution III in  $S_3$  gives the additional designs shown in Table 5 along with some of the full subset designs. These designs are quite competitive in terms of  $D$ -efficiency and  $S_4 + \frac{1}{2}S_3 + 4S_0$  performs well for estimating the parameters individually. It does, however, have the feature typical of  $D$ -optimal designs that it is somewhat poor for estimating the quadratic parameters and it allows only 4 degrees of freedom for pure error. These considerations led to the decision to use  $S_4 + 2S_1 + 4S_0$  for the experiment being confirmed.

For larger experiments, another modification of subset designs is possible, using the same idea as larger Box-Behnken designs. Since the subset  $S_r$  contains a  $2^r$  factorial design in each set of  $r$  factors, the number of runs can be reduced by using an incomplete subsubset of  $S_r$ . For example, with four factors,  $S_2$  consists of  $2^2$  factorial designs in each pair of factors. If only the subsubset of pairs  $\{(1, 2), (1, 3), (2, 4), (3, 4)\}$  is used, an

Table 5. Designs and their properties for Example 1.

Design	$\beta_i$	Variances $\times \frac{n}{\sigma^2}$		$D$
		$\frac{1}{2}\beta_{ii}$	$\beta_{ij}$	
$S_4 + 2S_1 + 4S_0$	1.8	1.668	2.25	13.94
$S_3 + 4S_0$	1.5	1.125	2.25	15.64
$S_4 + \frac{1}{2}S_4^{IV} + S_1 + 4S_0$	2.385	3.473	1.695	14.92
$S_4 + \frac{1}{2}S_4^{III} + S_1 + 4S_0$	1.385(1) 1.549(3)	3.473	1.5(3) 1.664(3)	14.95
$S_4 + \frac{1}{2}S_3 + 4S_0$	1.286(1) 1.369(3)	2.388	1.573	16.88
$\frac{1}{2}S_4^{IV} + S_2 + 4S_0$	1.8	1.027	5.392	12.10
$\frac{1}{2}S_4^{III} + S_2 + 4S_0$	1.8(1)	2.446(3) 1.027	3(3) 4.082(3)	12.79

(m) indicates that  $m$  of the effects have this variance.

incomplete subset design is obtained. There is little advantage in doing this here, but for more factors, useful designs can be discovered. Orthogonality is retained and, if the subsubset forms a balanced incomplete block design, then factorwise balance is retained.

#### 4.3. Two-Level Factors

Sometimes, as in Example 2, it is necessary to include one or more two-level factors, along with several three-level factors in a response surface design. The inclusion of a few two-level factors in response surface designs was discussed by Draper and John<sup>9</sup> and Wu and Ding<sup>32</sup> who gave several examples constructed in an *ad hoc* fashion and, of course, it is possible to use a search routine to find a  $D$ - or weighted- $A$ -optimal design. However, two-level factors fit very easily into the framework of subset designs. Within each subset, the points with level 0 of the two-level factors are simply deleted. Then, the subsets, or fractional subsets, or incomplete subsets, are combined as before to generate a number of candidate designs.

#### Example 2 (cont'd)

For five factors, one at two levels, in 40 runs, the redefined subsets which might be useful are  $\frac{1}{2}S_5^V$  with 16 runs,  $\frac{1}{4}S_5^{III}$  with 8 runs,  $\frac{1}{2}S_4^{IV}$  with 32 runs,  $\frac{1}{2}S_3^{III}$  with 24 runs,  $S_2$  with 16 runs and  $S_1$  with 2 runs. The candidate designs and their properties are shown in Table 6. Only the first three designs seem to be worth serious consideration. Although  $\frac{1}{2}S_5 + S_2 + 4S_1$  has the lowest  $D$ -efficiency among these three designs, it has 8 degrees of freedom for pure error, where the other two designs do not allow pure error to be estimated. This was the design used in the experiment and is shown in Table 7.

Table 6. Designs and their properties for Example 2.

Design	Variances $\times \frac{n}{\sigma^2}$ (min-max)				$D$
	$\beta_i$	$\frac{1}{2}\beta_{ii}$	$\beta_{ij}$		
$\frac{1}{2}S_5 + \frac{1}{2}S_3$	1.136–1.531	1.198	1.531–2.045		20.20
$\frac{1}{2}S_5 + S_2 + 4S_1$	1–2	1.964	2–2.5		16.01
$\frac{1}{4}S_5 + \frac{1}{2}S_4$	1.125–1.451	2.5	3.028–4.495		17.60
$\frac{1}{4}S_5 + \frac{1}{2}S_3 + 4S_1$	2.051–7.286	1.094	5.384–6.916		12.38
$\frac{1}{4}S_5 + S_2 + 8S_1$	1.25–10	1.991	6–25		7.23
$\frac{1}{2}S_3 + S_2$	2.5–3.429	6.667	3.429–12.5		9.15

$S_r$  has points with factor  $X_1$  at 0 deleted

## 5. Unit Structures

It is possible to define any block structure, or indeed any unit covariance structure, and arrange the treatment design in runs as efficiently as possible. However, there are advantages in using simple orthogonal block structures if possible. A *trivial block structure* is a set of equal-sized blocks. Two block factors are *nested* if the blocks defined by one of them contain complete sets of blocks defined by the other, e.g. in a block design, taking the runs as a blocking factor, the runs are nested within the blocks. We denote this by Blocks/Runs. Two block factors are *crossed* if each level of one appears with each level of the other, e.g. in a row-column design, rows and columns are crossed. We denote this by Rows\*Columns. Nelder<sup>25</sup> defined a *simple orthogonal block structure* to be one that can be obtained from trivial block structures by any combination of nesting and crossing.

Thus simple orthogonal block structures include any block design with equal sized blocks, complete row-column designs, resolved block designs, split-plot designs and many more complex structures. In general, we use the crossing and nesting operators defined above to denote the unit structure as  $(U_{11} * \dots * U_{1u_1}) / \dots / (U_{s1} * \dots * U_{su_s})$ . This corresponds exactly to the Wilkinson-Rogers notation used by packages such as R, Splus and GenStat.

Given a treatment set, chosen using the methods in Section 4 or by any other means, and a block structure, the problem is to allocate the treatments to the runs, i.e. decide which treatment should be applied to which run. This will be done by choosing an efficient design and then randomizing it by randomly relabelling the blocks for each blocking factor independently. Thus for a block design, the blocks are randomly relabelled and then the runs within each block are randomly relabelled. For a row-column design

the rows are randomly relabelled and the columns are randomly relabelled. One advantage of simple orthogonal block structures is that, if they are randomized in this way, the linear mixed model for analysis is justified by the randomization alone, without any further assumption about a population from which the runs are sampled.

Table 7. Design used for Example 2.

Factors				
$X_1$	$X_2$	$X_3$	$X_4$	$X_5$
-1	-1	-1	-1	-1
-1	-1	-1	1	1
-1	-1	1	-1	1
-1	-1	1	1	-1
-1	1	-1	-1	1
-1	1	-1	1	-1
-1	1	1	-1	-1
-1	1	1	1	1
1	-1	-1	-1	1
1	-1	-1	1	-1
1	-1	1	-1	-1
1	-1	1	1	1
1	1	-1	-1	-1
1	1	-1	1	1
1	1	1	-1	1
1	1	1	1	-1
-1	-1	0	0	0
-1	1	0	0	0
1	-1	0	0	0
1	1	0	0	0
-1	0	-1	0	0
-1	0	1	0	0
1	0	-1	0	0
1	0	1	0	0
-1	0	0	-1	0
-1	0	0	1	0
1	0	0	-1	0
1	0	0	1	0
-1	0	0	0	-1
-1	0	0	0	1
1	0	0	0	-1
1	0	0	0	1
-1	0	0	0	0
1	0	0	0	0
-1	0	0	0	0
1	0	0	0	0
-1	0	0	0	0
1	0	0	0	0
-1	0	0	0	0
1	0	0	0	0

### 5.1. Trivial Blocking

We consider first a single blocking factor, so that the  $n$  runs are split into  $n/n_b$  blocks, each of size  $n_b$ . The aim is to allocate the treatments chosen to these runs in such a way that the properties of the treatment design are preserved as well as possible. This is most easily achieved if it is possible to construct an orthogonally blocked response surface design, in which there is no correlation between the estimated block effects and any estimated parameter in the second order model.

It is occasionally possible to construct an orthogonally blocked design using the methods of Section 4.3. There, if we restrict attention to designs made up of full factorial sets, all parameters corresponding to two-level factors are estimated orthogonally to all other parameters. Therefore, the two-level factors can be treated as dummy factors for a blocking factor, e.g. if there are two two-level factors, the four combinations of their levels can be used to define four blocks. The same can be done with fractional subset designs of Resolution V and higher for up to four blocks and for Resolution IV fractional subset designs for two blocks. Thus, for example, the design obtained for Example 2 could be used to study four three-level factors in two blocks each of size 20.

However, it is easily seen that this approach is of use in only rare situations. In particular, it can only be applied when the number of blocks is a power of 2 and then only when a good design for a two-level factor exists. Box and Draper<sup>4</sup> gave a few other examples of orthogonally blocked central composite and Box-Behnken designs, but these exist only for a few particular block sizes. Other special cases were given by Edmondson<sup>10</sup>. In Example 1 there are to be six blocks and it is not one of the special cases, so a different approach is required. No combinatorial method of constructing good designs is known, so an algorithmic approach is usually used.

The algorithmic approach to arranging designs in blocks requires two important features to be defined:

- a criterion for discriminating between designs;
- a search algorithm for generating candidate designs to be compared.

Several published algorithms use computational tricks which are specific to the particular criterion the authors are using. However, the overall structure of any of the algorithms can be used with any criterion, although they will run more slowly with some criteria than with others.

### 5.1.1. Criteria for Comparing Designs

The most commonly used criteria are variance based. Assume that the model for a blocked response surface design is

$$\mathbf{Y} = \mathbf{B}\boldsymbol{\alpha} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (3)$$

where  $\mathbf{Y}$  is the vector of responses,  $\mathbf{B}$  is a matrix whose columns are indicators for the blocks,  $\boldsymbol{\alpha}$  is the vector of block means,  $\mathbf{X}$  is the design matrix for the second order polynomial model,  $\boldsymbol{\beta}$  is the vector of parameters for the polynomial model and  $\boldsymbol{\epsilon}$  is the random error vector, having expectation  $\mathbf{0}$  and variance matrix  $\sigma^2\mathbf{I}$ . Although this model implies fixed block effects and we have recommended random block effects for the analysis, for design purposes it is reasonable to work with model (3). In the linear mixed model, the variances of parameter estimates depend on both the variance components and their estimates. It is possible to obtain optimal designs by using prior point estimates of the variance components, as was done by Goos and Vandebroek<sup>15</sup>. However, this leads to designs which are sensitive to these prior guesses. In the linear mixed model, the variances of parameter estimates increase as  $\sigma_B^2$ , the block variance component increases. Therefore, a robust design will be one that is good when  $\sigma_B^2 \rightarrow \infty$ , in which case the mixed model tends to model (3). Therefore most authors prefer to choose designs which are optimal for the fixed effects model, even if they will analyse the data using the mixed model. An alternative would be to take a fully Bayesian approach to both design and analysis, but that is not considered further here.

The most commonly used criterion is  $D_s$ -efficiency for the parameters  $\boldsymbol{\beta}$ , which minimizes the size of a joint confidence region for these parameters, given the treatment design. It is easily shown that the  $D_s$  criterion has a one-to-one relationship with  $|[\mathbf{B} : \mathbf{X}]'[\mathbf{B} : \mathbf{X}]|$ , which makes it very simple to calculate. Thus, it is possible to search over many more designs using the  $D_s$  criterion than using some of the other criteria to be discussed. Cook and Nachtsheim<sup>7</sup> and Goos and Vandebroek<sup>15</sup> used  $D_s$ -efficiency to arrange designs in blocks. Nevertheless  $D_s$ -efficiency is not the only useful criterion and others, more closely connected with the search for near-orthogonality, have been suggested.

Trinca and Gilmour<sup>29</sup> used the weighted mean efficiency factor, or  $M_w$ , criterion. Letting  $\boldsymbol{\beta}' = [\beta_1, \dots, \beta_p]$ , the efficiency factor for a parameter  $\hat{\beta}_i$

is

$$E_i = \frac{V(\hat{\beta}_i)/\sigma^2}{V_b(\hat{\beta}_i)/\sigma_b^2} \times 100\%,$$

where the subscript  $b$  indicates that the quantity is from the blocked design, the absence of this subscript denoting the unblocked design. Then the weighted mean efficiency factor criterion is

$$M_w = \frac{\sum_{i=1}^p w_i E_i}{\sum_{i=1}^p w_i},$$

where  $w_i$  is the weight given to  $\beta_i$ . This criterion is similar to the first part of the  $(M, S)$  criterion which is frequently used for blocking designs with unstructured treatments but which uses canonical efficiency factors rather than the efficiency factors for the parameters of interest - see, for example, John and Williams<sup>18</sup>.

One advantage of the  $M_w$  criterion is that it allows the relative information on different parameters in the treatment design to be retained in the blocked design. For example, it might be that in some experiment a design is chosen which has high variances for the interaction parameters relative to the linear and quadratic parameters, for example because prior knowledge indicates that they are likely to be unimportant. Then  $D_s$ -optimal blocking will tend to arrange the design in blocks in order to improve the estimation of the interaction terms as much as possible at the cost of estimating the linear and quadratic parameters less well. By measuring the efficiencies relative to the unblocked design, the  $M_w$  criterion allows the properties of the unblocked design to be retained.

Nguyen<sup>27</sup> used a different criterion to construct nearly orthogonal designs. Letting  $\mathbf{X}_c$  be the centered form of  $\mathbf{X}$  he chose designs that minimize the sum of squared elements of  $\mathbf{B}'\mathbf{X}_c$ . Although it is not directly variance based, this criterion is related to orthogonality and produces reasonable designs in many situations. It deserves further investigation.

### 5.1.2. Algorithms

The optimally blocked design can be found by evaluating the relevant criterion for every possible arrangement of the design in blocks. However, this is computationally impossible for all but the smallest problems. In Example 1, even taking advantage of the isomorphisms defined by the symmetry of factors, of blocks, of runs within blocks and of replicated treatments, there

are still approximately  $10^{16}$  possible designs. Even if this becomes computationally feasible in the future, slightly bigger problems produce numbers of candidate designs many orders of magnitude greater still. Nonetheless, for small experiments, especially with small block sizes, this approach is feasible and should be used.

For most experiments, however, a search procedure is still needed. Most of these are based on interchange algorithms, which take a starting design and (usually) systematically interchange treatments between blocks to try to find the best design. The simplest version is the *first ascent interchange algorithm*, which works as follows for a criterion  $C$  that must be maximized.

- (1) Calculate  $C$  for the starting design, which becomes the current design.
- (2) Let  $i = 1, j = 1, k = 2$  and  $l = 1$ .
- (3) Interchange the treatment in run  $j$  of block  $i$  with the treatment in run  $l$  of block  $k$  and calculate the criterion value  $C_{new}$  for the updated design. If  $C_{new} > C$ , then retain the new design as the current design and let  $C = C_{new}$ . Otherwise, reverse the interchange.
- (4) If  $l < n_b$ , let  $l = l + 1$  and go to step (3). Otherwise, let  $l = 1$ .
- (5) If  $j < n_b$ , let  $j = j + 1$  and go to step (3). Otherwise, let  $j = 1$ .
- (6) If  $k < \frac{n}{n_b}$ , let  $k = k + 1$  and go to step (3).
- (7) If  $i < \frac{n}{n_b} - 1$ , let  $i = i + 1$ , then let  $k = i + 1$  and go to step (3).
- (8) If the design has changed since step (2), go to step (2). Otherwise, the current design is the final design.

It is most common to choose a starting design at random, since a “good” starting design is often a local optimum, which does not allow the algorithm to improve it, but not a global optimum. Usually the algorithm is run several times, called *tries*, with a different randomly chosen starting design each time.

Several attempts to improve the basic algorithm can be made, such as the following.

- A *steepest ascent* algorithm can be used, in which all interchanges are checked before making the one which gives the greatest improvement.
- Use can be made of *simulated annealing*, in which an interchange which does not improve the design is accepted with a small probability, or other stochastic optimization techniques such as genetic algorithms.
- Interchanges can be restricted to being between blocks which seem to need improved, for example by using the *basic scores* (Trinca and Gilmour<sup>29</sup>).

Simulated annealing has the advantage that one try is much more likely to lead to the optimal design, but the disadvantage that each try takes a much longer time to run. The use of basic scores allows each try to run much more quickly, but has the disadvantage that each try is less likely to find the optimal design.

If the objective is to find the best design possible in a fixed time, it seems that different algorithms are better for different problems and it is difficult to predict which will work best in any situation. We usually prefer the first ascent algorithm, because of its simplicity, generality and flexibility. Unlike simulated annealing and the use of basic scores, there are no tuning constants to be set, so the time saved can be devoted to searching. Another advantage is that, if the  $M_w$  criterion is being used, designs can be generated quickly with a number of weight patterns to see what pattern of efficiency factors result. This is more difficult with simulated annealing, where each try takes considerably longer.

### **Example 1 (cont'd)**

The first ascent algorithm was used to arrange the design obtained in Section 4.1, namely  $S_4 + 2S_1 + 4S_0$ , in 6 blocks of size 6. Using weights  $w_i = 2$  for each linear effect and  $w_i = 1$  for each quadratic and interaction effect, with 50 tries, the design in Table 8 was obtained. This design has  $M_w = 94.46\%$  and the efficiency factors for each parameter are shown in Table 9. The pattern of efficiency factors was considered acceptable and this design was used in the experiment.

### **5.2. Crossed Blocking**

It might be reasonable to use two crossed blocking factors. In Example 1, the blocks corresponded to days. If there were also expected to be systematic differences between the times of the day, a second blocking factor could be used and a row and column design obtained. In this example, if the days were split into 6 times a  $6 \times 6$  row-column design with structure Days\*Times would be required. In fact, it is straightforward to extend the methods of Section 5.1 to this case.

The model will be

$$\mathbf{Y} = \mathbf{R}\boldsymbol{\alpha}_r + \mathbf{C}\boldsymbol{\alpha}_c + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where the columns of  $\mathbf{R}$  are row indicators and the columns of  $\mathbf{C}$  are column indicators. The  $D_s$  criterion is defined as for block designs, i.e. maximise

Table 8. Blocked design for Example 1.

Block	Factors			
	$X_1$	$X_2$	$X_3$	$X_4$
1	-1	-1	-1	-1
	1	-1	1	-1
	-1	0	0	0
	0	1	0	0
	0	0	-1	0
	0	0	0	1
2	-1	-1	-1	1
	1	-1	1	1
	1	0	0	0
	0	1	0	0
	0	0	1	0
	0	0	0	-1
3	-1	-1	1	-1
	-1	1	-1	1
	1	-1	-1	1
	1	1	1	-1
	0	0	0	0
	0	0	0	0
4	-1	-1	1	1
	-1	1	1	-1
	1	-1	-1	-1
	1	1	-1	1
	0	0	0	0
	0	0	0	0
5	-1	1	-1	-1
	-1	1	1	1
	1	0	0	0
	0	-1	0	0
	0	0	1	0
	0	0	0	-1
6	1	1	-1	-1
	1	1	1	1
	-1	0	0	0
	0	-1	0	0
	0	0	-1	0
	0	0	0	1

$|[\mathbf{R} : \mathbf{C} : \mathbf{X}]' [\mathbf{R} : \mathbf{C} : \mathbf{X}]|$ . For  $M_w$ , define the efficiency factor for  $\hat{\beta}_i$  to be

$$E_i = \frac{V(\hat{\beta}_i) / \sigma^2}{V_{rc}(\hat{\beta}_i) / \sigma_{rc}^2} \times 100\%,$$

where the subscript  $rc$  indicates that the quantity is from the row-column design, the absence of this subscript denoting the unblocked design. Then

Table 9. Efficiency factors for the design in Table 8.

Parameter	Efficiency Factor
$\beta_1$	96.23
$\beta_2$	96.30
$\beta_3$	96.36
$\beta_4$	96.23
$\beta_{11}$	99.72
$\beta_{22}$	99.72
$\beta_{33}$	99.72
$\beta_{44}$	99.72
$\beta_{12}$	91.23
$\beta_{13}$	74.36
$\beta_{23}$	100.00
$\beta_{14}$	100.00
$\beta_{24}$	91.23
$\beta_{34}$	74.36

the weighted mean efficiency factor criterion is

$$M_w = \frac{\sum_{i=1}^p w_i E_i}{\sum_{i=1}^p w_i},$$

as before.

The interchange algorithm works as before, except that interchanges between all pairs of runs, not just those in different blocks, must now be considered, since any pair of runs are in either different rows or different columns. Full details of this algorithm and examples were given by Gilmour and Trinca<sup>13</sup>. It is equally straightforward to extend the criteria and algorithms to any number of crossed blocking factors, although more than two are rare in practice.

### 5.3. Nested Blocking

The need to use nested block structures is rather rare in experiments in bioprocessing unlike, for example, in agricultural field and greenhouse experiments, where it is rather common. However, it might arise occasionally and a discussion of nested structures is a useful pre-requisite for the discussion of multi-stratum structures in Section 6.

Nested blocks might arise in an experiment like Example 1 if the days themselves could be grouped into superblocks, e.g. if two days were used in each of three weeks, and most of the differences between days were expected to be between weeks. Then the structure would be Weeks/Days/Runs, with

3 weeks, 2 days in each week and 6 runs in each day. The most important decision is to choose a good block design, exactly as in Section 5.1. However, since the linear mixed model will be used, the variances of parameter estimates also depend on the arrangement of blocks to superblocks. This is particularly important if the superblocks variance component is large and the blocks variance component is small.

A special case of the algorithm proposed by Trinca and Gilmour<sup>30</sup> can be used here. Use the model with superblocks, but not blocks, i.e.

$$\mathbf{Y} = \mathbf{S}\boldsymbol{\gamma} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where  $\mathbf{S}$  is the matrix of indicator variables for superblocks and  $\boldsymbol{\gamma}$  is the vector of superblock parameters, to define the weighted mean efficiency factor criterion  $M_w^s$ , as before. With the block design fixed, the arrangement of blocks to superblocks is chosen to maximize  $M_w^s$ . This might be done using an interchange algorithm to interchange blocks between superblocks. However in some cases a complete search is possible. For example, in the modification of Example 1 described above, there are only 15 different allocations of blocks to superblocks possible.

Note that, using this algorithm, no damage is done to the block design, i.e. we will always have an optimal block design ignoring superblocks. Since the only cost in having superblocks is to reduce the residual degrees of freedom in the blocks stratum, perhaps they should be used more often if there is an expected pattern in the block means.

#### 5.4. General Structures

The methods of the last three subsections can be combined to deal with any simple orthogonal block structure that arises. For example, if we have a row-column structure with more than one run within each row×column combination, i.e. (Rows\*Columns)/Runs, we can choose an optimal block design for the blocks defined by the row×column combinations, then try to improve the design with respect to the model without block effects by interchanging blocks between rows and columns. It is quite straightforward to modify the algorithms for particular structures, but more difficult to write an all-encompassing algorithm which will cover every case.

The algorithms described in this section are not limited to simple orthogonal block structures and can easily be adapted for unequally sized blocks, for example. Producing designs in such cases is not difficult, but believing the assumptions which are required to make them useful might

be. Since there is no randomization justification for the model, the assumption that the runs are a sample from a population having responses which are independent with constant variance is necessary. The assumption of constant variance for blocks of unequal sizes might be difficult to accept.

## 6. Multi-Stratum Designs

In many experiments, including many in bioprocessing such as Example 3, some factors are more difficult to change than others. If the randomization is restricted so that these factors are varied less often than others, a split-plot, or more generally a multi-stratum, structure arises. In Chapter 5 Ganju and Lucas discuss inadvertent split-plots. Here we consider how to design the experiment in advance, allowing for split-plots. In the general case, there are  $s$  strata, containing nested blocking factors  $U_1 / \cdots / U_s$ , with different factors randomized in different strata. In Example 3 we have Mainunits/Runs, with Pressure randomized to main units and the other four factors randomized to runs.

With response surface treatment designs it is usually impossible to use the standard balanced split-plot designs, so an algorithmic approach is required. Trinca and Gilmour<sup>30</sup> described a general algorithm for multi-stratum structures, which we will summarize here. Goos and Vandebroek<sup>16</sup> described an algorithm for constructing  $D$ -optimal split-plot designs when the variance components are known. Again the algorithm of Trinca and Gilmour produces designs which are optimal as the ratios of variance components tend to infinity and so are robust to the least favorable situations, whereas the  $D$ -optimal designs are sensitive to the prior guesses used for the variance components and, indeed, to the estimated variance components that arise.

The general idea of Trinca and Gilmour's multi-stratum algorithm is that a design is chosen separately for the factors randomized to each stratum, using the methods of Sections 4 and 5.1, and then they are combined in a way that takes account of the multi-stratum structure. Assume that stratum  $i$  has  $n_i$  experimental units within each unit of stratum  $i - 1$  and that there are  $f_i$  factors randomized to stratum  $i$ . Then, for stratum  $i$ , do the following:

- (1) use the methods of Section 4 to choose a treatment design for  $f_i$  factors in  $n_1 \times \cdots \times n_i$  runs;
- (2) treating  $U_{i-1}$  as blocks, use the methods of Section 5.1 to arrange the treatment design into  $n_1 \times \cdots \times n_{i-1}$  blocks of size  $n_i$ .

Having obtained designs for each stage, we combine them in a way that retains the structures already obtained while being as efficient as possible for estimating the second order model. In a two-stratum structure, assume that we have chosen a treatment design for the  $f_1$  factors in main units, with no blocking, and a design for the  $f_2$  factors in runs, with main units as blocks. Then we choose the final design by allocating the blocks of the second component design to the runs of the first component design to optimise a criterion such as weighted- $A$ - or  $D_s$ -efficiency for the second order model with  $f_1 + f_2$  factors. This can be done either by using a complete search or using an interchange algorithm. It is this step that ensures that the full treatment set chosen takes account of the multi-stratum structure of the experiment.

If there are more than two strata, the above step is repeated, starting in the highest stratum, to combine each stratum with those above it. If there are strata without treatment factors, e.g. Blocks/Mainunits/Runs, an additional step can be run using the methods of Section 5.3. In this case, having obtained the combined design, we interchange blocks from the Runs stratum between main units having the same treatment design for the factors applied to main units, to see if a better design can be obtained with regard to Blocks if the main units variance component is zero. Clearly, there will often be limited interchanges that can be done at this stage, e.g. if there are few replicates in the treatment design for main unit factors. Further details of this algorithm were given by Trinca and Gilmour<sup>30</sup>.

### **Example 3 (cont'd)**

The experimenters considered it desirable to set the pressure only once for each level, whereas statistically it would be desirable to set it for each of the 30 runs. The best practical solution that could be managed was to set the pressure six times, with other factors being set five times within each of the settings of pressure. Thus the structure is Mainunits/Runs, with  $n_1 = 6$ ,  $n_2 = 5$ ,  $f_1 = 1$  and  $f_2 = 4$ .

We start by choosing a design for each stratum separately. For main units, the only sensible design is to have two runs at each level of pressure. There is no higher stratum, so they will be completely randomized. For the runs, the treatment design  $S_4 + S_1 + 6S_0$ , i.e. a central composite design, was chosen. This was then arranged into 6 blocks of 5 using the algorithm described in Section 5.1.2 using the  $M_w$  criterion with equal weights.

The final stage is to combine these designs, i.e. to decide which block in the lower stratum becomes which main unit. Since there are only 90

candidate designs, a complete search is straightforward. This was done and the design with the highest weighted  $A$ -efficiency is shown in Table 10. This was the design actually used in the experiment.

Table 10. Design for Example 3.

Main Unit	Factors				
	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$
1	-1	-1	-1	-1	1
	-1	-1	1	1	-1
	-1	1	-1	1	-1
	-1	1	1	-1	1
	-1	0	0	0	0
2	-1	-1	-1	1	-1
	-1	-1	1	-1	-1
	-1	1	-1	-1	1
	-1	1	1	1	1
	-1	0	0	0	0
3	0	-1	-1	-1	-1
	0	-1	1	1	1
	0	1	-1	1	1
	0	1	1	-1	-1
	0	0	0	0	0
4	0	-1	-1	1	1
	0	-1	1	-1	1
	0	1	-1	-1	-1
	0	1	1	1	-1
	0	0	0	0	0
5	1	-1	0	0	0
	1	0	1	0	0
	1	0	0	1	0
	1	0	0	0	1
	1	0	0	0	0
6	1	1	0	0	0
	1	0	-1	0	0
	1	0	0	-1	0
	1	0	0	0	-1
	1	0	0	0	0

## 7. Data Analysis

The data arising from the designs described here can be analysed using standard methods. Linear mixed models, with random effects for each blocking factor and the polynomial response surface model parameters as fixed effects, are appropriate for all of these structures. Residual maximum likelihood (REML) has become the standard method for fitting these models

and is available in statistical packages such as R, Splus, GenStat and SAS. For block designs, the model with fixed block effects will give similar estimates to the mixed model, but Gilmour and Trinca<sup>12</sup> showed that there is some benefit to be gained from using the mixed model. Of course, for multi-stratum structures, the mixed model is essential.

Having obtained parameter estimates from the mixed model, these can be interpreted in the usual way. For example, surface and contour plots can be drawn, canonical analysis can be performed and the location of the optimal combination of factor levels can be estimated. The point estimates can be interpreted in the usual way, although it has to be recognised that predictions are subject to variation between blocks as well as between runs within blocks. See Gilmour and Trinca<sup>12</sup> for more discussion of prediction from blocked response surface designs.

## 8. Final Comments

The methods outlined in this chapter allow the integration of the best of response surface methodology with the best of the classical design principles of randomization, replication and blocking, for experiments on processes with high variation. The design methodologies outlined here comprise a complete solution to the problem of choosing designs for these experiments. The analysis of data is standard. We hope that more statisticians and experimenters will employ these methods and find them useful in practice.

There might be scope for exploring Bayesian methods for these problems. Bayesian analysis could be used to include prior information about treatment effects and variance components. The latter could be particularly useful for small multi-stratum designs like Example 3. If this analysis seems useful, then Bayesian designs could be explored to see whether they are substantially different from those chosen here. We have not explored Bayesian methods in this context.

## References

1. Atkinson, A. C. and Donev, A. N. (1989) The construction of exact  $D$ -optimum experimental designs with application to blocking response surface designs. *Biometrika*, **76**, 515–526.
2. Bailey, R. A. (1981) A unified approach to design of experiments. *Journal of the Royal Statistical Society, Series A*, **144**, 214–223.
3. Bailey, R. A. (1998) Statistics and mathematics: the appropriate use of mathematics within statistics (with discussion). *Journal of the Royal Statistical Society, Series D*, **47**, 261–271; discussion 273–290.

4. Box, G. E. P. and Draper, N. R. (1987) *Empirical Model Building and Response Surfaces*. New York: Wiley.
5. Box, G. E. P. and Wilson, K. B. (1951) On the experimental attainment of optimum conditions (with discussion). *Journal of the Royal Statistical Society, Series B*, **13**, 1–45.
6. Coleman, D. E. and Montgomery, D. C. (1993) A systematic approach to planning for a designed industrial experiment (with discussion). *Technometrics*, **35**, 1–27.
7. Cook, R. D. and Nachtsheim, C. J. (1989) Computer-aided blocking of factorial and response-surface designs. *Technometrics*, **31**, 339–346.
8. Davis, T. P. and Draper, N. R. (1995) A note on remnant three-level second order designs. *Technical Report No. 954*, Department of Statistics, University of Wisconsin, Madison.
9. Draper, N. R. and John, J. A. (1988) Response-surface designs for quantitative and qualitative variables. *Technometrics*, **30**, 423–428.
10. Edmondson, R. N. (1994) Fractional factorial designs for factors with a prime number of quantitative levels. *Journal of the Royal Statistical Society, Series B*, **56**, 611–622.
11. Gilmour, S. G. (2005) Response surface designs for experiments in bioprocessing. To appear in *Biometrics*.
12. Gilmour, S. G. and Trinca, L. A. (2000) Some practical advice on polynomial regression analysis from blocked response surface designs. *Communications in Statistics - Theory and Methods*, **29**, 2157–2180.
13. Gilmour, S. G. and Trinca, L. A. (2003) Row-column response surface designs. *Journal of Quality Technology*, **35**, 184–193.
14. Gilmour, S. G., Pardo, J. M., Trinca, L. A., Nirajan, K. and Mottram, D. S. (2000) A split-unit response surface design for improving aroma retention in freeze dried coffee. *Proceedings of the 6th European Conference on Food-Industry and Statistics*, Pau, France, 18.0–18.9.
15. Goos, P. and Vandebroek, M. (2001) D-optimal response surface designs in the presence of random block effects. *Computational Statistics and Data Analysis*, **37**, 433–453.
16. Goos, P. and Vandebroek, M. (2001) Optimal split-plot designs. *Journal of Quality Technology*, **33**, 436–450.
17. Hoke, A. T. (1974) Economical second order designs based on irregular fractions of the  $3^n$  factorial. *Technometrics*, **16**, 375–384.
18. John, J. A. and Williams, E. R. (1995) *Cyclic and Computer Generated Designs*, 2nd edition. London: Chapman & Hall.
19. Khuri, A. I. and Cornell, J. A. (1996) *Response Surfaces*, 2nd edition. New York: Dekker.
20. Mead, R. (1988) *The Design of Experiments*. Cambridge: Cambridge University Press.
21. Mead, R. (1990) The nonorthogonal design of experiments (with discussion). *Journal of the Royal Statistical Society, Series A*, **153**, 151–201.
22. Montgomery, D. C. (1991) *Design and Analysis of Experiments*, 3rd edition. New York: Wiley.

23. Myers, R. H. and Montgomery, D. C. (2002) *Response Surface Methodology*, 2nd edition. New York: Wiley.
24. Myers, R. H., Montgomery, D. C., Vining, G. G., Borror, C. M. and Kowalski, S. M. (2004) Response surface methodology: a retrospective and literature survey. *Journal of Quality Technology*, **36**, 53–77.
25. Nelder, J. A. (1965) The analysis of randomized experiments with orthogonal block structure I: block structure and the null analysis of variance. *Journal of the Royal Society of London, Series A*, **283**, 147–162.
26. Nelder, J. A. (1977) A reformulation of linear models (with discussion). *Journal of the Royal Statistical Society, Series A*, **140**, 48–76.
27. Nguyen, N. K. (2001) Cutting experimental designs into blocks. *Australian and New Zealand Journal of Statistics*, **43**, 367–374.
28. Rosenthal, A., Pyle, D. L., Nirajan, K., Gilmour, S. and Trinca, L. (2001) Combined effect of operational variables and enzyme activity on aqueous enzymatic extraction of oil and protein from soybean. *Enzyme and Microbial Technology*, **28**, 499–509.
29. Trinca, L. A. and Gilmour, S. G. (2000) An algorithm for arranging response surface designs in small blocks. *Computational Statistics and Data Analysis*, **33**, 25–43; Erratum (2002), **40**, 475.
30. Trinca, L. A. and Gilmour, S. G. (2001) Multi-stratum response surface designs. *Technometrics*, **43**, 25–33.
31. Varnalis, A. I., Brennan, J. G., MacDougall, D. B. and Gilmour, S. G. (2004) Optimisation of high temperature puffing of potato cubes using response surface methodology. *Journal of Food Engineering*, **61**, 153–163.
32. Wu, C. F. J. and Ding, Y. (1998) Construction of response surface designs for qualitative and quantitative factors. *Journal of Statistical Planning and Inference*, **71**, 331–348.

## CHAPTER 3

### RANDOM RUN ORDER, RANDOMIZATION AND INADVERTENT SPLIT-PLOTS IN RESPONSE SURFACE EXPERIMENTS

Jitendra Ganju

*Chiron Corporation  
4560 Horton Street, M/S U-140  
Emeryville, CA 94608  
E-mail: jitendra\_ganju@chiron.com*

James M. Lucas

*J.M. Lucas and Associates  
5120 New Kent Road  
Wilmington, DE 19808*

The meaning and implementation of the physical act of randomization has been inadequate for a certain class of experiments. Factorial and response surface experiments that require resetting of factors are experiments that belong to this class. Such experiments use a random run order but the factor levels are not independently reset if successive runs require the same level of that factor. Proper randomization also requires independent resetting of factor levels for every run. We discuss the consequences of not independently resetting factor levels. The disadvantages are that normal theory-based tests are misleading and the diagnostic power to detect the restriction in randomization is poor to nil for experiments that are common in practice. The advantage, when precision of estimates is not of paramount concern, is that not resetting is faster and cheaper. We discuss the properties of experiments that do not reset factors, discuss when they should be used and show where they can be improved upon by using split-plot experiments. We describe some important unanswered questions and give examples of successful split-plot experiments.

## 1. Introduction

We give two examples to illustrate that the meaning of randomization provided in the literature is inadequate.

**Example 1:** Joiner and Campbell<sup>1</sup> discuss randomizing the two levels (A and B) of one factor, the height of a burner. For an eight-run experiment they say that a randomized experiment that results in the sequence AAAABBAA has a deficiency because it has only one set-up of burner height.

**Example 2:** Mason, Gunst and Hess<sup>2</sup> discuss a half-fraction of an experiment with 6 factors each having 2 levels. They provide a randomized sequence of the 32 treatment combinations and note that for one of the factors there are 20 resettings.

If the experiments above were to be properly randomized, then burner height in Example 1 would have to be set-up independently 8 times and in Example 2, all 6 factors would need to be set-up independently 32 times.

See Ganju and Lucas<sup>3,4</sup> for more details of these 2 examples and for other examples from the literature. We also recommend reviewing the numerous books on experimental design; note that the emphasis is on selecting a run order at random, not on resetting factor levels for each run. An exception to this is the book by Box, Hunter and Hunter<sup>5</sup> who emphasize resetting to obtain "a genuine run replicate." They say "this problem of wrongly assessing experimental errors...has led...to gross underestimates..." A large fraction of industrial experiments are random run order experiments. For an indication of the prevalence of randomizing the run order but not independently resetting, the second author asked the attendees at his recent talks three questions: 1) "How many of you are involved with running experiments?" Almost all of the attendees raised their hands. (Over 500 attendees have been asked.) 2) "How many of you randomize to guard against trends or other unexpected events?" Virtually the same set of hands was raised. 3) "If the same level of a factor such as temperature is required on successive runs, how many of you set that factor to a neutral level and then reset it?" Very few hands were raised. This convenience sample, combined with our literature survey, demonstrates that there has been a widespread lack

of understanding about "randomization." Later in this chapter we note that awareness of this topic has been increasing.

When experiments are conducted using only a random run order we refer to them as RRO experiments. These experiments are also "inadvertent split-plots" because the split-plotting was not intentional. A whole plot is formed each time successive runs of a factor are not reset. In general, the whole plots will be unbalanced. When factor levels are not reset the measurements within each whole plot are correlated. These measurements would not be considered correlated if the factor levels were independently reset from run to run.

In some experiments factor levels are hard to change (HTC) as opposed to easy to change (ETC). A HTC factor is one whose levels are not reset for successive runs because it is physically difficult, time consuming or expensive to reset. However, because of the lack of emphasis on the need for resetting, even factors that are ETC may not be reset. It is our experience that HTC factors tend to have a larger variance component associated with not resetting. The situation when there is one HTC or one ETC factor has been studied in detail especially for  $2^K$  experiments (Anbari<sup>6,7</sup>, Anbari and Lucas<sup>8,9</sup>, Ganju and Lucas<sup>3,4,10</sup>, and Ju and Lucas<sup>11</sup>). For  $2^K$  experiments with one HTC or one ETC factor, it is much better to run a blocked (split-plot) experiment instead of an RRO experiment. The experiment with split-plot blocking will dominate a randomized experiment (all factors reset each time) and an RRO experiment by being cheaper to run and by having a smaller variance of prediction. Webb, Lucas and Borkowski<sup>12</sup> note that for an experiment with 7 or fewer factors having 3 or more HTC factors, there is no good blocking procedure so an RRO experiment may be cost effective.

In Section 2 we will take an example to describe the consequence of not independently resetting factor levels from run to run. In Section 3 we get into technical matters such as the bias over all randomizations. In Section 4 we show why the diagnostic power to detect the split-plotting effect is so poor. In Section 5 we discuss the properties of RRO experiments. Section 6 discusses cost and efficiencies of experiments and shows that split-plot experiments dominate RRO experiments and randomized experiments on the basis of cost and variance of prediction.

In Section 7 we note that awareness on this topic is increasing and we suggest an area for further research.

## **2. Why Statistical Tests from RRO Experiments are Misleading**

Ganju and Lucas<sup>4</sup> consider a 3-factor (A, B, and C) 2 level (- and +) experiment for a main effects model. Suppose we selected 2 run orders at random from the 8! possible run orders. Call them RRO 1 and RRO 2. These 2 run orders are shown in Table 1. Suppose that factor A is not reset independently from run to run. Then A changes levels 3 times (or is set-up 4 times) and B and C change levels 7 times (or are set-up 8 times) for both RRO 1 and 2. Thus, the physical set-up required for both experiments is the same. Both experiments have formed 4 whole plots because A is not reset for runs 1 and 2, runs 3 and 4, runs 5 and 6 and runs 7 and 8. Let  $\sigma_A$  denote the standard deviation of the error associated with the whole plot formed due to factor A. Let  $\sigma_e$  denote the standard deviation of errors from other sources.

For RROs 1 and 2 we generated normally distributed errors with  $\lambda = \sigma_A/\sigma_e = 1$  and 3 and with no active effects. A main effects model was fit to each of the 10,000 datasets generated so the error degrees of freedom (df) = 4. The model was not adjusted for split-plotting because experimenters believe RRO 1 and RRO 2 to be randomized experiments. The Type I error rate was calculated as the percent of time the *t*-statistic exceed the 5% level. This is the percent of time the absolute value of the t-test exceeded  $t_{.975,4} = 2.78$ . Table 1 shows the empirical Type I error rates for RRO 1 and RRO 2.

Table 1. The Type I error rate for two RROs. Normally distributed split-plot and whole plot errors generated for the simulation when  $\lambda = \sigma_A/\sigma_e = 1$  and 3. Simulation size is 10,000.

RRO1	1	2	3	4	5	6	7	8	Type I Rate %	
RRO2	6	2	3	4	5	1	7	8	RRO1 $\lambda = 1, 3$	RRO2 $\lambda = 1, 3$
A	-	-	+	+	-	-	+	+	11.3, 19.6	12.5, 29.6
B	-	-	-	-	+	+	+	+	5.4, 2.8	12.8, 29.7
C	+	-	+	-	-	+	-	+	2.4, 0.5	3, 1.2

Table 2. EMS for RROs 1 and 2. The  $Q$ 's denote the quadratic form associated with main effects.

	RRO 1	RRO 2
Intercept	$2\sigma_A^2 + \sigma_e^2 + Q_1$	$2\sigma_A^2 + \sigma_e^2 + Q_1$
A	$2\sigma_A^2 + \sigma_e^2 + Q_2$	$2\sigma_A^2 + \sigma_e^2 + Q_2$
B	$\sigma_A^2 + \sigma_e^2 + Q_3$	$2\sigma_A^2 + \sigma_e^2 + Q_3$
C	$\sigma_e^2 + Q_4$	$\sigma_e^2 + Q_4$
AB	$\sigma_A^2 + \sigma_e^2$	$2\sigma_A^2 + \sigma_e^2$
AC	$\sigma_e^2$	$\sigma_e^2$
BC	$\sigma_A^2 + \sigma_e^2$	$\sigma_e^2$
ABC	$\sigma_A^2 + \sigma_e^2$	$\sigma_e^2$
Error	$0.75\sigma_A^2 + \sigma_e^2$	$0.5\sigma_A^2 + \sigma_e^2$

For null main effects, the Type I error rates for A, B and C should be 5%. However, because factor A is not independently reset, the Type I error rates for all factors is either less than or greater than 5%. The Type I error rate for factor A is greater than 5% whereas for factors B and C, the Type I error rates are less than or greater than 5%. Note also that even though RRO 1 and RRO 2 required the same physical set-up, the Type I error rate for B is materially different for the two run orders.

In Table 2 we show the expected mean squares (EMS) for an ordinary least squares (OLS) analysis of RROs 1 and 2 when A is not reset. The EMS for the error term is the average of the EMS of the 4 interaction terms not included in the main effects model. This table shows what the  $F$ -test is actually testing. For example, with RRO 2 the  $F$ -test for factor B is  $2\sigma_A^2 + \sigma_e^2 + Q_3 / 0.5\sigma_A^2 + \sigma_e^2$  whereas for factor C is  $\sigma_e^2 + Q_4 / 0.5\sigma_A^2 + \sigma_e^2$  (where the  $Q$ 's denote quadratic forms associated with the main effects).

The EMS explain why the Type I error rate for factor A will for all run orders always exceed 5%. But will the EMS explain why for factors B or C the Type I error rates may be less than or greater than 5%? The EMS explain RRO 2 completely because it has the blocking structure

$I = A = B = AB$ . Table 2 shows that the EMS for A and B are identical and Table 1 shows that the error rates are also the same. The EMS, however, do not explain RRO 1. Table 2 indicates that the  $F$ -test for factor B is  $\sigma_A^2 + \sigma_e^2 + Q_3 / 0.75\sigma_A^2 + \sigma_e^2$ ; this indicates that as  $\lambda$  increases, the Type I error rate should increase. However, Table 1 shows that the Type I error rate for factor B decreases from 5.4% to 2.8% as  $\lambda$  increases from 1 to 3.

To explain this anomaly, consider the extreme case of no split-plot error. Let the whole plot error caused by factor A for runs 1 – 8 be denoted as  $\varepsilon_1, \varepsilon_1, \varepsilon_2, \varepsilon_2, \varepsilon_3, \varepsilon_3, \varepsilon_4$  and  $\varepsilon_4$ . Runs 1 and 2, 3 and 4, 5 and 6, and 7 and 8 have the same realization of error because factor A with runs  $--++--++$  is not reset. Because of the null B effect, the B effect can be estimated as  $2(\varepsilon_4 - \varepsilon_2)$  which also equals the estimate of the effect for the AB interaction. The AB term, however, is contained in the error for factor B; thus the numerator and denominator of the  $F$ -test for factor B are correlated. Because of this correlation,  $F$  will always be  $\leq 4$ . This is less than the 7.71 critical value for a 5% level test with 4 df. This means that for RRO 1, when there is only whole plot error, the critical value will never be exceeded.

We will explain in the next section why over *all* randomizations the factor that is not reset independently will always exceed the nominal error rate. The factors that are independently reset will average out and preserve approximately the nominal error rate.

### 3. Examining the Split-Plotting Effect over All Randomizations

In Ganju and Lucas<sup>10</sup> we examine what happens over all randomizations; we augment the fixed effects model by adding a term for the whole plots that are randomly formed. Then we derive an expression for the residual error variance as a linear combination of the split-plot errors and the whole plot errors.

To demonstrate how the model is augmented we take an example of a  $2^4$  factorial design with factors A, B, C and D. Suppose A is not independently reset and B, C and D are independently reset each time. Suppose that the model that will be fit to the data is a main effects and a 2-factor interaction model written in notation as

$$y_{16 \times 1} = X_{16 \times 11} \beta_{11 \times 1} + \varepsilon_{16 \times 1} \quad (1)$$

where the matrix  $X$  consists of known values and  $\beta$  is a vector of fixed effects. As is typical when randomization restrictions are not recognized, it is assumed that the error terms are independently and identically distributed with common variance  $\sigma^2$ .

Now suppose that the run order selected at random gave the following sequence for A: 1 1 -1 -1 -1 -1 1 1 1 1 -1 -1 1 1 -1 -1. Then a model that recognizes that A has not been independently reset is given by

$$y_{16 \times 1} = X_{16 \times 11} \beta_{11 \times 1} + Z_{16 \times 6} \gamma_{6 \times 1} + \xi_{16 \times 1} \quad (2)$$

where the matrix  $Z$  is given by  $diag(J_2, J_4, J_4, J_2, J_2, J_2)$  where  $J_n$  represents a unit vector of length  $n$  and  $diag(J_n)$  means a diagonal matrix with the  $i^{th}$  diagonal element equal to  $J_n$ .  $\gamma$  is a random vector representing the whole plot error. Let  $\sigma_A^2$  denote the whole plot error due to factor A and  $\sigma_e^2$  denote the error from other sources. We refer to (1) as F (for fixed effects) and (2) as M (for mixed effects).

Although the above description was for the  $2^4$  factorial design we continue using the above notation more generally for response surface experiments. For any given run order  $\sigma^2 = \sigma_e^2 + t\sigma_A^2$  where  $t = trace[HZZ']/\{N - r(X)\}$  and where N denotes the number of treatment combinations,  $r(\cdot)$  denotes the rank of the matrix contained in the argument and  $H = I - X(X'X)^{-1}X'$  (the details are in Ganju and Lucas<sup>10</sup>). Note that when factor A is reset independently each time  $t = 1$  because  $trace[HZZ'] = N - r(X)$  giving  $\sigma^2 = \sigma_e^2 + \sigma_A^2$ .

Over all random run orders then

$$\bar{\sigma}^2 = \sigma_e^2 + \bar{t}\sigma_A^2. \quad (3)$$

For complete  $L^K$  factorial designs used to estimate all  $L - 1$  terms for the factor not reset (i.e., factor A), Ganju and Lucas<sup>10</sup> prove that the average variance in (3) can be written as

$$\bar{\sigma}^2 = \sigma_e^2 + (1 - p)\sigma_A^2 \quad (4)$$

where  $p$  denotes the proportion of times the  $(i, j)$  element equals 1 in the matrix  $ZZ'$  when averaged over all possible run orders. The closed form expression for  $p$  for complete  $L^K$  factorial designs was derived by Ju and Lucas<sup>9</sup>:

$$p = \frac{2}{L^{K-1}(L-1) + 2}. \quad (5)$$

The average variance of the regression coefficients over all run orders is  $\bar{V}(\hat{\beta}) = \bar{\sigma}^2(X'X)^{-1}$ . From (4) this can be written as

$$\bar{V}(\hat{\beta}) = \left\{ \sigma_e^2 + (1-p)\sigma_A^2 \right\} (X'X)^{-1} \quad (6)$$

Since model F does not recognize the restrictions in randomization, the average variance is the same for all coefficients. Under model M we get

$$\begin{aligned} \bar{V}(\hat{\beta}) &= (X'X)^{-1} X' V(y) X (X'X)^{-1} \\ &= (X'X)^{-1} X' (\sigma_e^2 I + \sigma_A^2 \bar{Z}\bar{Z}') X (X'X)^{-1} \\ &= \sigma_e^2 (X'X)^{-1} + \sigma_A^2 (X'X)^{-1} X' \bar{Z}\bar{Z}' X (X'X)^{-1} \end{aligned}$$

Let  $\hat{\beta}$  be partitioned as  $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$  with the terms representing, respectively, the intercept, the factor not reset and the factors reset independently. Similarly, let  $(X_0, X_1, X_2)$  represent a conformable partition for  $X'$ . Then following the proof in Ganju and Lucas<sup>10</sup> we see that

$$\bar{V}(\hat{\beta}_1) = (X_1' X_1)_*^{-1} \left[ \sigma_e^2 + \sigma_A^2 \left\{ 1 + p(L^{K-1} - 1) \right\} \right] \quad (7)$$

and

$$\bar{V}(\hat{\beta}_2) = (X_2' X_2)_*^{-1} \left\{ \sigma_e^2 + (1-p)\sigma_A^2 \right\} \quad (8)$$

where  $(X_1' X_1)_*^{-1}$  and  $(X_2' X_2)_*^{-1}$ , of dimension  $L-1 \times L-1$  and  $L^K - L \times L^K - L$ , respectively, represent inverses of  $(X_1' X_1)^{-1}$  and

$(X_2'X_2)^{-1}$ , in the partitioned  $X'X$  matrix. Comparing (6) with (7) and (8) we see that the average variances are the same for the factors reset independently but the average variance is larger for the  $L-1$  estimated main effects of the factor that is not reset. Even though  $\bar{V}(\hat{\beta}_2)$  is not biased, this finding is of little comfort. As we have shown in Section 2, for any given run order, the Type I error rate is either less than or greater than the nominal rate. Equations (6) and (8) indicate that over all run orders, the error rates average out to the nominal rate.

Ganju and Lucas<sup>10</sup> show that for incomplete factorials the variances are biased even for the factors that are reset independently. They show that for a four-factor Box-Behnken design<sup>13</sup> the variance associated with the linear (quadratic) terms of the factors independently reset is less (more) than the variance obtained when the model ignores the randomization restriction.

#### 4. The Diagnostic Power to Retrospectively Detect the Randomization Restriction

In Ganju and Lucas<sup>3</sup> we examine the diagnostic power to detect the effect of using an RRO and analyzing the data ignoring the correlation between measurements. We first consider the case where the experimenter incorrectly believes the experiment to be properly randomized so ordinary least squares (OLS) is used for analysis. Then we consider the case where knowledge of the randomization restriction is known after the experiment is run and investigate how much better we can do with this additional information.

In the absence of suspecting inadvertent split-plotting, the residuals of model F are examined for any unusual patterns. In this regard it is common to graphically examine the residuals for any patterns and test for the linearity between the normal quantiles and ordered OLS residuals (Johnson and Wichern<sup>14</sup>). We examine this correlation for a complete  $2^5$  factorial experiment under a main effects and 2-factor interaction model. This model leaves 16 of the 32 degrees of freedom (df) for error; this is considered to be sufficiently large for diagnostic purposes (Draper and Smith<sup>15</sup>). We conducted a simulation to examine the correlation between the normalized scores and the OLS residuals for different values of

whole plot to split-plot errors. We present results only for large to extremely large values of  $\lambda$ , where  $\lambda = \sigma_A^2 / \sigma_e^2$ . See Table 3 for the summary statistics;  $\bar{r}$  refers to the average of the 1000 correlations,  $sd(r)$  refers to the standard deviation of  $r$ ,  $\bar{\sigma}^2$  refers to the average of 1000 model F mean square errors (MSEs) and  $sd(\hat{\sigma}^2)$  denotes the standard deviation of the MSE. The critical value of the test for normality at the 0.05 level is 0.964. Regardless of the value of  $\lambda$  in Table 3 the hypothesis that the residuals are normal could not be rejected for more than 95% of the 1000 correlations generated by the simulation. We note that the inflation in MSE observed in Table 3 as  $\lambda$  increases was demonstrated for the general case in Section 2 by (4) with  $p$  equaling 1/9 for the  $2^5$  factorial experiment.

We also visually examined several plots to detect patterns that would not be discerned by the summaries presented in Table 3. For small to moderately large values of  $\lambda$  no unusual patterns were discerned. Only when  $\lambda$  was very large ( $\geq 100$ ) the residuals started to cluster; the clusters were around the  $45^\circ$  line so the correlation did not increase. The simulation results and visual inspection of plots show that if data are collected from an RRO experiment and analyzed by OLS, then usual residual diagnostics will almost certainly not reveal the restriction on randomization.

Table 3. Correlation between OLS residuals and normalized scores based on a simulation study. Main effects and 2-factor interaction model for a  $2^5$  factorial.

$\lambda$	$\bar{r}$	$sd(\bar{r})$	$\bar{\sigma}^2$	$sd(\hat{\sigma}^2)$
10	0.9854519	0.0073	10.07	4.93
100	0.9854678	0.0074	92.64	46.41
1000	0.9855751	0.0076	908.37	454.52

We now consider what we might gain if we retrospectively knew about the factor whose levels were not independently reset. We will demonstrate that in many practical cases it will still not be possible to detect the effect of the restriction in randomization.

From the partitioning of sum of squares we know that the error sum of squares of model F equals the sum of squares for the whole plots and

the error sum of squares of model M. The expected mean squares for the whole plots is  $\sigma_e^2 + m\sigma_A^2$  (see Searle<sup>16</sup>), where

$$m = \frac{\text{trace}\{ZZ'(I - X(X'X)^{-1}X')\}}{r(W) - r(X)} \text{ with } W = X|Z. \quad (9)$$

Assuming normality for both the split- and whole plot errors, we can test the hypothesis  $H_O : \sigma_A^2 = 0$  with the  $F$  test

$$F_{r(W)-r(X), N-r(W)} = \frac{MS(\text{whole plots})}{MSE}$$

where  $N$  denotes the total number of observations. Given the fixed effect terms to be included in the model, the error df of model M,  $N - r(W)$ , or equivalently the whole plot df,  $r(W) - r(X)$ , is determined entirely by the run order selected. For many run orders, the  $F$  test does not exist because the error df of model M,  $N - r(W)$  equals 0. The problem of assessing the ability to detect randomization restrictions can be reduced to determining the probability mass function of  $r(W) - r(X)$  or equivalently  $N - r(W)$ .

Any permutation of a run order within a whole plot, or permutation of whole plots, will yield the same df because the matrix  $Z$  does not change. Run orders that result in the same value of df are called redundant run orders. Consider again a  $2^3$  design that has  $8!$  or 40,320 run orders. Let the 8 treatment combinations be represented as  $a_1, a_2, a_3, a_4, b_1, b_2, b_3, b_4$  where the  $a$ 's represent the lower level of factor A (-) and the  $b$ 's represent the higher level of A (+). The df for the  $F$  test obtained for the run order  $a_1, a_2|b_1, b_2|a_3, a_4|b_3, b_4$  will equal the value obtained for  $b_2, b_1|a_3, a_4|b_3, b_4|a_1, a_2$ . Such run orders are redundant. Run orders that may potentially yield different df are called non-redundant run orders. Ganju and Lucas<sup>3</sup> demonstrate how to obtain the number of non-redundant run orders. In particular they show that the number of non-redundant run orders for  $2^3$  and  $2^4$  designs are, respectively, 197 and 12,694,301.

For each non-redundant run order we calculate the error df of model M (or equivalently, the whole plot df) and the probability of occurrence of each run order. Table 4 shows the probability of error df for the  $2^3$  (main effects model),  $2^4$  (main effects and 2 factor interaction model) and  $2^5$  (main effects and 2 factor interaction model) complete factorial designs. The probability mass function (pmf) is exact for the  $2^3$  design but is estimated for  $2^4$  and  $2^5$  designs because complete enumeration of the non-redundant set of run orders was not practical.

For the  $2^3$  design (main effects model), the error df equal 0 for 30% of run orders. For almost 3% of run orders, the error df equal 4, or equivalently, the whole plot df equals 0. For such run orders the whole plots are completely confounded with the factor A main effect. For the remaining run orders resulting 1, 2 or 3 error df, the whole plot variance component would need to be very large to declare the variance estimate statistically larger than 0. In other words, for 33% of run orders the test for split-plotting does not exist and for the remaining 67% the whole plot variance would need to be very large relative to the split-plot variance.

For the  $2^4$  design (main effects and 2 factor interaction model) the error df equal 0 for 82% of run orders. For the remaining 17% of run orders the power to test for split-plotting caused by factor A is very low.

Table 4. Probability mass functions for the error df of model M. Results are exact for the  $2^3$  design and estimated via simulation (20,000 iterations) for the  $2^4$  and  $2^5$  designs.

$2^3$ design	pmf	$2^4$ design	pmf	$2^5$ design	pmf
0	0.3000	0	0.8186	0	0.3844
1	0.31429	1	0.1323	1	0.1735
2	0.23651	2	0.0403	2	0.1475
3	0.12063	3	0.0070	3	0.1175
4	0.02857	4	0.0080	4	0.0830
		5	0.0010	5	0.0484
				6	0.0282
				7	0.0142
				8	0.0061
				9	0.0022
				10	0.0008
				11	0.0002

For the  $2^5$  design (main effects and 2 factor interaction model) the error df equal 0 for 38% of run orders. The error df equal 1, 2 or 3 approximately 44% of the time; the whole plot variance would need to be very large to detect the split-plotting effect. The probability of 12, 13, 14, 15 and 16 error df were estimated to be almost zero. Model F has half the total number of df available for error (16 out of 32). This is considered adequate for diagnostic purposes. However, we see that for 82% of run orders, there is either no test for split-plotting or the whole plot variance would need to be large relative to the split-plot variance for detecting the split-plot effect.

## 5. The Expected Covariance Matrix for RRO Experiments

Because so many industrial experiments are RRO experiments, it is worthwhile to understand their properties. In this section we describe the covariance matrix of RRO experiments. A use of this matrix is to find the expected variance of prediction for a RRO experiment; this value is used in the comparisons in the next section. This covariance matrix was provided in Webb, Lucas and Borkowski<sup>12</sup> and generalizes the work of Ju and Lucas<sup>11</sup>. Practical implications are discussed in Ju and Lucas<sup>11</sup>, Anbari and Lucas<sup>9</sup> and in the next section.

The expected covariance matrix for an  $L^K$  factorial RRO experiment can be written as

$$V = \sigma_e^2 + \sum_{i=1}^M (1-p)\sigma_{Wi}^2 + p \sum_{i=1}^M \sigma_{Wi}^2 Z_i Z_i'$$

where  $M$  denotes the number of HTC factors ( $1 \leq M \leq K$ ) and  $p$  as defined by (6).

## 6. G-Efficiencies and Cost When an RRO is Used

In this section we discuss efficiencies (using the maximum predicted variance, a value inversely proportional to the G-efficiency) and cost of using an RRO versus randomization for  $2^4$  experiments. We show that a split-plot experiment can dominate both RRO and randomized

Table 5. Comparison of alternatives for the selection of blocks in a  $2^4$  design. Main effects and 2-factor interaction model.

Design	Variance multiplier		Cost multiplier HTC
	ETC ( $\sigma_e^2$ )	HTC ( $\sigma_A^2$ )	
Completely randomized	11	11	16
Random run order	11	12	9*
Blocked design, block size = 4	11	8	4
Blocked design, block size = 8	11	16	2

\*Expected number of resets for the HTC factor

experiments on the basis of variance of prediction and cost. This means that split-plot experiment should be used more often than RRO and randomized experiments.

Table 5 is from Anbari<sup>6,7</sup> and Anbari and Lucas<sup>8,9</sup> and shows the results for one HTC factor (factor A) when the model includes main effects and 2 factor interactions. They show that with one HTC factor, minimum prediction variance is achieved with a split-plot experiment using 4 blocks (with I = A = BCD). This split-plot experiment is both less expensive and more precise than an RRO or a randomized experiment. The split-plot experiment has a maximum variance of prediction of  $(11\sigma_e^2 + 8\sigma_A^2)/16$  while a completely randomized experiment has  $(11\sigma_e^2 + 11\sigma_A^2)/16$ . The split-plot also requires only 4 resets while a completely randomized experiment requires 16. Table 5 also shows that the four-block split-plot experiment dominates a RRO experiment on a cost and variance of prediction basis. The experiment that uses the cost of information will either use four blocks or two blocks. Anbari<sup>6,7</sup> and Anbari and Lucas<sup>8,9</sup> show that the optimum block size is either 4 or 8. It depends on the size of the variance components, the costs of changing the factors and on the model: for a main effects model, block sizes 2, 4 or 8 can be considered. A table of optimum blocking for 3 to 7 factors is in Anbari and Lucas<sup>9</sup>.

Similar domination is achieved by a split-plot when there are two HTC factors. In our  $2^4$  example with a second HTC factor (B) we simply block also block on B within each A block. We obtain 8 B blocks in addition to the 4 A blocks and achieve a split-split-plot with three

error terms. The maximum variance of prediction is  $(11\sigma_e^2 + 8\sigma_A^2 + 10\sigma_B^2)/16$  while the completely randomized design has a larger maximum variance of prediction of  $(11\sigma_e^2 + 11\sigma_A^2 + 11\sigma_B^2)/16$ . Details are shown in Webb<sup>17</sup> and a catalogue of blocking with two HTC factors is in preparation.

Table 5 also indicates how the results will generalize when none of the factors are reset. The expected number of factor resets will drop from  $2^4 = 16$  to  $.5 \times 16 + 1 = 9$  while the maximum expected variance over the experimental region increases from  $11\sigma_e^2 + 11\sum_{i=1}^4 \sigma_{wi}^2$  to  $11\sigma_e^2 + 12\sum_{i=1}^4 \sigma_{wi}^2$ , where  $\sigma_{wi}^2$  denotes the whole plot variance component associated with factor  $i$ . Therefore, if scientific hypothesis testing is not critical, it may be reasonable to use an RRO and not randomize. We note that the maximum variance increases more slowly than the cost drops. Experimenters have had many successes when an RRO is used. While we know of no situations where there was an explicit recommendation to use an RRO, Webb, Lucas, and Borkowski<sup>12</sup> note that with three or more H-T-C factors that there is no good blocking procedure (for seven or fewer factors) so a RRO can be cost effective. While this result may generate some controversy, we do not consider it surprising that there are situations where it is very reasonable to use an RRO.

We believe that it is very important to recognize and understand the difference between RRO and randomized experiments so that the appropriate experimental procedure is used. Tiahrt and Weeks<sup>18</sup> and Tiahrt<sup>19</sup> give run orders for two level factorials that require changes in only one factor from one run to the next. Webb and Lucas<sup>20</sup> are evaluating these procedures in their evaluation of the properties of low cost experimental procedures. When the purpose of the experiment is scientific hypothesis testing, complete randomization or appropriate blocking should be required.

## 7. Remarks

We have discussed RRO experiments that are not reset when successive runs have the same level. This design gives biased test results and after the fact it is difficult to determine whether the experiment was RRO as

opposed to proper randomization. When the run order is available a mixed model may help in extracting more information from the data. We recommend that editors require more information from authors about the experimental protocol (in particular, the randomization and resetting procedure). Ganju and Lucas<sup>21</sup> analyze data from a  $3^2$  factorial experiment (first analyzed by Khuri<sup>22</sup>) and speculate why the experiment may not have been completely randomized. We have shown why a split-plot experiment can dominate RRO and randomized experiments from a variance of prediction and cost basis. In summary, these results indicate that split-plot experiments should be conducted more frequently than RRO experiments.

Awareness of the need for resetting factors is increasing. Trinca and Gilmour<sup>23</sup> give the first published algorithm for designing unbalanced split-plot experiments, and end their discussion of randomization with: "If the randomization turns out to give the order shown, or any other order with the same level of feed position on consecutive days, the experimenter should take apart the equipment and reassemble it with the feed in the same position. This is necessary to ensure that the effect being measured is really due to the feed position and not due to differences in assembly, or degradation after assembly. This issue was discussed in detail by Ju and Lucas (1992) and Ganju and Lucas (1997, 1999)." Arvidsson and Gremyr<sup>24</sup> note that a comprehensive definition of randomization provided by Ganju and Lucas<sup>10</sup> includes RRO and independent resetting of factor levels. Books by Ryan<sup>25</sup> and Goos<sup>26</sup> have drawn attention to the recent work in this area. Mason, Gunst and Hess<sup>2</sup> in their 2<sup>nd</sup> edition cite work by Ju and Lucas<sup>11</sup> and revise the definition of randomization (compared with their 1<sup>st</sup> edition published in 1989).

One area of further work is the development of response surface designs for estimating a quadratic model when there is one (or more) hard to change factor. These would replace Box-Behnken and composite designs when there are HTC and ETC factors. Because of the need to estimate two (or more) variance components as well as all terms in a quadratic model the designs will probably require a few more experimental runs than Box-Behnken or composite designs. Goos and Vandebroek<sup>27</sup> have given some preliminary results but a good set of

response surface designs for experiments with 2 to 7 factors remains to be developed.

## Acknowledgments

The first author thanks Lynn Jacklevich for her expert typesetting help.

## References

1. B. L. Joiner and C. Campbell, *Technometrics*, **18**, 249-259 (1976).
2. R. L. Mason, R. F. Gunst and H. L. Hess, *Statistical Design and Analysis of Experiments with Applications to Engineering and Science*. (New York, Wiley and Wiley-Interscience, 1989 and 2003), Chapter 10.
3. J. Ganju and J. M. Lucas, *J. Stat. Plan. Infer.* **80**, 129-140 (1999).
4. J. Ganju and J. M. Lucas, *J. Stat. Plan. Infer.* **133**, 199-210 (2005).
5. G. E. P. Box, W. G. Hunter and J. S. Hunter, *Statistics for Experimenters* (Wiley, New York, 1978), p. 319.
6. F. T. Anbari, *Experimental Designs for Quality Improvement when there are Hard-to-Change Factors and Easy-to-Change Factors*. (Unpublished Ph.D. Dissertation, Drexel University, Philadelphia, PA 1993.)
7. F. T. Anbari, *ASA Joint Statistical Meetings* [CD] 2004.
8. F. T. Anbari and J. M. Lucas, *ASQC Ann. Qual. Cong. Trans.*, 852-863 (1994).
9. F. T. Anbari and J. M. Lucas, Designing and Running Super Efficient Experiments: Optimum Blocking with One Hard to Change Factor. (2005). *Submitted for publication*.
10. J. Ganju and J. M. Lucas, *Commun Stat A - Theor*, **26**, 47-63 (1997).
11. H. L. Ju and J. M. Lucas, *J. Qual. Technol.* **34**, 411-421 (2002).
12. D. F. Webb DF, J. M. Lucas and J. J. Borkowski, *J. Qual. Technol.* **36**, 1-11 (2004).
13. G. E. P. Box and D. W. Behnken., *Technometrics*, **2**, 455-475 (1960).
14. R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*. (Prentice-Hall, New Jersey 1988).
15. N. R. Draper and H. Smith, *Applied Regression Analysis*. (Wiley, New York 1981).
16. S. R. Searle, *Linear Models*. (Wiley, New York 1971), p. 445
17. D. F. Webb, *Randomizing Run Orders Without Resetting Factors*, (Unpublished Ph.D. Dissertation, Montana State University, 1999).
18. K. J. Tiahrt and D. L. Weeks, *Technometrics*, **12**, 471-483 (1970).
19. K. J. Tiahrt, *J. Qual. Technol.* **3** 120-128 (1971).
20. D. F. Webb and J. M. Lucas (2005). Benefits and Drawbacks to Factorial Experiments with a Minimal Number of Factor Level Changes. *In preparation*.
21. J. Ganju and J. M. Lucas, *Am. Stat.* **54**, 5-11 (2000).
22. A. I. Khuri, *Technometrics*, **34**, 26-37 (1992).
23. L. A. Trinca and S. G. Gilmour, *Technometrics*, **43**, 25-33 (2001).
24. M. Arvidsson and I. Gremyr I, *Qual Reliab Eng Inter*, **19**, 87-99 (2003).

25. T. P. Ryan, *Statistical Methods for Quality Improvement* (Wiley-Interscience 2000).
26. P. Goos, *The Optimal Design of Blocked and Split-Plot Experiments*. (Springer 2002).
27. P. Goos and M. Vandebroek, *J. Qual. Technol.* **36**, 12-26 (2004).

## **CHAPTER 4**

### **STATISTICAL INFERENCE FOR RESPONSE SURFACE OPTIMA**

Dennis K. J. Lin

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

John J. Peterson

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

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

#### **1. Introduction**

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

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

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

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

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

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

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

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

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

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

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

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \varepsilon \quad (3)$$

and

$$\begin{aligned} y = & \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \beta_{11} x_1^2 + \dots + \beta_{kk} x_k^2 \\ & + \beta_{12} x_1 x_2 + \dots + \beta_{k-1,k} x_{k-1} x_k + \varepsilon. \end{aligned} \quad (4)$$

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

$$y = \beta_0 + \mathbf{x}' \boldsymbol{\beta} + \mathbf{x}' \mathbf{B} \mathbf{x} + \varepsilon, \quad (5)$$

where  $\mathbf{x} = (x_1, \dots, x_k)', \boldsymbol{\beta} = (\beta_0, \dots, \beta_k)',$  and  $\mathbf{B}$  is a  $k \times k$  symmetric matrix with diagonal elements equal to  $\beta_{ii}$  ( $i = 1, \dots, k$ ) and off-diagonal elements equal to  $\frac{1}{2}\beta_{ij}$  ( $i < j, j = 2, \dots, k$ ). This would usually be fitted by least squares in the coded form. Substitution of the coding formulas into (3) or (4) enables the  $\alpha$ 's to be expressed in terms of the  $\beta$ 's, if desired.

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

## 2. Statistical Inference for the Path of Steepest Ascent

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

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

Box and Draper<sup>4</sup> show how to obtain a  $100(1-\alpha)\%$  confidence region for the  $\delta_i$ 's as follows. Suppose there are  $k$  design variables. The coefficients  $b_1, b_2, \dots, b_k$  provide estimates of the relative movement of variables along the path. For the first-order model, the true path is defined by parameters  $\beta_1, \beta_2, \dots, \beta_k$ . Furthermore,  $E(b_i) = \beta_i$ , for  $i=1, \dots, k$ , and the true coefficients are proportional to the relative movement along the path, which implies  $\beta_i = \gamma\delta_i$ , where the  $\delta_i$ 's represent the direction cosines of the path. In other words,  $\delta_1, \delta_2, \dots, \delta_k$

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

Another way to view this problem is to consider the confidence region to be derived by inverting a hypothesis test about the direction cosines. If  $\delta_{10}, \delta_{20}, \dots, \delta_{k0}$  are the specified direction cosines, which equal the true direction cosines, then it follows that the null hypothesis,  $H_0: \beta = \gamma\delta_0$ , is true, where  $\delta_0 = (\delta_{10}, \dots, \delta_{k0})'$ . Hence the confidence region for the true  $\delta$  value can be taken as the set of all  $\delta_0$  values such that  $H_0: \beta = \gamma\delta_0$  is not rejected at level  $\alpha$ . That is (assuming a two-level orthogonal design) for those  $\delta_i$ 's which satisfy

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

where  $s_b^2$  is an estimate of  $\sigma_b^2$  and  $v_b$  is the number of degrees of freedom on which this estimate  $s_b^2$  is based. Here,

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

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

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

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

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

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

Carter, Chinchilli, Myers, and Campbell<sup>10</sup> created a confidence interval for an eigenvalue of  $\mathbf{B}$  for rotatable or nonrotatable

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

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

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

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

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

#### 4. Statistical Inference for the Optimal Mean Response

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

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

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

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

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

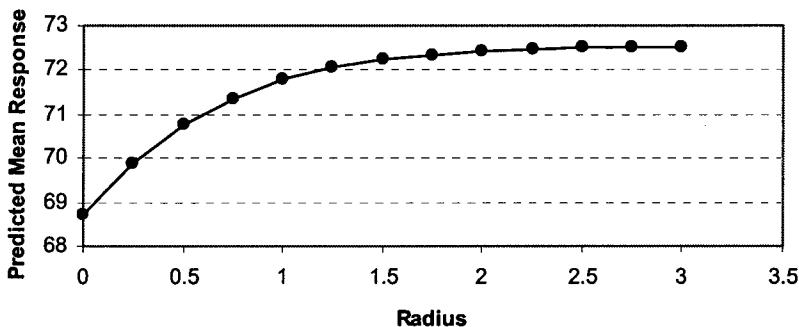


Fig. 1. The optimal mean response plot for the ridge analysis of the five-factor response surface example.

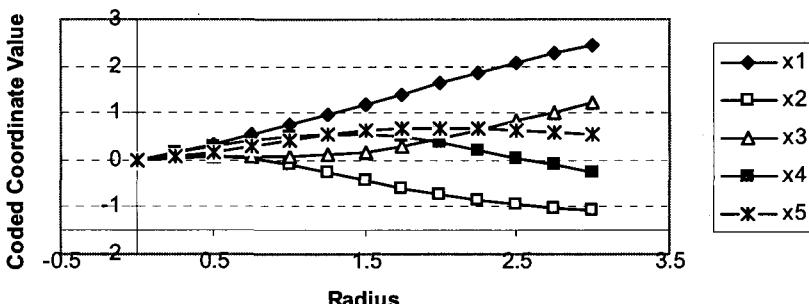


Fig. 2. The optimal coordinate plot for the ridge analysis of the five factor response surface example.

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

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

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

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

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

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

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

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

## 5. A Confidence Region for the Optimizing Factor Levels

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

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

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

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

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

## 6. Multiple Comparison for Response Surface Contrasts

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

$$\delta(\mathbf{x}) = E(Y|\mathbf{x}) - E(Y|\mathbf{x} = \mathbf{0}),$$

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

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

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

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

$$\delta(\mathbf{x}) = E(Y|\mathbf{x}_0) - E(Y|\mathbf{x}),$$

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

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

## 7. Statistical Inference for Multiple-Response Surfaces

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

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

where  $\mathbf{Y}$  is a  $q \times 1$  vector of  $q$  response types,  $\mathbf{B}$  is now a  $q \times p$  matrix of regression coefficients and  $\mathbf{z}(\mathbf{x})$  is a  $p \times 1$  vector-valued function of  $\mathbf{x}$ . The vector  $\boldsymbol{\varepsilon}$  has a multivariate normal distribution with mean vector  $\mathbf{0}$  and variance-covariance matrix,  $\boldsymbol{\Sigma}$ . Typically in response surface analysis, the model in (7) takes the form of  $q$  quadratic models for each mean response.

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

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

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

Cahya, del Castillo, and Peterson<sup>45</sup> have proposed a large sample approach to a confidence region for the optimal factor levels,  $x_0$ , for a desirability function. Here, they take a log or logit transformation of the desirability function to obtain a function,  $g(B;x)$ , of the regression model parameters in (7) and the factor levels,  $x$ . They then approximate  $g(B;x)$  by a first-order Taylor series about  $B$  and apply the methodology of Peterson, Cahya, and del Castillo<sup>29</sup> to obtain the confidence region for  $x_0$ .

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

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

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

more simulation work needs to be done here to examine more examples and to see how small the sample sizes can be and still provide reasonable asymptotic results.

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

$$Pr(Y \in A | x), Pr(D(Y) \geq D^* | x), \text{ or } Pr(Q(Y) \leq Q^* | x), \quad (9)$$

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

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

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

## 8. Recent Results and Some Future Directions for Research

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

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

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

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

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

$$MSE(x) = E_z [m(x, z; \phi) - T]^2, \quad (11)$$

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

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

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

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

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

These ridge analysis procedures involve maximizing nonlinear functions of the factors. For each  $r$ , these functions may have several local optima on the constraint set,  $S = \{x: x'x = r^2\}$ . For  $x = (x_1, \dots, x_k)'$ , the spherical constraint set,  $S$ , can be transformed into a  $(k-1)$  dimensional rectangle by the use of polar coordinates. Here,  $x = rt(a)$ , where  $a$  is a  $(k-1) \times 1$  vector of (radian) angles in the set

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

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

An additional area for future research work deals with the issue of obtaining reliable inferences with nonstandard regression models. Based on the empirical work of Lewis, Montgomery, and Myers<sup>61</sup>, PCD have suggested that their confidence region method for the optimal  $x$ -point ought to work reasonably well for generalized linear models but the simulation work still needs to be done. Likewise, one could ask if some of the methodologies above could be adapted to rank-based regression<sup>62</sup> or nonparametric regression<sup>63</sup>. It would not be surprising to be able to

make such adaptations successful with large sample sizes, but the real challenge would be to see how small the experimental designs could be and still provide useful inferences about response surface optima.

## References

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

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

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

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

## CHAPTER 5

### A SEARCH METHOD FOR THE EXPLORATION OF NEW REGIONS IN ROBUST PARAMETER DESIGN

Guillermo Miró-Quesada

*Bioprocess Research and Development, Lilly Technical Center-North,  
Indianapolis, IN 46221*

Enrique del Castillo

*Department of Industrial and Manufacturing Engineering,  
The Pennsylvania State University, University Park, PA 16802*

An approach for the exploration of new operating regions of a process in which noise factors are present is proposed and studied. The search method considers the curvature in the variance function, following a dual-response approach to process optimization in which a second order model for the process variance is utilized. The method assumes a "Target is best" objective for the process mean response, and proposes to conduct a second order search that minimizes the variance function while keeping feasibility for the mean target constraint. The method is illustrated in detail using simulated experiments.

#### 1. Introduction

When Response Surface Methodology (RSM) is used to improve the robustness of a process or product and the final solution is found at the border of the experimental region, the experimenter will probably like to explore outside the original experimental region. In the initial phase of RSM, this exploration is performed by conducting a search in the steepest ascent/descent direction estimated through a first order model. This classical application of Steepest Ascent assumes no noise factors are present, and instead, assumes all factors are controllable and affect the process mean response only. Furthermore, such search evidently does not consider curvature in the responses.

A useful approach to handle noise factors in a Taguchi-like robust pa-

rameter design (RPD) scenario (i.e., when noise factors may be present) is to model the mean and the variance of the responses separately as two *dual responses*. Box and Jones <sup>4</sup> and Myers <sup>8</sup> show how to obtain these two responses from an unreplicated experiment. In this approach, the response surface for the variance consists of a second degree polynomial with respect to the controllable factors. This contrasts with the first order models used in the usual initial phase of RSM, and points to the need for conducting searches under second order information, since searches based on the steepest ascent/descent direction are well known to be inefficient due to excessive zigzagging <sup>2</sup>. The possibility of zigzagging is specially harmful in RSM since a new factorial design must be run and a first order model fit each time a new search direction is needed.

In this paper, we develop a methodology to explore regions outside an initial experimental region in the context of an RPD study. We do this when the interest is to solve a *Target-is-Best* type of problem. The methodology takes into account the curvature present in the response surface for the variance, since when this surface is obtained through the dual response approach it is inherently quadratic. The methodology keeps the mean on target while minimizing the variance, solving a constrained optimization problem as the search proceeds.

The remainder of this paper is as follows. In section 2 we discuss the usual dual response approach to RPD. Then, in Section 3 we consider search directions for 2nd order models. The particular search direction we recommend for dual response optimization is explained and justified in Section 4. The estimation of this search direction from experimental data is described in Section 5, and the estimation of the variance response in the search direction is explained in Section 6. Section 7 summarizes the procedure in a step-by-step manner. Finally, sections 8 and 9 present two fully worked out examples based on simulated responses that illustrate the use of the proposed methodology.

## 2. An Initial Phase for the Dual Response Approach

To keep an approach equivalent to the usual initial phase in RSM, let us assume, as it is done in traditional RSM, that a first order model constitutes a good local approximation around the current operating conditions. However, in order for the variance to be a function of the controllable factors we need to have significant control-noise interactions. Therefore we will assume the following model:

$$Y|\mathbf{z} = \beta_0 + \mathbf{x}'\boldsymbol{\beta} + \mathbf{z}'\boldsymbol{\gamma} + \mathbf{x}'\boldsymbol{\Delta}\mathbf{z} + \epsilon \quad (1)$$

where  $\mathbf{x}$  is a  $k \times 1$  vector of controllable factors and  $\mathbf{z}$  is an  $r \times 1$  vector of noise factors. This model was also considered by Box and Jones<sup>4</sup>.

The solution to the RPD problem for the *Target-is-Best* case is given by the solution to the following mathematical program:

$$\begin{aligned} \min_{\mathbf{x}} \text{Var}_{\mathbf{z}}(Y) &= (\boldsymbol{\gamma} + \boldsymbol{\Delta}'\mathbf{x})'\boldsymbol{\Sigma}_{\mathbf{z}}(\boldsymbol{\gamma} + \boldsymbol{\Delta}'\mathbf{x}) + \sigma_{\epsilon}^2 \\ \text{subject to : } \\ E_{\mathbf{z}}(Y) &= \beta_0 + \mathbf{x}'\boldsymbol{\beta} = T, \\ \mathbf{x} &\in R \end{aligned} \quad (2)$$

where  $\boldsymbol{\Sigma}_{\mathbf{z}}$  is the covariance matrix of  $\mathbf{z}$ ,  $T$  is the target for the mean response and  $R$  is the region where the model is valid, usually defined by the experimental region. While we derive search directions for this formulation, which require the process to be exactly on target, the implementation of the solution, as will be shown in section 6, allows for some flexibility of the form  $T - L \leq E_{\mathbf{z}}(Y) \leq T + U$ . Other formulations, such as minimizing the mean square deviation from target, are possible (see Conclusions section).

Our goal is to solve problem (2) experimentally. For this purpose, let  $\bar{\mathbf{x}}$  denote its solution. If  $\bar{\mathbf{x}}$  is an interior point of  $R$ , then no further improvement can be obtained by changing the settings of the controllable factors  $\mathbf{x}$ . However, if the solution is on the boundary of  $R$ , then further improvement could be attainable by using settings such that  $\mathbf{x} \notin R$ . In this case, as it is done in the steepest ascent procedure for the mean response, we would like to explore regions not too far away from the boundaries of  $R$  by conducting a line search in a suitable direction.

### 3. A Search Direction for Second Order Functions in RSM

In this section we provide the rationale for a search direction used to optimize a second order model in general, for which the variance in model (2) is an instance. In section 4 we describe how to find such direction in practice for problem (2).

The recommended search direction in the initial phase of RSM is the steepest ascent direction<sup>3,7,10</sup>. This recommendation is made under the assumption that the underlying response surface can be appropriately modelled by a first order model or hyperplane<sup>5</sup>. Since the response surface for the variance in the dual response approach is given by a second degree polynomial, using the steepest ascent search will contradict the assumption of a locally valid first order approximation.

The method of Newton is a procedure that deflects the steepest ascent direction by premultiplying it by the inverse of the Hessian matrix. The second order Taylor approximation of the true response surface  $\eta(\mathbf{x})$  at a given point  $\bar{\mathbf{x}}$  is given by:

$$\eta(\mathbf{x}) \approx \eta(\bar{\mathbf{x}}) + \nabla\eta(\bar{\mathbf{x}})'(\mathbf{x} - \bar{\mathbf{x}}) + \frac{1}{2}(\mathbf{x} - \bar{\mathbf{x}})' \mathbf{H}(\bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}}) \quad (3)$$

where  $\nabla\eta(\bar{\mathbf{x}})$  denotes the  $k \times 1$  vector of first derivatives or gradient and  $\mathbf{H}(\bar{\mathbf{x}})$  the  $k \times k$  matrix of second derivatives or Hessian, both evaluated at  $\bar{\mathbf{x}}$ , the solution to (2). If  $\bar{\mathbf{x}}$  lies on the boundary of the experimental region  $R$ , then it constitutes a suitable point to start a line search to explore new regions outside  $R$ .

Assuming the Hessian is invertible, the method of Newton specifies the following direction<sup>a</sup> to perform a line search:

$$\mathbf{d}_N = -\mathbf{H}(\bar{\mathbf{x}})^{-1}\nabla\eta(\bar{\mathbf{x}}) \quad (4)$$

The method of Newton can be viewed as the logical extension of the steepest ascent direction under second order information<sup>2</sup>. This, together with its simplicity, makes it an attractive choice for new region exploration in RPD. However, in order to be implemented, an invertible Hessian is needed, and, in order to assure convergence (to a point of zero gradient, see Bazaraa et al.<sup>2</sup>), positive definiteness of the Hessian must also hold (i.e., all eigenvalues of  $\mathbf{H}(\bar{\mathbf{x}})$  must be positive).

For the case when the Hessian is not positive definite, Joshi et al.<sup>6</sup> proposed the following direction for performing a line search whenever curvature becomes evident in an RSM study:

$$\mathbf{d}_J = - \left[ \sum_{i \in I_+} \frac{\boldsymbol{\nu}_i \boldsymbol{\nu}'_i}{\lambda_i} \right] \nabla\eta(\bar{\mathbf{x}}) \quad (5)$$

where  $\lambda_i, \forall i = 1, \dots, k$  are the  $k$  eigenvalues of  $\mathbf{H}(\bar{\mathbf{x}})$ , with  $\boldsymbol{\nu}_i, \forall i = 1, \dots, k$  being the correspondent eigenvectors, and  $I_+ = \{i : \lambda_i > 0 \text{ and } \boldsymbol{\nu}'_i \nabla\eta(\bar{\mathbf{x}}) \neq 0\}$ , so that the sum is made only over the eigenvectors for which the correspondent eigenvalues are positive. This ensures that  $\mathbf{d}_J$  is an *improving* direction, since it forms an obtuse angle with the gradient:

$$\mathbf{d}'_J \nabla\eta(\bar{\mathbf{x}}) = - \sum_{i \in I_+} \frac{(\boldsymbol{\nu}'_i \nabla\eta(\bar{\mathbf{x}}))^2}{\lambda_i} < 0$$

<sup>a</sup>To simplify the expressions we will use nonunit vectors to represent a direction. These can be easily transformed into unit vectors by multiplying them by the inverse of their norms.

We note that whenever  $\mathbf{H}(\bar{\mathbf{x}})$  is positive definite, then  $\lambda_i > 0$ ,  $\forall i = 1, \dots, k$ , and the direction proposed by Joshi et al.<sup>6</sup> becomes the direction of the method of Newton, since then we can take  $I_+ = \{1, \dots, k\}$  and:

$$\begin{aligned}\mathbf{d}_J &= -\mathbf{V}'(\Lambda)^{-1}\mathbf{V}\nabla\eta(\bar{\mathbf{x}}) \\ &= -\mathbf{V}'(\mathbf{V}'\Lambda)^{-1}\nabla\eta(\bar{\mathbf{x}}) \\ &= -(\mathbf{V}'\Lambda\mathbf{V})^{-1}\nabla\eta(\bar{\mathbf{x}}) \\ &= -\mathbf{H}(\bar{\mathbf{x}})^{-1}\nabla\eta(\bar{\mathbf{x}}) \\ &= \mathbf{d}_N\end{aligned}$$

where  $\Lambda$  is a diagonal matrix of the eigenvalues and  $\mathbf{V} = [\mathbf{v}_1 \dots \mathbf{v}_k]'$ .

#### 4. A Search Direction for the Dual Response Approach

In order to apply the search direction (5), we first need to obtain the Hessian and gradient of the objective function in problem (2). Towards this end, we note that this objective function can be simplified to get:

$$Var_{\mathbf{z}}(Y) = \gamma'\Sigma_{\mathbf{z}}\gamma + 2\mathbf{x}'(\Delta\Sigma_{\mathbf{z}}\gamma) + \mathbf{x}'\Delta\Sigma_{\mathbf{z}}\Delta'\mathbf{x} + \sigma_\epsilon^2$$

from which we have that the gradient is given by  $\nabla\eta(\bar{\mathbf{x}}) = 2(\Delta\Sigma_{\mathbf{z}}\gamma + \Delta\Sigma_{\mathbf{z}}\Delta'\bar{\mathbf{x}})$  and the Hessian matrix by  $\mathbf{H}(\bar{\mathbf{x}}) = 2\Delta\Sigma_{\mathbf{z}}\Delta'$ . Since the objective function is quadratic, the Hessian is constant. Recall that a symmetric matrix  $\mathbf{A}$  is positive definite if  $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$ ,  $\forall \mathbf{x} \neq \mathbf{0}$ . The Hessian  $\mathbf{H}(\bar{\mathbf{x}})$  will be at least positive semi-definite because the positive definiteness of  $\Sigma_{\mathbf{z}}$  ensures that  $\mathbf{x}'\mathbf{H}(\bar{\mathbf{x}})\mathbf{x} = 0$  if and only if  $\Delta'\mathbf{x} = \mathbf{0}$ :

$$\mathbf{x}'\mathbf{H}(\bar{\mathbf{x}})\mathbf{x} = 2\mathbf{x}'\Delta\Sigma_{\mathbf{z}}\Delta'\mathbf{x} = 2(\Delta'\mathbf{x})'\Sigma_{\mathbf{z}}(\Delta'\mathbf{x}) = \mathbf{0} \Leftrightarrow \Delta'\mathbf{x} = \mathbf{0}.$$

If  $k > r$ , then  $\exists \mathbf{x} \neq \mathbf{0}$  such that  $\Delta'\mathbf{x} = \mathbf{0}$  (i.e., the left nullspace of  $\Delta$  is not empty, since it is of dimension equal to  $\text{rank}(\Delta) - k$ , and  $\text{rank}(\Delta) \leq \min(k, r) = r < k$ ). However, if  $k \leq r$  and  $\Delta$  is full rank, then  $\nexists \mathbf{x} \neq \mathbf{0}$  such that  $\Delta'\mathbf{x} = \mathbf{0}$ , since  $\text{rank}(\Delta) = \min(k, r) = k$ , which implies the nullspace is of dimension 0. That is, the *strict* convexity of  $Var_{\mathbf{z}}(Y)$  is assured whenever the number of noise factors is greater or equal than the number of controllable factors and  $\Delta$  is a full rank matrix.

Therefore we have to consider three cases:

Case 1:  $k = r$ , unique solution for  $\Delta'\mathbf{x} = \mathbf{0}$  and strict convexity.

Case 2:  $k < r$ , no solution for  $\Delta'\mathbf{x} = \mathbf{0}$  and strict convexity.

Case 3:  $k > r$ , multiple solutions for  $\Delta'\mathbf{x} = \mathbf{0}$  and *no* strict convexity.

We now study each case separately.

Case 1:  $k = r$

In this case the Hessian  $\mathbf{H}(\bar{\mathbf{x}})$  and  $\Delta$  are invertible and the directions of Newton and Joshi et al. are equal. Let  $\mathbf{d}_N^{(1)}$  denote this direction given by equation (4) when  $k = r$ :

$$\mathbf{d}_N^{(1)} = -(\Delta \Sigma_z \Delta')^{-1} (\Delta \Sigma_z \gamma + \Delta \Sigma_z \Delta' \bar{\mathbf{x}}) = -(\Delta')^{-1} \gamma - \bar{\mathbf{x}}.$$

Notice that if the current solution is  $\bar{\mathbf{x}} = \mathbf{0}$ , then this yields the well-known result (see Myers and Montgomery<sup>10</sup>, p. 495) for the unconstrained minimum, where the variance contribution of the noise factors is made equal to zero at:

$$\mathbf{x}^* = -(\Delta')^{-1} \gamma.$$

Now, moving in the direction defined by  $\mathbf{d}_N^{(1)}$  disregards the constraint  $\mathbf{x}'\beta = T$ . Since  $\bar{\mathbf{x}}$  solves problem (2), then we must have that  $\bar{\mathbf{x}}'\beta = T$ . Thus, *to retain feasibility we need to move in a direction orthogonal to  $\beta$ .* The following matrix projects onto the subspace orthogonal to  $\beta$ :

$$\mathbf{P} = \mathbf{I} - \frac{\beta\beta'}{\beta'\beta}. \quad (6)$$

Moving along the direction  $\mathbf{P}\mathbf{d}_N^{(1)}$  will maintain feasibility. If  $\lambda$  is the step size taken along  $\mathbf{d}_N^{(1)}$ , then feasibility for the process mean can be shown because:

$$\beta_0 + \beta'(\lambda \mathbf{P}\mathbf{d}_N^{(1)} + \bar{\mathbf{x}}) = \beta_0 + \lambda \left( \beta' - \frac{\beta'\beta\beta'}{\beta'\beta} \right) \mathbf{d}_N^{(1)} + \beta' \bar{\mathbf{x}} = \beta_0 + \lambda \mathbf{0}'\mathbf{d}_N^{(1)} + \beta' \bar{\mathbf{x}} = T.$$

The idempotency of  $\mathbf{P}$  makes it positive semi-definite and hence:

$$\nabla\eta(\bar{\mathbf{x}})' \mathbf{P}\mathbf{d}_N^{(1)} = -\nabla\eta(\bar{\mathbf{x}})' \mathbf{P} (\Delta \Sigma_z \Delta')^{-1} \nabla\eta(\bar{\mathbf{x}}) \leq 0 ,$$

implying that moving in the direction indicated by  $\mathbf{P}\mathbf{d}_N^{(1)}$  will not worsen the variance objective, (it could either reduce it or leave it unchanged).

The variance will not change only if  $\nabla\eta(\bar{\mathbf{x}}) = \lambda\beta$  ( $\lambda \in \mathbb{R}^1$ ), since  $\mathbf{P}\mathbf{d}_N^{(1)}$  is orthogonal to  $\beta$  by definition, and so  $\nabla\eta(\bar{\mathbf{x}})' \mathbf{P}\mathbf{d}_N^{(1)} = 0$ . There will be small improvements when the angle between  $\beta$  and  $\nabla\eta(\bar{\mathbf{x}})$  is small. This can be determined before spending resources in further experimentation. Figure 1 contains a sketch describing the projection. No reduction in the variance objective will be observed whenever the angle denoted by  $\theta$  in this figure is zero.

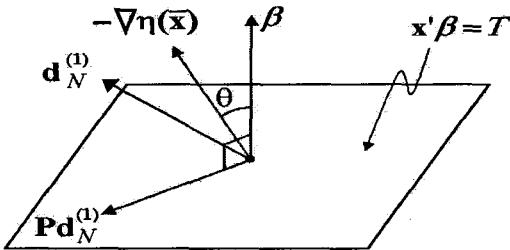


Fig. 1. Projection of the Newton Direction,  $\mathbf{d}_N^{(1)}$ , onto the subspace orthogonal to  $\beta$ .  $\theta$  denotes the angle between  $\beta$  and  $\nabla\eta(\bar{x})$ .

In view of the preceding discussion, when  $k = r$ , we propose using the following direction  $\mathbf{d}^{(1)}$  to perform a line search when  $k = r$ :

$$\mathbf{d}^{(1)} = - \left( \mathbf{I} - \frac{\beta\beta'}{\beta'\beta} \right) ((\Delta')^{-1}\gamma + \bar{x}). \quad (7)$$

#### Case 2: $k < r$

For Case 2 the direction of Newton is given by:

$$\mathbf{d}_N^{(2)} = -(\Delta\Sigma_z\Delta')^{-1} (\Delta\Sigma_z\gamma + \Delta\Sigma_z\Delta'\bar{x}) = -(\Delta\Sigma_z\Delta')^{-1} \Delta\Sigma_z\gamma + \bar{x}.$$

To retain feasibility the same projection matrix of equation (6) should be used, and since  $\Delta\Sigma_z\Delta'$  is also invertible, we get that the search direction  $\mathbf{Pd}_N^{(2)}$  will not worsen the objective but that it will not improve it only when  $\nabla\eta(\bar{x}) = \lambda\beta$ ,  $\lambda \in \mathbb{R}^1$ . Therefore, we recommend using the following search direction when  $k < r$ :

$$\mathbf{d}^{(2)} = - \left( \mathbf{I} - \frac{\beta\beta'}{\beta'\beta} \right) ((\Delta\Sigma_z\Delta')^{-1} \Delta\Sigma_z\gamma + \bar{x}). \quad (8)$$

#### Case 3: $k > r$

For this case the direction of Newton is not available because  $\mathbf{H}(\bar{x})$  is not invertible. In this case we propose to use the direction proposed by Joshi et al. from equation (5):

$$\mathbf{d}_J^{(3)} = - \left[ \sum_{i \in I_+} \frac{\nu_i \nu'_i}{\lambda_i} \right] (\Delta\Sigma_z\gamma + \Delta\Sigma_z\Delta'\bar{x})$$

where  $\lambda_i$  and  $\nu_i$ ,  $\forall i = 1, \dots, k$  are, respectively, the eigenvalues and eigenvectors of  $\mathbf{H}(\bar{x}) = \Delta\Sigma_z\Delta'$ .

The search direction for case 3, when  $k > r$ , is:

$$\mathbf{d}^{(3)} = - \left( \mathbf{I} - \frac{\boldsymbol{\beta}\boldsymbol{\beta}'}{\boldsymbol{\beta}'\boldsymbol{\beta}} \right) \left[ \sum_{i \in I_+} \frac{\boldsymbol{\nu}_i \boldsymbol{\nu}'_i}{\lambda_i} \right] (\Delta \boldsymbol{\Sigma}_{\mathbf{z}} \boldsymbol{\gamma} + \Delta \boldsymbol{\Sigma}_{\mathbf{z}} \Delta' \bar{\mathbf{x}}). \quad (9)$$

The variance will not change in direction  $\mathbf{d}^{(3)}$  if  $\nabla \eta(\bar{\mathbf{x}}) = \lambda \boldsymbol{\beta}$  or if  $\nabla \eta(\bar{\mathbf{x}})$  is a linear combination of  $\boldsymbol{\nu}_i$ ,  $i \notin I_+$ , so that it is orthogonal to all the rows of  $\sum_{i \in I_+} \boldsymbol{\nu}_i \boldsymbol{\nu}'_i / \lambda_i$ . This is explained in detail in Appendix 1.

## 5. Estimation of the Search Directions

In practice all the parameters in equations (7), (8) and (9) have to be estimated from data. Let  $\hat{\boldsymbol{\beta}}$ ,  $\hat{\Delta}$  and  $\hat{\boldsymbol{\gamma}}$  denote the corresponding OLS estimators. Under normality, these are also the ML estimators, so we can get ML estimators of the directions by simply “plugging in” the OLS estimators into the direction expressions. However, the nice properties of MLE’s are mostly asymptotic, and our concern here is with estimation from small samples, for which bias is a major consideration.

If we assume that a balanced orthogonal design was used to estimate the parameters, then we have that all these parameter estimates are unbiased, independent and have the same variance, namely  $\sigma_{\beta}^2 = \sigma_{\epsilon}^2/N$ , where  $N$  is the number of experiments used. However, finding small sample unbiased estimators of the search directions (7)-(9) is not possible. We will settle for at least estimating unbiasedly the projection matrix that allows us to maintain feasibility in problem (2). Since multiplying  $\mathbf{P}$  by a scalar does not change the projection properties of the resulting product, we can use instead  $(\boldsymbol{\beta}'\boldsymbol{\beta})\mathbf{P}$  as a projection matrix, which can be estimated unbiasedly.

An unbiased estimator of  $(\boldsymbol{\beta}'\boldsymbol{\beta})\mathbf{P}$  is:

$$\widehat{(\boldsymbol{\beta}'\boldsymbol{\beta})\mathbf{P}} = \left( \widehat{\boldsymbol{\beta}}'\widehat{\boldsymbol{\beta}} - \widehat{\sigma}_{\epsilon}^2(k-1)/N \right) \mathbf{I} - \widehat{\boldsymbol{\beta}}\widehat{\boldsymbol{\beta}}' \quad (10)$$

which can be shown to be unbiased because:

$$\begin{aligned} E_{\widehat{\boldsymbol{\beta}}}(\widehat{(\boldsymbol{\beta}'\boldsymbol{\beta})\mathbf{P}}) &= E \left[ \left( \widehat{\boldsymbol{\beta}}'\widehat{\boldsymbol{\beta}} - \widehat{\sigma}_{\epsilon}^2(k-1)/N \right) \mathbf{I} - \widehat{\boldsymbol{\beta}}\widehat{\boldsymbol{\beta}}' \right] \\ &= \left( E(\widehat{\boldsymbol{\beta}}'\widehat{\boldsymbol{\beta}}) - E(\widehat{\sigma}_{\epsilon}^2)k/N \right) \mathbf{I} - \left( E(\widehat{\boldsymbol{\beta}}\widehat{\boldsymbol{\beta}}') - E(\widehat{\sigma}_{\epsilon}^2)\mathbf{I}/N \right) \\ &= (\boldsymbol{\beta}'\boldsymbol{\beta} + \sigma_{\epsilon}^2 k/N - \sigma_{\epsilon}^2 k/N) \mathbf{I} - (\boldsymbol{\beta}\boldsymbol{\beta}' + \sigma_{\epsilon}^2 \mathbf{I}/N - \sigma_{\epsilon}^2 \mathbf{I}/N) \\ &= \boldsymbol{\beta}'\boldsymbol{\beta} \left( \mathbf{I} - \frac{\boldsymbol{\beta}\boldsymbol{\beta}'}{\boldsymbol{\beta}'\boldsymbol{\beta}} \right) \\ &= (\boldsymbol{\beta}'\boldsymbol{\beta})\mathbf{P}. \end{aligned}$$

Any multiple of  $\mathbf{P}$  will project into the same subspace than  $\mathbf{P}$  itself. Therefore, we could use  $(\beta' \beta) \mathbf{P} \mathbf{d}_N$  for  $k \leq r$  and  $(\beta' \beta) \mathbf{P} \mathbf{d}_J$  for  $k > r$ . For  $\mathbf{d}_N$  or  $\mathbf{d}_J$  we use plug-in estimates of the parameters but use the unbiased estimator (10) for  $(\beta' \beta) \mathbf{P}$ . This gives the following direction estimates which are “unbiasedly orthogonal” to  $\beta$ :

$$\hat{\mathbf{d}}^{(1)} = - \left( (\hat{\beta}' \hat{\beta} - \hat{\sigma}_\epsilon^2 (k-1)/N) \mathbf{I} - \hat{\beta} \hat{\beta}' \right) \left( (\hat{\Delta}')^{-1} \hat{\gamma} + \bar{\mathbf{x}} \right) \quad (11)$$

$$\hat{\mathbf{d}}^{(2)} = - \left( (\hat{\beta}' \hat{\beta} - \hat{\sigma}_\epsilon^2 (k-1)/N) \mathbf{I} - \hat{\beta} \hat{\beta}' \right) \left( (\hat{\Delta} \Sigma_{\mathbf{z}} \hat{\Delta}')^{-1} \Delta \Sigma_{\mathbf{z}} \hat{\gamma} + \bar{\mathbf{x}} \right) \quad (12)$$

$$\hat{\mathbf{d}}^{(3)} = - \left( (\hat{\beta}' \hat{\beta} - \hat{\sigma}_\epsilon^2 (k-1)/N) \mathbf{I} - \hat{\beta} \hat{\beta}' \right) \left[ \sum_{i \in I_+} \frac{\hat{\nu}_i \hat{\nu}'_i}{\hat{\lambda}_i} \right] (\hat{\Delta} \Sigma_{\mathbf{z}} \hat{\gamma} + \hat{\Delta} \Sigma_{\mathbf{z}} \hat{\Delta}' \bar{\mathbf{x}}). \quad (13)$$

Given the lack of bias of the estimated projection matrix, these directions will allow us to keep feasibility for the mean response, i.e., keep  $E_{\mathbf{z}}(Y) = T$ .

## 6. Estimation of $Var_{\mathbf{z}}(Y)$ Along the Search Direction

In the dual approach to RPD the variance of the response,  $Var_{\mathbf{z}}(Y)$ , is not directly observed. Therefore, techniques for indirectly evaluating this variance along the search direction are needed. One scheme to estimate  $Var_{\mathbf{z}}(Y)$  during the line search is to perform a 2-level factorial, possibly fractional, design on the noise factors only, without changing the setting of the controllable factors. Then instead of investing a large budget in fitting a complex model like the one in equation (1), we propose to fit a *main effects only* model on the noise factors:

$$Y_t | \mathbf{z} = \beta_{0,t} + \mathbf{z}' \boldsymbol{\gamma}_t + \epsilon \quad (14)$$

where the subscript “ $t$ ” indicates the model for the  $t^{th}$  step on the line search.

For this model we have an unconditional variance given by:

$$Var_{\mathbf{z}}(Y_t) = \boldsymbol{\gamma}'_t \Sigma_{\mathbf{z}} \boldsymbol{\gamma}_t + \sigma_\epsilon^2 \quad (15)$$

and an unconditional mean given by:

$$E_{\mathbf{z}}(Y_t) = \beta_{0,t}. \quad (16)$$

Then, as the line search in direction  $\hat{\mathbf{d}}^{(i)}$  is conducted ( $i \in \{1, 2, 3\}$ ), we should observe that the values of  $Var_{\mathbf{z}}(Y_t)$  diminish while  $E_{\mathbf{z}}(Y_t)$  remain relatively constant.

To obtain an unbiased estimator of  $E_{\mathbf{z}}(Y_t)$  is sufficient to substitute the OLS estimator  $\beta_{0,t}$  in equation (16):

$$\hat{E}_{\mathbf{z}}(Y_t) = \hat{\beta}_{0,t}. \quad (17)$$

However, for  $Var_{\mathbf{z}}(Y_t)$ , this substitution yields:

$$\widehat{Var}_{\mathbf{z}}(Y_t) = \hat{\gamma}'_t \Sigma_{\mathbf{z}} \hat{\gamma}_t + \hat{\sigma}_{\epsilon}^2 \quad (18)$$

which is easy to show to be a biased estimator. An unbiased estimator can be obtained by using a correction factor. However, the correction factor may result in a negative estimate. But since the correction factor will be a constant along the search, it will not have any effect on the performance of the search. Furthermore, in Appendix 2 we show that equation (18) is an unbiased estimator of the variance of the *predictions* given by the fitted model. Therefore, we suggest using equation (18) to estimate the variance of the response along the search direction.

It is also necessary to have available a suitable stopping rule to determine when to stop the search. In the present case a stop of the search should be triggered by two different events. One is an increase in the variance of the response and the second one is a deviation of the mean from the target value. It is desired to keep the mean response on-target throughout the search process because, in case it becomes largely off-target, the task of making the mean on-target again could be as complex as minimizing the variance. Thus, we should accomplish the minimization of variance without changing the mean too far from its target.

### 6.1. Relaxation of equality constraint

Since model (1) is a local approximation of the true response surface, and there are sampling errors associated with the estimates  $\hat{\beta}_{0,t}$ , deviations from target will be observed as the line searches are conducted. Therefore, it is necessary to determine a tolerance range for the values of  $\hat{E}_{\mathbf{z}}(Y_t)$ . This is also justified since some “bias” in the process may provide a decrease in variance not attainable if strict equality is always desired.

The tolerance range needs to be defined by the user based on the context of the experiment; such tolerance will most likely be based on specifications for the mean response. The search will be stopped whenever the mean response estimate falls outside the tolerance range or if a drop in  $\widehat{Var}_{\mathbf{z}}(Y_t)$  is observed.

## 7. Summary of Proposed Methodology

In order to aid understanding of the different steps in the proposed approach, we present in the following an algorithmic description of the procedure. We point out that blindly following this algorithmic description is not suggested; carefully conducting the DOE's and analyzing their results is always necessary.

- (1) Let  $i = 0$ .
- (2) Conduct a DOE in the  $k$  controllable factors suitable to fit model (1) around the current operating conditions,  $\mathbf{x}_i = \mathbf{x}_0$ .
- (3) Solve problem (2) for the desired target  $T$ , taking  $R$  to be the current experimental region. If the solution to (2),  $\mathbf{x}_{i+1} = \bar{\mathbf{x}}$  is inside of  $R$ , stop. If no feasible solution exists, change the target and/or the region  $R$ . Otherwise,  $\mathbf{x}_{i+1}$  is on the boundary, so we explore outside the initial region doing the following:
- (4) Let  $i \leftarrow i + 1$ . Select a step size  $\lambda$  (typically,  $0 < \lambda < 1$ ) and move the setting of the controllable factors to:

$$\mathbf{x}_{i+1} = \mathbf{x}_i + \lambda \widehat{\mathbf{d}}^{(j)}$$

where  $j = 1$  if  $k = r$  (equal no. of controllable and noise factors, use eq. 11);  $j = 2$  if  $k < r$  (use eq. 12); and  $j = 3$  if  $k > r$  (use eq. 13);

- (5) While keeping the controllable factors fixed at  $\mathbf{x}_{i+1}$ , conduct a DOE in the  $r$  noise factors suitable to fit model (14). Compute estimates of  $E_{\mathbf{z}}(Y_t)$  and  $Var_{\mathbf{z}}(Y_t)$  using (17) and (18).
- (6) If  $\widehat{E}_{\mathbf{z}}(Y_t)$  exceeds the tolerances (specs) for the process mean, stop searching in the direction used thus far, make  $\mathbf{x}_0$  equal to the current operating conditions and go to step 1. Otherwise, go to step 4.

The procedure ends when no significant variance reductions are observed. In the next two sections we illustrate the procedure in detail with two simulated examples.

An initial experimental region is assumed. How to determine this region depends on practical considerations that are context-dependent. As in any optimization problem, it should be clear that an initial region that is too far from the global optimum of problem (2) will require more experiments than one closer to the optimum.

## 8. A Simulation Example for $k = 3$ and $r = 3$

### 8.1. "True" function and its global optima

In this section we use simulated data to give an illustrative example of the application of the techniques developed in sections 2, to 6. For this purpose we use the following test function:

$$\eta(\mathbf{x}, \mathbf{z}) = \beta_0 + \boldsymbol{\beta}' \mathbf{x} + \mathbf{x}' \mathbf{B} \mathbf{x} + \boldsymbol{\rho}' \mathbf{x}^{(4)} + \mathbf{x}' \boldsymbol{\Delta} \mathbf{z} + \boldsymbol{\gamma}' \mathbf{z} + \mathbf{z}' \mathbf{C}(\mathbf{x}) \mathbf{z} + \epsilon \quad (19)$$

where :

- $\beta_0 = 0$ .
- $\mathbf{x} = [x_1 \ x_2 \ x_3]'$ , are the controllable factors.
- $\boldsymbol{\mu_s} = [10 \ -5 \ 25]'$ .
- $\mathbf{x}_s^{(4)} = [x_1^4 \ x_2^4 \ x_3^4]'$ .
- $\boldsymbol{\beta} = 10 [10 \ -15 \ -30]'$ .
- $\mathbf{B} = \begin{bmatrix} 9/2 & -3/2 & 1 \\ -3/2 & -6 & 0 \\ 1 & 0 & -5 \end{bmatrix}$ .
- $\boldsymbol{\rho} = \frac{1}{400} [3 \ 16 \ 24]'$ .
- $\mathbf{z} = [z_1 \ z_2 \ z_3]'$ , are the noise factors.
- $\boldsymbol{\gamma} = 10 [-20 \ 40 \ 20]'$ .
- $\boldsymbol{\Delta} = 10 \begin{bmatrix} 25 & 60 & -40 \\ 45 & 85 & 80 \\ 75 & -90 & 85 \end{bmatrix}$ .
- $\mathbf{C}(\mathbf{x}) = \begin{bmatrix} 0.196x_1^2 + 0.6 & 0.14x_1x_2 & 0.112x_3x_1 + 0.2 \\ 0.056x_1x_2 & 0.016x_2^2 + 0.8 & 0.032x_2x_3 + 0.3 \\ 0.112x_3x_1 + 0.2 & 0.032x_2x_3 + 0.3 & 0.064x_3^2 + 2.5 \end{bmatrix}$ .
- $\epsilon$  denotes an additive error such that  $\epsilon \sim N(0, \sigma_\epsilon^2)$ .

Notice that the number of controllable factors is the same as the number of noise factors and therefore we will use the search direction obtained under case 1 in section 5.

This form of test function was chosen because it differs significantly from the local models that are fitted during the search (equations 1 and 14), but at the same time allows taking the variance and mean operators in closed form. Therefore, for this test function we are able to obtain the true response surfaces for the mean and variance in closed form. Obtaining the true response surfaces in closed form allows determining the *global* solution to the RPD problem and compare this global optimum to the final solution obtained using the proposed search method. In Appendix 3 we give the details on how the test functions were obtained. The true response surfaces for the variance and mean are given following:

$$Var_{\mathbf{z}}(\eta(\mathbf{x}, \mathbf{z})) = (\Delta' \mathbf{x}_s + \gamma)' \Sigma_{\mathbf{z}} (\Delta' \mathbf{x}_s + \gamma) + 2tr(\mathbf{C}(\mathbf{x}) \Sigma_{\mathbf{z}} \mathbf{C}(\mathbf{x}) \Sigma_{\mathbf{z}}) + \sigma_{\epsilon}^2 \quad (20)$$

$$E_{\mathbf{z}}(\eta(\mathbf{x}, \mathbf{z})) = \beta_0 + \beta' \mathbf{x}_s + \mathbf{x}'_s \mathbf{B} \mathbf{x}_s + \rho' \mathbf{x}_s^{(4)} + tr(\mathbf{C}(\mathbf{x}) \Sigma_{\mathbf{z}}). \quad (21)$$

To take the variance and mean operators it was assumed that  $\mathbf{z} \sim N(\mathbf{0}, \Sigma_{\mathbf{z}})$ , with:

$$\Sigma_{\mathbf{z}} = \begin{bmatrix} 3 & -1 & -2 \\ -1 & 4 & 0 \\ -2 & 0 & 4 \end{bmatrix}.$$

We note that both response surfaces are fourth order polynomials on the controllable factors. The target value selected was  $T = 5$ .

Equations (20) and (21) can be used to solve the global RPD problem. This is given by the following mathematical program:

*min :*

$$Var_{\mathbf{z}}(\eta(\mathbf{x}, \mathbf{z})) = (\Delta' \mathbf{x}_s + \gamma)' \Sigma_{\mathbf{z}} (\Delta' \mathbf{x}_s + \gamma) + 2tr(\mathbf{C}(\mathbf{x}) \Sigma_{\mathbf{z}} \mathbf{C}(\mathbf{x}) \Sigma_{\mathbf{z}}) + \sigma_{\epsilon}^2$$

*subject to :*

$$E_{\mathbf{z}}(\eta(\mathbf{x}, \mathbf{z})) = \beta_0 + \beta' \mathbf{x}_s + \mathbf{x}'_s \mathbf{B} \mathbf{x}_s + \rho' \mathbf{x}_s^{(4)} + tr(\mathbf{C}(\mathbf{x}) \Sigma_{\mathbf{z}}) = 5.$$

(22)

This problem was solved using MatLab's fmincon routine. Five hundred initial solutions were randomly selected inside the cube  $C$  given by:

$$C = \{\mathbf{x} : -50 < x_1 < 50, -50 < x_2 < 50, -50 < x_3 < 50\}.$$

The algorithm converged to the same point,  $\mathbf{x}_g = [-0.27 \ -1.48 \ 1.03]$ , in each of the 500 trials, for which  $Var_{\mathbf{z}}(\eta(\mathbf{x}_g, \mathbf{z})) = 877.8$ .

## 8.2. Illustration of the methodology of section 7

To start the search, a point located at a distance of 20 units from  $\mathbf{x}_g$  and for which  $E_{\mathbf{z}}(\eta(\mathbf{x}, \mathbf{z})) = 5$  was selected as the starting point. For this example

the selected point is  $\mathbf{x}_0 = [15.53, 10.47, 1.71]$ . To simulate values of the response  $Y$  using equation (19) we set  $\sigma_\epsilon^2 = 1$ . Hence, the search was started at a point for which the true mean response is on target. In practice, if such a starting point is not available, traditional RSM tools for the mean response could be used to drive the process to target, and RSM books<sup>3,7,10</sup> discuss these methods in detail.

The model on equation (1) was fitted using simulated responses from equation (19) and a  $2_{VI}^{6-1}$  fractional factorial, with five center points, centered at  $\mathbf{x}_0$ . This design was used because a quarter fraction has only resolution IV and 2-factor interactions are aliased with one another. After fitting the model, problem (2) was solved. The experimental region  $R$  was taken to be a sphere of radius 2. The solution to (2) was used to start a line search in the direction given by equation (11). The step length,  $\lambda$ , along the search direction was set to 0.5. i.e. the mean and variance was estimated at points 0.5 units apart. The estimation was done using a  $2^3$  full factorial design on the noise factors, as described in section 6. The tolerance range for the estimated mean,  $\hat{E}_{\mathbf{z}}(Y_t)$ , was set to  $5 \pm 2.5$ . Whenever an increase in variance or an estimated mean outside of this range was observed the search was stopped and the location at the previous step was used as the center of a new  $2_{VI}^{6-1}$  design with which a new model (1) was estimated and the search was repeated. The complete multidimensional search was stopped when the stopping point of the line search was at less than 3 units of distance from the global optimum,  $\|\mathbf{x}_t - \mathbf{x}_g\|$ .

The results of the first  $2_{VI}^{6-1}$  experimental design, centered at  $\mathbf{x}_0$ , are given in Table 1. Model (1) was fitted using these results. The fitted model is given by:

$$Y = 4.3599 + [\mathbf{x}_1 \mathbf{x}_2 \mathbf{x}_3] \begin{bmatrix} 3.32 \\ -1.45 \\ -3.01 \end{bmatrix} + [\mathbf{z}_1 \mathbf{z}_2 \mathbf{z}_3] \begin{bmatrix} 96.8 \\ 170.8 \\ 37.8 \end{bmatrix} + [\mathbf{x}_1 \mathbf{x}_2 \mathbf{x}_3] \begin{bmatrix} 2.48 & 6.24 & -3.94 \\ 4.37 & 8.49 & 8.18 \\ 7.54 & -8.88 & 8.35 \end{bmatrix} \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \\ \mathbf{z}_3 \end{bmatrix} \quad (23)$$

The fit of the model was very good, with an  $R^2$  statistic of 99.9% and a residual mean squared error of  $\hat{\sigma}_\epsilon^2 = 0.9926$ , which agrees with the value of  $\sigma_\epsilon^2 = 1$  used to generate the simulated responses. The replicated center points were used to conduct an  $F$  test for lack of fit (see Neter et al.<sup>11</sup> pp. 115-124). The test was insignificant, with a p-value of 0.148.

Table 1. First  $2^{6-1}_{VI}$  experimental design for the example using simulated data.

$x_1$	$x_2$	$x_3$	$z_1$	$z_2$	$z_3$	$Y$
-1	-1	-1	-1	-1	-1	-267.3
1	-1	-1	-1	-1	1	-235.4
-1	1	-1	-1	-1	1	-229.9
1	1	-1	-1	-1	-1	-314.2
-1	-1	1	-1	-1	1	-204.1
1	-1	1	-1	-1	-1	-290.0
-1	1	1	-1	-1	-1	-332.0
1	1	1	-1	-1	1	-233.6
-1	-1	-1	1	-1	1	-51.4
1	-1	-1	1	-1	-1	-96.4
-1	1	-1	1	-1	-1	-129.6
1	1	-1	1	-1	1	-54.4
-1	-1	1	1	-1	-1	-90.6
1	-1	1	1	-1	1	-17.0
-1	1	1	1	-1	1	-3.4
1	1	1	1	-1	-1	-114.0
-1	-1	-1	-1	1	1	114.5
1	-1	-1	-1	1	-1	85.3
-1	1	-1	-1	1	-1	52.5
1	1	-1	-1	1	1	139.5
-1	-1	1	-1	1	-1	7.3
1	-1	1	-1	1	1	96.8
-1	1	1	-1	1	1	112.9
1	1	1	-1	1	-1	18.9
-1	-1	-1	1	1	-1	226.5
1	-1	-1	1	1	1	293.9
-1	1	-1	1	1	1	317.5
1	1	-1	1	1	-1	266.3
-1	-1	1	1	1	1	285.9
1	-1	1	1	1	-1	234.4
-1	1	1	1	1	-1	207.3
1	1	1	1	1	1	342.2
0	0	0	0	0	0	3.9
0	0	0	0	0	0	5.2
0	0	0	0	0	0	4.6
0	0	0	0	0	0	5.2
0	0	0	0	0	0	4.1

Using the parameter estimates of the fitted model given in equation (23) the dual approach problem (2) can be stated as:

$$\begin{aligned} \min \widehat{\text{Var}}_{\mathbf{z}}(Y) = & \\ & \left( \begin{bmatrix} 96.8 \\ 170.8 \\ 37.8 \end{bmatrix} + \begin{bmatrix} 2.48 & 4.37 & 7.54 \\ 6.24 & 8.49 & -8.88 \\ -3.94 & 8.18 & 8.35 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \right)' \begin{bmatrix} 3 & -1 & -2 \\ -1 & 4 & 0 \\ -2 & 0 & 4 \end{bmatrix} \\ & \times \left( \begin{bmatrix} 96.8 \\ 170.8 \\ 37.8 \end{bmatrix} + \begin{bmatrix} 2.48 & 4.37 & 7.54 \\ 6.24 & 8.49 & -8.88 \\ -3.94 & 8.18 & 8.35 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \right) \end{aligned}$$

subject to :

$$\widehat{E}_{\mathbf{z}}(Y) = 4.3599 + [x_1 \ x_2 \ x_3] \begin{bmatrix} 3.32 \\ -1.45 \\ -3.01 \end{bmatrix} = 5$$

$$\mathbf{x}'\mathbf{x} \leq 2.$$

The previous problem was solved using MatLab's nonlinear programming routine "fmincon". The solution obtained is  $\bar{\mathbf{x}} = [-0.0579, -1.3607, 0.3810]$  for which  $\widehat{\text{Var}}_{\mathbf{z}}(Y_t) = 8.62 \times 10^4$ . Notice that the fitted model and the dual response problem are both "local", that is, the coordinate system is centered at the design center. The search direction obtained using equation (11) is:

$$\begin{aligned} \widehat{\mathbf{a}}^{(1)} = & - \left( \left( \begin{bmatrix} 3.32 \\ -1.45 \\ -3.01 \end{bmatrix}' \begin{bmatrix} 3.32 \\ -1.45 \\ -3.01 \end{bmatrix} - \frac{0.9926(3-1)}{37} \right) \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 3.32 \\ -1.45 \\ -3.01 \end{bmatrix} \begin{bmatrix} 3.32 \\ -1.45 \\ -3.01 \end{bmatrix}' \right) \\ & \times \left( \begin{bmatrix} 2.48 & 4.37 & 7.54 \\ 6.24 & 8.49 & -8.88 \\ -3.94 & 8.18 & 8.35 \end{bmatrix}^{-1} \begin{bmatrix} 96.8 \\ 170.8 \\ 37.8 \end{bmatrix} + \begin{bmatrix} -0.0579 \\ -1.3607 \\ 0.3810 \end{bmatrix} \right) \propto \begin{bmatrix} -0.62 \\ -0.67 \\ -0.41 \end{bmatrix}, \end{aligned}$$

where the last right-hand side has been normalized to a unit vector. The variance and mean of the response was estimated at different steps in this direction using the procedure of section 6 with a  $2^3$  design on the noise factors. Table 2 contains the design used, the response values generated and the estimates of the mean,  $\widehat{E}_{\mathbf{z}}(Y_t)$ , and variance,  $\widehat{\text{Var}}_{\mathbf{z}}(Y_t)$ , for the complete line search. Notice that the search starts at  $t = 2$  because  $t = 0$  corresponds to the center of the previous  $2^{6-1}$  design and  $t = 1$  corresponds to  $\bar{\mathbf{x}}$ . We observe in this Table how the variance estimates start to reduce along the search while the mean estimates start to deviate from the target. The search is stopped at  $t = 10$  because the estimated mean equals 8.90,

Table 2. Evaluation of the expected variance and mean using  $2^3$  experimental designs during the first line search, first example.

2 <sup>3</sup> Design			Values of the Response at Each Step $t$ .								
$z_1$	$z_2$	$z_3$	$Y_{t=2}$	$Y_{t=3}$	$Y_{t=4}$	$Y_{t=5}$	$Y_{t=6}$	$Y_{t=7}$	$Y_{t=8}$	$Y_{t=9}$	$Y_{t=10}$
-1	-1	-1	-265.80	-252.70	-243.84	-233.44	-224.08	-212.29	-202.89	-194.07	-181.27
1	-1	-1	-88.00	-82.76	-79.52	-77.39	-75.80	-70.27	-70.16	-67.24	-65.11
-1	1	-1	40.73	46.03	49.31	52.65	58.36	62.91	68.17	72.78	77.74
1	1	-1	214.16	215.17	215.39	209.03	206.73	204.08	201.78	197.69	198.00
-1	-1	1	-211.16	-205.01	-201.12	-198.97	-195.35	-190.37	-184.95	-183.62	-177.98
1	-1	1	-26.21	-33.34	-38.93	-40.20	-45.34	-49.86	-53.37	-59.06	-61.46
-1	1	1	97.25	93.67	90.62	88.25	88.42	86.46	84.09	82.32	82.53
1	1	1	280.97	266.97	256.48	248.42	236.45	227.23	215.84	208.57	198.78
$E_z(Y_t) = \hat{\beta}_{0,t}$			5.24	6.00	6.05	6.04	6.17	7.24	7.32	7.17	8.90
$\hat{\gamma}_t$			$\begin{bmatrix} 89.99 \\ 153.03 \\ 29.97 \end{bmatrix}$	$\begin{bmatrix} 85.51 \\ 149.46 \\ 24.57 \end{bmatrix}$	$\begin{bmatrix} 82.31 \\ 146.90 \\ 20.71 \end{bmatrix}$	$\begin{bmatrix} 78.92 \\ 143.54 \\ 18.33 \end{bmatrix}$	$\begin{bmatrix} 74.34 \\ 141.31 \\ 14.87 \end{bmatrix}$	$\begin{bmatrix} 70.56 \\ 137.93 \\ 11.13 \end{bmatrix}$	$\begin{bmatrix} 66.21 \\ 135.16 \\ 8.09 \end{bmatrix}$	$\begin{bmatrix} 62.82 \\ 133.17 \\ 4.88 \end{bmatrix}$	$\begin{bmatrix} 58.65 \\ 130.36 \\ 1.56 \end{bmatrix}$
$Var_z(Y_t) =$ $\hat{\gamma}_t \Sigma_z \hat{\gamma}_t$			8.32	7.97	7.74	7.40	7.19	6.89	6.64	6.49	6.26

which is larger than the upper bound of the tolerance range,  $5 \pm 2.5$ . We point out that the variance estimate at the start of the search is 33% larger than at the end of the search. Therefore, the robustness has been improved substantially with minimal deviations from target. Since the best solution found during the search was found for  $t = 9$ , a new  $2_{VI}^{6-1}$  design is carried at this location after which the complete process is repeated. The results for the rest of the search are given in Tables 3 and 5. In addition, Figure 2 contains the path of the multidimensional search, while Figures 3 and 4 contain the estimated variances and means, respectively, along the multidimensional search.

In Tables 3 and 5 we observe how the dual response problem and the line searches work together to systematically reduce the variance. Although in this example all of the dual response problems were feasible, this might not occur in general, since there might not exist an  $\mathbf{x} \in R$  such that  $\beta_0 + \beta' \mathbf{x} = T$ . One reason for this infeasibility would be that the tolerance range given for the estimated mean in the previous line search was too large and, therefore, the dual response problem is not able to drive the response on target again. In case such infeasibility occurs, the user might relax the region  $R$  in order to obtain a feasible on-target solution.

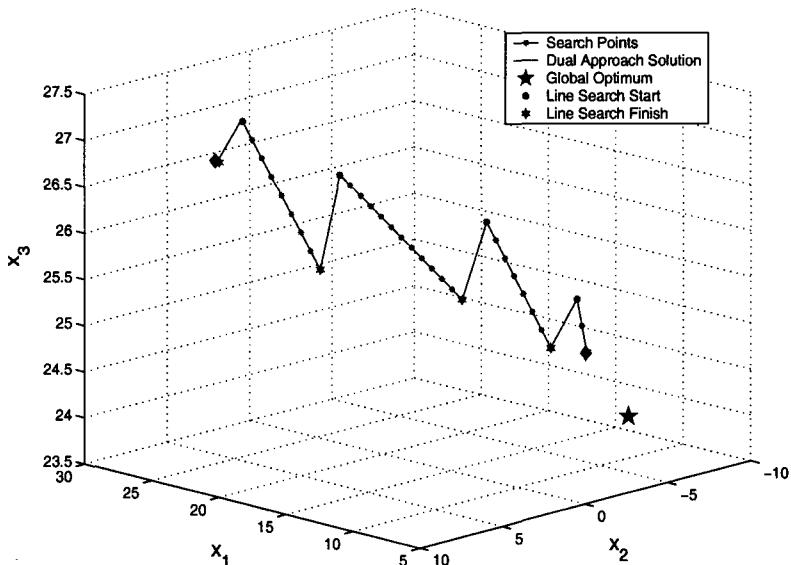


Fig. 2. Path of the  $\mathbf{x}_t$  points during the multidimensional search for the first example ( $k = 3$  and  $r = 3$ ).

Table 3. Simulation example of the new region exploration for RPD methodology, first example.

<b>t</b>	<b>x<sub>t</sub></b>	$\widehat{Var}_{\mathbf{z}}(Y_t) \times 10^{-4}$	$\widehat{E}_{\mathbf{z}}(Y_t)$	<b>DOE ran</b>	<b>Action</b>	$\ \mathbf{x}_t - \mathbf{x}_g\ $
0	[15.53, 10.47, 1.71]			$2_{VI}^{6-1}$	Start	20.00
1	[15.48, 9.11, 2.09]	8.62	5.00	Solution to problem 2		19.23
2	[15.17, 8.77, 1.89]	8.32	5.24	$2^3$	Continue	18.76
3	[14.87, 8.43, 1.68]	7.97	6.00	$2^3$	Continue	18.29
4	[14.56, 8.09, 1.48]	7.74	6.05	$2^3$	Continue	17.83
5	[14.26, 7.75, 1.27]	7.40	6.04	$2^3$	Continue	17.36
6	[13.96, 7.41, 1.07]	7.19	6.17	$2^3$	Continue	16.90
7	[13.65, 7.07, 0.86]	6.89	7.24	$2^3$	Continue	16.44
8	[13.35, 6.73, 0.66]	6.64	7.32	$2^3$	Continue	15.99
9	[13.04, 6.39, 0.46]	6.49	7.17	$2^3$ and $2_{VI}^{6-1}$	Continue	15.53
10	[12.74, 6.05, 0.25]	6.26	8.90	$2^3$	Stop, Mean Out	15.08
11	[12.26, 5.80, 1.48]	4.75	5.00	Solution to problem 2		14.71
12	[11.89, 5.49, 1.36]	4.38	4.57	$2^3$	Continue	14.22
13	[11.52, 5.17, 1.25]	4.13	4.77	$2^3$	Continue	13.73
14	[11.15, 4.86, 1.14]	3.88	5.03	$2^3$	Continue	13.24
15	[10.78, 4.54, 1.02]	3.57	5.42	$2^3$	Continue	12.75
16	[10.40, 4.23, 0.91]	3.34	4.98	$2^3$	Continue	12.26
17	[10.03, 3.91, 0.80]	3.11	5.11	$2^3$	Continue	11.77
18	[9.66, 3.60, 0.68]	2.86	6.21	$2^3$	Continue	11.28

Table 4. Simulation example of the new region exploration for RPD methodology, first example (cont.)

t	$\mathbf{x}_t$	$\widehat{Var}_{\mathbf{z}}(Y_t) \times 10^{-4}$	$\widehat{E}_{\mathbf{z}}(Y_t)$	DOE ran	Action	$\ \mathbf{x}_t - \mathbf{x}_g\ $
19	[9.29, 3.29, 0.57]	2.63	5.54	$2^3$	Continue	10.80
20	[8.92, 2.97, 0.46]	2.41	6.09	$2^3$	Continue	10.31
21	[8.55, 2.66, 0.34]	2.19	6.73	$2^3$	Continue	9.83
22	[8.17, 2.34, 0.23]	2.01	6.96	$2^3$	Continue	9.35
23	[7.80, 2.03, 0.12]	1.87	7.45	$2^3$ and $2_{VI}^{6-1}$	Continue	8.87
24	[7.43, 1.71, 0.00]	1.68	7.92	$2^3$	Stop, Mean Out	8.40
25	[7.15, 1.08, 0.94]	1.03	5.00	Solution to problem 2		8.08
26	[6.75, 0.85, 0.74]	0.96	5.07	$2^3$	Continue	7.60
27	[6.35, 0.62, 0.55]	0.87	6.08	$2^3$	Continue	7.11
28	[5.95, 0.39, 0.36]	0.77	5.16	$2^3$	Continue	6.63
29	[5.55, 0.16, 0.17]	0.72	6.80	$2^3$	Continue	6.16
30	[5.14, -0.07, -0.02]	0.66	6.58	$2^3$	Continue	5.68
31	[4.74, -0.30, -0.21]	0.59	6.72	$2^3$	Continue	5.21
32	[4.34, -0.53, -0.40]	0.59	7.49	$2^3$ and $2_{VI}^{6-1}$	Continue	4.75
33	[3.94, -0.76, -0.59]	0.54	7.77	$2^3$	Stop, Mean Out	4.29
34	[3.24, -1.23, 0.13]	0.22	5.00	Solution to problem 2		3.70
35	[2.82, -1.19, -0.14]	0.21	5.56	$2^3$	Continue	3.23
36	[2.41, -1.14, -0.41]	0.20	5.20	$2^3$	Continue	2.76
37	[1.99, -1.09, -0.68]	0.22	6.37	$2^3$	Stop, Var Higher	2.31

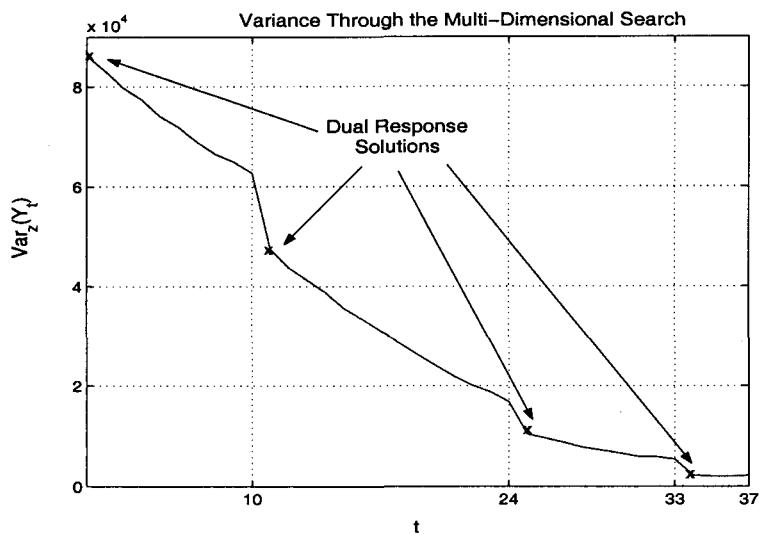


Fig. 3. Values of  $\widehat{Var}_z(Y_t)$  during the multidimensional search for the first example ( $k = 3$  and  $r = 3$ ).

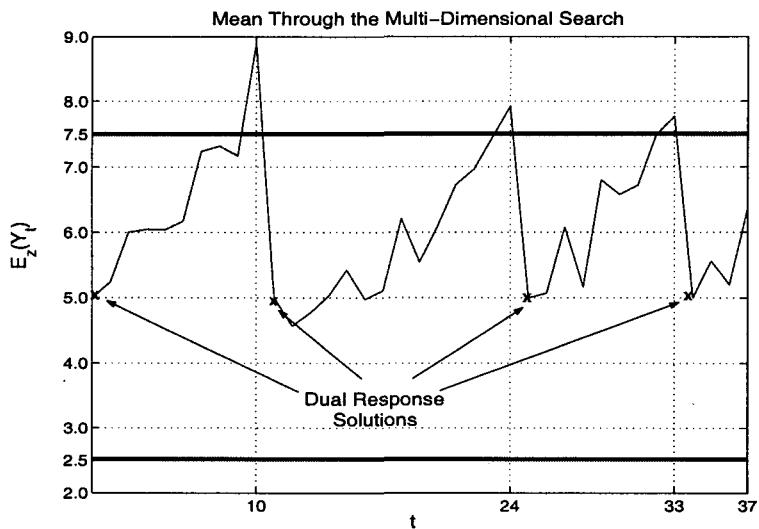


Fig. 4. Values of  $\widehat{E}_z(Y_t)$  during the multidimensional search for the first example ( $k = 3$  and  $r = 3$ ).

The path of the multidimensional search shown in Figure 2 indicates that the search approaches the global optimum during the line searches. The apparent zigzagging between the line searches and the dual response problems is because the dual response problems are driving the response back to the target. The systematic reduction of the variance is shown in Figure 3. It is seen that most of the reduction in variance is obtained during the line searches. The estimated means are given in Figure 4, where it is seen how the solutions to the dual response problems are setting the expected response on target again.

## 9. A Simulation Example for $k = 3$ and $r = 2$

### 9.1. “True” function and its global optimum

The test function for this second example was obtained from the one in section 8 by discarding the last element of  $\gamma$ , the last column of  $\Delta$  and the last column and row of  $\Sigma_z$  and  $C(x)$ . For this example a target value,  $T = -20$  was selected. The tolerance range, however, was kept equal, and, therefore, a line search was stopped if an increase in variance was observed or if the estimated mean was outside the  $-20 \pm 2.5$  range. The solution for the global dual response problem (equivalent to problem (22) in section 8) is  $x_g = [-10.30 \ -8.39 \ 1.19]$ , for which  $Var_z(\eta(x_g, z)) = 89.28$ .

### 9.2. Application of the methodology of section 7

The model in (1) was fitted using a  $2^{5-1}$  fractional factorial design, while the estimation of the variance and mean along the line searches was done using a  $2^2$  factorial on the two noise factors. All the other parameters used in the previous example were kept the same. The starting point was randomly selected from the subset of the sphere of radius 20 centered at  $x_g$  for which  $E_z(\eta(x, z)) = -20$ . Since for this case neither the Hessian,  $H(x) = 2\Delta\Sigma_z\Delta$  nor  $\Delta$  are invertible, the search direction given in equation (13) was used to calculate the search direction. We detail the calculation of this direction for the first line search. We skip the results of the first  $2^{5-1}_{VI}$  design and just state the parameter estimates obtained:

$$\hat{\beta}_0 = -20.26.$$

$$\hat{\beta} = \begin{bmatrix} 0.1011 \\ -1.719 \\ -2.627 \end{bmatrix}$$

$$\hat{\gamma} = \begin{bmatrix} -18.00 \\ -227.28 \end{bmatrix},$$

$$\hat{\Delta} = \begin{bmatrix} -2.594 & 5.655 \\ 4.559 & 8.566 \\ 7.534 & -9.142 \end{bmatrix}.$$

The model fit was, as in the previous example, very good, with an  $R^2$  statistic of 99.9% and a  $\hat{\sigma}_\epsilon^2 = 1.057$ .

With these parameter estimates the dual response problem was solved. The optimal solution, in local units, is  $\bar{x} = [0.5373, 1.0588, -0.7682]$ , for which  $Var_z(\eta(x_g, z)) = 1.669 \times 10^5$ . The estimated Hessian matrix was also computed with these parameter estimates:

$$\mathbf{H}(\bar{x}) = 2\hat{\Delta}\Sigma_z\hat{\Delta}'$$

$$= \begin{bmatrix} 237.6 & 362.5 & -334.1 \\ 362.5 & 555.5 & -466.1 \\ -334.1 & -466.1 & 1284.6 \end{bmatrix}.$$

The eigenvectors of  $\mathbf{H}(\bar{x})$  are [1662, 0, 415]. The columns of  $\mathbf{V}$  contain the corresponding normalized eigenvectors:

$$\mathbf{V} = \begin{bmatrix} -0.311 & -0.848 & 0.429 \\ -0.454 & 0.529 & 0.717 \\ 0.835 & -0.028 & 0.549 \end{bmatrix}.$$

Therefore  $I_+ = \{1, 3\}$  and

$$\sum_{i \in I_+} \frac{\nu_i u'}{\lambda_i} = \frac{1}{1662} \begin{bmatrix} -0.311 \\ -0.454 \\ 0.835 \end{bmatrix} \begin{bmatrix} -0.311 \\ -0.454 \\ 0.835 \end{bmatrix}' + \frac{1}{415} \begin{bmatrix} 0.429 \\ 0.717 \\ 0.549 \end{bmatrix} \begin{bmatrix} 0.429 \\ 0.717 \\ 0.549 \end{bmatrix}'$$

$$= 10^{-4} \begin{bmatrix} 5.02 & 8.26 & 4.11 \\ 8.26 & 13.61 & 7.20 \\ 4.11 & 7.20 & 11.46 \end{bmatrix}$$

To compute the search direction we also need the following:

$$(\hat{\beta}'\hat{\beta} - \hat{\sigma}_\epsilon^2(k-1)/N) \mathbf{I} - \hat{\beta}\hat{\beta}' = \begin{bmatrix} 9.76 & 0.17 & 0.27 \\ 0.17 & 6.81 & -4.52 \\ 0.27 & -4.52 & 2.86 \end{bmatrix},$$

$$\hat{\Delta}\Sigma_z\hat{\gamma} = 10^4 \begin{bmatrix} -4.59 \\ -6.84 \\ 9.45 \end{bmatrix}$$

Then, the search direction is given by:

$$\begin{aligned}\hat{\mathbf{d}}^{(3)} &= 10^4 \begin{bmatrix} 9.76 & 0.17 & 0.27 \\ 0.17 & 6.81 & -4.52 \\ 0.27 & -4.52 & 2.86 \end{bmatrix} \begin{bmatrix} 5.02 & 8.26 & 4.11 \\ 8.26 & 13.61 & 7.20 \\ 4.11 & 7.20 & 11.46 \end{bmatrix} \\ &\times 10^4 \left( \begin{bmatrix} -4.59 \\ -6.84 \\ 9.45 \end{bmatrix} + \begin{bmatrix} 237.6 & 362.5 & -334.1 \\ 362.5 & 555.5 & -466.1 \\ -334.1 & -466.1 & 1284.6 \end{bmatrix} \begin{bmatrix} 0.5373 \\ 1.0588 \\ -0.7682 \end{bmatrix} \right) \\ &\propto \begin{bmatrix} 0.478 \\ 0.743 \\ -0.467 \end{bmatrix}.\end{aligned}$$

The results for the complete multidimensional search are given in Tables 5 and 5. It is seen how the computed search directions are useful in minimizing the variance of the response at the same time that the deviations of the mean response are small. In the first line search the variance is reduced from  $16.69 \times 10^4$  to  $10.02 \times 10^4$ , a reduction of 40%. In addition, we observe how the solution to the second dual response problem at  $t = 10$  incurs in an increase in variance compared to the last step of the previous search.

In Figure 5 we observe how the procedure takes an indirect path towards approaching the global minimum. This behavior is better observed in Figure 6 which is a projection of the path in Figure 5 onto the  $x_1 - x_2$  plane. This indirect path is taken to maintain the mean between the tolerance range specified. However, we observe that as the global optimum is approached, the search directions are not improving the objective. This is true for the search directions obtained from the dual response problems solved at  $t = 39$  and  $t = 41$ . The reason for this is that the two variance estimates being compared are not much different, and that the difference is similar to the uncertainty with which they are being estimated. The true value of the variance at  $\mathbf{x}_{39} = [-7.56, 6.67, 1.39]$ , is  $Var_{\mathbf{z}}(\eta(\mathbf{x}_{39}, \mathbf{z})) = 0.107 \times 10^4$  while at  $\mathbf{x}_{40} = [-8.05, 6.60, 1.37]$  is  $Var_{\mathbf{z}}(\eta(\mathbf{x}_{40}, \mathbf{z})) = 0.0977 \times 10^4$ . Then, the direction selected was indeed an improving direction. The problem is that the improvement was small enough to be confounded with the uncertainty in the two variance estimates. Notice that  $Var_{\mathbf{z}}(\eta(\mathbf{x}_{39}, \mathbf{z}))$  is actually a prediction from the variance model used to solve the correspondent problem (2) while  $Var_{\mathbf{z}}(\eta(\mathbf{x}_{40}, \mathbf{z}))$  is obtained from the estimates of a  $2^2$  design which are also subject to sampling variability.

Table 5. Simulation example of the new region exploration for RPD methodology, second example.

$t$	$\mathbf{x}_t$	$\widehat{Var}_{\mathbf{z}}(Y_t) \times 10^{-4}$	$\widehat{E}_{\mathbf{z}}(Y_t)$	DOE ran	Action	$\ \mathbf{x}_t - \mathbf{x}_g\ $
0	$[-11.84, -10.34, 8.04]$			$2_{VI}^{5-1}$	Start	20.00
1	$[-11.31, -9.28, 7.27]$	16.69	-20.00		Solution to Problem 2	18.71
2	$[-11.07, -8.91, 7.04]$	15.69	-19.81	$2^2$	Continue	18.27
3	$[-10.83, -8.54, 6.80]$	14.48	-20.66	$2^2$	Continue	17.84
4	$[-10.59, -8.17, 6.57]$	13.77	-18.89	$2^2$	Continue	17.41
5	$[-10.35, -7.80, 6.34]$	12.65	-18.76	$2^2$	Continue	16.98
6	$[-10.11, -7.43, 6.10]$	11.73	-18.11	$2^2$	Continue	16.56
7	$[-9.87, -7.05, 5.87]$	10.89	-17.76	$2^2$	Continue	16.14
8	$[-9.63, -6.68, 5.63]$	10.02	-17.74	$2^2$ and $2_V^{5-1}$	Continue	15.73
9	$[-9.39, -6.31, 5.40]$	9.11	-17.00	$2^2$	Stop, Mean Out	15.32
10	$[-9.30, -5.41, 6.15]$	9.28	-20.00		Solution to Problem 2	14.69
11	$[-9.07, -4.97, 6.05]$	8.66	-20.23	$2^2$	Continue	14.27
12	$[-8.85, -4.54, 5.95]$	7.88	-21.36	$2^2$	Continue	13.85
13	$[-8.63, -4.10, 5.85]$	7.49	-19.92	$2^2$	Continue	13.43
14	$[-8.40, -3.66, 5.75]$	6.77	-20.16	$2^2$	Continue	13.02
15	$[-8.18, -3.23, 5.65]$	6.22	-19.93	$2^2$	Continue	12.62
16	$[-7.96, -2.79, 5.56]$	5.73	-20.04	$2^2$	Continue	12.23
17	$[-7.73, -2.35, 5.46]$	5.21	-20.52	$2^2$	Continue	11.84
18	$[-7.51, -1.92, 5.36]$	4.66	-20.32	$2^2$	Continue	11.46
19	$[-7.28, -1.48, 5.26]$	4.28	-19.89	$2^2$	Continue	11.09
20	$[-7.06, -1.05, 5.16]$	3.92	-20.84	$2^2$	Continue	10.73
21	$[-6.84, -0.61, 5.06]$	3.45	-19.54	$2^2$	Continue	10.39

Table 6. Simulation example of the new region exploration for RPD methodology, second example (cont.)

t	$\mathbf{x}_t$	$\widehat{Var}_{\mathbf{z}}(Y_t) \times 10^{-4}$	$\widehat{E}_{\mathbf{z}}(Y_t)$	DOE ran	Action	$\ \mathbf{x}_t - \mathbf{x}_g\ $
22	[-6.61, -0.17, 4.96]	3.07	-21.09	$2^2$	Continue	10.05
23	[-6.39, 0.26, 4.86]	2.76	-21.04	$2^2$	Continue	9.73
24	[-6.17, 0.70, 4.76]	2.39	-20.29	$2^2$	Continue	9.43
25	[-5.94, 1.14, 4.67]	2.11	-20.90	$2^2$	Continue	9.14
26	[-5.72, 1.57, 4.57]	1.82	-21.22	$2^2$	Continue	8.87
27	[-5.50, 2.01, 4.47]	1.56	-21.63	$2^2$ and $2_V^{5-1}$	Continue	8.63
28	[-5.27, 2.44, 4.37]	1.37	-22.78	$2^2$	Stop, Mean Out	8.40
29	[-5.20, 3.02, 3.52]	0.791	-20.00	Solution to Problem 2		7.76
30	[-5.33, 3.44, 3.29]	0.664	-20.03	$2^2$	Continue	7.32
31	[-5.45, 3.87, 3.06]	0.491	-21.15	$2^2$	Continue	6.88
32	[-5.57, 4.29, 2.83]	0.434	-19.71	$2^2$	Continue	6.46
33	[-5.70, 4.72, 2.59]	0.315	-19.94	$2^2$	Continue	6.05
34	[-5.82, 5.15, 2.36]	0.246	-19.71	$2^2$	Continue	5.65
35	[-5.94, 5.57, 2.13]	0.199	-19.79	$2^2$	Continue	5.27
36	[-6.07, 6.00, 1.90]	0.174	-20.22	$2^2$	Continue	4.91
37	[-6.19, 6.42, 1.66]	0.154	-19.95	$2^2$ and $2_V^{5-1}$	Continue	4.58
38	[-6.31, 6.85, 1.43]	0.177	-19.41	$2^2$	Stop, Var Higher	4.28
39	[-7.56, 6.67, 1.39]	0.103	-20.00	Solution to Problem 2 and $2_V^{5-1}$		3.24
40	[-8.05, 6.60, 1.37]	0.114	-19.44	$2^2$	Stop, Var Higher	2.88
41	[-8.83, 7.23, 1.17]	0.081	-20.00	Solution to Problem 2		1.86
42	[-8.69, 7.11, 1.17]	0.098	-19.45	$2^2$	Stop, Var Higher	1.38

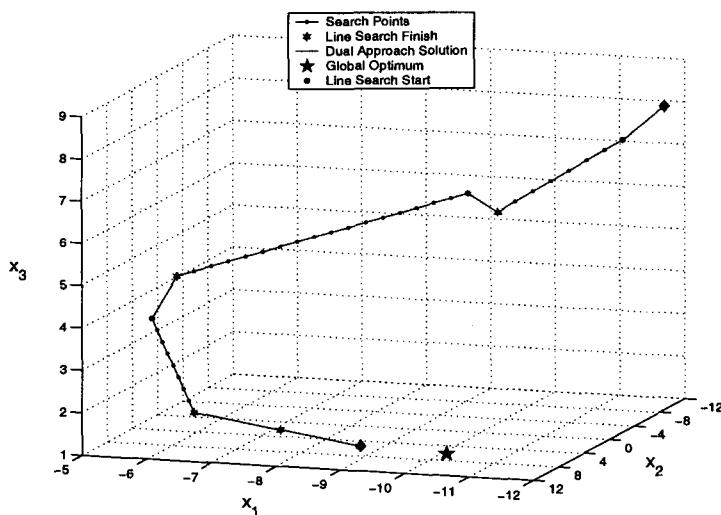


Fig. 5. Path of the  $\mathbf{x}_t$  points during the multidimensional search for the second example ( $k = 3$  and  $r = 2$ ).

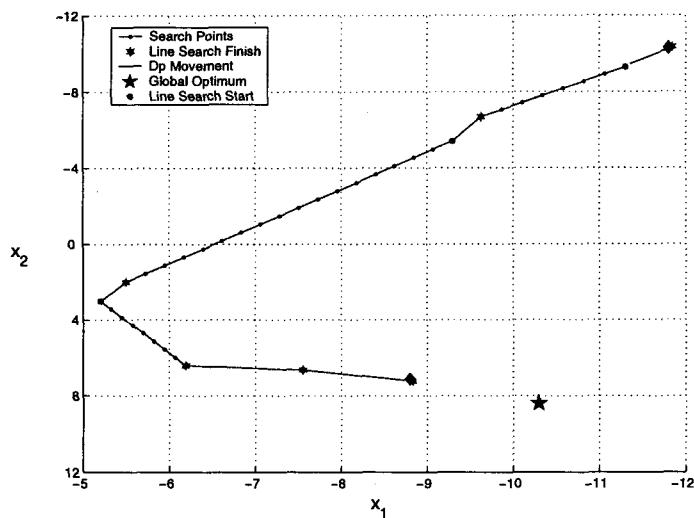


Fig. 6. Projection of the path in Figure 5 onto the  $x_1 - x_2$  plane, second example.

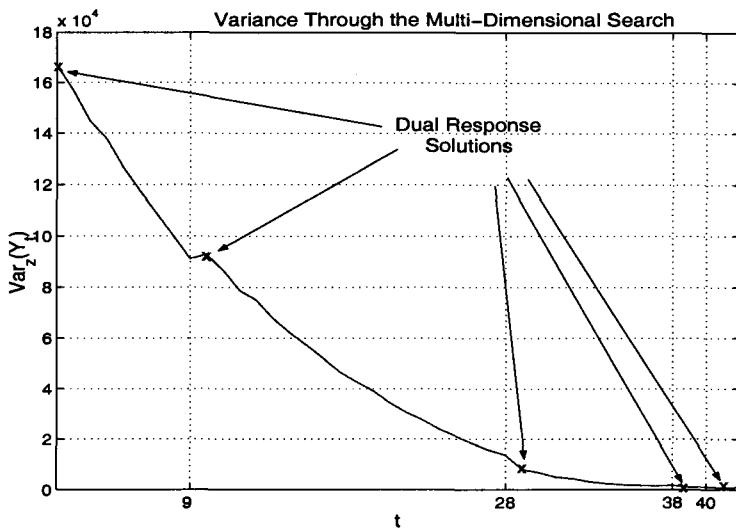


Fig. 7. Values of  $\widehat{Var}_z(Y_t)$  during the multidimensional search for the second example ( $k = 3$  and  $r = 2$ ).

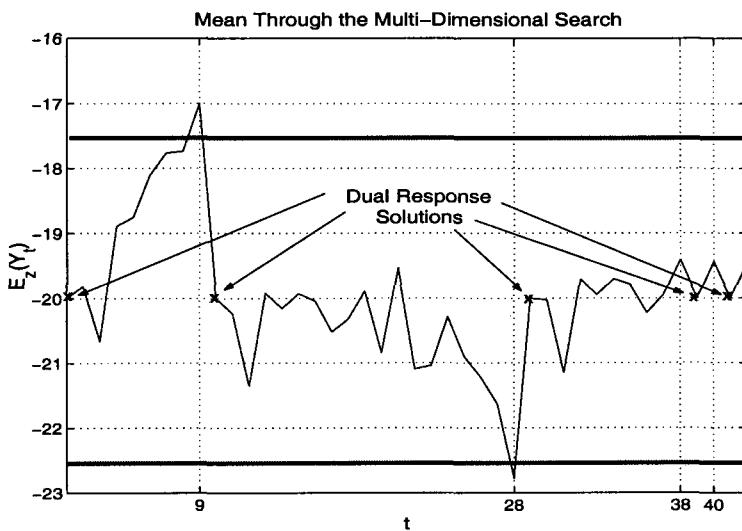


Fig. 8. Values of  $\widehat{E}_z(Y_t)$  during the multidimensional search for the second example ( $k = 3$  and  $r = 2$ ).

A similar situation happened in the search started at  $\mathbf{x}_{41} = [-8.83, 7.23, 1.17]$ . The “goodness” of the direction can be seen in the fact that the distance to the true optimum is reducing, as it can be seen in the last column of Tables 5 and 5. The procedure was stopped when no further reductions in variance were observed.

## 10. Conclusions and Discussion

A novel methodology was presented for exploring new regions when noise factors are present. Search directions were proposed for a “target is best”-type problem. It was demonstrated by example that the methodology is of practical value when an experimenter wishes to minimize the variability of a response for the common case when some tolerance is specified for the response.

As pointed out by a referee, a modification to the approach presented in this paper is to minimize the mean square deviation, i.e.,  $\min \text{MSE}(Y) = \text{Var}_z(Y) + (E_z(Y) - T)^2$  subject to  $\mathbf{x} \in R$ , instead of solving (2). The objective is still quadratic for the assumed models, and the search directions would be simple modifications of the ones presented in this paper. The problem of how to estimate these directions remains. As mentioned in section 5, ML estimators could be used in such case. The small sample properties of such directions remains a matter of further research.

## References

1. S.F. Arnold. *Mathematical Statistics*. Prentice Hall, Englewood Cliffs, NJ, 1990.
2. M. S. Bazaraa, H. D. Sherali, and C. M. Shetty. *Nonlinear Programming: Theory and Algorithms*. John Wiley & Sons, New York, 1993.
3. G.E.P. Box and N.R. Draper. *Empirical Model Building and Response Surfaces*. John Wiley & Sons, New York, 1987.
4. G.E.P. Box and S. Jones. Designing products that are robust to the environment. *Total Quality Management*, 3, 1-18, 1992.
5. G.E.P. Box and K.B. Wilson. On the experimental attainment of optimal conditions. *Journal of the Royal Statistical Society, Series B*, 13:1-45, 1951.
6. S. Joshi, H.D. Sherali, and J.D. Tew. An enhanced response surface methodology algorithm using gradient deflection and second-order search strategies. *Computers in Operations Research*, 25:531-541, 1998.
7. A.I. Khuri and J. A. Cornell. *Response Surfaces: Designs and Analyses*. Marcel Dekker, New York, 2nd edition, 1996.
8. R.H. Myers. Response surface methodology in quality improvement. *Communications in Statistics, Theory and Methods*, 20:457-476, 1991.

9. R.H. Myers. Response surface methodology: Current status and future directions. *Journal of Quality Technology*, 31(1):30–44, 1999.
10. R.H. Myers and D.C. Montgomery. *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. Wiley, New York, NY, 2002.
11. J. Neter, M.H. Kutner, and C.J. Nachtsheim. *Applied Linear Statistical Models*. WCB-McGraw-Hill, New York, NY, 2002.
12. S.R. Searle. *Linear Models*. Wiley, New York, NY, 1971.

### **Appendix 1. Cases When $\mathbf{d}^{(3)}$ Will Not be an Improving Direction**

Let us denote by  $\mathbf{J}$  the following matrix:

$$\mathbf{J} = \sum_{i \in I_+} \frac{\boldsymbol{\nu}_i \boldsymbol{\nu}'_i}{\lambda_i}$$

This is the matrix that pre-multiplies the gradient in the direction proposed by Joshi et al. (equation 5).

Assuming that the  $r \times k$  matrix  $\Delta$  has full row rank (recall that for this case  $k > r$ ), we have that  $\text{rank}(\mathbf{H}(\bar{\mathbf{x}}) = \Delta \Sigma_z \Delta') = r$  with all the non-zero eigenvalues being positive. Therefore, there are  $r$  elements in  $I_+$  and  $\text{rank}(\mathbf{J}) = r$  because it is made of the sum of the outer-product of  $r$  orthogonal vectors. Whenever the gradient is a linear combination of the eigenvectors corresponding to the eigenvalues not included in  $I_+$ , that is:

$$\nabla \eta(\bar{\mathbf{x}}) = \sum_{i \notin I_+} a_i \boldsymbol{\nu}_1 ,$$

then:

$$\mathbf{J} \nabla \eta(\bar{\mathbf{x}}) = \mathbf{0}$$

and  $\mathbf{d}^{(3)}$  will not be an improving direction. Of course if  $\mathbf{J} \nabla \eta(\bar{\mathbf{x}}) \neq \mathbf{0}$  but  $\nabla \eta(\bar{\mathbf{x}}) = \lambda \boldsymbol{\beta}$ , then the resulting projected direction  $\mathbf{d}^{(3)}$  will also be zero.

### **Appendix 2. Analysis of the Estimation of $\text{Var}_{\mathbf{z}}(Y)$ Along the Search**

The estimator in equation (18) is biased:

$$\begin{aligned} E_{\gamma, \hat{\sigma}_\epsilon^2} \left( \widehat{\text{Var}}_{\mathbf{z}}(Y_t) \right) &= E_{\gamma, \hat{\sigma}_\epsilon^2} (\hat{\gamma}'_t \Sigma_z \hat{\gamma}_t + \hat{\sigma}_\epsilon^2) \\ &= \boldsymbol{\gamma}' \Sigma_z \boldsymbol{\gamma} + \text{tr}(\mathbf{V} \Sigma_z) , \end{aligned}$$

where  $\mathbf{V} = Var(\hat{\gamma})$ , the covariance matrix of  $\hat{\gamma}$ , which is available from the regression model. Now, denoting the predictions from this model as  $\hat{Y}_t = \hat{\gamma}'\mathbf{z} + \epsilon$ , we obtain that the variance of these predictions, with respect to the residual noise  $\epsilon$ , the parameter estimates  $\gamma$  and the noise factors  $\mathbf{z}$  is given by:

$$\begin{aligned} Var_{\hat{\gamma}, \mathbf{z}, \epsilon} (\hat{Y}_t) &= Var_{\hat{\gamma}, \mathbf{z}, \epsilon} (\hat{\gamma}'\mathbf{z} + \epsilon) \\ &= Var_{\hat{\gamma}, \mathbf{z}} (\hat{\gamma}'\mathbf{z}) + \sigma_\epsilon^2 \\ &= Var_{\mathbf{z}} (E_{\hat{\gamma}} (\hat{\gamma}'\mathbf{z}|\mathbf{z})) + E_{\mathbf{z}} (Var_{\hat{\gamma}} (\hat{\gamma}'\mathbf{z}|\mathbf{z})) \\ &= Var_{\mathbf{z}} (\gamma'\mathbf{z}) + E_{\mathbf{z}} (\mathbf{z}'\mathbf{V}\mathbf{z}) \\ &= \gamma' \Sigma_{\mathbf{z}} \gamma + tr (\mathbf{V} \Sigma_{\mathbf{z}}). \end{aligned}$$

Hence:

$$E_{\gamma, \hat{\sigma}_\epsilon^2} (\widehat{Var}_{\mathbf{z}} (Y_t)) = Var_{\hat{\gamma}, \mathbf{z}, \epsilon} (\hat{Y}_t),$$

that is, the estimator in equation (18) is unbiased in estimating the variance in the predictions. However, if the same design is used in every step along the search, the term  $tr (\mathbf{V} \Sigma_{\mathbf{z}})$  remains constant. Therefore, correcting by it will not have an effect on when to stop the search or on the assessment if the variance is still increasing.

### **Appendix 3. Response Surfaces for the Mean and Variance of the Test Function**

The test function used in the simulation to study the methodology for new region exploration in RPD is given by:

$$\eta(\mathbf{x}, \mathbf{z}) = \beta_0 + \beta' \mathbf{x}_s + \mathbf{x}_s' \mathbf{B} \mathbf{x}_s + \rho' \mathbf{x}_s^{(4)} + \mathbf{x}_s' \Delta \mathbf{z} + \gamma' \mathbf{z} + \mathbf{z}' \mathbf{C}(\mathbf{x}) \mathbf{z} + \epsilon \quad (.1)$$

where it is assumed that  $\epsilon \sim N(0, \sigma_\epsilon^2)$  and  $\mathbf{z} \sim N(\mathbf{0}, \Sigma_{\mathbf{z}})$  each independent of the other.

To obtain the mean value of the true response over the noise variables,  $E_{\mathbf{z}}(\eta(\mathbf{x}, \mathbf{z}))$ , we simply use the usual formula for the expected value of a quadratic form:

$$\begin{aligned} E_{\mathbf{z}}(\eta(\mathbf{x}, \mathbf{z})) &= E_{\mathbf{z}}(\beta_0 + \beta' \mathbf{x}_s + \mathbf{x}_s' \mathbf{B} \mathbf{x}_s + \rho' \mathbf{x}_s^{(4)} + \mathbf{x}_s' \Delta \mathbf{z} + \gamma' \mathbf{z} + \mathbf{z}' \mathbf{C}(\mathbf{x}) \mathbf{z} + \epsilon) \\ &= E_{\mathbf{z}}(\beta_0 + \beta' \mathbf{x}_s + \mathbf{x}_s' \mathbf{B} \mathbf{x}_s + \rho' \mathbf{x}_s^{(4)}) + E_{\mathbf{z}}(\mathbf{x}_s' \Delta \mathbf{z}) + E_{\mathbf{z}}(\mathbf{z}' \mathbf{C}(\mathbf{x}) \mathbf{z}) \end{aligned}$$

Since the first term in the right-hand side does not have any random variables, we have that:

$$E_{\mathbf{z}} (\beta_0 + \boldsymbol{\beta}' \mathbf{x}_s + \mathbf{x}'_s \mathbf{B} \mathbf{x}_s) = \beta_0 + \boldsymbol{\beta}' \mathbf{x}_s + \mathbf{x}'_s \mathbf{B} \mathbf{x}_s .$$

The second term in the right hand side is zero because:

$$E_{\mathbf{z}} (\mathbf{x}'_s \Delta \mathbf{z}) = \mathbf{x}' \Delta E_{\mathbf{z}} (\mathbf{z}) = \mathbf{x}' \Delta \mathbf{0} = 0 .$$

The last term implies taking the expectation of a quadratic form (see for example Arnold<sup>1</sup>, p. 204):

$$\begin{aligned} E_{\mathbf{z}} (\mathbf{z}' \mathbf{C}(\mathbf{x}_s) \mathbf{z}) &= E_{\mathbf{z}} (\mathbf{z}') \mathbf{C}(\mathbf{x}_s) E_{\mathbf{z}} (\mathbf{z}) + \text{tr} (\mathbf{C}(\mathbf{x}) \Sigma_{\mathbf{z}}) \\ &= \mathbf{0}' \mathbf{C}(\mathbf{x}_s) \mathbf{0} + \text{tr} (\mathbf{C}(\mathbf{x}) \Sigma_{\mathbf{z}}) \\ &= \text{tr} (\mathbf{C}(\mathbf{x}) \Sigma_{\mathbf{z}}) . \end{aligned}$$

The response surface for the mean is given by:

$$E_{\mathbf{z}} (\eta(\mathbf{x}, \mathbf{z})) = \beta_0 + \boldsymbol{\beta}' \mathbf{x}_s + \mathbf{x}'_s \mathbf{B} \mathbf{x}_s + \rho' \mathbf{x}_s^{(4)} + \text{tr} (\mathbf{C}(\mathbf{x}) \Sigma_{\mathbf{z}}) .$$

To take the variance operator on  $\eta(\mathbf{x}, \mathbf{z})$  over the noise factors  $\mathbf{z}$  we should first take common factor:

$$\begin{aligned} Var_{\mathbf{z}} (\eta(\mathbf{x}, \mathbf{z})) &= Var_{\mathbf{z}} (\beta_0 + \boldsymbol{\beta}' \mathbf{x}_s + \mathbf{x}'_s \mathbf{B} \mathbf{x}_s + \rho' \mathbf{x}_s^{(4)}) \\ &\quad + Var_{\mathbf{z}} ((\mathbf{x}'_s \Delta + \gamma)' \mathbf{z} + \mathbf{z}' \mathbf{C}(\mathbf{x}) \mathbf{z} + \epsilon) + \sigma_{\epsilon}^2 . \end{aligned}$$

The first term in the right-hand side is equal to zero because neither the controllable factors  $\mathbf{x}$ , nor the parameters  $\beta_0, \boldsymbol{\beta}, \mathbf{B}$ , and  $\rho$  are random variables. Therefore:

$$\begin{aligned} Var_{\mathbf{z}} (\eta(\mathbf{x}, \mathbf{z})) &= Var_{\mathbf{z}} ((\mathbf{x}'_s \Delta + \gamma)' \mathbf{z}) + Var_{\mathbf{z}} (\mathbf{z}' \mathbf{C}(\mathbf{x}) \mathbf{z}) \\ &\quad + 2 \text{Cov} ((\mathbf{x}'_s \Delta + \gamma)' \mathbf{z}, (\mathbf{z}' \mathbf{C}(\mathbf{x}) \mathbf{z})) + \sigma_{\epsilon}^2 . \end{aligned}$$

However, the covariance term (third on the right-hand side) only contains odd order moments zero-mean normally distributed random variables, and hence they are all zero. Then the first term on the right-hand side becomes:

$$Var_{\mathbf{z}} ((\mathbf{x}'_s \Delta + \gamma)' \mathbf{z}) = (\mathbf{x}'_s \Delta + \gamma)' \Sigma_{\mathbf{z}} (\mathbf{x}'_s \Delta + \gamma) .$$

For the second term we need to use Theorem 1 from Searle<sup>12</sup> (p. 55) for taking the variance of a quadratic form. This also requires the normality of  $\mathbf{z}$ :

$$\begin{aligned} Var_{\mathbf{z}}(\mathbf{z}' \mathbf{C}(\mathbf{x}) \mathbf{z}) &= 4E_{\mathbf{z}}(\mathbf{z}') \mathbf{C}(\mathbf{x}) \Sigma_{\mathbf{z}} \mathbf{C}(\mathbf{x}) E_{\mathbf{z}}(\mathbf{z}) + 2tr(\mathbf{C}(\mathbf{x}) \Sigma_{\mathbf{z}} \mathbf{C}(\mathbf{x}) \Sigma_{\mathbf{z}}) \\ &= 2tr(\mathbf{C}(\mathbf{x}) \Sigma_{\mathbf{z}} \mathbf{C}(\mathbf{x}) \Sigma_{\mathbf{z}}) \end{aligned}$$

since the expected value of the noise factors is zero. Then the response surface for the variance is given by:

$$Var_{\mathbf{z}}(\eta(\mathbf{x}, \mathbf{z})) = (\Delta' \mathbf{x}_s + \gamma)' \Sigma_{\mathbf{z}} (\Delta' \mathbf{x}_s + \gamma) + 2tr(\mathbf{C}(\mathbf{x}) \Sigma_{\mathbf{z}} \mathbf{C}(\mathbf{x}) \Sigma_{\mathbf{z}}) + \sigma_{\epsilon}^2.$$

This page is intentionally left blank

## **CHAPTER 6**

### **RESPONSE SURFACE APPROACHES TO ROBUST PARAMETER DESIGN**

Timothy J. Robinson

*Department of Statistics, University of Wyoming*

Shaun S. Wulff

*Department of Statistics, University of Wyoming*

In manufacturing, quality is essential for the survival of the process. Assuming an appropriate response for identifying quality can be obtained, the goal of the researcher is to find operating conditions in which the response is minimized, maximized, or reaches some target. In the 1980's Genichi Taguchi, a Japanese quality consultant, popularized the notion that quality is also a function of the variability of the process. As such, he noted that frequently, most of the variability associated with the response can be attributed to the presence of a set of factors known as noise factors (parameters). To study the influence of these variables on the process, Taguchi encouraged the use of these variables in the experimental design stage of process improvement. During the last 15 years, a great deal of attention in the literature has been given to the use of response surface techniques for studying and minimizing the impact of noise variables on manufacturing processes. The topic is commonly known as robust parameter design (RPD). The phrase, robust parameter design, implies the design of a process to be robust to the inevitable changes of the noise parameters. In this chapter, we provide a historical account of the application of response surface methodology to the RPD problem.

## 1. Introduction

Before describing the applications of response surface methodology to the RPD problem, it is helpful to reflect on the historical background of RPD. Formalization of RPD in the U.S. began with Japanese quality consultant Genichi Taguchi in the mid 1980's. Parameter design is an engineering methodology intended as a cost-effective approach for improving the quality of products and processes. The assumption is that factors in a system can be divided into two groups: control factors (the vector is commonly denoted by  $x$ ) and noise factors (the vector is commonly denoted by  $z$ ). Control factors are those factors whose levels remain unchanged in the process once their levels are selected, whereas the levels of noise factors, which are hard or expensive to control in the production setting, are assumed to vary at random within the process. Examples of noise factors include: environmental conditions such as temperature and humidity, variation in raw materials, product aging, and tolerances in the control factors. The lack of stability of the noise factors in the process can transmit unwanted variation to the response. The phrase *robust parameter design* (RPD refers to the *design of a process* (not to be confused with experimental design) in such a way as to operate at levels of the control factors that make the process *robust* or insensitive to the inevitable changes in the noise factors.

Taguchi not only called attention to the philosophy of RPD but he also proposed experimental design and analysis strategies for identifying the settings of  $x$  that would achieve robust performance. The experimental designs proposed by Taguchi are known as orthogonal arrays. Orthogonal arrays cross an orthogonal design in  $x$  (known as the 'control' or 'inner array') with an orthogonal design involving  $z$  (the 'outer' or 'noise array').

We illustrate the crossed array with an example that will be used throughout the chapter. Engel<sup>1</sup> reports an experiment intended to improve an injection molding process. The goal of the experimenter was to determine the process variable settings for which percentage shrinkage would be closest to a target value. Seven control factors and three noise factors were under investigation. Table 1 lists the factors and the corresponding data. The experimental design in the control factors is a

Table 1. Injection Molding Data from Engel<sup>1</sup>

A	B	C	D	E	F	G	% shrinkage (M,N,O)			
							-1	-1	1	1
							-1	1	-1	1
-1	-1	-1	-1	-1	-1	-1	2.2	2.1	2.3	2.3
-1	-1	-1	1	1	1	1	0.3	2.5	2.7	0.3
-1	1	1	-1	-1	1	1	0.5	3.1	0.4	2.8
-1	1	1	1	1	-1	-1	2.0	1.9	1.8	2.0
1	-1	1	-1	1	-1	1	3.0	3.1	3.0	3.0
1	-1	1	1	-1	1	-1	2.1	4.2	1.0	3.1
1	1	-1	-1	1	1	-1	4.0	1.9	4.6	2.2
1	1	-1	1	-1	-1	1	2.0	1.9	1.9	1.8

$2^{7-4}$  factorial design and the experimental design in the noise factors is a  $2^{3-1}$  factorial design. The 'crossing' of these two orthogonal designs implies that every control factor setting occurs with every noise factor setting and vice-versa, producing 32 runs.

Taguchi proposed the crossed orthogonal array due to the assumptions that the process mean and variance at the  $i^{\text{th}}$  setting of the control factors is the same over all  $j$  settings of the noise factors. Consequently, the responses at the  $i^{\text{th}}$  setting of the control factors can be treated as 'replicates' which could be used to estimate the process variance at the  $i^{\text{th}}$  setting of the control factors. These summary statistics then provide information about the process mean and variance. Taguchi proposed combining the sample mean and variance at each of the control factor settings into a single performance measure known as the signal-to-noise ratio (SNR). The statistical analysis would then be done on the SNR. The formulation of the SNR depends on the goal of the researcher. Taguchi claims that, in general, the researcher's goals for the response variable fall into one of three possible categories: minimizing, maximizing, or obtaining a target. If for example, the goal is for the

process mean to reach a targeted value, a proposed formulation of the SNR would be

$$-10 \log s_i^2 \quad (1)$$

where,  $s_i^2$  denotes the sample variance of the runs at the  $i^{th}$  setting of the control factors. This formulation of the SNR is based upon a squared error loss criterion,

$$E_z(y - T)^2 = Var(y) + (E[y - T])^2$$

where  $T$  is the desired target of the mean response. The formulation in (1) assumes that the user can identify sufficient *adjustment or tuning factors*. Adjustment factors are those control factors that allow the analyst to alter the mean but allow the variance to be unchanged. If such factors exist, then the squared error loss criterion reduces to  $Var(y)$  and the formulation of the SNR in (1) is intuitive and by analyzing the sample means and SNR, the user can determine which factors influence the mean and which influence the variance. Assuming the presence of adjustment factors, the analysis is a two-step process: Select the levels of the adjustment factor(s) which bring the mean to target, and then select levels of the other control factors which maximize the SNR. Details of Taguchi's methods of analysis of the SNR are provided in Taguchi<sup>2</sup>.

As an illustration of Taguchi's approach to RPD, consider the data in Table 1. The main effect plots of the sample means and the SNRs versus the levels of the control factors are provided in Figures 1 and 2, respectively. In this data set, the goal is to achieve a target reduction percentage of 1.5. It is also noteworthy that the variance is a function of the mean. Taguchi's SNR for settings when the goal is 'target is best' and the variance depends on the mean is  $10 \log \left( \frac{\bar{y}_i^2}{s_i^2} \right)$ . When the

variance is a function of the mean, adjustment factors are sought which bring the mean to target but leave the coefficient of variation unaffected. The user then seeks to find the settings of  $x$  that maximize the SNR. From Figures 1 and 2, factors  $A$ ,  $D$ , and  $G$  appear to influence the mean while factor  $F$  appears to influence the SNR. Consequently, factors  $A$ ,  $D$

and  $G$  would be used to bring the mean to target and factor  $F$  would be held at its low level in order to maximize the SNR.

Taguchi's contributions to parameter design can be summarized into three areas—quality philosophy and practice, experimental design, and data analysis. No one will argue that quality philosophy must incorporate both the mean and variance of a quality measure. However, since the late 1980's, many important works have been published which more effectively address the quality philosophy in terms of experimental design and data analysis. Nair<sup>3</sup> brought together leading researchers in RPD in a panel discussion to summarize the directions of research in RPD that would shape the work in this area until the present time.

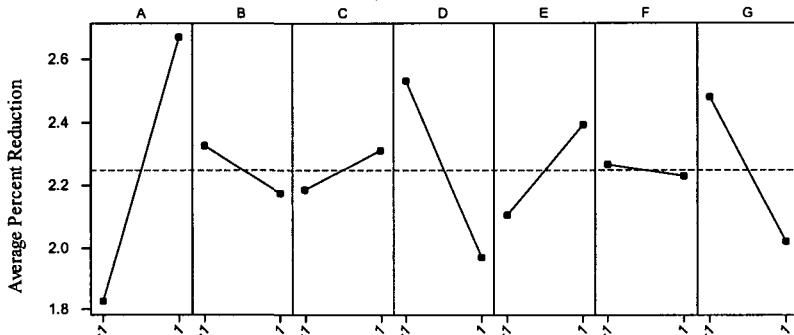


Fig. 1. Main effect plots of the sample means vs. control factors.

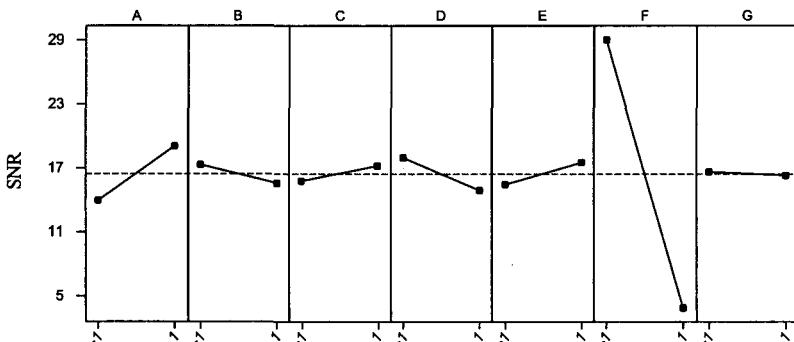


Fig. 2. Main effect plots of SNRs vs. control factors.

Several criticisms of the Taguchi approach are worth noting as they provide motivation for the response surface approach to robust design. Taguchi's approach is a 'one-shot' experiment with the goal being to pick the optimal setting of  $x$ . Many would argue that process optimization is overrated and that it is more important to understand the system as a whole [see for instance Box<sup>4</sup>, Nair and Pregibon<sup>5</sup>, and Myers and Montgomery<sup>6</sup> (henceforth referred to as MM)]. The use of the SNR does not guarantee that the researcher will be able to uncouple those control factors which influence the mean and which control factors influence the variance (*dispersion effects*). As an illustration, consider the SNR given in expression (1). It is clear the SNR confounds location and dispersion since maximization of (1) does not distinguish which control factors influence the mean and which influence the variance. A more informative approach would be one that could uncouple those effects influencing the mean and those influencing the variance. Leon, Shoemaker, and Kacker<sup>7</sup> suggested the analysis of performance measures independent of adjustment (PerMIA) instead of the traditional SNRs. Logothetis<sup>8</sup> used transformations to uncouple mean and dispersion effects.

In determining which control factors are adjustment factors (those influencing only the mean), Taguchi utilizes main effect plots of the  $\bar{y}_i$  and  $\text{SNR}_i$  versus the individual control factors. This approach assumes 'main-effect' only models relating the mean (or variance) and the control factors. Taguchi rarely considers interactions among the control factors. This is evident in the non-estimability of these interactions in many of the crossed array designs proposed by Taguchi. Box and Jones<sup>9,10</sup> point out that many crossed arrays are often run as split-plot designs, but this is also not accounted for in Taguchi's analysis.

Many of the criticisms of Taguchi are addressed by focusing on model building for the mean and variance. The concept of model building lends itself to the use of response surface methodology (RSM) as an appealing approach to the robust design problem. The use of RSM also brings to issue the types of designs proposed by Taguchi. RSM designs have been effectively utilized in the RPD problem and will be the topic of the third section. The various types of response surface

models that have been proposed for the robust design problem will be the topic of the next section.

## 2. Response Surface Models for Robust Design

As we have mentioned, understanding a system involves understanding which variables impact the mean of the system and which variables impact the variance of the system (some variables affect both). The concept of response surface methodology (RSM) is that of sequential experimentation for building appropriate models that enable one to understand the engineering system. Box and Wilson<sup>11</sup> are considered the pioneers of RSM. Fundamental to response surface methodology are model(s) that relate the response of interest to a set of independent variables. Many RSM models for robust design have been proposed over the last decade, depending on the type of design (crossed vs. combined array, complete randomization or split-plot randomization), the distribution of the data, whether the noise effects are fixed or random, whether the mean and variance are assumed a function of only control factors or both control and noise factors, and whether the noise factors are continuous or categorical. These considerations will be addressed by the following subsections: the dual response models for crossed arrays with normal responses, the response method approach for combined arrays, the use of generalized linear models for non-normal data, and a section on miscellaneous RSM models for RPD.

### 2.1. Dual Response Models for the Crossed Array with Normal Responses

In the RSM approach to solve RPD problems, a designed experiment is performed which enables one to model the relationship between the control and noise factors and the response mean and variance. When the experimental design is a crossed array, Box<sup>4</sup>, and Vining and Myers<sup>12</sup> (henceforth referred to as VM), proposed separate regression models for the process mean and variance where the sample means and variances at each setting of  $\mathbf{x}$  would serve as the data for the models. Bartlett and

Kendall<sup>13</sup> were the first to study variance modeling, proposing the following model

$$\log(s_i^2) = \mathbf{x}_i^\top \boldsymbol{\gamma} + \varepsilon_i^*, \quad (2)$$

where,  $s_i^2$  is the sample variance taken from  $n$  replications at the  $i^{th}$  level of the control factors ( $\mathbf{x}_i$ ) and  $\boldsymbol{\gamma}$  is a  $k \times 1$  vector of dispersion effects. The log transformation of the sample variances is a natural transformation in order to assure positive variance estimates. Assuming that the errors in the means model are independent and normal and that their variance can be modeled as  $\log(\sigma_i^2) = \mathbf{x}_i^\top \boldsymbol{\gamma}$ , Bartlett and Kendall pointed out that the errors in (2) are normal with constant variance. Under these assumptions, the maximum likelihood estimates of the dispersion effects can be calculated via ordinary least squares. The means model is assumed to be

$$y_{ij} = \mathbf{x}_{ij}^\top \boldsymbol{\beta} + \varepsilon_{ij} \quad (3)$$

where  $y_{ij}$  is the observation at the  $i^{th}$  setting of the control factors and the  $j^{th}$  setting of the noise factors,  $\varepsilon_{ij} \sim N(0, \sigma_i^2)$ , and  $\boldsymbol{\beta}$  is the  $l \times 1$  vector of location effects. Mean and variance model parameter estimates can be obtained via the following estimated weighted least squares (EWLS) algorithm.

Step 1. Using some function,  $g$ , of the sample variances (the  $s_i^2$ ) as data for the variance model, fit the variance model

$$g(s_i^2) = \mathbf{u}_i^\top \boldsymbol{\gamma} + \varepsilon_i^*.$$

Here,  $\mathbf{u}_i^\top \boldsymbol{\gamma}$  is used instead of  $\mathbf{x}_i^\top \boldsymbol{\beta}$  to allow for the fact that not all control factors which influence the mean, influence the variance.

Step 2. Use  $\hat{v}_i = g^{-1}(\mathbf{u}_i^\top \hat{\boldsymbol{\gamma}})$  as estimated weights to compute  $\hat{\mathbf{V}} = \text{diag}\{\hat{v}_1, \hat{v}_2, \dots, \hat{v}_n\}$  where  $\hat{\boldsymbol{\gamma}}$  comes from Step 1.

Step 3. Use  $\hat{\mathbf{V}}$  as the estimated weight matrix to fit the mean model,  $E(y_i) = \mathbf{x}_i^\top \boldsymbol{\beta}$ , where  $\hat{\boldsymbol{\beta}} = (\mathbf{X}' \hat{\mathbf{V}}^{-1} \mathbf{X})^{-1} \mathbf{X}' \hat{\mathbf{V}}^{-1} \mathbf{y}$  is the EWLS estimator of  $\boldsymbol{\beta}$ .

To illustrate use of this algorithm, we return to the data provided in Table 1. We begin by obtaining a variance model using the natural log of the sample variances as responses. Since there are seven control factors and only eight distinct levels of the control factors, the design is saturated and so a normal probability plot is utilized to determine important main effects. Based on a normal probability plot, holding time ( $F$ ) is the only control factor influencing process variance. Consequently, the variance model

$$\ln(s_i^2) = \gamma_o + \gamma_F F + \varepsilon_i^*$$

is fit via OLS, producing the estimated variance model

$$s_i^2 = \hat{\sigma}_i^2 = \exp(-2.23 + 2.86F). \quad (4)$$

From (4), half of the data is assumed to display a process variance of  $e^{0.63} = 1.877$  ( $F=1$ ) and the other half a process variance of  $e^{-5.09}=0.006$  ( $F=-1$ ). Consequently, using  $\hat{w}_i^{-1} = \exp(2.23 - 2.86F_i)$  as weights, half of the observations will be weighted by a value of  $1/0.006 = 162.39$  and the other half by a value of  $1/1.877 = 0.53$  in an EWLS fit to the process mean model

$$y_{ij} = \mathbf{x}_i' \boldsymbol{\beta} + \varepsilon_{ij},$$

where  $\mathbf{x}_i'$  contains the linear main effects of the seven control factors. The resulting estimate of the process mean is given by

$$\widehat{E(y_{ij})} = 2.27 + 0.19A - 0.36D + 0.21E. \quad (5)$$

Consequently, cycle time ( $A$ ), holding pressure ( $D$ ), and injection speed ( $E$ ) are determined to affect the mean. The main effect plots in Figure 1, suggest that cycle time ( $A$ ), holding pressure ( $D$ ), and gate size ( $G$ ) influence the mean. The difference in conclusions between observing the main effect plots in Figure 1 and the EWLS analysis above could be due to the fact that the non-constant variance is accounted for in EWLS but not in the observance of main effect plots. Note that holding time ( $F$ )

appeared to be a dispersion effect in Figure 2 and the EWLS analysis confirms this fact.

A variety of forms of the above algorithm have been proposed over the years, mostly due to differing methodologies for variance estimation. Engel<sup>1</sup> addressed the uncoupling of mean and dispersion effects by extending the ideas of Logothetis<sup>8</sup>. Specifically, Engel assumed that the process variance could be written as the product of two components:  $Var(y_i) = \phi_i V(\mu_i, \theta)$ . Here,  $\phi_i$  denotes an overdispersion parameter, which is assumed to be a function of those control factors that only influence the variance (and thus independent of the mean), and  $V(\mu_i, \theta)$  is the variance function, assumed to have the form  $V(\mu_i, \theta) = \mu_i^\theta$ . The utility of this approach is that the mean is extracted from the variance via  $V(\mu_i, \theta)$ . This ‘power of the mean’ variance function is a special case of Davidian and Carroll<sup>14,15</sup>. Engel estimates the mean and variance model parameters via an EWLS algorithm that is similar to the one given above and details may be found in Engel. The analysis yields similar estimates of the mean and variance as those previously illustrated with the VM approach.

MM present an appealing approach to modeling the sample variances which exploits the assumption that  $s_i^2(n_i - 1) / \sigma_i^2 \sim \chi_{n_i-1}^2$ . Since the  $\chi_{df}^2$  is a special case of a gamma distribution with dispersion parameter  $2/df$ , the sample variances can be modeled as function of the  $\mathbf{u}_i$  using generalized linear models (GLM) [McCullagh and Nelder<sup>16</sup>]. A popular link function for modeling sample variances is the log link. The EWLS algorithm above can then be revised in Step 1 by using a gamma GLM on the  $s_i^2$  with log link and dispersion parameter  $2/df$  to fit the variance model  $\ln(s_i^2) = \mathbf{u}_i' \boldsymbol{\gamma}$ . The modeling of GLMs has become more feasible over the last decade with the advances in software. The SAS procedure GENMOD is useful for modeling almost any GLM. Other examples of variance modeling in crossed-array experiments can be found in Grego<sup>17</sup> and Vining and Bohn<sup>18</sup>. Vining and Schaub<sup>19</sup> discuss strategies for replicating runs in designs so as to obtain enough information for estimating a variance model.

Models for the mean and variance are more appealing than the main effects analyses of the sample means and SNRs for a variety of reasons. Models not only help one more fully understand the underlying process, but they also enable one to identify which control factors affect the mean, which affect the variance, and which affect both. The models in (2) and (3) also enable one to study possible interactions among the control factors, assuming a suitable experimental design. It should also be pointed out that the EWLS algorithm above enables one to estimate the mean while accounting for the heterogeneity of variances present at each of the settings of the control factors.

With an estimate of the mean and variance via regression modeling, VM appealed to the dual response problem (see Myers and Carter<sup>20</sup>) to find optimal operating conditions for the control factors. The dual response problem involves determining the set of operating conditions  $\mathbf{x} = \mathbf{x}^*$  which optimize a primary response  $y_p$  subject to  $y_s = \theta$ , where  $\theta$  is some acceptable value of the secondary response,  $y_s$ . Since robust design involves  $\mu_y$  and  $\sigma_y^2$ , the determination of which characteristic is the primary response depends on the ultimate goal of the experiment. If sufficient adjustment factors can be determined (those factors whose coefficients are significant in (3), but not in (2), and whose levels can move the mean to target), the two-step method by Taguchi is approached by minimizing  $\hat{\sigma}_y^2$  subject to the constraint  $\hat{\mu}_y = T$ . After estimating the process mean and variance from the models in (2) and (3), Lagrangian multipliers are used to find the settings of  $\mathbf{x}$  which optimizes the primary response  $y_p$  subject to  $y_s = \theta$ .

The constraints in the optimization problem utilized by VM all involve equalities. Del Castillo and Montgomery<sup>21</sup> [henceforth referred to as DM] pointed out that the technique of VM does not always produce local optima and propose the use of nonlinear programming for determining optimal operating conditions. DM proposed the use of a generalized reduced gradient (GRG) algorithm to optimize VM's problem with inequality constraints. Nonlinear programming is appealing for a variety of reasons: 1. Constraints can be formulated that are appropriate for the experimental design region, 2. Many software packages accommodate it (including SAS and EXCEL), 3. Nonlinear programming can accommodate a range of possible values for the

secondary response. Fathi<sup>22</sup> also proposed using nonlinear programming techniques to solve the dual response problem.

To illustrate the approach of DM, we again refer to the RPD data in Table 1 where the goal is to obtain operating conditions  $\mathbf{x}=\mathbf{x}^*$  such that the mean is close to a target,  $T=1.5$  and the variance is minimized. Using the estimate of the mean given in (5), the constraint

$$\hat{\mu} - T = \left[ (2.27 + 0.19A - 0.36D + 0.21E) - 1.5 \right] \quad (6)$$

is formulated and the generalized reduced gradient method is utilized to minimize

$$\hat{\sigma}_i^2 = \exp(-2.23 + 2.86F). \quad (7)$$

The resulting optimal settings are  $A=-1$ ,  $D=1$ ,  $E=-1$ , and  $F=-1$ , yielding an estimated process mean of 1.51 (recall  $T=1.5$ ) and process variance of 0.006.

Instead of subjecting the primary response to a specific value and then minimizing or maximizing the other response, Lin and Tu<sup>23</sup> [henceforth referred to as LT] proposed a different formulation. LT point out that subjecting the primary response to a specific value essentially admits no bias. If one is willing to admit a little bias in the response, it is often the case that a substantial reduction in response variability can be obtained. As a result, the following solution is proposed: 1. Find models for the mean and variance. 2. Find  $\mathbf{x} = \mathbf{x}^*$  such that the estimated MSE,  $\widehat{MSE(y)} = (\hat{\mu}_y - T)^2 + \hat{\sigma}_y^2$ , is minimized. In many cases, the approach by LT has been demonstrated to provide superior solutions to that of DM. For the data set in Table 1, DM and LT provide the same optimal solution.

A criticism of LT is that minimizing mean squared error places no restriction on how far the resulting value of  $\mu_y$  might be from the target,  $T$ . In settings where it is crucial to keep the mean on target, the method of LT may not be sufficient. Copeland and Nelson<sup>24</sup> [henceforth referred to as CN] address this concern by suggesting to formulate the problem as follows: minimize  $\hat{\sigma}_y$  such that  $(\hat{\mu}_y - T) \leq \Delta^2$ . In this case, one is interested in obtaining a solution for  $\mu_y$  that is within some specified  $\Delta$

from target. For instance, if it is necessary that  $\mu_y$  be within  $\Delta^*$  of  $T$  with probability  $1-\alpha$  and  $\hat{\mu}_y \pm \delta$  is a  $(1-\alpha)$  100% confidence interval for  $\mu_y$ , then one would use  $\Delta = \Delta^* - \delta$ . To find optimal operating conditions, CN propose using the Nelder-Mead simplex procedure for direct minimization of the function  $\hat{\sigma}_y + \varepsilon$  where

$$\varepsilon = \begin{cases} (\hat{\mu}_y - T)^2 & \text{if } (\hat{\mu}_y - T)^2 > \Delta^2 \\ 0 & \text{if } (\hat{\mu}_y - T)^2 \leq \Delta^2 \end{cases} . \quad (8)$$

Objective functions are also proposed for the ‘larger is better’ and ‘smaller is better’ scenarios. CN demonstrate their methodology to be equally as effective as that proposed by LT.

## 2.2. Response Method Approaches for Combined Arrays with Normal Responses

Crossed arrays have been criticized for two major reasons. First, if there are numerous control and noise factors, the design, even if highly fractionated, often requires too many runs to be of practical use. Second, the often highly fractionated nature of these designs does not enable the user to study two factor interactions among control factors. To address these issues, Welch, Yu, Kang and Sacks<sup>25</sup> proposed the use of a *combined array*, a single experimental design for both control and noise factors. Myers, Khuri, and Vining<sup>26</sup> [henceforth referred to as MKV] suggested the following model for the response at the  $i^{th}$  setting of the control factors  $\mathbf{x}_i$  and the  $j^{th}$  setting of the noise factors  $\mathbf{z}_j$

$$y_{ij} = y(\mathbf{x}_i, \mathbf{z}_j) = \beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta} + \mathbf{x}_i^\top \mathbf{B} \mathbf{x}_i + \mathbf{z}_j^\top \boldsymbol{\gamma} + \mathbf{x}_i^\top \boldsymbol{\Delta} \mathbf{z}_j + \varepsilon_{ij} . \quad (9)$$

Here,  $\beta_0$  is the intercept,  $\boldsymbol{\beta}$  ( $r_x r_x I$ ) and  $\boldsymbol{\gamma}$  ( $r_z r_z I$ ) are vectors of coefficients for the control and noise variable main effects, respectively,  $\mathbf{B}$  ( $r_x r_x r_x$ ) is a matrix whose diagonals are the coefficients for the pure quadratic effects of the control factors and whose off-diagonals are one-half of the interaction effects of the control factors, and  $\boldsymbol{\Delta}$  ( $r_x r_z r_z$ ) is the matrix of control factor by noise factor interaction effects, and the  $\varepsilon_{ij}$  are assumed *i.i.d.*  $N(0, \sigma^2)$ .

The model for the conditional mean response is given by

$$E(y_{ij} | \mathbf{z}_j) = \beta_0 + \mathbf{x}_i' \boldsymbol{\beta} + \mathbf{x}_i' \mathbf{B} \mathbf{x}_i + (\boldsymbol{\gamma}' + \mathbf{x}_i' \boldsymbol{\Delta}) \mathbf{z}_j. \quad (10)$$

Assuming  $E(\mathbf{z}_j) = 0$ , an expression for the process mean is obtained by taking the expectation operator across the equation in (10), resulting in

$$E_{\mathbf{z}_j}[y_{ij}] = \beta_0 + \mathbf{x}_i' \boldsymbol{\beta} + \mathbf{x}_i' \mathbf{B} \mathbf{x}_i. \quad (11)$$

A general form for the process variance can be obtained by writing

$$Var_{\mathbf{z}_j}(y_{ij}) = Var_{\mathbf{z}_j}[E(y_{ij} | \mathbf{z}_j)] + E_{\mathbf{z}_j}[Var(y_{ij} | \mathbf{z}_j)] \quad (12)$$

where  $Var_{\mathbf{z}_j}(y_{ij} | \mathbf{z}_j)$  is the residual variation,  $\sigma^2$ , and  $E_{\mathbf{z}_j}[Var(y_{ij} | \mathbf{z}_j)]$  denotes the variance of the conditional mean. Taking the variance operator across the equation in (10), the process variance is given by

$$Var_{\mathbf{z}_j}(y_{ij}) = (\boldsymbol{\gamma}' + \mathbf{x}_i' \boldsymbol{\Delta}) Var_{\mathbf{z}}(\mathbf{z}_j) (\boldsymbol{\gamma}' + \mathbf{x}_i' \boldsymbol{\Delta})' + \sigma^2. \quad (13)$$

Shoemaker, Tsui, and Wu<sup>27</sup> and MKV pointed out that the process variation given in (13) is not only a function of the noise variables, but also of the control variables through the control by noise factor interaction parameters in  $\boldsymbol{\Delta}$ . Note that  $(\boldsymbol{\gamma}' + \mathbf{x}_i' \boldsymbol{\Delta})$  is the vector of partial derivatives of  $y(\mathbf{x}, \mathbf{z})$  with respect to  $\mathbf{z}$  and the larger the derivatives, the greater the process variance. This is consistent with the nature of control by noise interaction plots. A flat control by noise interaction plot is a result of small derivatives and indicates small process variance. Essentially the robust design problem consists of finding the value of  $\mathbf{x} = \mathbf{x}^*$  such that the mean is at some desirable value and  $(\boldsymbol{\gamma}' + \mathbf{x}_i' \boldsymbol{\Delta}) \approx \mathbf{0}$ . Lee and Nelder<sup>28</sup> [henceforth referred to as LN] suggest making the conditional mean equal to the process mean by finding settings of  $\mathbf{x}$  which force  $(\boldsymbol{\gamma}' + \mathbf{x}_i' \boldsymbol{\Delta}) = \mathbf{0}$  and then utilizing adjustment factors to bring the mean to target. If sufficient adjustment factors do not exist to bring the mean to target or a solution is not feasible, the approach of LT can be used. Myers, Kim, and Griffiths<sup>29</sup> develop methodology that allows for a

confidence region on the location of the control factors that yield minimum process variance. The methodologies utilize the concept of ridge analysis. If there are not too many control factors, the optimal operating conditions may be visualized by overlaying the response surfaces for the mean and variance models. Some software packages will construct plots [*propagation of error* (POE) plots] of the square root of the variance response surface.

The RPD approach outlined above is known as the *response method approach*. This approach differs from those discussed in the previous section where two separate variance models were fit. Here, only a model for the response is fit and then expressions for the estimated process mean and variance are obtained by taking the expectation and variance operators across the response model and replacing the coefficients with the estimated coefficients from the fitted response model.

We illustrate the response method approach by analyzing the data from Table 1. Steinberg and Bursztyn<sup>30</sup> [henceforth referred to as SB] point out that the data in Table 1, while collected from a  $2^{7-4}$  orthogonal array in the control factors, crossed with a  $2^{3-1}$  design in the noise factors, can be viewed as a  $2^{10-5}$  combined array. The choice of defining contrasts enables estimation of all main effects ( $7+3=10$ ) as well as all two-factor control by noise interactions ( $7 \times 3 = 21$ ). Since these 31 effects exhaust all of the 31 orthogonal contrasts in the design, a normal probability plot of these effects is used to determine the important effects. SB reported that three factors have large effects: cycle time ( $A$ ), holding pressure ( $D$ ), and gate size ( $G$ ). In addition, there are two large interactions: moisture content by cavity thickness ( $NC$ ) and moisture content by injection speed ( $NE$ ). As a result, we fit the following response model

$$y = \beta_0 + \beta_A A + \beta_C C + \beta_D D + \beta_E E + \beta_G G + \beta_N N + \beta_{NC} NC + \beta_{NE} NE + \varepsilon . \quad (14)$$

In the analysis, all coefficients are significant at the 0.05 level except for those for  $C$  and  $N$ . However, these main effects will be included in the final model since their interaction,  $CN$ , is significant. The final estimated response model is then

$$\hat{y} = 2.25 + 0.43A + 0.06C - 0.28D + 0.14E - 0.46G \\ + (1.18C - 1.11E)N \quad (15)$$

with an estimated residual variation (the mean squared error) of  $\hat{\sigma}^2 = 0.94$ . The main effect for  $N$  is missing from the expression in (15) since its estimated slope is close to zero. Assuming zero expectation for the noise factors, the estimated process mean is obtained as follows

$$\widehat{E_N(y)} = 2.25 + 0.43A + 0.06C - 0.28D + 0.14E - 0.46G . \quad (16)$$

Assuming that moisture content ( $N$ ) was coded in the design as  $\pm\sigma$ , the estimated process variance can be obtained as demonstrated in expression (13), yielding

$$\begin{aligned} \widehat{Var(y_{ij})} &= (\boldsymbol{\gamma}' + \mathbf{x}_i' \boldsymbol{\Delta}) \widehat{Var_z(\mathbf{z}_j)} (\boldsymbol{\gamma}' + \mathbf{x}_i' \boldsymbol{\Delta})' + \hat{\sigma}^2 \\ &= (1.18C - 1.11E)^2 + 0.94 . \end{aligned} \quad (17)$$

From the expression in (17), the process variance ranges from 0.9449 when  $(C, E) = (1, 1)$  or  $(-1, -1)$  to 6.1841 when  $(C, D) = (1, -1)$  or  $(-1, 1)$ . Constraining the estimated mean in (17) to be 1.5 and minimizing the variance, we obtain the optimal operating conditions  $A = -0.94$ ,  $C = -0.94$ ,  $D = -1$ ,  $E = -1$ , and  $G = 0.93$  which results in a process variance of 0.94. It is easy to see from (17) that the process variation can be reduced to residual variation when  $C = (1.11/1.18)E = 0.94E$ .

Recall from section 2.1 that the dual response approach to the data estimated the process variance to be from 0.006 to 1.877 whereas the response method approach estimates the process variance to be from 0.9449 to 6.1841. SB discuss possible reasons for the differences in results between the dual response and response method approaches. McShane Vaughn<sup>31</sup> presents theoretical comparisons between the two approaches in terms of the efficiency of estimation of the process mean and variance.

Thus far, in discussing the response method approach to RPD, the residual variation has been assumed constant, that is,  $Var(\varepsilon) = \sigma^2$ . Engel and Huele<sup>32</sup> point out that if a noise variable is not included in the

experiment, it will result in non-constant residual variation. To accommodate this assumption, the response model in (9) is assumed to have independent residual errors distributed as  $\varepsilon_{ij} \sim N(0, \sigma_i^2)$ . The expression for the process mean is still given by the expression in (11) and the expression for the process variance remains the same as in (13) but with  $\sigma^2$  replaced with  $\sigma_i^2$ . When non-constant residual variation exists, we must decide on a methodology for estimating the values of  $\sigma_i^2$ . It is often the case that the residual variation is a function of the design variables and some form of variance modeling required (see previous discussion in Section 2.1).

Nelder and Lee<sup>33</sup> and MKV were the first to suggest applying generalized linear models (GLM's) to modeling the variance in robust design. Engel and Huele<sup>32</sup> suggest using a GLM to model the residual variation as a function of the design variables and using iteratively reweighted least squares (IRLS) for fitting the response model given in (9). The proposed model for residual variance is

$$\sigma_i^2 = \exp \left\{ \mathbf{x}_i^\top \boldsymbol{\gamma} \right\}. \quad (18)$$

The exponential model for the variance is the model used by Grego<sup>17</sup> and Chan and Mak<sup>34</sup>. Aitkin<sup>35</sup> demonstrated that when one assumes a normal distribution and identity link for the mean and a gamma distribution with log link for the variance, an IRLS algorithm using the squared residuals from the means fit as data for the variance model provides maximum likelihood estimates of the mean and variance model parameters. Lee and Nelder<sup>36</sup> pointed out that the maximum likelihood estimation of  $\boldsymbol{\gamma}$  does not allow for the estimation of the mean in the formation of the residuals. Consequently, the estimates of the parameters in  $\boldsymbol{\gamma}$  are biased when the number of mean parameters increases with sample size. They point out that restricted maximum likelihood (REML) estimates of  $\boldsymbol{\gamma}$  help correct for the bias problem. MM outline an IRLS algorithm similar to the one presented above, but adapted for REML.

### 2.3. Generalized Linear Models (GLM) for Non-Normal Responses

The response models considered thus are appropriate when one can assume the data follows a normal distribution. In many industrial experiments, the responses are Poisson (count data), exponential or gamma (time-to-failure data), or Bernoulli (defective/non-defective). Many other examples exist and the reader is referred to Hamada and Nelder<sup>37</sup>, Myers and Montgomery<sup>38</sup>, and Myers, Montgomery and Vining<sup>39</sup>. We have already mentioned the utility of the GLM in modeling the variance and in this section we discuss the use of GLMs in the joint modeling of the mean and variance. Nelder and Lee<sup>33</sup> were the first to propose the joint modeling of the mean and variance in RPD using GLMs and considered the case where both the mean and variance were function of the control variables. Lee and Nelder<sup>36</sup> discuss the use of GLMs when the mean is a function of both control and noise variables and the variance is a function of only the control variables and LN extend the GLM framework to RPD situations in which the process mean and variance are both functions of control and noise factors. LN write the process variance as

$$\text{Var}(y_{ij}) = \text{Var}\{E(y_{ij} | \mathbf{z}_j)\} + E\{\phi_{ij}V(\mu_{ij})\} \quad (19)$$

where, from Section 2.1,  $\phi_{ij}$  denotes an overdispersion parameter which is independent of the mean and  $V(\mu_{ij})$  represents that part of the variance which changes with the mean due to the distributional form of the response.

Models for the process mean and variance utilizing the GLM structure can be formulated from the response model for assumed linear models given in expression (9). If  $\mu_{ij}$  is the mean response at the  $i^{th}$  setting of the control factors  $\mathbf{x}_i$  and the  $j^{th}$  setting of the noise factors  $\mathbf{z}_j$ , the GLM response model is written as

$$g(\mu_{ij}) = \eta_{ij} = \beta_0 + \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{x}_i^T \mathbf{B} \mathbf{x}_i + \mathbf{z}_j^T \boldsymbol{\gamma} + \mathbf{x}_i^T \Delta \mathbf{z}_j \quad (20)$$

where  $\eta_{ij}$  is the linear predictor and  $g$  is the link function. Myers, Brenneman and Myers<sup>40</sup> [henceforth referred to as MBM] demonstrate

that an approximate expression for the process mean can be derived from (20) via a second-order Taylor series expansion of (20) about  $\eta_{i0} = E_z[\eta_i]$ , the expectation of the linear predictor with respect to the noise factors. The approximate expression for the process mean (dropping subscripts) is then given by

$$E_z[\mu(\eta)] \approx \mu(\eta_0) + \frac{1}{2}\mu'(\eta_0)Var_z[\eta] \quad (21)$$

where  $\mu'(\eta_0) = \left[ \frac{\partial^2 \mu}{\partial \eta^2} \right]_{\eta=\eta_0}$  and  $Var_z[\eta]$  is the variance of the linear predictor.

and Note that if it is possible to select levels of the control variables which make  $(\gamma' + \mathbf{x}'\Delta) = 0$ , the second term in (21) can be eliminated and the process mean becomes a function of only the control variables.

To develop an expression for the process variance, MBM utilize the expression in (19). Using the delta method, the variance of the conditional mean,  $Var_z[E(y_{ij} | \mathbf{z}_j)]$ , can be written as

$$\begin{aligned} Var_z[\mu] &= \left[ \frac{\partial \mu}{\partial \eta} \right]_{\eta=\eta_0} Var_z[\eta] \left[ \frac{\partial \mu}{\partial \eta} \right]_{\eta=\eta_0}' \\ &= \left[ \frac{\partial \mu}{\partial \eta} \right]_{\eta=\eta_0} (\gamma' + \mathbf{x}'\Delta) Var_z(\mathbf{z}) (\gamma' + \mathbf{x}'\Delta)' \left[ \frac{\partial \mu}{\partial \eta} \right]_{\eta=\eta_0}' . \end{aligned} \quad (22)$$

Note that if levels of the control factors are selected such that  $(\gamma' + \mathbf{x}'\Delta) = 0$ ,  $Var_z[\mu] = 0$  and the process variance in (19) reduces to  $E\{\phi_{ij}V(\mu_{ij})\}$ .

To estimate the process mean and variance, LN propose the fitting of joint GLMs. Joint GLMs are composed of three parts: 1. a model for the process variance  $Var(y_{ij}) = \phi_{ij}V(\mu(\eta_i))$ ; 2. a GLM for the process mean  $\eta_{ij} = g(\mu_{ij}) = \beta_0 + \mathbf{x}_i'\boldsymbol{\beta} + \mathbf{x}_i'\mathbf{B}\mathbf{x}_i + \mathbf{z}_j'\gamma + \mathbf{x}_i'\Delta\mathbf{z}_j$ ; and 3. a GLM for the

overdispersion  $\xi_{ij} = f(\phi_{ij}) = \beta_0^* + \mathbf{x}_i^* \boldsymbol{\beta}^* + \mathbf{x}_i^* \mathbf{B}^* \mathbf{x}_i^* + \mathbf{z}_j^* \boldsymbol{\gamma} + \mathbf{x}_i^* \Delta^* \mathbf{z}_j$ , where the '\*' superscript in the overdispersion GLM implies that it is possible that not all control factor terms in the means model will exist in the overdispersion model. Given dispersion parameters,  $\phi_{ij}$ , as prior weights, the regression parameters in the means model are estimated via iterative reweighted least squares (IRLS). The estimated GLM from the means model is used to create responses  $d_{ij}^*$ , which are in turn used for fitting the GLM in the overdispersion model. The reciprocals of the fitted values from the dispersion GLM provide new weights for the mean GLM to be re-estimated. LN outline an extended REML algorithm for estimating the parameters in both GLMs and a set of Genstat 5.4.1 procedures for fitting the models is available from LN. MBM point out that the assumption of LN in finding levels of the control factors such that  $(\boldsymbol{\gamma} + \mathbf{x}^* \Delta) = \mathbf{0}$  is often impractical and may provide control factor settings that are inferior to others which are selected using squared error loss approach for optimization.

It is helpful to illustrate the use of joint GLMs in RPD with an example. Schmidt and Launsby<sup>41</sup> present data from a solder experiment that was used to find conditions that minimize the number of solder defects. The experimental design crosses a  $2^{5-2}$  array in the control factors (labeled a-e) with a  $2^{3-1}$  array in the noise factors (labeled o-q). LN assume  $Var(y_{ij}) = \phi_{ij} \mu(\eta_{ij})^2$  and then fit GLMs for the mean (Poisson with log link) and dispersion (gamma with log link) yielding

$$\widehat{g(\mu_{ij})} = \widehat{\ln(\mu_{ij})} = 5.29 - 0.07a - 0.005b + 0.28c + 0.09e - (0.07 + 0.05b)o \quad (23)$$

for the estimated mean and

$$\widehat{f(\phi_{ij})} = \widehat{\ln(\phi_{ij})} = -4.39 - 0.69a - 1.04c - (1.39 + 1.85a)o \quad (24)$$

for the dispersion. By setting  $b = -0.07/0.05 = -1.4$  in (23) and by setting  $a = -1.39/1.85 = -0.75$ , the mean and variance models become

robust to changes in the noise variable ( $o$ ). LN report the solution  $a = .75$ ,  $b = -1.4$ ,  $c = -1$ , and  $e = -1$  for minimizing the mean number of defects while at the same time minimizing the variance. MBM note that the optimal level for  $b$  is beyond the  $\pm 1$  design region and thus may not be useful to the experimenter. As an alternative solution to the RPD problem, MBM propose the use of unconstrained optimization of the estimated mean squared error. The estimated MSE can be obtained by combining the expressions for the approximate mean given in (21) and that of the process variance given in (19). Here,  $\mu(\eta) = \exp(\eta)$ , resulting in the following expression for the estimated process mean using the expression in (21)

$$\begin{aligned}\widehat{E_z[\mu(\eta)]} &\approx \exp(\hat{\eta}_0) + \frac{1}{2} \exp(\hat{\eta}_0) \widehat{Var_z[\eta_0]} \\ &= \exp(5.29 - 0.07a - 0.005b + 0.28c + 0.09e) \\ &\quad \times \left[ 1 + \frac{1}{2} (0.07 + 0.05b)^2 \sigma_o^2 \right].\end{aligned}\quad (25)$$

The estimated process variance is given by

$$\begin{aligned}\widehat{Var_z(y)} &\approx \widehat{\exp(2\eta_0) Var(\eta)} + E_z \left[ \widehat{\phi_{ij}} V(\widehat{\mu_{ij}}) \right] \\ &= \exp(2[5.29 - 0.07a - 0.005b + 0.28c + 0.09e]) (0.07 + 0.05b)^2 \sigma_o^2 \\ &\quad + E_z \left[ \exp(-4.39 - 0.69a - 1.04c - (1.39 + 1.85a)o) \right. \\ &\quad \left. \times \exp(2[5.29 - 0.07a - 0.005b + 0.28c + 0.09e - (0.07 + 0.05b)o]) \right].\end{aligned}\quad (26)$$

Expanding the second component in (26) about  $o = E(o)$  in a second order Taylor series expansion, the expression in (26) is approximated by

$$\begin{aligned}\widehat{Var_z(y)} &\approx \exp(2[5.29 - 0.07a - 0.005b + 0.28c + 0.09e]) \\ &\quad \times (0.07 + 0.05b)^2 \sigma_o^2 + \exp(6.19 - 0.83a - 0.01b - 0.48c + 0.18e) \\ &\quad \times \left( 1 + \frac{1}{2} (1.46 + 1.85a + 0.05b)^2 \sigma_o^2 \right).\end{aligned}$$

Using the expressions for the estimated process mean and variance and developing an MSE criterion for a target=0, MBM demonstrate that the optimal settings of the control factors are:  $a=1$ ,  $b=-0.6$ ,  $c=1$ , and  $e=-1$ . This solution results in an estimated MSE of 18304 whereas the solutions given by LN provide an estimated MSE of 22386.

It should be noted that the determination of optimal operating conditions for nonlinear objective functions with nonlinear constraints can be a difficult problem. Brinkley, Meyer, and Lu<sup>42</sup> discuss a case study where the responses are Poisson and demonstrate the use of nonlinear programming in finding optimal operating conditions. Paul and Khuri<sup>43</sup> develop ridge analysis for GLM responses and demonstrate the use of ridge analysis in finding optimal operating conditions.

#### ***2.4. Other Applications of RSM in RPD***

In many applications, mixture experiments are useful. Mixture experiments involve the blending or mixing of two or more ingredients to form a product. In these types of experiments, it is of interest to find the component proportions which lead to the optimal product. Since the sum of the mixture proportions must equal 1, the feasible design region is a simplex. In many mixture settings, there are also process variables. Process variables encompass any factor that is not a mixture factor, but are influential in the response of interest. For an overview of mixture experiments, see Cornell<sup>44</sup>. Steiner and Hamada<sup>45</sup> [henceforth referred to as SH] consider the situation when some of the process variables are noise variables and the goal of the mixture experiment is to find mixture proportions and process settings that lead to a high quality product that is robust to noise. A quadratic loss objective function is developed and the use of LaGrangian multipliers is employed for determining optimal operating conditions. Goldfarb, Borror, and Montgomery<sup>46</sup> [henceforth referred to as GBM] consider the situation studied by SH, but propose the use of the Derringer and Suich<sup>47</sup> desirability function in the optimization algorithm. The model used by SH involves interactions among the noise variables whereas, GBM do not include these terms. SH consider only noise variables that are assumed to be independent of one another whereas GBM allow for dependence among noise variables. An

additional interesting application considered by SH is when the researcher has additional information on some of the noise variables and can use the means of those noise variables as optimization variables along with the design factors. Related ideas in non-mixture experiments are considered by Pledger<sup>48</sup> and Joseph<sup>49</sup>.

Much of the literature on robust design has been developed under the assumption that the noise variables are independent and normally distributed with mean zero and variance  $\sigma_{z_i}^2$ , where  $z_i$  denotes the  $i^{th}$  noise variable. However, in many industrial applications, the noise variables are categorical in nature (e.g., different suppliers, different operators, different brands of equipment). Brenneman and Myers<sup>50</sup> point out that the variation between levels of the categorical noise variable can be modeled via a multinomial distribution. Expressions for the mean and variance are developed under the assumption of a multinomial distribution and details of the two-step approach to solving the RPD problem are provided for a package-sealing example where the noise factor is 'supplier of raw material'.

Thus far, discussion of RPD has involved control and noise factors. Khuri<sup>51</sup> and Wolfinger and Tobias<sup>52</sup> discuss the role of random factors in RPD. Random factors are those whose levels are randomly selected from a hypothetical infinite population of levels, which are usually assumed to be normal with mean zero and unknown variance. Thus, observations with the same level of the random factor are potentially correlated. Examples of random factors include batches of material, blocks, replicates, and whole plots. Khuri<sup>53</sup> first discussed the application of RSM when there are random factors and described the use of linear mixed models for fitting these models. Linear mixed models can be fit using the MIXED procedure in SAS. In finding optimal settings of the control factors, it is desired to choose levels that result in minimum variation in the random factor. All control factor by random effect interactions are random and Khuri<sup>53,51</sup> and Wolfinger and Tobias<sup>52</sup> discuss inferential procedures regarding their associated variance components. Future research may involve the analysis of non-normal responses in the presence of random effects utilizing generalized linear mixed models (GLMM).

### 3. Experimental Designs for Robust Parameter Design

The experimental design is an important consideration in robust parameter design problems. The experimental design dictates the number of required experimental runs and the effects that can be fit or estimated by the modeling procedure. Many of the designs used in RPD are adapted from response surface designs. MM provide a nice summary of experimental designs used in RSM and RPD. Robinson, Borror, and Myers<sup>54</sup> also review various designs for use in RPD. This section presents designs for fitting first-order models with interactions, designs for fitting second-order models, design optimality and computer generated designs, and designs accommodating random factors.

#### 3.1. *Designs for Fitting First-Order Models with Interactions*

A *first-order model with interactions* contains main effects and two-way interactions involving the control and noise factors. Such models for RPD can be obtained from (9) by removing the quadratic terms involving the control factors ( $\mathbf{x}_i^T \mathbf{B} \mathbf{x}_i$ ). When there are  $k$  control and noise factors, a *two-level factorial design* (or  $2^k$  design) can be used to fit the model. These orthogonal designs are also a useful building block for other designs used in RPD. MM (p. 557) present a  $2^4$  design for the RPD problem in the pilot plant experiment where there are three control factors and one noise factor.

The  $2^k$  design is a standard design in RSM, but it can require a large number of runs when  $k$  is large. For the example presented in section 1,  $k = 10$  and the number of required runs would be 1024. *Two-level fractional factorial designs* (or  $2^{k-p}$  designs) can be used to reduce the number of runs at the expense of fitting particular effects. In fractional factorial designs, some of the effects are *confounded* or *aliased* as they cannot be separated. The idea is to confound the effects of interest with those higher order effects that are not expected to be important. The aliasing structure can be characterized through *design resolution*. Keuhl<sup>55</sup> (p. 398) indicates that a factorial design has resolution  $R$ , if no  $p$ -factor effect is aliased with another effect containing less than  $R - p$  factors.

A  $2^{k-p}$  design which has resolution  $R$  will be denoted  $2_R^{k-p}$ . Common design resolutions include III, IV, and V.

Factorial and fractional factorial designs are frequently used in RPD to form crossed arrays. Shoemaker, Tsui, and Wu<sup>27</sup> [henceforth referred to as STW] and MM (p. 587) discuss an example where there are 3 control factors  $x_1, x_2, x_3$  and 3 noise factors  $z_1, z_2, z_3$ . Using the defining relations  $I = x_1x_2x_3$  and  $I = z_1z_2z_3$ , the outer and inner arrays both consist of  $2_{\text{III}}^{3-1}$  designs. The crossed array can be viewed as a combined array through the defining relation  $I = x_1x_2x_3 = z_1z_2z_3 = x_1x_2x_3z_1z_2z_3$ . This is a  $2^{6-2}$  design where no interactions in the control factors can be estimated and the six main effects are estimable only if no two factor interactions are important. The remaining six degrees of freedom consist of interactions between control and noise variables at the expense of other interactions that might be of interest, such as interactions between control variables. An alternative  $2^{6-2}$  fractional factorial design would have the defining relation  $I = x_1x_2x_3z_1 = z_1z_2z_3 = x_1x_2x_3z_2z_3$ . This design has resolution III with regard to noise interactions and resolution IV with regard to other interactions. This is an example of a design in which the combined array has *mixed resolution*. The three control-by-control interactions can be estimated assuming  $x_1z_1$ ,  $x_2z_1$ , and  $x_3z_1$  are negligible. Thus, the experimenter has the option to exchange control-by-noise interactions for control-by-control interactions when the appropriate assumptions are met. Through this example, STW demonstrate the flexibility and usefulness of the combined array for choosing a crossed array with mixed resolution which allows for estimation of the effects of interest in RPD.

### 3.2. Designs for Fitting Second-Order Models

A full *second-order model* contains main effects, quadratic effects, and two-way interactions involving the control and noise factors. However, in RPD, quadratic noise effects are usually not of interest, and the resulting second-order model is given in (9). In order to fit a quadratic effect, the design must accommodate at least three levels for that effect. One approach is to use *three-level factorial designs* (or  $3^k$  designs). When there is only interest in quadratic effects for the control factors,

$3^k$  designs can be run where  $k$  denotes the number of control factors. The number of runs for  $3^k$  designs becomes quite large when  $k > 3$ . In order to reduce the number of runs, a *three-level fractional factorial design* (or  $3^{k-p}$  design) can be used. MM (p. 589) consider a 36-run crossed array with a  $3_{III}^{3-1}$  design for the inner array, consisting of 3 control factors, and a  $2^2$  design for the outer array, consisting of 2 noise factors.

The *central composite design* (CCD) is one of the most commonly used response surface designs for fitting second-order models. A central composite design consists of  $F$  factorial points,  $2k$  axial points ( $\pm\alpha$ ), and  $n_c$  center points. The factorial points are formed from a complete  $2^k$  design or a  $2_V^{k-p}$  design. The factorial portion is used to fit all linear and interaction terms. The axial points provide additional levels of the factor for purposes of estimation of the quadratic terms. The choice of  $\alpha$  depends upon the design region of interest. For a spherical design region, common choices are the spherical distance  $\alpha = \sqrt{k}$  or  $\alpha = \sqrt[4]{F}$ . Multiple runs at the center of the design region allows for estimation of pure error and also contributes to the estimation of quadratic terms. The value of  $n_c$  can be chosen by the experimenter. For spherical design regions, MM (p. 337) indicate  $n_c$  should be from 3 to 5. A special case of the CCD can be used for cuboidal design regions where the levels of the factors must lie within the coded values of -1 to +1. The CCD, for this case, is called the *face-centered cube design* (FCCD) and has the axial distance set at  $\alpha = 1$ . For cuboidal design regions, MM (p. 337) recommend  $n_c$  should be from 1 to 2 for the FCCD, but others, such as Lucas<sup>56</sup> and Borkowski and Lucas<sup>57</sup> (henceforth referred to as BL) do not always include center runs. A standard CCD (or FCCD) for two control factors and two noise factors is given in Table 2. When  $n_c = 3$ , this design requires 27 runs, which is much smaller than the  $3^4$  design.

The CCD can also become burdensome to implement when  $k$  is large. Since there is usually no interest in quadratic terms for the noise factors in RPD, a *modified CCD* (MCCD) can be used to reduce the number of runs by ignoring the axial points for the noise factors. The modified CCD, when  $k = 4$ , can be obtained from the standard CCD in Table 2 by removing the second set of axial points. For this example, the MCCD results in a reduction of 4 runs from the standard CCD.

Table 2. The CCD and SCD for 2 Control and 2 Noise Factors.

		CCD				SCD			
Points		x <sub>1</sub>	x <sub>2</sub>	z <sub>1</sub>	z <sub>2</sub>	x <sub>1</sub>	x <sub>2</sub>	z <sub>1</sub>	z <sub>2</sub>
Factorial	-1	-1	-1	-1	-1	-1	-1	-1	1
	-1	-1	-1	-1	1	1	-1	-1	-1
	-1	-1	1	-1	-1	-1	1	-1	-1
	-1	-1	1	1	-1	-1	-1	1	1
	-1	1	-1	-1	1	1	1	-1	1
	-1	1	-1	1	1	1	-1	1	-1
	-1	1	1	-1	-1	-1	1	1	-1
	-1	1	1	1	1	1	1	1	1
	1	-1	-1	-1					
	1	-1	-1	1					
	1	-1	1	-1					
	1	-1	1	1					
	1	1	-1	-1					
	1	1	-1	1					
	1	1	1	-1					
	1	1	1	1					
Axial	{	$\alpha$	0	0	0	$\alpha$	0	0	0
		$-\alpha$	0	0	0	$-\alpha$	0	0	0
		0	$\alpha$	0	0	0	$\alpha$	0	0
		0	$-\alpha$	0	0	0	$-\alpha$	0	0
Delete	{	0	0	$\alpha$	0	0	0	$\alpha$	0
in		0	0	$-\alpha$	0	0	0	$-\alpha$	0
Modified	{	0	0	0	$\alpha$	0	0	0	$\alpha$
		0	0	0	$-\alpha$	0	0	0	$-\alpha$
Center	{	0	0	0	0	0	0	0	0
		0	0	0	0	0	0	0	0
		0	0	0	0	0	0	0	0

Another alternative for reducing the number of required runs is the *small composite design* (SCD) presented by Hartley<sup>58</sup> and also discussed by MM. SCDs differ from CCDs with respect to the factorial portion. In the SCD, the factorial portion is not a  $2^k$  or  $2^{k-p}$  design, but a resolution III\* fraction where the defining relation does not consist of any four-way

interactions. In addition, a modified SCD (MSCD) can be obtained in the same way as a MCCD). A potential problem with the MSCD is that there are not always enough runs to accommodate all the interaction terms of interest. However, suitable design points can be used to augment these designs to fit additional terms. A SCD is also given in Table 2 for two control factors ( $x_1, x_2$ ) and two noise factors ( $z_1, z_2$ ). The defining relation for the factorial portion is  $I = x_1 x_2 z_2$ .

The concept of mixed resolution can also be used for fitting quadratic effects in the control variables using composite mixed resolution designs (CMRD) developed by Lucas<sup>59,60</sup> and BL. The CMRD is a special type of CCD with a generator in the factorial portion chosen such that among the control factors the design has resolution V, among the noise factors the design has at least resolution III, and two-way interactions involving control and noise factors are not aliased with any main effects or any two-way interactions. These designs provide a nice alternative to MCCDs and MSCDs. For example, suppose there are 3 control factors  $x_1, x_2, x_3$  and 2 noise factors  $z_1, z_2$ . The associated CMRD requires 22 runs consisting of the factorial portion given by a  $2^{5-1}$  design having the defining relation  $I = x_1 x_2 x_3 z_1 z_2$ , six axial points for the control factors, and no center runs.

### **3.3. Optimal Design Criteria**

The previous sections presented various experimental designs for RPD for fitting first and second order models with emphasis on the number of required runs and the terms that can be modeled. For a particular RPD problem, there may be many designs to choose from. Optimal design theory provides additional criteria to help in choosing from among these candidate designs. MM (Section 8.2.1) present various optimality criteria which depend upon the *moment matrix* or the quantity  $M = X'X/n$  where  $n$  denotes the total number of design runs. Note that this moment matrix is only applicable when the design is completely randomized and constant error variance is assumed. Under the assumptions of the response method approach discussed in section 2.2, the inverse of the moment matrix is proportional to the matrix containing the variances and covariances of the estimates of regression coefficients in (9).

One of the most commonly used criteria is the *D-criterion* given by  $d_\xi = |\mathbf{X}_\xi' \mathbf{X}_\xi| / n^p = |\mathbf{M}|$  where  $\xi$  denotes a given design and  $p$  is the number of terms to be fit in the model. It is desirable to choose a design  $\xi$  such that  $d_\xi$  is large as this means the set of regression coefficients can be estimated more precisely under  $\xi$ . The *D-optimal* design is one in which  $d_\xi$  is maximized for all designs. Supposing  $\xi$  is the D-optimal design, the *D-efficiency* of a design  $\xi$  is  $D_{\text{eff}} = (d_\xi / d_{\xi^*})^{1/p}$ . A second commonly used criterion is based upon the *scaled prediction variance* at the point  $\mathbf{x}$  which is given by  $v(\mathbf{x}) = \mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x} = \mathbf{x}'\mathbf{M}^{-1}\mathbf{x}$ . Since the prediction variance depends upon the location in the design space, the *G-criterion*, for a design  $\xi$ , is defined as  $g_\xi = \max_{\mathbf{x}} v(\mathbf{x})$ . Thus, it is desirable to choose a design  $\xi$  such that  $g_\xi$  is small since this means that the highest possible prediction variance in the design region will be small. For cuboidal and spherical design regions,  $g_\xi \geq p$  for all designs  $\xi$ . Thus, the *G-efficiency* of designs  $\xi$  for these regions is  $G_{\text{eff}} = p / g_\xi$ .

The  $2^k$  and  $2^{k-p}$  designs are optimal for fitting first order models with interactions. For complete second order models, the CCD and FCCD are rarely optimal, but these designs are good (MM p. 399). Lucas<sup>56</sup> provides tables of D and G efficiencies for common response surface designs which accommodate second order models. For example, consider a complete second order model for the CCD in Table 2. For a spherical design region, where  $\alpha = 2$ , the standard CCD has  $D_{\text{eff}} = 98.9\%$  and  $G_{\text{eff}} = 95.2\%$ . For the cuboidal design region, where  $\alpha = 1$  and there are no center runs, the standard CCD has  $D_{\text{eff}} = 93.6\%$  and  $G_{\text{eff}} = 81.1\%$ .

As mentioned earlier, quadratic effects in the noise variables terms are often not of interest in RPD. For two control factors and two noise factors, the associated model in (9) would have  $p = 13$  terms. The OPTEX procedure in SAS can be used to calculate the D-criterion values  $d_\xi^{1/13}$  for the second-order designs in section 3.2. Assuming a cuboidal design region and no center runs, these values are 48.17% for the 243 run  $3^4$  design, 57.03% for the 24 run CCD, 59.16% for the 20 run MCCD, and 34.03% for the 16 run SCD. It should be noted that the CMRD is equivalent to the MCCD in this case. Clearly, the  $3^4$  design is too large and does not even have the highest value for the D-criterion. The CCD and MCCD are the two best designs in this comparison. There is a

substantial drop in the D-criterion for the SCD. The MSCD could not be used since it cannot fit the interaction effect  $x_1x_2$ .

It is important to remember that the design criteria are evaluated with respect to the model that is of interest in the RPD problem. Once again, suppose there are 3 control variables and 2 noise variables for the cuboidal design region. Consider the crossed array  $3_{\text{III}}^{3-1} \times 2^2$  discussed in section 3.2. As BL indicate, this array is optimal for the model

$$\begin{aligned} y = & \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{33} x_3^2 \\ & + \delta_1 z_1 + \delta_2 z_2 + \delta_{11} x_1^2 z_1 + \delta_{11} x_2^2 z_2 + \varepsilon. \end{aligned}$$

On the other hand, for the model in (9), this design has 0% efficiency since it cannot estimate all the terms. However, BL point out that the associated CMRD is a good design with  $D_{\text{eff}} = 88.9\%$  and  $G_{\text{eff}} = 87.1\%$ .

### 3.4. Computer Generated Designs

Many computer packages, such as the OPTEX procedure in SAS, are available which can select a design based upon optimality criteria such as those described in section 3.3. These *computer-generated designs* (CGDs) are useful in RSM for several reasons: (1) obtaining a good design for a given model, a set of candidate points, given ranges on the design variables, and other possible constraints (2) determining the best design points for augmenting a current design, and (3) replacing runs in a current design. Role (1) is helpful for RPD since the models of interest, such as (9), are not standard. In addition, CGDs can be obtained when the noise variables are categorical. Role (2) is important in RPD because it may be necessary to add runs to small designs, such as MSCDs or mixed resolution designs, in order to estimate effects of special interest. Role (3) is helpful when there are practical limitations on the design region.

For instance, consider augmenting the MSCD obtained from the SCD in Table 2 so that the additional runs allow for fitting the interaction effect  $x_1x_2$ . The MSCD, without center runs, consists of 12 runs while the SCD consists of 16 runs. Using the OPTEX procedure in SAS, the CGD will be conditionally D-optimal as it includes the original 12 runs

and the remaining runs chosen from among the design points of the  $3^4$  design. For an augmented MSCD with 13 runs,  $d_{\xi}^{1/13} = 34.23$  and the chosen point is an axial point for a noise factor. In addition, for 16 runs, the CGD has  $d_{\xi}^{1/13} = 35.71$  and the chosen points consist of two replicated factorial points and both axial points for one of the noise factors. Thus, the CGD is not the same as the SCD and the D-criterion value is slightly better than that associated with the SCD in Table 2.

### 3.5. Randomized Block and Split Plot Designs

In many experimental settings in RPD, it may not be possible to perform all design runs under like conditions. *Blocks* refer to those units under which the experimental conditions are homogeneous. Examples of blocks could include batches of material, time periods, or operators. The use of blocking can also allow for more precise estimation of the effects associated with control and noise factors. A common design for incorporating blocks is the *randomized complete block design* (RCBD) where every combination of the control and noise variables occur in each block. The blocks can be fixed or random factors. The blocks are fixed when the experiment uses only the specific blocking units of interest. On the other hand, random blocks are those in which the blocks used for the experiment are sampled from some large population. Examples of the RBCD for RPD when the blocks are random are presented by Wolfinger and Tobias<sup>52</sup>. Since every treatment combination must occur in every block, the number of experimental runs for the RBCD can be quite large. In order to reduce the number of runs, a *randomized incomplete block design* (RICBD) can be used (Keuhl<sup>55</sup> Chapter 11). In order to create RICBDs, treatments that are not of interest are confounded with blocks. Goos<sup>61</sup> discusses D-optimal designs in the presence of random blocks.

Frequently, in RPD, the control factors are easy to change and the noise factors are hard to change. Thus, to save time and money, the noise factors are held fixed while the levels of the control factors are varied. Such a restriction in the randomization results in a *split plot design* (SPD) where each setting of the noise variable corresponds to a *whole plot*. The whole plots are random factors and runs in the same whole plot are potentially correlated. Box and Jones<sup>9</sup> show that SPDs can be of

significant value in RPD because these designs are easier to conduct and result in efficient estimation of the control by noise interactions. In their work, three experimental scenarios were considered: (1) noise factors contained in whole plots and control factors contained in subplots (as mentioned above), (2) control factors contained in whole plots while noise factors are contained in subplots, and (3) strip-block designs where subplot factors are randomly assigned in strips across each whole plot. The authors recommend using (1) since this option provides the most information about the control factors. These authors also indicate that case (3) can be easier to implement in some experiments. Bingham and Sitter<sup>62</sup> also discuss *fractional factorial split-plot designs* (FFSP) for cases (1) and (2). A catalog of designs are provided for 16 and 32 runs. For example, suppose there are two noise variables  $z_1, z_2$  run as whole plot factors and four control variables  $x_1, x_2, x_3, x_4$  run as subplot plot factors. Bingham and Sitter recommend the 16 run  $2^{(2+4)-(0+2)}$  FFSP with 4 whole plots, using the defining relation  $I = z_1 z_2 x_1 x_3 = z_1 x_2 x_3 x_4$  or  $x_3 = z_1 z_2 x_1$  and  $x_4 = z_1 x_2 x_3$ . Kowalski and Vining<sup>63</sup> provide an overview of the use of SPDs in industry. Goos and Vanderbroek<sup>64,65</sup> and Goos<sup>61</sup> discuss issues and methods for finding D-optimal designs for SPDs.

#### 4. Conclusions

This chapter has served to review the applications of response surface methods to robust parameter design. Many of the common response surface methods have greatly advanced RPD in terms of modeling, optimization, and choice of experimental design. Topics involving non-normal responses, optimal design choice, categorical noise factors, dependent noise factors, the presence of random effects, and cost issues deserve further attention. Cost issues deserve further attention.

#### References

1. Engel, J. Modeling Variation in Industrial Experiments. *Applied Statistics* 1992, **41**: 579-593.
2. Taguchi, G. *System of Experimental Design: Engineering Methods to Optimize Quality and Minimize Cost*. White Plains, NY: UNIPUB/Kraus International, 1987.

3. Nair, V.N. Taguchi's Parameter Design: A Panel Discussion. *Technometrics* 1992; **34**:127-161.
4. Box, G.E.P. Signal-to-Noise Ratios, Performance Criteria, and Transformations. *Technometrics* 1988, **30**: 1-17.
5. Nair, V.N. and Pregibon, D. Analyzing Dispersion Effects From Replicated Factorial Experiments. *Technometrics* 1988, **30**: 247-257.
6. Myers, R.H. and Montgomery, D.C. *Response Surface Methodology: Process and Product Optimization Using Designed Experiments (2<sup>nd</sup> Edition)*. Wiley: New York, 2002.
7. Leon, R.L.; Shoemaker, A.C.; and Kacker, R.N. Performance Measures Independent of Adjustment. *Technometrics* 1987, **29**: 253-265.
8. Logothetis. Box-Cox Transformations and the Taguchi Method. *Applied Stat.* 1990, **39**: 31-48.
9. Box, G. E. P. and Jones, S. Split-Plot Designs for Robust Product Experimentation. *Journal of Applied Statistics* 1992, **19**: 3-26.
10. Box and Jones. Split Plots for Robust Product and Process Experimentation. *Quality Engineering* 2000, **13**: 127-134.
11. Box, G.E.P. and Wilson, K.B. On the Experimental Attainment of Optimum Conditions. *Journal of the Royal Statistical Society, Series B* 1951, **13**: 1-45.
12. Vining, G.G. and Myers, R.H. Combining Taguchi and Response Surface Philosophies: A Dual Response Approach. *Journal of Quality Technology* 1990, **22**: 38-45.
13. Bartlett, M.S. and Kendall, D.G. The Statistical Analysis of Variance Heterogeneity and the Logarithmic Transformation. *Journal of the Royal Statistical Society, Series B* 1946, **8**: 128-150.
14. Davidian and Carroll. Variance Function Estimation. *Journal of the American Statistical Association* 1987, **82**: 1079-1091.
15. Davidian and Carroll. A note on extended quasi-likelihood. *Journal of the Royal Statistical Society B* 1988, **50**: 74-82.
16. McCullagh and Nelder. *Generalized Linear Models (2<sup>nd</sup> Edition)*. London: Chapman and Hall, 1992.
17. Grego, J.M. Generalized Linear Models and Process Variation. *Journal of Quality Technology* 1993, **25**: 288-295.
18. Vining, G.G. and Bohn, L.L. Response Surfaces for the Mean and Variance Using a Nonparametric Approach. *Journal of Quality Technology* 1998, **30**: 282-291.
19. Vining, G.G. and Schaub, D. Experimental Designs for Estimating Both the Mean and Variance Functions. *Journal of Quality Technology* 1996, **29**: 135-147.
20. Myers, R.H. and Carter, W.H. Jr. Response Surface Techniques for Dual Response Systems. *Technometrics* 1973, **15**: 301-317.
21. Del Castillo, E. and Montgomery, D.C. A Nonlinear Programming Solution to the Dual Response Problem. *Journal of Quality Technology* 1993, **25**: 199-204.
22. Fathi, Y. A nonlinear programming approach to the parameter design problem. *European Journal of Operational Research* 1991, **53**: 371-381.
23. Lin, D.K.J. and Wanzhu, T. Dual Response Surface Optimization. *Journal of Quality Technology* 1995, **27**: 34-39.

24. Copeland, K.A.F. and Nelson, P. R. Dual Response Optimization via Direct Function Minimization. *Journal of Quality Technology* 1996, **28**: 331-336.
25. Welch, W.J.; Yu, T.K.; Kang, S.M.; and Sacks, J. Computer Experiments for Quality Control by Parameter Design". *Journal of Quality Technology* 1990, **22**: 15-22.
26. Myers, R.H.; Khuri, A.I.; and Vining, G.G. Response Surface Alternatives to the Taguchi Robust Parameter Design Approach. *The American Statistician* 1992, **46**: 131-139.
27. Shoemaker, A.C.; Tsui, K.L. and Wu, C.F.J. Economical Experimentation Methods for Robust Parameter Design. *Technometrics* 1991, **33**: 415-427.
28. Lee, Y. and Nelder, J.A. Robust Design via Generalized Linear Models. *Journal of Quality Technology* 2003, **35**: 2-12.
29. Myers, R.H.; Kim, Y.; and Griffiths, K.L. Response Surface Methods and the Use of Noise Variables. *Journal of Quality Technology* 1997, **29**: 429-440.
30. Steinberg, D.M. and Bursztyn, D. Dispersion Effects in Robust-Design Experiments with Noise Factors. *Journal of Quality Technology* 1994, **26**: 12-20.
31. McShane Vaughn. The Response Model Method in Robust Design. *Unpublished PhD Dissertation* 1998, Georgia Institute of Technology.
32. Engel, J. and Huele, A.F. A Generalized Linear Modeling Approach to Robust Design. *Technometrics* 1996, **38**: 365-373.
33. Nelder, J.A. and Lee, Y. Generalized Linear Models for the Analysis of Taguchi-Type Experiments". *Applied Stochastic Models and Data Analysis* 1991, **7**:107-120.
34. Chan, L.K. and Mak, T.K. A Regression Approach for Discovering Small Variation Around a Target". *Applied Statistics* 1995, **44**: 369-377.
35. Aitkin, M. Modeling Variance Heterogeneity in Normal Regression Using GLIM. *Applied Statistics* 1987, **36**: 332-339.
36. Lee, Y. and Nelder, J.A. Joint Modeling of Mean and Dispersion. *Technometrics* 1998, **40**: 168-175.
37. Hamada, M. and Nelder, J.A. Generalized Linear Models for Quality-Improvement Experiments. *Journal of Quality Technology* 1997, **29**: 292-304.
38. Myers, R.H. and Montgomery, D.C. A Tutorial on Generalized Linear Models. *Journal of Quality Technology* 1997, **29**: 274-291.
39. Myers, R.H.; Montgomery, D.C.; and Vining G.G. *Generalized Linear Models: With Applications in Engineering and the Sciences*. Wiley: New York, 2002.
40. Myers, Brenneman, Myers. A Dual Response Approach to Robust Parameter Design for a Generalized Linear Model. *Journal of Quality Technology* 2005, **37**: 130-138.
41. Schmidt and Launsby. Understanding Industrial Designed Experiments. Air Academy Press 1990, Colorado Springs, CO.
42. Brinkley, P.A.; Meyer, K.P.; and Lu, J.C. Combined Generalized Linear Modelling / Non-Linear Programming Approach to Robust Process Design: a Case-study in Circuit Board Quality Improvement. *Applied Statistics* 1996, **45**: 99-110.
43. Paul, S. and Khuri, A.I. Modified Ridge Analysis Under NonStandard Conditions. *Communications in Statistics-Theory and Methods* 2000, **29**: 2181-2200.
44. Cornell, J.A. *Experiments With Mixtures: Designs, Models, and the Analysis of Mixture Data (3<sup>rd</sup> edition)*, Wiley: New York, 2002.

45. Steiner, S.H. and Hamada, M. Making Mixtures Robust to Noise and Mixing Measurement Errors. *Journal of Quality Technology* 1997, **29**: 441-450.
46. Goldfarb, H.; Borror, C.M.; and Montgomery, D.C. Mixture-Process Experiments Involving Noise Variables. *Journal of Quality Technology* 2004, **36**: 245-262.
47. Deringer, G. and Suich, R. Simultaneous Optimization of Several Response Variables. *Journal of Quality Technology* 1980, **12**: 214-219.
48. Pledger, M. Observable Uncontrollable Factors in Parameter Design. *Journal of Quality Technology* 1996, **28**: 153-162.
49. Joseph, R.V. Robust Parameter Design with Feed-Forward Control. *Technometrics* 2003, **45**: 284-292.
50. Brenneman, W.A.; and Myers, W.R. Robust Parameter Design with Categorical Noise Variables. *Journal of Quality Technology* 2003, **35**: 335-341.
51. Khuri, A.I. Response Surface Models with Mixed Effects. *Journal of Quality Technology* 1996, **28**: 177-186.
52. Wolfinger, R.D. and Tobias, R.D. Joint Estimation of Location, Dispersion, and Random Effects in Robust Design. *Technometrics* 1998, **40**: 62-71.
53. Khuri, A.I. Response Surface Models with Random Block Effects. *Technometrics* 1992, **34**: 26-37.
54. Robinson, T.J.; Borror, C.M.; Myers, R.M. Robust Parameter Design: A Review. *Quality and Reliability Engineering International* 2004, **20**: 81-101.
55. Keuhl, R.O. *Design of Experiments: Statistical Principles of Research Design and Analysis (2<sup>nd</sup> Edition)* Duxbury: New York, 1994.
56. Lucas, J.M. Which Response Surface is Best. *Technometrics* 1976, **18**: 411-417.
57. Borkowski, J.J. and Lucas, J.M. Designs of Mixed Resolution for Process Robustness Studies. *Technometrics* 1997, **39**: 63-70.
58. Hartley, H. O. Smallest Composite Design for Quadratic Response Surfaces. *Biometrics* 1959, **15**: 611-624.
59. Lucas, J. M. Achieving a Robust Process Using Response Surface Methodology in *Proceedings of the American Statistical Association, Sesquicentennial Invited Paper Sessions* 1989, pp. 579-593.
60. Lucas, J.M. How to Achieve a Robust Process Using Response Surface Methodology. *Journal of Quality Technology* 1994, **25**: 248-260.
61. Goos, P. *The Optimal Design of Blocked and Split-Plot Experiments*. Springer: New York, 2002.
62. Bingham, D. and Sitter, R. R. Fractional Factorial Split-Plot Designs for Robust Parameter Experiments. *Technometrics* 2003, **45**: 80-89.
63. Kowalski, S.M. and Vining, G.G. Split-Plot Experimentation for Process and Quality Improvement, in H. Lenz (ed.), *Frontiers in Statistical Quality Control 6*, Heidelberg: Springer: New York, 335-350, 2001.
64. Goos, P. and Vandebroek, M. Optimal Split-Plot Designs. *Journal of Quality Technology*. 2001, **33**: 436-450.
65. Goos, P. and Vandebroek, M. D-Optimal Split-Plot Designs With Given Numbers and Sizes of Whole Plots. *Journal of Quality Technology*. 2003, **45**: 235-245.

This page is intentionally left blank

## **CHAPTER 7**

### **RESPONSE SURFACE METHODS AND THEIR APPLICATION IN THE TREATMENT OF CANCER WITH DRUG COMBINATIONS: SOME REFLECTIONS**

Kathryn Dawson<sup>1</sup>, T.J. Eller<sup>2</sup>, and Walter H. Carter, Jr.<sup>3</sup>

<sup>1</sup>*Novartis Oncology, 180 Park Ave., Building 105*

*Florham Park, NJ 07932-0675*

<sup>2</sup>*SunTrust Bank, Mail Code CS-RVW-7900, 1001 Semmes Ave.*

*Richmond, VA 23224*

<sup>3</sup>*Department of Biostatistics, Virginia Commonwealth University*

*P.O. Box 980032, Richmond, VA 23298-0032*

The goal of many Phase I clinical studies in cancer treatments is the determination of treatment levels that balances the potential benefit and harm associated with the treatment. In particular, increasing dose levels of the cytotoxic drugs used in cancer therapy are simultaneously associated with increased efficacy and an increase in toxic side effects. In Phase I studies, a maximum tolerated dose (MTD) is determined. This dose is then applied in subsequent efficacy studies. In this discussion, a within-patient evolutionary operations procedure (EVOP) is applied to guide a within-patient titration to the determination of the MTD. This approach can be used in studies involving one or more drugs and can incorporate multiple patient responses. While estimating the MTD associated with an underlying dose-response surface, this method does not require the specification of the functional form of the surface.

#### **1. Introduction**

The determination of the treatment level of a drug for the treatment of cancer occurs in three phases. The challenge is to determine a dose of a cytotoxic agent that maximizes efficacy but simultaneously controls the toxicity at a manageable level. Typically, a Phase I study is performed to estimate the maximum tolerated dose (MTD) while recording any exploratory evidence of efficacy. Since higher doses are thought to be associated with increased efficacy, the study design involves increasing

the dose until the level of toxicity offsets any benefit. One well accepted definition of the maximum tolerated dose (MTD) is "...the dose associated with serious but reversible side effects in a sizeable proportion of patients and the one that offers the best chance for a favorable therapeutic ratio."<sup>1</sup> The objective of the following Phase II studies is the determination of therapeutic effect of the new treatment at its MTD. If, as the result of a Phase II study, it is determined that the drug at the MTD is efficacious in the treatment of disease, it then becomes important to compare the results of treatment with the new drug to those associated with the current therapy. These randomized comparative studies comprise Phase III clinical trials. The use of this three tiered approach to evaluate and determine new therapies for cancer is widely accepted and is vigorously defended by cancer researchers. However, the methods currently in use generally require the specification of the functional form of the dose response relationship and do not adequately consider interaction effects of drugs in combination.

In the classic Phase I trial design, before the trial begins, the researcher specifies the enrollment protocol, consisting of the cohort size, the starting dose, all possible subsequent doses, the dose escalation rules, the trial stopping rules, and the definitions of the dose limiting toxicities (DLTs). The DLTs are generally categorical variables that define a given clinical endpoint, such as suppression of neutrophils/granulocytes (ANC/AGC), into 5 grades. Generally, these categories are further collapsed into DLT categories such as "Yes/No" (e.g., only a grade 5 reaction would be considered a DLT) or "Mild/Moderate/Severe" (e.g., grades 1-2 would be characterized as Mild, 2-3, as Moderate, and 5 as Severe, or dose limiting).

The traditional design is generally defined as one in which groups of three patients are treated. If no DLTs are observed at the first dose level then three more patients are treated at the next highest dose. Otherwise, three more patients are treated at the same dose. If only one in six patients experience a DLT, the dose is escalated again. If more than one in six patients experience a DLT, then the trial is stopped and either the stopping dose, or the next lower dose is declared to be the MTD. In the traditional design, each patient gets treated at one specific dose level and

the reactions of that patient are categorized as whether or not the reactions are associated with a DLT.

The classic Phase 1 design seems to be a reasonable way to explore a dose response relationship with the goal of finding the MTD for a single drug. When multiple drugs are used in combination, the possibility of drug interactions must be considered when estimating the MTD. Procedures currently in use to evaluate drugs in combination are largely *ad hoc*. Often the doses of all but one of the drugs is fixed across the trial, and the dose of only one drug is escalated during the trial. There is a need for more rigorously developed study designs with known properties.

Response surface methods have been used to illuminate the search for appropriate levels of the components in a drug combination. Methods for estimating optimal dose levels for a combination given according to a fixed schedule of administration have been developed.<sup>2</sup> Techniques for testing for drug activity and therapeutic synergism based on confidence regions about the location and response at the optimal treatment levels have also been developed.<sup>3</sup> In another application, a robust parameter design approach to determining desirable treatment levels in the presence of noise factors was derived.<sup>4</sup> None of these methods fit neatly into the three phase paradigm that is so firmly entrenched within the cancer research community of scientists. Further, each of the response surface methods referred to above employs an empirical modeling approach. The inferences drawn using this approach are sometimes criticized even after the adequacy of the model has been demonstrated. Possibly, for these reasons, the response surface methods mentioned above have not been well accepted as a tool for evaluating cancer therapies.

In the usual three phase approach applied to a combination drug therapy, if superior results in the Phase II and III trials are observed at the MTD, the dose combination will be adopted. Otherwise, it is often discarded. One cannot conclude; however, that the dose combination included in the experimental therapy is not effective if the therapy is discarded. Rather the new therapy may have been ineffective because a sub-optimal dose combination was considered. However, because of the structure of the usual Phase I trials little information is provided that can be used to determine a possibly effective and tolerable dose

combination of the agents considered. A more efficient Phase I study design could incorporate an exploration of the assumed underlying dose-response surface. Box was early to suggest that an evolutionary operations approach may be helpful.<sup>5</sup> He suggested that such methods

“...might in fact be used to get maximum information from the normal treatment of patients by practicing practitioners. It would be necessary for a central agency to obtain agreement that doctors, in using a particular therapy in normal practice (as contrasted with special research studies), would vary the therapy slightly in accordance with a prescribed plan. With a suitable statistical plan, differences arising from small deliberate changes in the therapy can be detected when the information is collected. In this way a steady evolution in medical practice might be set in motion to augment more specialized research studies.”

In the following section the application of this type of study design will be applied to the determination of the MTD in cancer studies. A new Evolutionary Operations Procedure (EVOP) approach to determining the MTD is proposed. The EVOP approach can estimate the MTD of a single drug within a patient, or the MTD of a drug combination. In contrast to current classical practice, the EVOP approach evaluates toxicity continuously using any number of clinical endpoints, the methodology does not require any specification of the dose response surface, and the drug(s) are titrated within each patient in order to get a MTD for the individual. Additionally, because intrapatient dose titration is used, every patient has the potential to benefit from treatment with the drug combination under study. This is different from current Phase I trial design where patients who enter the trial early are treated at lower doses which are less likely to offer therapeutic benefit. EVOP methodology uses the patient's own multiple responses to one or more drugs and intrapatient dose titration to explore unknown dose response surfaces without specification of the dose response surface. A further assumption,

not incorporated in classic Phase 1 trial design, is that this relationship may differ among patients.

Desirability functions allow the physician's own priorities and desires about the clinical endpoints to be built into the improvement of the desirability of the drug dose.<sup>6</sup> Search procedures guide the intrapatient dose titration to more toxic drug outcomes. There are many direct search methods to guide the intrapatient titration to the maximum composite desirability, D. In the single drug case, bracketing will find the dose associated with the maximum desirability.<sup>7</sup> In the multi-drug case, the Nelder-Mead search procedure can be employed.<sup>7,8</sup>

## **2. Estimation of the Within Patient MTD Using Evolutionary Operative Procedure (EVOP)**

Shih has used the Evolutionary Operations Procedure (EVOP) approach in the therapeutic context.<sup>9</sup> Specifically, in Shih's work the goal was to determine a dose combination that resulted in beneficial responses within patients. As an extension of Shih's work, the EVOP approach can be applied to the determination of the maximum tolerated dose (MTD) in a single or combination Phase I drug study when intrapatient dose titration is possible. Study designs incorporating intrapatient dosing have been considered. In particular, Simon *et al.*, describes an intrapatient dose titration design for Phase I studies of cytotoxic drugs.<sup>10</sup> In this design the next dose is increased if no serious toxic events were observed at the prior dose. The EVOP technique generalizes this approach by not only evaluating toxicity using any number of clinical endpoints and any number of drugs.

In the protocol development stage of the EVOP Phase I trial estimating a MTD, the physician defines the drugs that compose the treatment, the starting dose(s) and the clinical outcomes that would be followed in the patient. Included in the clinical outcomes are measures of one or more toxicities associated with the treatment. For example, in treatments of cancers involving head and neck, difficulty with swallowing and dryness of the mouth may be two relevant toxicity outcomes. Desirability functions for each clinical endpoint are then defined.<sup>6,11,9</sup> The form of the desirability function as a function of the

clinical endpoint is determined by the physician in conjunction with the statistician. The response of the desirability function ranges from 0, representing a completely undesirable toxicity level, to 1, representing a completely desirable or ideal response. In cytotoxic drug studies, higher doses are associated with increased efficacy. However, as dose increases the toxicity increases. Therefore, a given toxicity desirability function will typically increase until toxicity levels deemed too severe are reached at which point the desirability decreases. In conjunction with the toxicity desirability functions, other relevant clinical desirability functions may also be defined.

During the trial the unknown within-patient dose-response surface is searched for the dose that optimizes the over-all desirability associated with the treatment. During a given dose round a specified dose combination is administered and the clinical outcome observed. The actual number of dose rounds administered to the patient will depend on the time needed to optimize the overall desirability. This process can be defined as follows.

For a given patient let

$M$  = number of drugs included in the treatment

$C$  = number of clinical endpoints to be observed

$x_j$  =  $M \times 1$  vector of dose levels in the  $j^{\text{th}}$  dose round

$e_j$  =  $C \times 1$  vector of clinical endpoints observed in the  $j^{\text{th}}$  dose round.

For each clinical endpoint a desirability function  $d = g_i(e)$ ,  $i=1,2,\dots,C$  is defined. The overall desirability<sup>6</sup> associated with the  $j^{\text{th}}$  dose round is given by

$$D_j = \left( \prod_{i=1}^C g_i(e_{ij}) \right)^{1/C} \quad (1)$$

and is considered the response in the underlying within patient dose-response relationship. The response surface can be written as

$$D = \left( \prod_{i=1}^C g_i \{f_i(x)\} \right)^{1/C} \quad (2)$$

where  $e_i = f_i(x)$ ,  $i=1,2,\dots,C$  are unknown functions of the dose combination  $x$ . Since the functional relationship between dose and outcome,  $e_i = f_i(x)$  is unknown, the patient's clinical responses at a given dose will be observed. In particular, after administering dose combination  $x_j$  in the  $j^{\text{th}}$  dose round, the clinical endpoints,  $e_{1j}, e_{2j}, \dots, e_{Cj}$ , are observed and  $D_j$  determined based on (1). This value of  $D_j$  is a realization of a point on the overall desirability dose-response surface (2) at dose  $x_j$ . A search algorithm using the observed points from all dose rounds  $\{(x_i, D_i), i=1,2,\dots,j\}$  can be used to determine  $x_{j+1}$ , a dose combination directed to the maximum desirability on the within patient dose-response surface.

To more clearly illustrate the dose-response structure just defined, a hypothetical example is described in the next section.

### 3. Hypothetical Example

Suppose a Phase I two drug, three clinical endpoint study is conducted to determine the MTD within a given patient. Let  $(x_1, x_2)$  denote the doses of two drugs where  $0 \leq x_1 \leq 5$ ,  $0 \leq x_2 \leq 10$ . The hypothetical continuous clinical outcomes are platelet count ranging from 0 to 600, a vomiting severity score (0 – 200) and a tumor response score (0 – 1500) measuring efficacy of the treatment. The first rows of Figures 1-3 contain the desirability functions which were theoretically derived based on physician preference and clinical knowledge. The desirability associated with platelet count is maximized at a clinically optimal platelet count (Figure 1c). Values below and above this value are considered less desirable. Vomiting is an anticipated outcome associated with increased doses of cytotoxic drug levels. It is thought that lower levels of this toxicity may indicate higher doses can be tolerated. Therefore, the desirability function associated with the vomiting severity score increases at lower levels of this toxicity (Figure 2a). However, when the vomiting score reaches a level considered too severe the desirability

decreases. Lastly, since increasing tumor response scores are associated with increased efficacy, the desirability function associated with this outcome is nondecreasing (Figure 3a).

For this example, each  $f_i(x)$ ,  $i = 1, 2, 3$  is assumed to be known and is plotted in the second row of Figures 1-3. While these functions will typically be unknown, the form of the underlying surface associated with these particular outcomes may be similar to the figures. For example, Figure 1b indicates that platelet count decreases with increasing dose. The vomiting severity score (Figure 2b) and the tumor response score (Figure 3b) each increase with increasing dose. Since each  $f_i(x)$ ,  $i = 1, 2, 3$  is given, the desirability of each outcome as a function of dose can be determined and are shown in the last row of Figures 1-3. These plots illustrate the desirability of a given dose relative to a fixed outcome. Using (2) the overall desirability dose-response surface can be determined (Figure 4). The dose combination associated with the maximum overall desirability is the MTD. In this example, the Nelder-Mead search procedure may be employed.<sup>8</sup> The search process can be stopped based on many criteria. For example, titration might be stopped when there are no clinically differences in the responses under consideration. Alternatively, a fixed number of dose rounds may be specified in the study design or the search may be stopped based on physician judgment.

For this example, the MTD is (2.6, 7.4). At this dose combination the platelet count is 202 with desirability 0.95, the vomiting severity score was 117 with desirability 0.99, the tumor response score was 524 with desirability 0.29 and the overall desirability at this dose combination is 0.65. While the desirability associated with tumor score is low, indicating higher doses may more beneficial, the toxicity associated with these higher doses offsets the benefit. In application, the response surfaces illustrated in Figure 1 can not be determined. However, the patient's responses at a given dose provide estimates of points on the overall desirability dose-response surface, which can than be searched with direct search algorithms.

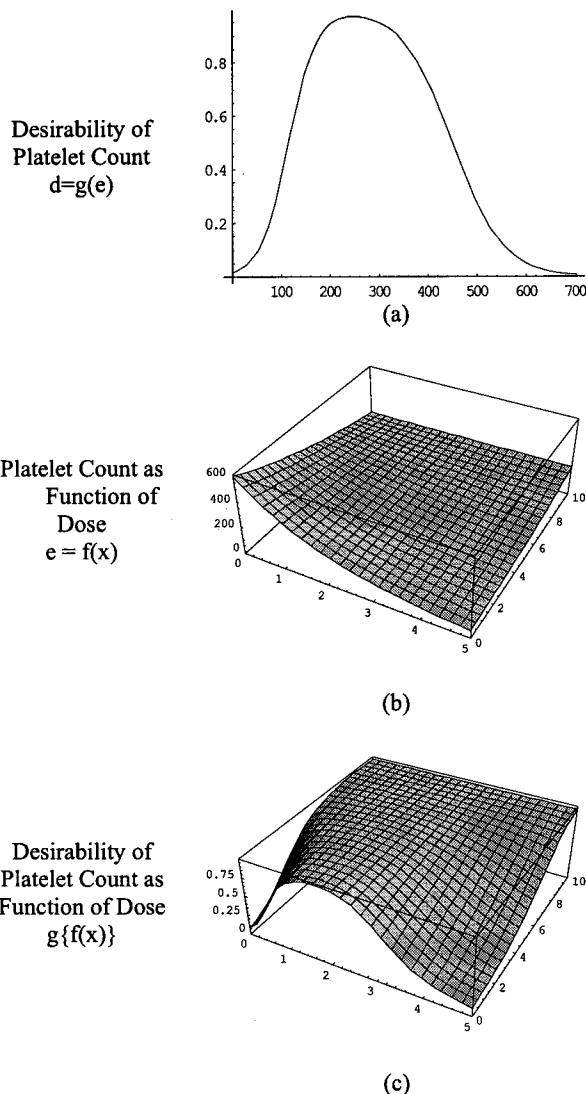


Fig. 1. Desirability of a Hypothetical Clinical Outcome for Platelet Count.

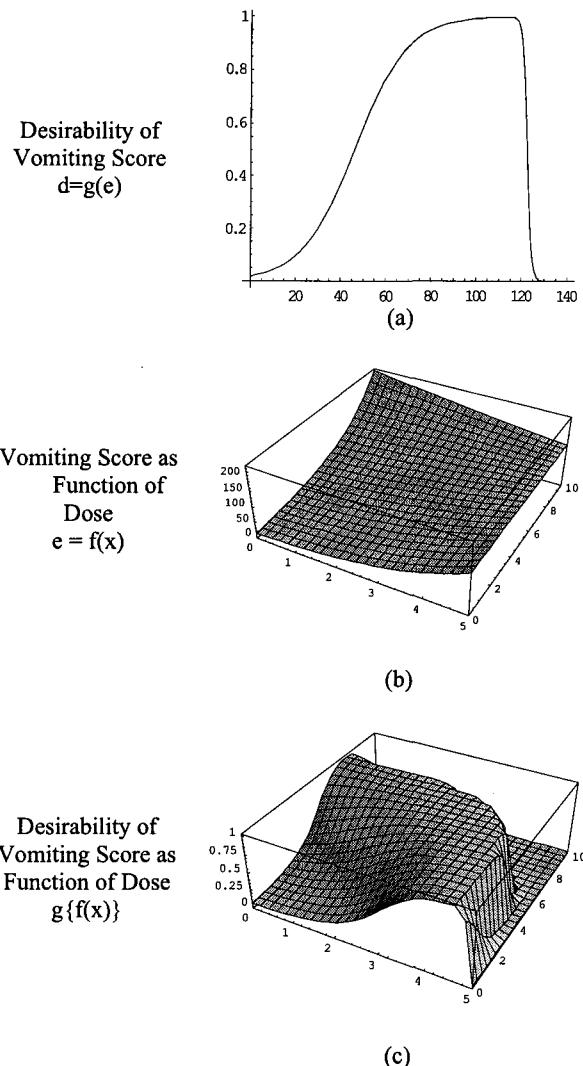


Fig. 2. Desirability of a Hypothetical Clinical Outcome for Vomiting Severity Score.

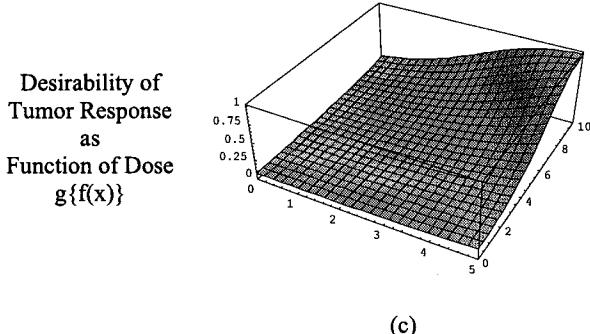
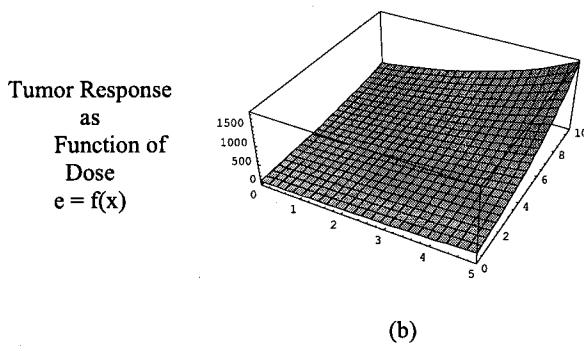
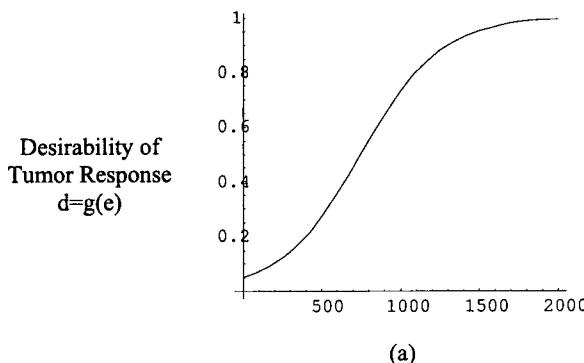


Fig. 3. Desirability of a Hypothetical Clinical Outcome for Tumor Response.

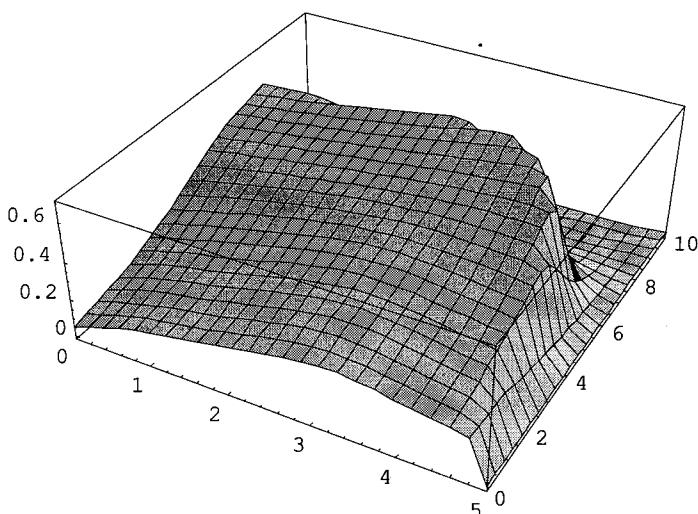


Fig. 4. Overall Desirability as a function of Dose – A Hypothetical Example.

#### 4. Conclusions

When the dose titration procedure within the  $i^{\text{th}}$  patient is completed the  $i^{\text{th}}$  patient's maximum tolerated dose for that patient,  $\text{MTD}_i$ , is estimated as the dose combination associated with the highest overall desirability. Assume a  $M$  drug study includes  $n$  patients. The set  $\{\text{MTD}_1, \text{MTD}_2, \dots, \text{MTD}_n\}$  of  $n$  points in the  $M$  dimensional dose space can be used to estimate an MTD for the population. A confidence ellipsoid placed about this estimate can be used to draw inference about the variability associated with MTD caused both by random error as well as within patient differences.<sup>9</sup> Furthermore, within patient variables can be considered as possible contributors to variability among the estimated MTDs. The physician can use these results to define one or more drug combinations to study in Phase II and III studies.

While an underlying within-patient dose-response surface is assumed, using the EVOP approach the actual functional form of the surface is not specified. Rather, using the patient observed responses at a set of given doses a search algorithm can be employed to determine the next dose

most likely to move in the direction of the optimal response. In this approach, the response is a function of the desirabilities associated with one or more clinical outcomes. By incorporating toxicity outcomes, the MTD of a combination of cytotoxic drugs can be estimated.

## References

1. Piantadosi, S., (1997). Clinical trials: A methodologic perspective, *John Wiley and Sons, Inc.*
2. Stablein, D. M., Carter, W. H., Jr., and Wampler, G.L. (1980). Survival analysis of drug combinations using a hazards model with time-dependent covariates, *Biometrics*, 36, pp. 537-546.
3. Carter, W. H., Jr., Wampler, G. L., Stablein, D. M., and Campbell, E. D. (1982). Drug activity and therapeutic synergism in cancer treatment, *Cancer Research*, 42, pp. 2963-2971.
4. Kuhn, A. M., Carter, W. H., Jr., and Myers, R. H. (2000). Incorporating noise factors into experiments with censored data, *Technometrics*, 42, pp. 376-383.
5. Box, G. E. P. (1958). In discussion of experimental design in combination chemotherapy, *Annals of the New York Academy of Sciences*, 76, pp. 909-931.
6. Harrington, E. C. (1965). The desirability function, *Industrial Quality Control*, 21, pp. 494-498.
7. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P. (1992). Recipes in C, *The Art of Scientific Computing*, Cambridge University Press.
8. Lagarias, J. C., Reeds, J. A., Wright, M. H. and Wright, P. E. (1998). Convergence properties of the Nelder-Mead simplex method in low dimensions, *SIAM Journal of Optimization*, Vol. 9, No. 1, pp. 112-147.
9. Shih, M., Gennings, C., Chinchilli, V. M., and Carter, W. H., Jr. (2003). Titrating and evaluating multi-drug regimens within subjects, *Statistics in Medicine*, 22, pp. 2257-2279.
10. Simon, R. *et.al.* (1997). Accelerated titration designs for Phase I clinical trials in oncology, *Journal of the National Cancer Institute*, August 6, Vol. 89, No. 15, pp. 1138-1147.
11. Gibb, R. D. (1998). Optimal treatment combination estimation for univariate and multivariate response surface applications. Ph.D. Dissertation, Department of Biostatistics, Medical College of Virginia, Virginia Commonwealth University, Richmond, Virginia.

This page is intentionally left blank

## CHAPTER 8

# GENERALIZED LINEAR MODELS AND RESPONSE TRANSFORMATION

Anthony C. Atkinson

*Department of Statistics*

*The London School of Economics, London WC2A 2AE, UK*

*E-mail: a.c.atkinson@lse.ac.uk*

Response surface designs are found for models in which there is a mean-variance relationship. This may arise because generalized linear models are appropriate or because the response requires transformation to have a constant variance. Transformation for both sides of a mechanistic model is introduced. The special problem of design for nonlinear response surface models is resolved through the use of structured parameters.

### 1. Introduction

Response surface models are used when the response  $y$  is a smooth function of parameters and explanatory variables. We write

$$y = h(x, \theta, \epsilon),$$

where  $y$  is a response, usually univariate,  $x$  is a vector of  $m$  explanatory variables or covariates and  $\epsilon$  is an error term. In most of the large literature on designs for response surfaces from Box and Draper<sup>1</sup> to Myers *et al.*<sup>2</sup> it is assumed that the model for the  $i$ th observation can be written in the simpler form

$$y_i = \eta(x_i, \theta) + \epsilon_i, \quad (i = 1, \dots, n) \quad (1)$$

where the additive errors  $\epsilon_i$  are independent and identically distributed with constant variance  $\sigma^2$ . As a consequence, the variance does not depend on the mean of  $y_i$ . This paper is concerned with situations where this simple assumption does not hold and the variance is indeed a function of the mean.

We explore the consequences for experimental design of three ways in which mean variance relationships can be modelled. The paper begins in §2 with a brief introduction to generalized linear models and to the theory of optimum experimental design. In §3 this theory is applied to designs for gamma models, which can be identical to standard response surface designs, and, in §4, to designs for binary responses, which are usually very different. In these numerical examples there are two explanatory variables, whereas most papers in the statistical literature consider designs with a single factor; references are in §4.3. The efficiency of standard designs for generalized linear models is investigated in §5.

An alternative to the mean-variance relationships provided by generalized linear models is the normal-theory response model with parameterized variance functions leading to weighted regression; optimum designs are described in §6. A special case is when the variance is a function of the mean.

The third class of models, but when the variance increases with the mean, uses the family of power transformations introduced by Box and Cox<sup>3</sup> to make the variance approximately independent of the mean. A distinction from the designs of §6 is that now the response has a skewed distribution. Response surface designs for estimation of the transformation parameter are described in §7. If, however, there is a physical relationship between the response and the model in the absence of error, both sides have to be transformed so that the relationship is preserved. The theory of design for these transformations is in §8. Again, the examples in the literature have only one explanatory variable, often time. Section 9 extends the method to several variables through the use of "structured parameters".

The description of response surface models at the beginning of this section is clearly over-simplified. For example, in addition to the continuous variables  $x$  there may be blocking variables, or some of the variables  $x$  may be concomitant variables that are not experimentally controllable. The paper concludes in §10 with a brief discussion of some omitted topics.

## 2. Theory

### 2.1. Generalized Linear Models

In addition to the additive errors assumed in (1) it is customary to assume that the response surface model is linear in the parameters so that

$$y_i = \eta_i + \epsilon_i = f^T(x_i)\beta + \epsilon_i, \quad (2)$$

where  $f(x)$  is a  $p \times 1$  vector of powers and products of the  $m$  explanatory variables. Often the parameter sensitivities  $f(x)$  will be the terms of a second-order polynomial, which can be thought of as coming from Taylor series expansion of the nonlinear model (1). Of course, both the assumed constant variance of the  $\epsilon_i$  and the adequacy of the linear model  $f(x)$  need to be checked against the data.

If the errors  $\epsilon_i$  are normally distributed, (2) defines a linear multiple regression model which can be written as

$$E(y) = \mu = \eta = f^T(x)\beta, \quad (3)$$

where  $\mu$ , the mean of  $y$ , is equal to the linear predictor  $\eta$ . The generalized linear model extends (3) to any distribution belonging to the one-parameter exponential family. As well as the normal, this includes the gamma, Poisson and binomial distributions. A second extension is the introduction of a link function  $g(\mu) = \eta$ , relating the mean and the linear predictor. For the binomial data of §4 the link function is such that, however the values of  $x$  and  $\beta$  vary, the mean  $\mu$  satisfies the physically meaningful constraint that  $0 \leq \mu \leq 1$ . The seminal work on generalized linear models is McCullagh and Nelder<sup>4</sup>. Myers *et al.*<sup>5</sup> emphasize data analysis. Chapter 6 of Atkinson and Riani<sup>6</sup> gives a succinct introduction. None discusses design.

We are particularly interested in the relationship between the mean and the variance of the observations. Let the variance of  $y$  be  $\text{var}(y)$ . Then

$$\text{var}(y) = \phi V(\mu),$$

where  $\phi$  is the dispersion parameter, equal to  $\sigma^2$  for the normal distribution and one for the binomial. The variance function  $V(\mu)$ , which is specific to the error distribution, then determines the relationship between mean and variance.

Optimum experimental design for generalized linear models is aided by the simple form of the information matrix. Maximum likelihood estimation of the parameters  $\beta$  of the linear predictor reduces to iterative weighted least squares, with the weights for individual observations given by

$$w = V^{-1}(\mu) \left( \frac{d\mu}{d\eta} \right)^2. \quad (4)$$

It follows that the weights depend both on the distribution of  $y$  and on the link function.

## 2.2. Optimum Experimental Design

The algorithms of optimum experimental design provide a straightforward way of constructing designs for generalized linear models, which can then be compared with regression designs. Optimum experimental design is described in the books of Pukelsheim<sup>7</sup>, Fedorov and Hackl<sup>8</sup> and Atkinson and Donev<sup>9</sup>, where designs for generalized linear models are discussed in §22.5.

For the nonlinear regression model (1), the parameter sensitivity for observation  $i$  is the  $p \times 1$  vector

$$f(x_i, \theta) = \left\{ \frac{\partial \eta(x_i, \theta)}{\partial \theta_j} \right\}. \quad (j = 1, \dots, p) \quad (5)$$

These sensitivities reduce to those introduced in (2) for the linear regression model.

The contribution of observation  $i$  to the information matrix for weighted least squares estimation of the parameters is proportional to  $f(x_i, \theta)w_i f^T(x_i, \theta)$ . Experimental designs are sought which make large suitable functions of this matrix summed over all observations. The dependence of the information on the number of experimental trials  $n$  can be removed by standardization, working instead with the fractions of trials  $p_i = n_i/n$  at each of the  $k$  support points  $x_i$  of the design. More general mathematical results about designs are obtained on replacing the fractions by a measure  $\xi$ , yielding a ‘continuous’ design in which the  $n_i$  are no longer required to be integer. However practical designs do require that the  $n_i$  be integer.

For an experimental design represented as a measure  $\xi$  over the design region  $\mathcal{X}$ , the information matrix is

$$M(\xi, \theta) = \int_{\mathcal{X}} f(x, \theta) w f^T(x, \theta) \xi(dx) = E_{\xi} f(x) w f^T(x, \theta) = E_{\xi} I(x, \theta),$$

which is a function of  $\theta$ . Optimum design theory is concerned with minimization of the convex function  $\Psi\{M(\xi, \theta)\}$ . Because this is a well-behaved optimization problem, the optimality of any design  $\xi$  can be checked by using the directional derivative  $\phi(x, \xi, \theta)$ . For the optimum design  $\xi^*$ ,  $\phi(x, \xi^*, \theta)$  is zero at the design points and positive elsewhere. The equivalence theorems for various criteria follow from this result.

One frequently used design criterion, and that used in this chapter, is  $D$ -optimality in which the generalized variance of the parameter estimates is minimized, when

$$\Psi\{M(\xi, \theta)\} = -\log |M(\xi, \theta)|,$$

equivalent to maximizing  $|M(\xi, \theta)|$ . The equivalence theorem relates  $D$ -optimality to the variance of the prediction  $\hat{y}(x)$  for the design  $\xi$ , standardized by taking the error variance as one. Let

$$d(x, \xi, \theta) = wf^T(x, \theta)M^{-1}(\xi, \theta)f(x, \theta). \quad (6)$$

Then the derivative  $\phi(x, \xi, \theta) = p - d(x, \xi, \theta)$  and the maximum value,  $p$ , of  $d(x, \xi^*, \theta)$  occurs at the points of support of the design. This condition provides a method of checking the optimality of a purported optimum design.

$D$ -optimum designs for the linear regression model (2) with additive independent errors of constant variance do not depend on the value of  $\beta$ . However, the designs found in this chapter depend not only on the model and on the design region  $\mathcal{X}$ , but also on the value of the parameter  $\theta$ . We find locally optimum designs, that is designs that are optimum for a specified value  $\theta_0$  of  $\theta$ . Bayesian optimum designs where  $\theta$  has a prior distribution are briefly mentioned in §10.

### 3. Optimum Designs for Gamma Models

The gamma model is often an alternative to response transformation. In particular, with a log link, it is often hard to distinguish the gamma from a linear regression model with logged response. Examples are given in §7.5 of Myers *et al*<sup>5</sup>.

A useful, flexible family of links is the Box and Cox family

$$g(\mu) = (\mu^\lambda - 1)/\lambda = \eta, \quad (7)$$

which is continuous as  $\lambda \rightarrow 0$ , yielding the log link. Differentiation of (7) yields

$$\frac{d\eta}{d\mu} = \mu^{\lambda-1}. \quad (8)$$

Since the variance function for the gamma distribution is

$$V(\mu) = \mu^2,$$

the combination of (4) and (8) shows that the weights for the gamma distribution with this link family are

$$w = V^{-1}(\mu) \left( \frac{d\mu}{d\eta} \right)^2 = \mu^{-2\lambda} = 1/(1 + \lambda\eta)^2. \quad (9)$$

When  $\lambda = 0$ , that is for the log link, (9) shows that the weights are equal to one. It therefore follows that optimum designs for gamma models with this link are identical to optimum designs for the regression model (2), of course when the linear predictors  $\eta$  are the same. Comparisons of the analyses of data from designed experiments when the data are transformed or analysed using a gamma model with log link are given by Lewis *et al*<sup>10</sup>.

The analysis of regression data, when the model includes a constant term, is the same, for fixed  $\lambda$ , whether the Box and Cox transformation is used, or the straight power transformation (Schlesselman<sup>11</sup>). But the Box and Cox link (7) does not yield exactly the same designs as the power family of links  $g(\mu) = \mu^\lambda$ . Burridge and Sebastiani<sup>12</sup>, see §4.3, find optimum designs for the power link when the linear predictor is first order and  $\lambda > 0$ . They show that now

$$w = \mu^{-2\lambda} / \lambda^2 = 1/(\lambda\eta)^2. \quad (10)$$

The multiplying factor  $\lambda^{-2}$  applies to all design points equally and so does not affect the design, which is independent of  $\lambda$ . A consequence is that the optimum designs are discontinuous as  $\lambda \rightarrow 0$ .

To find designs that illustrate the difference between regression and the gamma GLM with the Box and Cox link, we need to use a value of  $\lambda \neq 0$ . We investigate designs for a two-factor example when the linear predictor is second-order.

Atkinson and Riani<sup>6</sup> §6.9 use a gamma model to analyse data from Nelson<sup>13</sup> on the degradation of insulation due to elevated temperature at a series of times. They do not consider design. These data do not yield a particularly clean model as there seem to be some identifiable subsets which do not completely agree with the fitted response-surface model. However, for our purposes, a second-order model is required in the two continuous variables and the gamma model fits best with a power link with  $\lambda = 0.5$ . A theoretical difficulty with such a value of  $\lambda$  is that  $\mu$  must be  $> 0$ , while  $\eta$  is, in principle, unconstrained.

We scale the variables so that the design region  $\mathcal{X}$  is the unit square with vertices  $(-1, -1)$ ,  $(-1, 1)$ ,  $(1, -1)$  and  $(1, 1)$ . The linear predictor is the quadratic

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2. \quad (11)$$

Then the standard  $D$ -optimum design for the normal theory regression model given in Table 6 has unequally weighted support at the points of the  $3^2$  factorial. This design is, from what was said above, also optimum for the gamma model with log link.

For other links the design will depend on the actual values of the parameters  $\beta$  in (11). Any design found will therefore be locally optimum. We take  $\beta$  to have the values given in Table 1, the set G1 being rounded from an analysis of Nelson's data.

The optimum design for G1 is in Table 2. This shows that, at the points of the design, the minimum value of  $\mu$  is 2.00 and the maximum 15.60. If these were normal responses that had to be non-negative, this kind of range would indicate a power transformation.

The optimum design was found by a grid search with steps of 0.05 in  $x_1$  and  $x_2$ . For each set of  $k$  support points the optimum design weights  $p_i$  were found by numerical search using the transformation to  $(k - 1)$ -dimensional polar co-ordinates described by Atkinson and Donev<sup>9</sup>, p.104. The variance of prediction  $d(x, \xi, \theta)$  (6) was calculated at each grid point and the point with the maximum value added to the design, whilst one point was deleted. The process continued until the point to be added was already in the design.

Table 1.  $D$ -optimum designs for two gamma models: parameter values for linear predictors, second-order in two variables.

Design	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_{11}$	$\beta_{22}$	$\beta_{12}$
G1	5.47	-0.92	-1.30	-0.38	-0.90	-1.14
G2	5.47	-0.46	-0.65	-0.19	-0.45	-0.57

Table 2.  $D$ -optimum designs for the parameter sets G1 and G2 of Table 1.

Design Point $i$	Design G1				Design G2			
	$x_{1i}$	$x_{2i}$	$p_i$	$\mu_i$	$x_{1i}$	$x_{2i}$	$p_i$	$\mu_i$
1	-1.00	-1.00	0.163	13.21	-1.00	-1.00	0.150	13.58
2	-1.00	0.05	0.016	15.60	-1.00	0.00	0.057	14.98
3	-1.00	1.00	0.162	12.08	-1.00	1.00	0.150	13.00
4	0.00	-1.00	0.003	15.48	0.00	-1.00	0.053	14.71
5	0.15	0.15	0.060	12.54	0.05	0.05	0.086	13.73
6	0.25	1.00	0.144	5.60	0.10	1.00	0.107	9.81
7	1.00	-1.00	0.163	14.86	1.00	-1.00	0.149	14.40
8	1.00	0.25	0.129	7.57	1.00	0.10	0.097	11.20
9	1.00	1.00	0.160	2.00	1.00	1.00	0.151	6.63

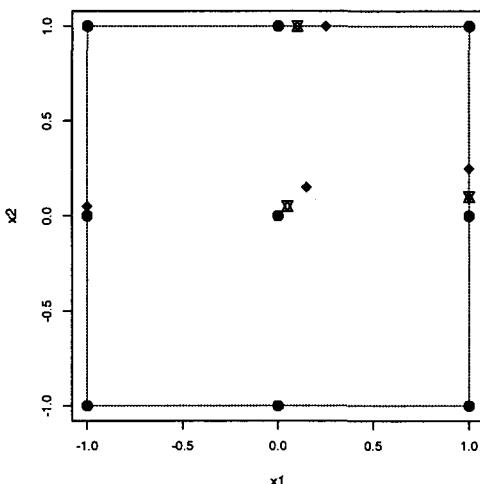


Fig. 1. Support points for  $D$ -optimum designs for gamma models in Table 2: ●, the points of the  $3^2$  factorial; rhombus, G1 and star, G2.

As Table 2 and Figure 1 show, the support points of the design are a slight, almost symmetrical, distortion of those of the  $3^2$  factorial. However, the design weights are not symmetrical, with a minimum weight of 0.003 on design point 4.

There are two sets of conditions under which the optimum design for the gamma model is identical to the optimum design for the second-order normal theory response surface model that has support at the points of the  $3^2$  factorial. One occurs when the log link is appropriate, whatever the values of the parameters  $\beta$  in the linear predictor. The other is when the effects of the variables are small, that is when the parameters in the linear predictor, other than  $\beta_0$ , tend to zero (Cox<sup>14</sup>). To illustrate this point we also found the  $D$ -optimum design for the set of parameter values G2 in Table 1 in which all parameters, other than  $\beta_0$ , have half the values they have for design G1. As Table 2 shows, the range of means at the design points is now 6.63 to 14.98, an appreciable reduction in the ratio of largest to smallest response. The support points of the design are shown in Figure 1 by open stars: all are moved in the direction of the factorial design when compared to the points of support of G1. The results in Table 2 show that the optimum design weights are now more nearly even, with a minimum value of 0.053 to be compared with the previous minimum of 0.003.

The relationship between these two-variable response surface designs for gamma models and the normal theory response surface design is explored further in §5, where we calculate the efficiency of standard designs, including the  $3^2$  factorial, for our non-standard problem.

## 4. Models and Designs for Binomial Data

### 4.1. One Variable

If designs for gamma models seem much like those for normal theory models, the same is not true for designs for binary data.

The greatest interest in the literature, see §4.3, has been in logistic models for binary data, particularly with one explanatory variable. This linear logistic model is that

$$\log\{\mu/(1-\mu)\} = \eta = \alpha + \beta x, \quad (12)$$

where  $\mu$  is the probability of failure, often death in, for example, insecticide studies. Other links include the probit, the complementary loglog and arcsine, which has good robustness properties. Discussion of these links and examples of their use in the analysis of data can be found in Atkinson and Riani<sup>6</sup> §§6.13-6.20.

For the logistic link in (12) the variance function is

$$V(\mu) = \mu(1-\mu).$$

Differentiation of the logistic link yields

$$\frac{d\mu}{d\eta} = \mu(1-\mu).$$

The iterative weight, for this link, is thus

$$w = \mu(1-\mu).$$

As might be expected, experiments with  $\mu$  near to zero or one are thus non-informative; a set of nearly all successes or failures does not provide good parameter estimates.

For a sufficiently large design region  $\mathcal{X}$ , the  $D$ -optimum design for  $\alpha = 0, \beta = 1$  puts half the trials at  $x = -1.543$  and half at  $x = 1.543$ . This is a special case of the result that  $p$ -point  $D$ -optimum designs for models

with  $p$  parameters put weight  $1/p$  at each design point. As Figure 1 shows,  $D$ -optimum designs may have more than this minimum number of support points; for the designs in the figure  $p = 6$  and  $k = 9$ .

#### 4.2. Response Surface Designs

The properties of designs for response surface models, that is with two or more explanatory variables, depend more strongly on the experimental region and on the values of the parameters than do those for the gamma model §3. Although it was assumed in the previous section that the experimental region  $\mathcal{X}$  was effectively unbounded, the design was constrained by the weight  $w$  to lie in a region in which  $\mu$  was not too close to zero or one. But with more than one explanatory variable constraints on the region are necessary. For example, for the two variable model

$$\log\{\mu/(1-\mu)\} = \eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2, \quad (13)$$

with  $\beta^T = (0, \gamma, \gamma)$ , all points for which  $x_1 + x_2 = 0$  yield a value of 0.5 for  $\mu$ , however extreme the values of  $x$ . We now explore designs for the linear predictor (13) with the logistic link for a variety of parameter values.

Four sets of parameter values are given in Table 3. In all cases we take the same design region as before, that is  $\mathcal{X}$  is the square with vertices  $\pm 1$ . The  $D$ -optimum designs for the sets B1 and B2 are listed in Table 4. The values of B1 (0, 1, 1) are smallest and the table shows that the design has support at the points of the  $2^2$  factorial, although the design weights are not quite equal, as they would be for the normal theory model and also for the logistic model as  $\beta_1$  and  $\beta_2 \rightarrow 0$ . At those factorial points for which  $x_1 + x_2 = 0$ ,  $\mu = 0.5$  since  $\beta_1 = \beta_2$ . At the other design points  $\mu = 0.119$  and 0.881, slightly more extreme values than the values of 0.176 and 0.824 for the experiment with a single  $x$ .

Table 3.  $D$ -optimum designs for four binomial models with logistic link: parameter values for linear predictors, first order in two variables.

Design	$\beta_0$	$\beta_1$	$\beta_2$
B1	0	1	1
B2	0	2	2
B3	2	2	2
B4	2.5	2	2

Table 4.  $D$ -optimum designs for binomial models with the parameter sets B1 and B2 of Table 3.

Design Point $i$	Design B1				Design B2			
	$x_{1i}$	$x_{2i}$	$p_i$	$\mu_i$	$x_{1i}$	$x_{2i}$	$p_i$	$\mu_i$
1	1	-1	0.296	0.500	-1.0	1.0	0.251	0.500
2	-1	-1	0.204	0.119	-1.0	0.1	0.142	0.142
3	1	1	0.204	0.881	-0.1	1.0	0.142	0.858
4	-1	1	0.296	0.500	0.1	-1.0	0.095	0.142
5					1.0	-0.1	0.095	0.858
6					1.0	-1.0	0.275	0.500

A most interesting feature of our example is that the number of support points of the design depends upon the values of the parameters  $\beta$ . From Carathéodory's Theorem (Silvey<sup>15</sup>, Appendix 2), the maximum number of support points required by an optimum design is  $p(p+1)/2$ . Our second set of parameters, B2 in which  $\beta^T = (0, 2, 2)$ , gives an optimum design with six support points, that is the value of this bound when  $p = 3$ . These points are places where  $\mu = 0.142, 0.5$  and  $0.858$ . Dr T.H. Waterhouse has commented that this six-point design is not unique. There are also two designs with unequal weights on four of the six support points with the same value of the optimality criterion as the design in Table 4. The six-point design given here is a non-unique convex combination of these four-point designs.

The relationship between the support points of the design and the values of  $\mu$  is highlighted in Figure 2 where the pale areas are regions in which

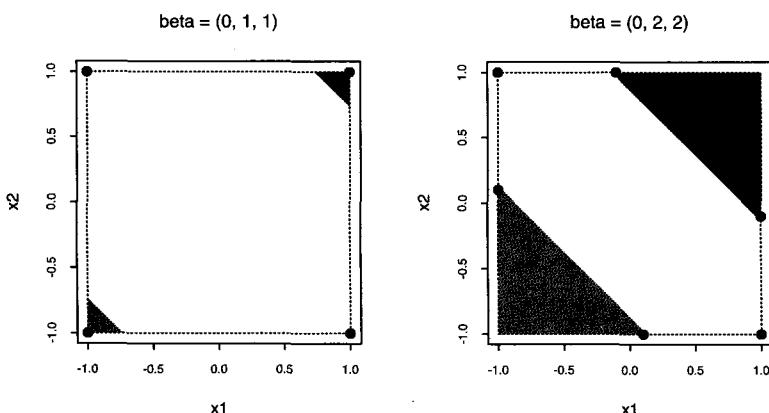


Fig. 2. Support points for  $D$ -optimum designs for binomial models B1 and B2 in Table 3. In the lightly shaded area  $\mu \leq 0.15$ , whereas, in the darker region,  $\mu \geq 0.85$ .

$\mu \leq 0.15$ , with the dark regions the complementary ones where  $\mu \geq 0.85$ . Apart from the design points where  $\mu = 0.5$ , all other design points are close to those boundaries of these regions where  $\mu$  is around 0.15 and 0.85.

The  $D$ -optimum designs for the two remaining sets of parameters in Table 3 are given in Table 5. These designs have respectively 4 and 3 points of support. When  $\beta^T = (2, 2, 2)$ , the design points are where  $\mu = 0.182$  and 0.818. For  $\beta^T = (2.5, 2, 2)$  the values are 0.182 and 0.832. For this three-point design for a three parameter model, the design weights  $p_i = 1/3$ .

The relationship between the design points and the values of  $\mu$  are shown, for these designs, in Figure 3. For  $\beta^T = (2, 2, 2)$  the design points lie slightly away from the boundaries of the regions of high and low values of  $\mu$ , as they do to a lesser extent in the right-hand panel of the figure. With  $\beta^T = (2.5, 2, 2)$  the minimum value of  $\mu$ , 0.182 at  $(-1, -1)$ , is sufficiently high that there are no experimental conditions for which  $\mu = 0.15$ : the panel of the figure contains only one shaded area.

Table 5.  $D$ -optimum designs for binomial models with the parameter sets B3 and B4 of Table 3.

Design Point $i$	Design B3				Design B4			
	$x_{1i}$	$x_{2i}$	$p_i$	$\mu_i$	$x_{1i}$	$x_{2i}$	$p_i$	$\mu_i$
1	-1.00	-0.75	0.168	0.182	-1.00	0.55	0.333	0.832
2	-1.00	0.75	0.332	0.818	-1.00	-1.00	0.333	0.182
3	-0.75	-1.00	0.168	0.182	0.55	-1.00	0.333	0.832
4	0.75	-1.00	0.332	0.818				

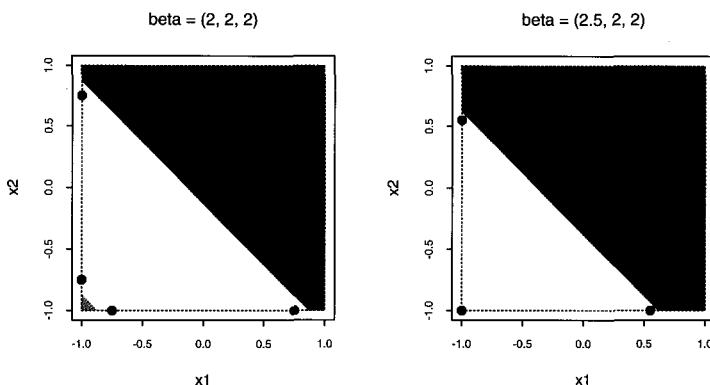


Fig. 3. Support points for  $D$ -optimum designs for binomial models B3 and B4 in Table 3. In the lightly shaded area  $\mu \leq 0.15$ , whereas, in the darker region,  $\mu \geq 0.85$ .

### 4.3. The Induced Design Region

Designs for the one-variable logistic model (12) were studied by Chaloner and Larntz<sup>16</sup>; their optimum design given in §4.1 for  $\alpha = 0, \beta = 1$  puts half the trials at  $x = -1.543$  and half at  $x = 1.543$ , where the expected response is 0.176 and 0.824. Although designs for other values of the parameters can likewise be found numerically, the  $D$ -optimum design in this problem puts the trials where the expected response has these values, the translation into the experimental variable  $x$  depending on the values of  $\alpha$  and  $\beta$ . This canonical form, explored by Ford *et al.*<sup>17</sup> comes from the estimation procedure for generalized linear models. Because weighted least squares is used, design for the logistic model (13) is equivalent to design for the linear model

$$\eta = \beta_0 \sqrt{w} + \beta_1 \sqrt{wx_1} + \beta_2 \sqrt{wx_2}, = \beta_0 z_0 + \beta_1 z_1 + \beta_2 z_2. \quad (14)$$

The design region  $\mathcal{X}$  is then replaced by the induced design region  $\mathcal{Z}$ , the space in which the values of  $z$  can fall as  $x$  varies. Since  $p = 3$ , the induced design space  $\mathcal{Z}$  is of dimension three. Two examples, projected onto  $z_1$  and  $z_2$  and so ignoring  $z_0 = \sqrt{w}$ , are given in Figure 4 for  $\mathcal{X}$  the unit square. In the left-hand panel of the figure  $\beta^T = (0, 2, 2)$  so that at the corner of  $\mathcal{X}$  for which  $x_1 = x_2 = 1$ ,  $\eta = 4$  and  $\mu = 0.982$ . This is well beyond the range for informative experiments and the projection of the induced design space appears to be folded over. As a consequence, experiments at extreme positions in  $\mathcal{Z}$  are not at extreme points in  $\mathcal{X}$ . The results in the other panel for  $\beta^T = (2, 2, 2)$  are similar, but more extreme. For both sets of parameter values the design points lie, as they should, on the boundary of the design region.

Ford *et al.*<sup>17</sup> used the induced design region to provide a canonical form for designs under a variety of optimality criteria with several links for binary data when there is one explanatory variable. Extensions to multi-factor first-order models for binary data are given by Sitter and Torsney<sup>18</sup> and Torsney and Gunduz<sup>19</sup>. A difficulty is that, although the support points of the design can be found from  $\mathcal{Z}$ , numerical calculation is required to find the optimum weights. Use of the technique by Burridge and Sebastiani<sup>12</sup>, mentioned in §3, for gamma models, led to conditions on the values of the parameters  $\beta$  under which the optimum designs have only one factor at a time not at the lower level and other conditions when two-level factorial designs are optimum. These latter conditions are a restatement of the results of Cox<sup>14</sup>.

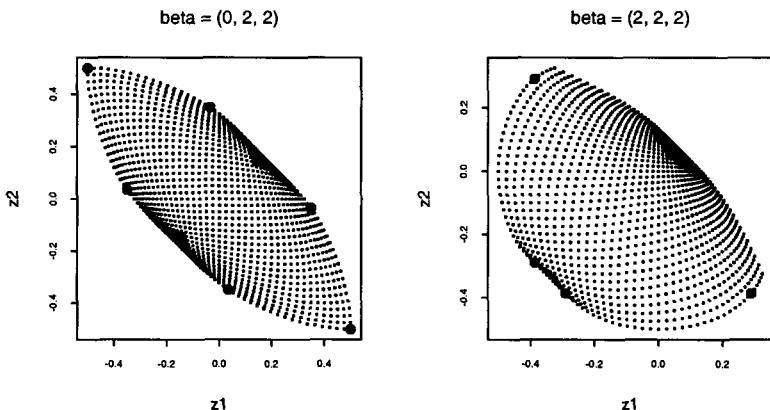


Fig. 4. Support points for  $D$ -optimum designs for binomial models B2 and B3 in Table 3 plotted in the induced design region  $\mathcal{Z}$ .

The extension of the procedure based on (14) to second-order models such as (11) is not obvious. The problem is the way in which the weights enter in the transformation from  $\mathcal{X}$  to  $\mathcal{Z}$ . If, as in (14),  $z_j = \sqrt{w}x_j$ , then, for example, the interaction term in the linear predictor  $\sqrt{w}x_jx_k$  is not equal to  $z_jz_k$ .

These examples show the importance of both the design region and the value of  $\mu$  in determining the optimum design. In order to reveal the structure of the designs as clearly as possible, the designs considered have all had  $\beta_1 = \beta_2$ , and so are symmetrical in  $x_1$  and  $x_2$ . Both the design region and the values of  $\mu$  are important in the asymmetric designs when the two parameter values are not equal. Asymmetric designs also arise with the log log and complementary log log links, since these links are not symmetrical.

In the next section we see how well our designs for the logistic link can be approximated by standard designs such as the  $2^2$  factorial.

## 5. Standard Designs and Generalized Linear Models

### 5.1. Gamma Models

The designs for second-order gamma models with the parameter sets G1 and G2 of Table 1 are both slight distortions of the  $3^2$  factorial. As the values of the parameters, apart from  $\beta_0$ , tend to zero, the design tends towards the  $D$ -optimum design for the second-order regression model. This design, given in Table 6 with minimal rounding of the weights, has unequal support at the points of the  $3^2$  factorial, the weights depending on the number of

Table 6.  $D$ -optimum designs for two-factor normal theory response surface model with design region the unit square.

$x_{1i}$	-1	-1	-1	0	0	0	1	1	1
$x_{2i}$	-1	0	1	-1	0	1	-1	0	1
$p_i$	$p^{(1)}$	$p^{(2)}$	$p^{(1)}$	$p^{(2)}$	$p^{(3)}$	$p^{(2)}$	$p^{(1)}$	$p^{(2)}$	$p^{(1)}$
where	$p^{(1)} = 0.1458$				$p^{(2)} = 0.0801$				$p^{(3)} = 0.0962$

non-zero co-ordinates of the design point (Farrell *et al.*<sup>20</sup>; Atkinson and Donev<sup>9</sup>, p.130). The simplest design with this support is the  $3^2$  factorial in which all weights are equal to 1/9.

We compare three designs for their efficiency relative to the  $D$ -optimum design of Table 2, that is the design for the more extreme parameter set G1 of Table 1. To calculate the efficiency of the designs let the determinant of the information matrix for the optimum design be  $|M(\xi^*)|$  and, for some other design be  $|M(\xi)|$ . Then the efficiency is

$$D_{\text{eff}} = \{|M(\xi)| / |M(\xi^*)|\}^{1/p} \quad (15)$$

with  $p = 6$ . The three designs for comparison are: the  $D$ -optimum design for the less extreme parameter set G2, also given in Table 2, the  $D$ -optimum regression design of Table 6 and the equi-replicated  $3^2$  factorial. The resulting efficiencies are in Table 7.

Table 7. Percentage efficiencies  $D_{\text{eff}}$  of approximations to the  $D$ -optimum design for the second-order response surface gamma model with parameters G1 of Table 1.

$D$ -optimum design G2, Table 2	98.11
$D$ -optimum normal theory, Table 6	95.42
Unweighted $3^2$ factorial	92.39

The main feature of these designs is how efficient they are for the gamma model, all efficiencies being greater than 90%. The designs in the table are ordered by their closeness to that for the gamma model. The design for parameter G2 is for a model with smaller effects than G1, so that the design is between that for G1 and the design for the normal theory model. The equi-weighted  $3^2$  factorial is furthest from the other designs, which all put greatest weight on the extreme points of the design region, that is at the points of the  $2^2$  factorial.

## 5.2. Binomial Models

An indication of the example with a gamma model is that standard designs may be satisfactory for gamma models. The same conclusion does not hold for our binomial example.

Table 8. Percentage efficiencies  $D_{\text{eff}}$  of two approximations to the  $D$ -optimum design for the first-order response surface binomial model. The design in which only one factor at a time is changed has three support points.

$\beta$	$2^2$ Factorial	One Factor
(0, 1, 1)	98.74	92.97
(0, 2, 2)	76.56	79.19
(2, 2, 2)	71.28	92.90
(2.5, 2, 2)	67.66	88.68

Efficiencies were calculated for the four parameter sets of Table 3 for two standard designs. One was the  $2^2$  factorial with weight 1/4 at the four design points  $(-1, -1), (-1, 1), (1, -1)$  and  $(1, 1)$ . The other design puts weight 1/3 at the first three of these support points. This “one factor at a time” design excludes the treatment combination when both factors are simultaneously at their high level. The efficiencies are calculated as in (15), but now  $p = 3$ .

The eight efficiencies in Table 8 are readily interpreted with the help of Figures 2 and 3. For  $\beta^T = (0, 1, 1)$  the optimum design is supported on the points of the  $2^2$  factorial. The unweighted  $2^2$  factorial has an efficiency of 98.74% for this parameter value, with the one factor at a time design, which has support at three of the four points of the optimum design, having a slightly lower efficiency of 92.97%. When  $\beta^T = (0, 2, 2)$ , the two standard designs share only two points of support with the optimum design. They have efficiencies of 76.56 and 79.19%. The efficiency for the one factor at a time design is slightly higher since it puts weight 2/3 on the common points, as opposed to 0.5 for the factorial design. For the other two designs, the one-factor design provides the more efficient approximation. This is because the support of both optimum designs is such that  $x_1 + x_2 \leq 0$ , a condition also satisfied by the points of the one-factor design. The unique point of the  $2^2$  factorial, that is  $(1, 1)$ , is virtually non-informative, so that one quarter of the experimental effort is wasted. Indeed, the efficiencies of the factorial design for  $\beta^T = (2, 2, 2)$  and  $(2.5, 2, 2)$  are both less than 75%. These values are to be compared with 92.9 and 88.68% for the one-factor design.

The results for the gamma and binomial models have something in common. When the parameter effects are small, the  $D$ -optimum designs are virtually those for the normal model, the unequally weighted  $3^2$  factorial of Table 6 for the second-order model or the  $2^2$  factorial for the first-order model. Their higher order analogues, the weighted  $3^m$  factorial and the unweighted  $2^m$  factorial are close to optimum when there are  $m$  factors. These designs are also optimum for the gamma model with the log link. Our results show that these factorial designs, as well as the unweighted  $3^m$  factorial, are efficient for the gamma model with a different link. However, for the binomial response, the design depends strongly on the values of the parameters, with regions in which the value of  $\mu$  is appreciably greater than 0.85 or less than 0.15 best avoided. Finding approximate designs for binomial models using regression analogues will therefore involve searching over irregular design regions where the response does not fall outside these limits. An example would be the hexagon formed by the six design points of the right-hand panel of Figure 2. Searching over the original design space to find the optimum design for the generalized linear model would both be easier and lead to a more efficient design than would trying to find such a regression approximation.

## 6. Structured Mean-Variance Relationships

We now return to designs for continuous responses. In §3 these responses had a gamma distribution with the variance proportional to the square of the mean. In this section we concentrate on normal, but heteroscedastic, responses, so that the distribution is symmetrical. In the remaining sections of the paper the emphasis is on response transformation, which arises from the attempt to normalize asymmetrical response distributions in which the variance is a function of the mean. In this section we outline the design consequences of a parameterized variance function. The details are given by Atkinson and Cook<sup>21</sup>.

Statistical models in which both means and variances are functions of explanatory variables have become increasingly important in quality control (Nelder and Lee<sup>22</sup>, Box<sup>23</sup>), although the design consequences have been less explored. The possibility of additive heteroscedastic errors, known up to a constant of proportionality, is routinely considered by, for example, Fedorov<sup>24</sup>. Here the model has the more general form

$$y = f^T(x)\beta + \sigma[v\{g^T(z)\alpha\}]^{1/2}\epsilon, \quad (16)$$

where  $x$  and  $z$  are design vectors of dimension  $m_x$  and  $m_z$  with  $f(x)$  and  $g(z)$  respectively  $p \times 1$  and  $q \times 1$  vectors of linearly independent continuous functions, as was  $f(x)$  in (2). The error term  $\epsilon$  is standardised to have expectation zero and unit variance. In order to derive information matrices it will, in addition, be taken to have a normal distribution. The unknown parameters are  $\alpha$ ,  $\beta$  and  $\sigma > 0$ . It follows from (16) that, at the point  $(x, y)$ ,  $E(y) = f^T(x)\beta$  and  $\text{var}(y) = \sigma^2[v\{g^T(z)\alpha\}]$ .

For applications it is often useful to take  $v$  to be the exponential function and then to work with a linear model for the logarithm of the variance

$$\log\{\text{var}(y)\} = \log\sigma^2 + g^T(z)\alpha. \quad (17)$$

Atkinson and Cook<sup>21</sup> identify two special cases of (16) that deserve attention. One is when the design variables influencing the mean are the same as those influencing the variance, that is  $x = z$ , so that (16) becomes

$$y = f^T(x)\beta + \sigma[v\{g^T(x)\alpha\}]^{1/2}\epsilon.$$

A further specialization is when the variance depends on  $x$  only through the mean so that

$$y = f^T(x)\beta + \sigma[v\{\nu f^T(x)\beta\}]^{1/2}\epsilon, \quad (18)$$

where  $\nu$  is an unknown real-valued parameter that allows for the strength of dependence of the variance function on the mean.

The structure of the information matrices reflects the contributions to the estimation of the parameters from information coming from the mean and from the variance. When  $\alpha = \alpha_0$  and  $\sigma^2 = \sigma_0^2$  are known the information per observation on  $\beta$  in (16) has the well-known form

$$I_m(x, z|\beta, \alpha_0, \sigma_0^2) = \frac{f(x)f(x)^T}{\sigma_0^2 v\{g^T(z)\alpha_0\}}, \quad (19)$$

leading to estimation by (non-iterative) weighted least squares. The information for known  $\beta = \beta_0$  can likewise be found and does not depend on these parameters for the linear model for the mean. The  $(p+q+1) \times (p+q+1)$  information matrix for all the parameters is therefore block diagonal. A consequence is that the  $D$ -optimum design criterion for  $\alpha$ ,  $\beta$  and  $\sigma^2$  is the product of the design criteria for the two individual sets of parameters.

For (18), in which the variance is a function of the mean, there are  $p + 2$  parameters. When the variance has the exponential form (17) the information matrix for one observation can be written as

$$I(x|\beta, \nu, \sigma^2) = z_1 z_1^T + z_2 z_2^T, \quad (20)$$

where

$$z_1^T(x|\beta, \nu, \sigma^2) = \{\nu f^T(x), f^T(x)\beta, \sigma^2\}/\sqrt{2}$$

and

$$z_2^T(x|\beta, \nu, \sigma^2) = \{f^T(x)\nu^{-1/2}\{\nu f^T(x)\beta\}, 0, 0\}/\sigma^2.$$

Comparison of (20) with (19) shows the extra precision that can be obtained when information about  $\beta$  comes both from the structure of both the mean and of the variance.

Atkinson and Cook<sup>21</sup> give examples of designs for a two-factor response surface model. For (16) the optimum design depends on all  $p + q + 1$  parameters in the model. Both locally optimum and Bayesian designs are found for several parameter values and variance structures. Although some of the designs are like slightly distorted versions of the  $3^2$  factorial, other designs can be very different, particularly when both the mean and variance are unknown and the variance changes appreciably over the experimental region. Similar derivations, a discussion of estimation and an example of a design for a single-factor nonlinear model are given by Downing *et al.*<sup>25</sup>. The extension to generalized linear models is in Rodrigues Pinto and Ponce de Leon<sup>26</sup>. They give a response surface example with a Poisson mean structure and a gamma structure for the variance.

## 7. Design for a Response Transformation

Often in modelling response surfaces the variance can be made independent of the mean by use of the power transformation of Box and Cox<sup>3</sup>. With  $y(\lambda)$  the scaled power transformation of the positive, univariate response variable  $y$ , the linear response surface model (2) becomes

$$y_i(\lambda) = \frac{(y_i^\lambda - 1)}{\lambda} = f^T(x_i)\beta + \epsilon_i, \quad (21)$$

for some unknown value of  $\lambda$ .

Usually experiments are designed to estimate the parameters  $\beta$ . Any transformation, if needed, is considered separately and does not affect the design. As an alternative, Atkinson and Cook<sup>27</sup> find  $D$ -optimum designs for simultaneous estimation of all the parameters  $\beta$ ,  $\sigma^2$  and  $\lambda$  as well as  $D_s$ -optimum designs for the subsets  $\beta$  and  $\lambda$ . Obtaining the design criterion is complicated by the nonlinear nature of the successive derivatives of  $y(\lambda)$  with  $\lambda$ , the expectations of which are needed to calculate the expected information matrix. Atkinson and Cook<sup>27</sup> use Taylor series expansions to obtain an approximate expected information matrix per observation,  $I_a(\theta)$ .

For the value of  $\lambda$  for which the transformation holds, let

$$\mu(x|\theta) = E(y^\lambda) = \lambda f^T(x)\beta + 1. \quad (22)$$

The constructed variable

$$v(\lambda) = \partial y(\lambda)/\partial \lambda$$

has expectation

$$E(v) = E\{v(\lambda)\} \approx \frac{\mu(x|\theta) \log\{\mu(x|\theta)\} - \mu(x|\theta) + 1}{\lambda^2}.$$

Then

$$I_a(\theta) = z_1 z_1^T + z_2 z_2^T, \quad (23)$$

where

$$\begin{aligned} z_1^T &= (f(x)/\sigma \quad 0 \quad -E(v)/\sigma), \\ z_2^T &= (0 \quad 1/(\sqrt{2}\sigma^2) \quad -\sqrt{2}\log(\mu)/\lambda), \end{aligned}$$

with a similar expression for  $\lambda = 0$ .

The structure of (23) is similar to that of (20) for heteroscedastic linear models. The first term in  $z_1$  is the information for  $\beta$  provided by the regression function. Atkinson and Cook<sup>27</sup> argue that there are two sources of information about the transformation. One comes from the constructed variable  $v(\lambda)$ , which is similar to the constructed variables used in the analysis of response transformations (Atkinson<sup>28</sup>; Cook and Weisberg<sup>29</sup>). The transformation information in the variance function comes from the logarithm of the regression function  $\log(\mu)$ , where  $\mu$  is defined in (22). The sum of squares of  $\log(\mu)$  over the design enters into the second factor on the

right of (23) through the last term of  $z_2$ , indicating a preference for designs with relatively large changes in the variance. This condition agrees with the common knowledge that response transformations are relatively well determined when the response ranges over several orders of magnitude.

The examples of designs given by Atkinson and Cook<sup>27</sup> include a second-order response surface in two variables. Many of the designs are slightly distorted versions of the  $3^2$  factorial. However, if the response has a sharp maximum away from these nine values, extra support points enter the design - one design in their Figure 2 has thirteen points of support.

## 8. Response Transformation in Mechanistic Models

### 8.1. Transforming Both Sides of a Mechanistic Model

If the model  $y = \eta(x, \theta)$  expresses a physical relationship, that relationship will be destroyed by power transformation of only the response as in (21). Instead, transformation of both sides of the model, described in Chapter 4 of Carroll and Ruppert<sup>30</sup>, is required in which (21) is replaced by

$$y^\lambda = \{\eta(x, \theta)\}^\lambda + \epsilon. \quad (24)$$

Atkinson<sup>31</sup> gives examples for nonlinear models from chemical kinetics with time the single explanatory variable; if a kinetic model is such that the concentrations of the components sum to one, the sum of the power-transformed components will not do so unless the model is also transformed.

The transformation can have an appreciable effect on the optimum design for estimation of  $\theta$ . A simple example is the nonlinear response model resulting from first-order decay



in which the concentration of chemical  $A$  at time  $t$  is given by the nonlinear function

$$[A] = \eta_A(t, \theta) = e^{-\theta t} \quad (\theta, t \geq 0), \quad (25)$$

if it is assumed that the initial concentration of  $A$  is 1. If the  $i$ th experiment consists of measuring the concentration of  $A$  at time  $t_i$ , the simple statistical model of the observations (1) is

$$y_i = \eta_A(t_i, \theta) + \epsilon_i,$$

where the errors  $\epsilon_i$  are independently distributed with zero mean and constant variance. The locally  $D$ -optimum design minimising the variance of

$\hat{\theta}$  (Box and Lucas<sup>32</sup>, Atkinson and Donev<sup>9</sup>) consists of taking all measurements where  $f(t, \theta)$  (5) is a maximum, that is at the time  $t^* = 1/\theta$ .

Now suppose that the model requires the log transformation to give constant variance. If  $[A]$  is measured, taking logarithms of both sides of (25), combined with additive errors, yields the statistical model

$$\log y_i = \log\{\eta_A(t_i, \theta)\} + \epsilon_i = -\theta t_i + \epsilon_i.$$

The log transformation thus results in a linear statistical model with response  $\log y$ , for which the parameter sensitivity is just the time  $t$ . The optimum design puts all observations at the maximum possible time; a seemingly slight assumption about the error distribution has had a huge effect on the optimum experimental design.

## 8.2. Parameter Sensitivities

The notation for the parameter sensitivities has to be extended to accommodate transformation. Let (5) be written

$$f_j(1; x, \theta) = \frac{\partial \eta(x, \theta)}{\partial \psi_j}. \quad (26)$$

The parameter sensitivities in the transformed model (24) are then

$$f_j(\lambda; x, \theta) = \frac{1}{\lambda} \frac{\partial \{\eta(x, \theta)\}^\lambda}{\partial \theta_j} = \{\eta(x, \theta)\}^{\lambda-1} \frac{\partial \eta(x, \theta)}{\partial \theta_j} = f_j(1; x, \theta) / \{\eta(x, \theta)\}^{1-\lambda}, \quad (27)$$

since multiplication by known  $\lambda$  does not affect the optimum design.

If  $\lambda < 1$ , the variance of the observations increases with the value of  $\eta(x, \theta)$ . Thus transformation of both sides for such values of  $\lambda$  will increase the relative value of the sensitivities for values of  $x$  where the response is small. We can expect that designs for  $\lambda < 1$  will include observations at lower concentrations than those when no transformation is needed.

## 9. Structured Parameters

### 9.1. Parameter Sensitivities

This section extends the parametrization of kinetic models, such as those in Atkinson<sup>31</sup>, to yield nonlinear response surface models. In general let

$$\theta_j = \phi_j \exp\{g(x, \psi_j)\}, \quad (28)$$

where  $g(x, \psi_j)$  is a function relating the kinetic parameter  $\theta_j$  to the vector of explanatory variables  $x$ . It is simplest to take  $g(\cdot)$  as a linear function. In the examples here it is also assumed that there is only one explanatory variable, other than time, and that all parameters are similarly influenced by  $x$ . Thus (28) becomes

$$\theta_j = \phi_j \exp(\psi x). \quad (29)$$

A medical example is where  $x$  is body weight. The rate of pharmacokinetic reactions depends on the concentration of drugs, so that, when the same dose is given to all patients, the reactions proceed more slowly for larger  $x$  and  $\psi$  would be negative. In chemical processes  $x$  might be a variable such as stirring rate, catalyst activity or feedstock purity, which would again affect all rates in the same way. The approximate rule from chemistry that a ten degree centigrade rise in temperature approximately doubles the rate of chemical reactions is also modelled by (29).

Design to estimate the parameters  $\theta$  in (24) is replaced by design to estimate the vector  $\phi$  and the scalar  $\psi$ . If the sensitivity

$$\frac{\partial \eta}{\partial \theta_j} = f_j^\theta$$

then, from the chain rule operating on (29),

$$f_j^\phi = \frac{\partial \eta}{\partial \phi_j} = \frac{\partial \eta}{\partial \theta_j} \frac{\partial \theta_j}{\partial \phi_j} = f_j^\theta e^{\psi x}. \quad (30)$$

The sensitivity  $f^\psi$  is found by differentiation of all  $p - 1$  elements of  $\theta$  to be

$$f^\psi = \sum_{j=1}^{p-1} \frac{\partial \eta}{\partial \theta_j} \frac{\partial \theta_j}{\partial \psi} = \sum_{j=1}^{p-1} f_j^\theta \phi_j x e^{\psi x} = x \sum_{j=1}^{p-1} f_j^\theta \phi_j. \quad (31)$$

Two simple examples of nonlinear models with this parametric structure are now considered. In both the response is a smooth, nonlinear function of two variables,  $x$  and the time  $t$ .

## 9.2. Exponential Decay

The model for exponential decay was introduced in §8.1, where designs were found when the response is  $[A]$ . In this section designs are found when the response is the concentration of  $[B]$ .

Since no material is lost during the reaction,  $[A] + [B] = 1$ . From (25) the concentration of  $B$  at time  $t$  is therefore

$$[B] = \eta_B(t, \theta) = 1 - e^{-\theta t} \quad (\theta, t \geq 0).$$

In the absence of transformation the sensitivity is minus that for  $[A]$  and the  $D$ -optimum design again puts all trials  $t = 1/\theta$ .

We start by incorporating the structured parameters of the previous section. With the single structured parameter of the form (29), the model for the concentration of  $B$  at time  $t$  becomes

$$[B] = \eta_B(t, \theta) = 1 - \exp(\phi t e^{-\psi x}). \quad (\phi, t \geq 0).$$

From (30) the sensitivity

$$f^\phi = t e^{-\theta t} e^{\psi x}.$$

Also, since  $\theta$  is scalar,

$$f^\psi = x \phi t e^{-\theta t} e^{\psi x}. \quad (32)$$

Usually, as in the earlier examples in this chapter, a numerical optimization would be required to find the  $D$ -optimum design over the design region of values of  $x$  and  $t$ . But here some analytical progress can be made.

First consider  $f^\phi$ . Although  $\theta$  is a function of  $\phi, \psi$  and  $x$ , it is not a function of  $t$ . So, for fixed  $x$ , the optimum value of  $t$  is found by differentiation to be, as before,

$$t^* = 1/\theta = e^{\psi x}/\phi. \quad (33)$$

This optimum value of  $t$  can now be substituted in (32) to give

$$f^\psi = x \phi t^* e^{-\theta t^*} e^{\psi x} = x e^{-1}.$$

Thus, provided the design region allows, the optimum design consists of trials at the upper and lower values of  $x$  at a time  $t^*$  given by (33). For the numerical example of this section the design region  $\mathcal{X}$  is  $0 \leq t \leq 20$  and  $-1 \leq x \leq 1$ . With  $\phi = 0.2$ ,  $t^* = 5$  when  $x = 0$ . The value of  $\psi = \log 2$  results in a doubling of the rate of reaction when  $x = 1$  and a halving when  $x = -1$ . In the absence of transformation the optimum design consists of equal numbers of trials at  $t = 2.5$ ,  $x = 1$  and  $t = 10$ ,  $x = -1$ .

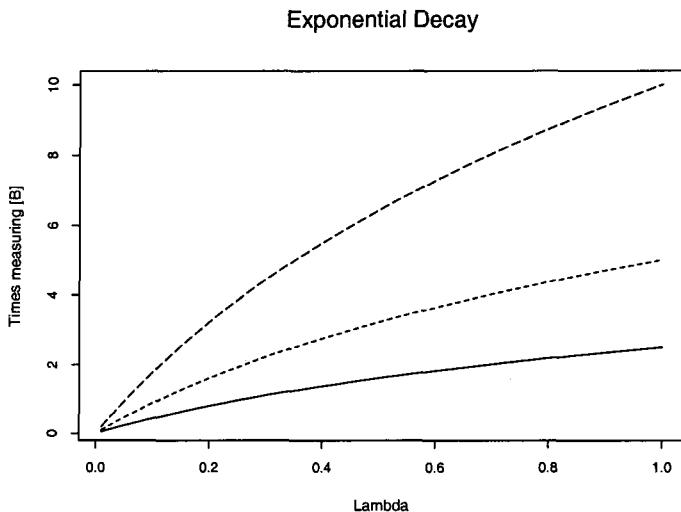


Fig. 5.  $D$ -optimum designs for exponential decay as  $\lambda$  varies. The central line is the optimum design for the simple model with parameter  $\theta = 0.2$ . The optimum design for the structured parameter puts  $n/2$  trials at twice this time and  $x = -1$  and the other  $n/2$  at half this time with  $x = 1$ .

Now suppose that the model needs to be transformed to give constant variance. From (27) the sensitivity for  $\theta$  is

$$f_j^\theta(\lambda; x, \theta) = t \exp(-\theta t) \{1 - \exp(-\theta t)\}^{\lambda-1},$$

which is maximized by the optimum time. The argument in the preceding paragraph shows that the optimum design puts half the trials at twice this optimum time and half at half the time. The design is shown in Figure 5. As  $\lambda \rightarrow 0$ , the optimum times do likewise.

### 9.3. Two Consecutive First-Order Reactions

As a second and more complicated example, locally  $D$ -optimum designs are found in this section for the model for two consecutive first-order reactions introduced by Box and Lucas<sup>32</sup>.

The two reactions can be written



When both reactions are first order, an explicit algebraic solution can be found for the concentrations of the reactants as a function of time. If the

initial concentration of  $A$  is one and that of  $B$  and  $C$  are zero,  $\eta_A(t, \theta)$  follows the exponential decay (25) with  $\theta = \theta_1$  and

$$\eta_B(t, \theta) = \frac{\theta_1}{\theta_1 - \theta_2} (e^{-\theta_2 t} - e^{-\theta_1 t}). \quad (34)$$

As the reaction proceeds,  $[B]$  rises from zero to a maximum and then decreases again to zero. With prior parameter values  $\theta_1 = 0.7$  and  $\theta_2 = 0.2$  and in the absence of transformation, the optimum design if just  $[B]$  is measured has two equally weighted support points at times of 1.230 and 6.858.

We now extend this model with the structured parameters (29). The effect on the concentration  $[B]$  is

$$\eta_B(t, x, \phi, \psi) = \frac{\phi_1}{\phi_1 - \phi_2} \{ \exp(-\phi_2 t e^{\psi x}) - \exp(-\phi_1 t e^{\psi x}) \}, \quad (35)$$

so that the effect of  $x$  can again be thought of as to multiply the time scale. With the values of  $x$  and  $\psi$  as in the previous example, the optimum points at the high level of  $x$  would be 0.615 and 3.429, whereas for the low level of  $x$  they would be 2.460 and 13.72. However, this model has three parameters, not four, so it is not clear that the optimum design will have exactly this structure.

Optimum designs for transformations are found numerically combining the sensitivities (30) and (31) for the structured parameters with those for transformations in (27). Table 9 gives optimum designs for several transformations when the maximum value of  $t$  is 20. The last two columns give the design for  $\lambda = 1$ . Although the design points are not far from the four suggested by elementary reasoning, the design weights are not equal. The weight on the two lower time points is 0.196, whereas the two higher times have weights of 0.304. For  $x = 0$  differentiation of (34) shows that the maximum response is at  $t = 2.506$ . The two lower time points in the design are below this value, the two upper ones above.

Table 9. Two consecutive first-order reactions with a structured parameter.  $D$ -optimum designs for four values of  $\lambda$ .

$x_i$	$\lambda = 0.25$		$\lambda = 0.5$		$\lambda = 0.75$		$\lambda = 1$	
	$t_i$	$p_i$	$t_i$	$p_i$	$t_i$	$p_i$	$t_i$	$p_i$
1	0.258	0.268	0.406	0.196	0.505	0.192	0.574	0.196
-1	1.018	0.084	1.621	0.177	2.019	0.192	2.299	0.196
1	10.75	0.327	5.642	0.314	4.154	0.308	3.475	0.304
-1	20.00	0.321	20.00	0.313	16.61	0.308	13.90	0.304

### Four Optimum Designs

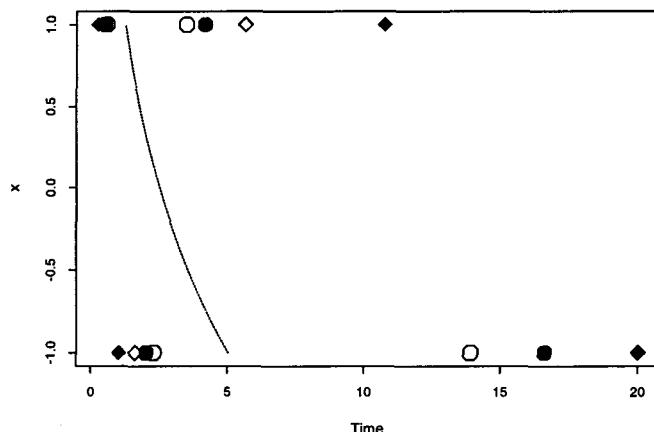


Fig. 6. Consecutive first-order reactions:  $D$ -optimum designs for four values of  $\lambda$ . Open circle,  $\lambda = 1$ , filled circle,  $\lambda = 0.75$ , open rhombus,  $\lambda = 0.5$  and filled rhombus,  $\lambda = 0.25$ ; dotted line, time of maximum yield.

The results of Table 9 show that, as  $\lambda$  decreases, the time points of the optimum designs move apart to regions of lower concentration. These results are illustrated for the four designs of the table in Figure 6. The dotted line in the figure shows the time of maximum yield  $t_{\max}$ , given from (35) by

$$t_{\max} = \frac{\log(\phi_1/\phi_2)}{\phi_1 - \phi_2} e^{-\psi_x}.$$

For the two lower values of  $\lambda$  and  $x = -1$ , the time of reading is 20, the maximum value in the design region. Hence the symbols are overplotted.

Although all designs in Figure 6 have four points of support, Table 9 shows that the weight on the lower time point with  $x = -1$  decreases as  $\lambda$  decreases. For lower values of  $\lambda$  a three-point design is optimum: for  $\lambda = 0.1$ , one of the support points of the design has the intermediate  $x$  value of  $-0.778$ .

#### 9.4. Extensions

This example shows that the structure of the response surface design for nonlinear models may change appreciably with the value of the transformation parameter. More detailed results on the variation of design with criterion are available for nonlinear models with the single variable time.

The variation seems to be greatest when more than one response is measured. However, although the designs may seem very different, each may be efficient for an appreciable range of values of  $\lambda$ . Bayesian designs, optimum for a specified prior distribution of  $\lambda$  values can then be found (Atkinson<sup>33</sup>). For some choices of prior the resulting designs are efficient for almost all values of  $\lambda$  in the range  $0 < \lambda \leq 1$ .

Transformations were introduced in this chapter to provide statistical models in which the observations had constant variance. An alternative (Bogacka and Wright<sup>34</sup>) is to use weighted least squares with weights proportional to  $E(y)^{-2(1-\lambda)}$ . The resulting parameter sensitivities, and so designs, are identical to those of §8.

## 10. Discussion

The emphasis in this paper is on locally  $D$ -optimum designs. Extensions would be to designs for multivariate responses, to other optimality criteria, such as the average variance of prediction, and to design when prior information is available about the parameters. Robinson and Khuri<sup>35</sup> use variance-dispersion graphs to evaluate designs for generalized linear models. Conventional response surface methodology contains a number of other strands. For example, Box and Draper<sup>1</sup> stress the importance of allowing for the biases that arise from an inadequate model and Box and Draper<sup>36</sup> list 14 desiderata of a good experimental design. The application of some of these ideas to designs for generalized linear models might, for example, help in understanding and improving designs for blocking that would be produced by optimum design algorithms.

## References

1. G. E. P. Box and N. R. Draper. The choice of a second order rotatable design. *Biometrika*, 50:335–352, 1963.
2. R. H. Myers, D. C. Montgomery, G. C. Vining, C. M. Borror, and S. M. Kowalski. Response surface methodology: a retrospective and literature survey. *Journal of Quality Technology*, 36:53–77, 2004.
3. G. E. P. Box and D. R. Cox. An analysis of transformations (with discussion). *Journal of the Royal Statistical Society, Series B*, 26:211–246, 1964.
4. P. McCullagh and J. A. Nelder. *Generalized Linear Models (2nd edition)*. Chapman and Hall, London, 1989.
5. R. H. Myers, D. C. Montgomery, and G. C. Vining. *Generalized Linear Models: with Applications in Engineering and the Sciences*. Wiley, New York, 2001.

6. A. C. Atkinson and M. Riani. *Robust Diagnostic Regression Analysis*. Springer-Verlag, New York, 2000.
7. F. Pukelsheim. *Optimal Design of Experiments*. Wiley, New York, 1993.
8. V. V. Fedorov and P. Hackl. *Model-Oriented Design of Experiments*. Lecture Notes in Statistics 125. Springer, New York, 1997.
9. A. C. Atkinson and A. N. Donev. *Optimum Experimental Designs*. Oxford University Press, Oxford, 1992.
10. S. L. Lewis, D. C. Montgomery, and R. H. Myers. Confidence interval coverage for designed experiments analyzed with glms. *Journal of Quality Technology*, 33:279–292, 2001.
11. J. Schlesselman. Power families: a note on the Box and Cox transformation. *Journal of the Royal Statistical Society, Series B*, 33:307–11, 1971.
12. J. Burridge and P. Sebastiani. D-optimal designs for generalised linear models with variance proportional to the square of the mean. *Biometrika*, 81:295–304, 1994.
13. W. Nelson. The analysis of performance-degradation data. *IEEE Transactions on Reliability*, R-30:149–155, 1981.
14. D. R. Cox. A note on design when response has an exponential family distribution. *Biometrika*, 75:161–164, 1988.
15. S. D. Silvey. *Optimum Design*. Chapman and Hall, London, 1980.
16. K. Chaloner and K. Larntz. Optimal Bayesian design applied to logistic regression experiments. *Journal of Statistical Planning and Inference*, 21:191–208, 1989.
17. I. Ford, B. Torsney, and C. F. J. Wu. The use of a canonical form in the construction of locally optimal designs for non-linear problems. *Journal of the Royal Statistical Society, Series B*, 54:569–583, 1992.
18. R. S. Sitter and B. Torsney. D-optimal designs for generalized linear models. In C. P. Kitsos and W. G. Müller, editors, *MODA 4 – Advances in Model-Oriented Data Analysis*, pages 87–102. Physica-Verlag, Heidelberg, 1995.
19. A. Torsney and N. Gunduz. On optimal designs for high dimensional binary regression models. In A. C. Atkinson, B. Bogacka, and A. Zhigljavsky, editors, *Optimal Design 2000*, pages 275–285. Kluwer, Dordrecht, 2001.
20. R. H. Farrell, J. Kiefer, and A. Walbran. Optimum multivariate designs. In *Proc. 5th Berkeley Symposium*, volume 1, pages 113–138, Berkeley, CA, 1968. University of California Press.
21. A. C. Atkinson and R. D Cook. D-optimum designs for heteroscedastic linear models. *Journal of the American Statistical Association*, 90:204–212, 1995.
22. J. A. Nelder and Y. Lee. Generalized linear models for the analysis of Taguchi-type experiments. *Applied Stochastic Models and Data Analysis*, 7:107–120, 1991.
23. G. E. P. Box. Quality improvement - the new industrial revolution. *International Statistical Review*, 61:1–19, 1993.
24. V. V. Fedorov. *Theory of Optimal Experiments*. Academic Press, New York, 1972.
25. D. Downing, V. V. Fedorov, and S. Leonov. Extracting information from the variance function: optimal design. In A. C. Atkinson, P. Hackl, and W. G.

- Müller, editors, *MODA 6 – Advances in Model-Oriented Design and Analysis*, pages 45–52. Physica-Verlag, Heidelberg, 2001.
- 26. E. Rodrigues Pinto and A. Ponce de Leon. Bayesian D-optimal designs for generalized linear models with a varying dispersion parameter. In A. Di Bucchianico, H. Läuter, and H. P. Wynn, editors, *MODA 7 – Advances in Model-Oriented Design and Analysis*, pages 143–151. Physica-Verlag, Heidelberg, 2004.
  - 27. A. C. Atkinson and R. D Cook. Designing for a response transformation parameter. *Journal of the Royal Statistical Society B*, 59:111–124, 1997.
  - 28. A. C. Atkinson. *Plots, Transformations, and Regression*. Oxford University Press, Oxford, 1985.
  - 29. R. D. Cook and S. Weisberg. *Residuals and Influence in Regression*. Chapman and Hall, London, 1982.
  - 30. R. J. Carroll and D. Ruppert. *Transformation and Weighting in Regression*. Chapman and Hall, London, 1988.
  - 31. A. C. Atkinson. Horwitz’s rule, transforming both sides and the design of experiments for mechanistic models. *Applied Statistics*, 52:261–278, 2003.
  - 32. G. E. P. Box and H. L. Lucas. Design of experiments in nonlinear situations. *Biometrika*, 46:77–90, 1959.
  - 33. A. C. Atkinson. Some Bayesian optimum designs for response transformation in nonlinear models with nonconstant variance. In A. Di Bucchianico, H. Läuter, and H. P. Wynn, editors, *MODA 7 – Advances in Model-Oriented Design and Analysis*, pages 13–22. Physica-Verlag, Heidelberg, 2004.
  - 34. B. Bogacka and F. Wright. A non-linear design problem in a chemical kinetic model with non-constant error variance. *Journal of Statistical Planning and Inference*, 128:633–648, 2005.
  - 35. K. S. Robinson and A. I. Khuri. Quantile dispersion graphs for evaluating and comparing designs for logistic regression models. *Computational Statistics and Data Analysis*, 43:47–62, 2003.
  - 36. G. E. P. Box and N. R. Draper. Robust designs. *Biometrika*, 62:347–352, 1975.

## CHAPTER 9

### GLM DESIGNS: THE DEPENDENCE ON UNKNOWN PARAMETERS DILEMMA

André I. Khuri<sup>a</sup> and Siuli Mukhopadhyay<sup>b</sup>

*Department of Statistics, University of Florida  
P.O. Box 118545, Gainesville, FL 32611-8545*

<sup>a</sup>*E-mail: ufakhuri@stat.ufl.edu*

<sup>b</sup>*E-mail: smukhopa@stat.ufl.edu*

The choice of design for a generalized linear model depends on the unknown parameters of the fitted model. This poses a difficult problem since the purpose of a design is to provide efficient estimates of the model's parameters. One approach to solving this problem uses the so-called quantile dispersion graphs (QDGs) of the mean-squared error of prediction (MSEP) associated with a given model. These are plots of the maxima and minima, over a parameter space, of the quantiles of the MSEP, which are obtained on concentric surfaces inside a region of interest. The plots provide a comprehensive assessment of the quality of prediction afforded by a given design. They also portray the dependence of the design on the model's parameters. The application of the QDGs is demonstrated using a model with a logarithmic link function and a Poisson-distributed response variable. Several variants of these conditions are considered, including a square root link in conjunction with the Poisson distribution, and several other combinations. The results indicate that the choice of the link function and/or the nature of the response distribution can have an effect on the shape of the QDGs for a given design.

#### 1. Introduction

The quality of prediction and the precision of estimating the parameters of a regression model depend to a large extent on the design used in a given experimental situation. By a design, we mean the specification of the settings of the various factors (control variables) that influence a response of interest. Such a design is called a *response surface design*. For brevity, we shall just use the word "design".

The choice of design is based on a postulated model deemed adequate to represent the mean response within a certain region of interest. In particular, if the model is linear in its parameters, then it is called a *linear model*. If, however, at least one parameter does not appear linearly in the model, then the model is said to be *nonlinear*. Linear models are typically used in a response surface investigation where little is known about the true relationship that may exist between the mean response and the control variables that affect the response. Such linear models are called *empirical*. For example, first-degree and second-degree linear models have traditionally been used in response surface methodology in order to arrive at optimum operating conditions on the control variables that maximize (or minimize) the predicted response [see Box and Draper<sup>1</sup>, Khuri and Cornell<sup>2</sup>, and Myers and Montgomery<sup>3</sup>]. Nonlinear models, on the other hand, tend to be used when the functional relationship between the mean response and the control variables is known, except for the actual values of the model's parameters. Knowledge of this relationship is possible if the experimenter is familiar with the particular mechanism that is inherent in the experimental process.

Whereas the statistical literature pertaining to designs for linear models is quite extensive, the same cannot be said about designs for nonlinear models. This is mainly due to the complexity associated with the construction of designs for the latter models. A major factor contributing to this complexity is the dependence of a nonlinear model design on the model's parameters, which are unknown. By contrast, designs for linear models are not afflicted with this dependence problem. Thus the construction of a design for a nonlinear model requires some knowledge of the model's parameters. This is an undesirable feature of nonlinear models which was most succinctly depicted in the following quotation by Cochran<sup>4</sup>: "You tell me the value of  $\beta$  and I promise to design the best experiment for estimating  $\beta$ ." The experimenter is therefore presented with a dilemma since the purpose of a design is to estimate the unknown parameters of the model using the data generated by the application of the design. To actually construct the best design, however, one needs to know the true values of the parameters.

Common approaches to solving this design dependence problem include the following:

- (i) the specification of initial values, or best "guesses", of the parameters involved, and the subsequent determination of the so-called

*locally-optimal design* using an appropriate design criterion such as *D-optimality* or *G-optimality*,

- (ii) the sequential approach which allows the user to obtain updated estimates of the parameters in successive stages, starting with the initial values used in the first stage,
- (iii) the Bayesian approach, where a prior distribution is assumed on the parameters, which is then incorporated into an appropriate design criterion by integrating it over the prior distribution.

In this chapter, we shall address the design dependence problem for a particular type of nonlinear models, namely, the so-called *generalized linear models*. The main focus of attention will be on the use of an alternative approach for evaluating and comparing designs for these models. This approach, which was recently introduced by Robinson and Khuri<sup>5</sup>, utilizes the so-called *quantile dispersion graphs* of the mean-squared error of prediction. The graphs provide an assessment of the overall prediction capability associated with a given design. They also give a clear depiction of the dependence of the design on the model's parameters. Similar plots were used by Khuri and Lee<sup>6</sup> for comparing designs for nonlinear models.

## 2. Generalized Linear Models

Generalized linear models (GLMs) were first introduced by Nelder and Wedderburn<sup>7</sup> as an extension of the class of linear models. They can be used to fit discrete as well as continuous data having a variety of parent distributions. Thus the traditional assumptions of normality and homogeneous variances, usually made in an analysis of variance (or regression) situation, are no longer needed. A classic book on GLMs is the one by McCullagh and Nelder<sup>8</sup>. In addition, the more recent books by Dobson<sup>9</sup>, Lindsey<sup>10</sup>, McCulloch and Searle<sup>11</sup>, and Myers, Montgomery, and Vining<sup>12</sup> provide added insight into the application and usefulness of GLMs.

The use of GLMs requires the specification of the following three components:

- (i) the data set under consideration consists of independent random variables,  $y_1, y_2, \dots, y_n$ , such that  $y_u$  has a distribution in the exponential family with a probability mass function (or a density function for a continuous data distribution) given by

$$\delta(y_u, \theta_u, \phi) = \exp[\phi\{y_u\theta_u - b(\theta_u) + c(y_u)\} + d(y_u, \phi)], \quad u = 1, 2, \dots, n, \quad (1)$$

where  $b(\cdot)$ ,  $c(\cdot)$  and  $d(\cdot, \cdot)$  are known functions,  $\theta_u$  is a canonical parameter, and  $\phi$  is a dispersion parameter, possibly unknown. The mean of  $y_u$  is  $\mu_u = \frac{db(\theta_u)}{d\theta_u}$ , and its variance is

$$\sigma_u^2 = \frac{1}{\phi} \frac{d^2 b(\theta_u)}{d\theta_u^2}, \quad u = 1, 2, \dots, n. \quad (2)$$

See McCullagh and Nelder<sup>8</sup>, pp. 28-29.

- (ii) a linear regression function,  $\eta$ , called a *linear predictor*, in  $k$  control variables,  $x_1, x_2, \dots, x_k$ , of the form

$$\eta(\mathbf{x}) = \mathbf{f}'(\mathbf{x})\boldsymbol{\beta}, \quad (3)$$

where  $\mathbf{f}(\mathbf{x})$  is a known  $p$ -component vector-valued function of  $\mathbf{x} = (x_1, x_2, \dots, x_k)'$ ,  $\boldsymbol{\beta}$  is a vector of  $p$  unknown parameters, and  $\mathbf{f}'(\mathbf{x})$  denotes the transpose of  $\mathbf{f}(\mathbf{x})$ ,

- (iii) a *link function*  $g(\mu)$  which relates  $\eta(\mathbf{x})$  to the mean response  $\mu(\mathbf{x})$  at a point  $\mathbf{x}$  in a region of interest,  $R$ ,

$$\eta(\mathbf{x}) = g[\mu(\mathbf{x})], \quad (4)$$

where  $g(\cdot)$  is a monotone differentiable function whose inverse is denoted by  $h(\cdot)$ . Thus

$$\mu(\mathbf{x}) = h[\eta(\mathbf{x})]. \quad (5)$$

Formula (5) indicates that the mean response is, in general, represented by a nonlinear model.

For example, the response  $y$  may be binary with a probability mass function of the form

$$\delta(y_u, \pi_u) = \exp[y_u \log\left(\frac{\pi_u}{1 - \pi_u}\right) + \log(1 - \pi_u)], \quad u = 1, 2, \dots, n, \quad (6)$$

$y_u = 1$  or  $0$ , and  $\pi_u$  is the probability that  $y_u = 1$  on the  $u$ -th trial ( $u = 1, 2, \dots, n$ ). Note that this distribution belongs to the exponential family with  $\phi = 1$ ,  $\theta_u = \log(\frac{\pi_u}{1 - \pi_u})$ ,  $b(\theta_u) = \log[1 + \exp(\theta_u)]$ ,  $c(y_u) = 0$ , and  $d(y_u, \phi) = 0$ ,  $u = 1, 2, \dots, n$ . In this case, the mean and variance of  $y_u$  are  $\mu_u = \pi_u$  and  $\sigma_u^2 = \pi_u(1 - \pi_u)$ , respectively,  $u = 1, 2, \dots, n$ .

The corresponding link function is the *logit link* (see McCullagh and Nelder<sup>8</sup>, p. 31), namely,

$$\eta(\mathbf{x}) = \log\left[\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})}\right], \quad (7)$$

and the linear predictor is of the form

$$\log \left[ \frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})} \right] = \mathbf{f}'(\mathbf{x})\boldsymbol{\beta}, \quad (8)$$

where  $\pi(\mathbf{x})$  denotes the probability of success (that is, achieving the value 1) at  $\mathbf{x}$ . The mean response,  $\pi(\mathbf{x})$ , at  $\mathbf{x}$  is then given by

$$\pi(\mathbf{x}) = \frac{\exp[\mathbf{f}'(\mathbf{x})\boldsymbol{\beta}]}{1 + \exp[\mathbf{f}'(\mathbf{x})\boldsymbol{\beta}]} \cdot \quad (9)$$

Model (9) is called the *logistic regression model*.

We note that the link function in (7) was obtained by equating  $\eta(\mathbf{x})$  to the canonical parameter  $\theta$ . This particular link function is called the *canonical link*, and its use leads to desirable statistical properties of the model, particularly in small samples (see McCullagh and Nelder<sup>8</sup>, p. 32).

## 2.1. Estimation of the Mean Response

Estimation of  $\boldsymbol{\beta}$  in (3) is based on the method of maximum likelihood, which is carried out using an iterative weighted least-squares procedure. An estimate of  $\eta(\mathbf{x})$  in (3) is then given by

$$\hat{\eta}(\mathbf{x}) = \mathbf{f}'(\mathbf{x})\hat{\boldsymbol{\beta}}, \quad (10)$$

and the corresponding estimate of the mean response  $\mu(\mathbf{x})$  in (5) is

$$\hat{\mu}(\mathbf{x}) = h[\mathbf{f}'(\mathbf{x})\hat{\boldsymbol{\beta}}], \quad (11)$$

where  $\hat{\boldsymbol{\beta}}$  is the maximum likelihood estimate of  $\boldsymbol{\beta}$ . The variance-covariance matrix of  $\hat{\boldsymbol{\beta}}$  is approximated by (see Robinson and Khuri<sup>5</sup>)

$$\text{Var}(\hat{\boldsymbol{\beta}}) \approx \frac{1}{\phi} (\mathbf{X}' \mathbf{W} \mathbf{X})^{-1}, \quad (12)$$

where  $\mathbf{X}$  is an  $n \times p$  matrix whose  $u$ th row is of the form  $\mathbf{f}'(\mathbf{x}_u)$ ,  $\mathbf{x}_u$  is the value of  $\mathbf{x}$  at the  $u$ th experimental run ( $u = 1, 2, \dots, n$ ), and  $\mathbf{W}$  is a diagonal matrix of the form

$$\mathbf{W} = \text{Diag}(w_1, w_2, \dots, w_n), \quad (13)$$

where  $w_u$  is given by

$$w_u = \frac{1}{v_u} \left( \frac{d\mu_u}{d\eta_u} \right)^2, \quad u = 1, 2, \dots, n. \quad (14)$$

In (14),  $\frac{d\mu_u}{d\eta_u}$  denotes the derivative of  $\mu(\mathbf{x})$  with respect to  $\eta(\mathbf{x})$  evaluated at  $\mathbf{x} = \mathbf{x}_u$ , and  $v_u = \frac{d\mu_u}{d\theta_u}$ . In addition, the variances of  $\hat{\eta}(\mathbf{x})$  and  $\hat{\mu}(\mathbf{x})$ , respectively, are approximated by (see Robinson and Khuri<sup>5</sup>)

$$\text{Var}[\hat{\eta}(\mathbf{x})] \approx \frac{1}{\phi} \mathbf{f}'(\mathbf{x}) [\mathbf{X}' \mathbf{W} \mathbf{X}]^{-1} \mathbf{f}(\mathbf{x}). \quad (15)$$

$$\text{Var}[\hat{\mu}(\mathbf{x})] \approx \frac{1}{\phi} \left[ \frac{d\mu(\mathbf{x})}{d\eta(\mathbf{x})} \right]^2 \mathbf{f}'(\mathbf{x}) [\mathbf{X}' \mathbf{W} \mathbf{X}]^{-1} \mathbf{f}(\mathbf{x}), \quad (16)$$

where  $\frac{d\mu(\mathbf{x})}{d\eta(\mathbf{x})}$  is the derivative of  $\mu(\mathbf{x})$  with respect to  $\eta(\mathbf{x})$  evaluated at a point  $\mathbf{x}$  in a region of interest,  $R$ . The variance of  $\hat{\mu}(\mathbf{x})$  is referred to as the *prediction variance*.

## 2.2. The Prediction Bias

The maximum likelihood estimate of  $\beta$  is, in general, a biased estimate of  $\beta$ . Cordeiro and McCullagh<sup>13</sup> developed an expression for the bias of  $\hat{\beta}$  for GLMs using the tensor methodology. Cadigan<sup>14</sup> presented a method for the computation of the bias which did not require using the tensor methodology. His expression for the bias is approximated by

$$\text{Bias}(\hat{\beta}) \approx -\frac{1}{2\phi} (\mathbf{X}' \mathbf{W} \mathbf{X})^{-1} \mathbf{X}' \mathbf{Z}_d \mathbf{F} \mathbf{1}_n, \quad (17)$$

where  $\mathbf{X}$  and  $\mathbf{W}$  are the same as in Section 2.1,  $\mathbf{Z}_d = \text{Diag}(z_{11}, z_{22}, \dots, z_{nn})$ , where  $z_{uu}$  is the  $u$ th diagonal element of  $\mathbf{Z} = \mathbf{X}(\mathbf{X}' \mathbf{W} \mathbf{X})^{-1} \mathbf{X}'$ ,  $\mathbf{F} = \text{Diag}(f_{11}, f_{22}, \dots, f_{nn})$ , where for  $u = 1, 2, \dots, n$ ,  $f_{uu}$  is given by

$$f_{uu} = \frac{1}{v_u} \left[ \frac{d^2 \mu_u}{d\eta_u^2} \right] \left[ \frac{d\mu_u}{d\eta_u} \right], u = 1, 2, \dots, n, \quad (18)$$

$v_u = \frac{d\mu_u}{d\theta_u}$ , and  $\mathbf{1}_n$  is a column vector of  $n$  ones. In (18),  $\frac{d^2 \mu_u}{d\eta_u^2}$  denotes the second derivative of  $\mu(\mathbf{x})$  with respect to  $\eta(\mathbf{x})$  evaluated at  $\mathbf{x}_u$  ( $u = 1, 2, \dots, n$ ). Cordeiro and McCullagh<sup>13</sup> (formula 6.3) showed that the bias of  $\hat{\beta}$  can also be derived approximately through a simple weighted linear regression computation of the form

$$\text{Bias}(\hat{\beta}) \approx (\mathbf{X}' \mathbf{W} \mathbf{X})^{-1} \mathbf{X}' \mathbf{W} \boldsymbol{\xi}, \quad (19)$$

where  $\boldsymbol{\xi} = -\frac{1}{2\phi} \mathbf{W}^{-1} \mathbf{Z}_d \mathbf{F} \mathbf{1}_n$  and  $\mathbf{W}$  plays the role of the weight matrix.

Using (19), the bias of  $\hat{\eta}(\mathbf{x})$  in (10) is approximated by

$$\text{Bias}[\hat{\eta}(\mathbf{x})] \approx \mathbf{f}'(\mathbf{x}) (\mathbf{X}' \mathbf{W} \mathbf{X})^{-1} \mathbf{X}' \mathbf{W} \boldsymbol{\xi}. \quad (20)$$

Furthermore the *prediction bias*, which is the bias of  $\hat{\mu}(\mathbf{x})$  in (11), is approximated by

$$\text{Bias}[\hat{\mu}(\mathbf{x})] \approx \text{Bias}[\hat{\eta}(\mathbf{x})] \frac{d\mu(\mathbf{x})}{d\eta(\mathbf{x})} + \frac{1}{2} \text{Var}[\hat{\eta}(\mathbf{x})] \frac{d^2\mu(\mathbf{x})}{d\eta^2(\mathbf{x})}, \quad (21)$$

where  $\frac{d^2\mu(\mathbf{x})}{d\eta^2(\mathbf{x})}$  is the second derivative of  $\mu(\mathbf{x})$  with respect to  $\eta(\mathbf{x})$  evaluated at a point  $\mathbf{x}$  in a region  $R$ , and  $\text{Var}[\hat{\eta}(\mathbf{x})]$  is approximated by formula (15) (see Robinson and Khuri<sup>5</sup>).

### 2.3. The Mean-Squared Error of Prediction

By definition, the *mean-squared error of prediction* at a point  $\mathbf{x}$  in a region  $R$ , denoted by  $\text{MSEP}(\mathbf{x})$ , is the mean-squared error of  $\hat{\mu}(\mathbf{x})$ , that is,

$$\text{MSE}[\hat{\mu}(\mathbf{x})] = E[\hat{\mu}(\mathbf{x}) - \mu(\mathbf{x})]^2. \quad (22)$$

The right-hand side of (22) can be partitioned into two components, namely,

$$\text{MSE}[\hat{\mu}(\mathbf{x})] = \text{Var}[\hat{\mu}(\mathbf{x})] + \{\text{Bias}[\hat{\mu}(\mathbf{x})]\}^2, \quad (23)$$

since  $\text{Bias}[\hat{\mu}(\mathbf{x})] = E[\hat{\mu}(\mathbf{x})] - \mu(\mathbf{x})$ . By combining formulas (16) and (21) we obtain

$$\begin{aligned} \text{MSE}[\hat{\mu}(\mathbf{x})] &\approx \frac{1}{\phi} \left[ \frac{d\mu(\mathbf{x})}{d\eta(\mathbf{x})} \right]^2 \mathbf{f}'(\mathbf{x}) [\mathbf{X}' \mathbf{W} \mathbf{X}]^{-1} \mathbf{f}(\mathbf{x}) \\ &+ \left\{ \text{Bias}[\hat{\eta}(\mathbf{x})] \frac{d\mu(\mathbf{x})}{d\eta(\mathbf{x})} + \frac{1}{2} \text{Var}[\hat{\eta}(\mathbf{x})] \frac{d^2\mu(\mathbf{x})}{d\eta^2(\mathbf{x})} \right\}^2, \end{aligned} \quad (24)$$

where  $\text{Var}[\hat{\eta}(\mathbf{x})]$  and  $\text{Bias}[\hat{\eta}(\mathbf{x})]$  are approximated by formulas (15) and (20), respectively.

## 3. Choice of Design

An important criterion for the choice of design for GLMs is the minimization of  $\text{MSEP}(\mathbf{x})$ , given in formula (24), over the region  $R$ . One major problem in doing so is the dependence of  $\text{MSEP}(\mathbf{x})$  on the unknown parameter vector  $\beta$  in model (3), since  $\eta(\mathbf{x})$ , and hence  $\mu(\mathbf{x})$ , depends on  $\beta$ . This is analogous to the design dependence problem mentioned earlier in Section 1 with regard to nonlinear models. The three approaches mentioned in that section can be applied to generate designs for GLMs.

In this section, we discuss a fourth approach proposed by Robinson and Khuri<sup>5</sup>, which is based on using the so-called *quantile dispersion graphs* (QDGs) of  $\text{MSEP}(\mathbf{x})$ . This approach is suited for comparing several candidate designs for fitting a generalized linear model, given a particular link function and a specified parent distribution for the data.

### 3.1. Quantile Dispersion Graphs

In this approach, quantiles of the distribution of  $\text{MSEP}(\mathbf{x})$  in (24) are obtained on several concentric surfaces inside the region  $R$  as follows: For a given design  $D$ , let  $Q_D(p, \beta, \nu)$  denote the  $p$ th quantile of the distribution of  $\text{MSEP}$  on  $R_\nu$ , which represents the surface of a region obtained by reducing  $R$  using a shrinkage factor  $\nu$ , and  $\beta$  is the vector of unknown parameters in model (3). By varying the value of  $\nu$ , one can cover the entire region  $R$ .

Now, in order to assess the dependence of the design  $D$  on  $\beta$ , a certain parameter space,  $C$ , to which  $\beta$  is assumed to belong, is specified. For a given  $\nu$ , the dependence of  $Q_D(p, \beta, \nu)$  on  $\beta$  is investigated by computing it for several values of  $\beta$  that form a grid  $C$  inside  $C$ . Subsequently, the minimum and maximum values of  $Q_D(p, \beta, \nu)$  over the values of  $\beta$  in  $C$  are obtained. This results in the following extrema of  $Q_D(p, \beta, \nu)$  for each  $\nu$  and a given  $p$ :

$$Q_D^{\min}(p, \nu) = \min_{\beta \in C} \{Q_D(p, \beta, \nu)\} \quad (25)$$

$$Q_D^{\max}(p, \nu) = \max_{\beta \in C} \{Q_D(p, \beta, \nu)\}. \quad (26)$$

Plotting these values against  $p$  results in the quantile dispersion graphs of the mean-squared error of prediction over the region  $R_\nu$ . By repeating the same process for several selected values of  $\nu$ , we obtain plots that portray the prediction capability associated with the design  $D$  throughout the region  $R$ . Such plots can be constructed for each of several candidate designs for the model. It should be noted that for a given  $\nu$ , a desirable feature of a design is to have close and small values of  $Q_D^{\min}$  and  $Q_D^{\max}$  over the range of  $p$  ( $0 \leq p \leq 1$ ). The smallness of  $Q_D^{\max}$  indicates small  $\text{MSEP}$  values on  $R_\nu$ , and the closeness of  $Q_D^{\min}$  and  $Q_D^{\max}$  indicates robustness to changes in the values of  $\beta$  that is induced by the design  $D$ .

There are several advantages to this approach, namely,

- (i) the performance of a design can be evaluated throughout the region  $R$ . Other design criteria base the evaluation of a design on a single measure, such as D-efficiency, but do not consider the quality of prediction inside  $R$ .
- (ii) unlike other variance-based design criteria, such as D-optimality or G-optimality, estimation bias is taken into account in the evaluation of a given design.
- (iii) the dependence of a design on the model's unknown parameters is clearly depicted by the QDGs throughout the region  $R$ .

- (iv) plotting the quantiles of the MSEP against  $p$  permits the consideration of models with several control variables, instead of just one, as is usually the case with other design criteria, such as D-optimality.

#### 4. Examples

The type of data we consider here is of counts. Such data appear in many applications, for example, the number of certain events within a fixed period of time (insurance claims, accidents, deaths, births, etc.), or the frequencies in each cell of a contingency table. There are several procedures for analyzing count data. In many situations, a Poisson distribution is assumed. Sometimes, the count data show more variability than this distribution allows. We can then introduce an additional overdispersion parameter to account for the extra variation. Quite often, however, the use of an overdispersion parameter is not of much help, and some other distributions, like the negative binomial or the gamma, may be more appropriate.

##### 4.1. Poisson Response with a Logarithmic Link

Here we consider a situation where the response follows a Poisson distribution. Hence, the probability mass function of  $y_u$  is given by

$$\delta(y_u, \lambda_u) = \exp[\phi\{(y_u \log \lambda_u - \lambda_u) - \log(y_u!)\}], u = 1, \dots, n. \quad (27)$$

By comparing (27) with (1) we note that the distribution of  $y_u$  belongs to the exponential family with  $\theta_u = \log \lambda_u$ ,  $b(\theta_u) = \lambda_u$ ,  $c(y_u) = -\log(y_u!)$ , and  $d(y_u, \phi) = 0$ . The mean and variance of  $y_u$  are  $E(y_u) = \lambda_u$  and  $\text{Var}(y_u) = \frac{\lambda_u}{\phi}$ ,  $u = 1, 2, \dots, n$ . The dispersion parameter  $\phi$  is unknown; it can be estimated or taken to be equal to 1. Nelder and Wedderburn<sup>7</sup> suggested taking  $\hat{\phi} = (n-p)/X^2$ , where  $X^2$  is a Pearson-type statistic and  $p$  is the number of elements in the parameter vector  $\beta$  (see Agresti<sup>15</sup>, p. 150).

Let  $\mathbf{x}$  be a vector of design settings in the region of interest,  $R$ . At the  $u$ th experimental run,  $\mathbf{x} = \mathbf{x}_u$ , the corresponding response value is  $y_u$  ( $u = 1, 2, \dots, n$ ). The mean response at  $\mathbf{x}$  is denoted by  $\lambda(\mathbf{x})$ . The canonical link function is the logarithmic (log) link (see McCullagh and Nelder<sup>8</sup>)

$$\eta(\mathbf{x}) = \log \lambda(\mathbf{x}) = \mathbf{f}'(\mathbf{x})\beta. \quad (28)$$

Hence,

$$\hat{\mu}(\mathbf{x}) = \hat{\lambda}(\mathbf{x}) = \exp[\mathbf{f}'(\mathbf{x})\hat{\beta}], \quad (29)$$

where  $\hat{\beta}$  is the maximum likelihood estimator (MLE) of  $\beta$ .

Using (16) and (21) it can be shown (see Appendix A) that the mean-squared error of prediction (MSEP) at  $\mathbf{x}$  is approximated by

$$\begin{aligned} \text{MSE}[\hat{\lambda}(\mathbf{x})] &\approx [\lambda(\mathbf{x})]^2 \mathbf{f}'(\mathbf{x}) \frac{(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}}{\phi} \mathbf{f}(\mathbf{x}) \\ &\quad + \{\mathbf{f}'(\mathbf{x}) \text{Bias}(\hat{\beta}) \lambda(\mathbf{x}) + \frac{1}{2} \mathbf{f}'(\mathbf{x}) \frac{(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}}{\phi} \mathbf{f}(\mathbf{x}) \lambda(\mathbf{x})\}^2, \end{aligned} \quad (30)$$

where

$$\text{Bias}(\hat{\beta}) \approx (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1} \mathbf{X}'\mathbf{W}\xi,$$

$\mathbf{W} = \text{Diag}(w_1, w_2, \dots, w_n)$  with  $w_u = \lambda_u$ , and  $\xi$  is an  $n \times 1$  vector whose  $u$ th element is  $-\frac{z_{uu}}{2\phi}$ , where  $z_{uu}$  is the  $u$ th diagonal element of  $\mathbf{Z} = \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'$ ,  $u = 1, 2, \dots, n$ . Here,  $\hat{\lambda}_u$  is given by

$$\hat{\lambda}_u = \exp[\mathbf{f}'(\mathbf{x}_u)\hat{\beta}], \quad u = 1, 2, \dots, n.$$

It should be noted that the MSEP at a point  $\mathbf{x}$  in  $R$  depends on the unknown parameter vector,  $\beta$ . Let us therefore denote the value of  $\text{MSE}[\hat{\lambda}(\mathbf{x})]$  by  $\tau_D(\mathbf{x}, \beta)$ .

#### 4.2. Poisson Response with a Square Root Link

Here, the relationship between the mean response and the linear predictor is

$$\eta(\mathbf{x}) = \sqrt{\lambda(\mathbf{x})} = \mathbf{f}'(\mathbf{x})\beta. \quad (31)$$

Using the MLE of  $\hat{\beta}$  we get the estimate of the mean response to be

$$\hat{\mu}(\mathbf{x}) = \hat{\lambda}(\mathbf{x}) = [\mathbf{f}'(\mathbf{x})\hat{\beta}]^2. \quad (32)$$

Equations (16) and (21) show (see Appendix B) that the mean-squared error of prediction (MSEP) at  $\mathbf{x}$  is approximated by

$$\begin{aligned} \text{MSE}[\hat{\lambda}(\mathbf{x})] &\approx 4\lambda(\mathbf{x}) \mathbf{f}'(\mathbf{x}) \frac{(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}}{\phi} \mathbf{f}(\mathbf{x}) \\ &\quad + \{2\mathbf{f}'(\mathbf{x}) \text{Bias}(\hat{\beta}) \sqrt{\lambda(\mathbf{x})} + \mathbf{f}'(\mathbf{x}) \frac{(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}}{\phi} \mathbf{f}(\mathbf{x})\}^2, \end{aligned} \quad (33)$$

where  $\text{Bias}(\hat{\beta})$  is approximated by formula (19),  $\mathbf{W} = \text{Diag}(w_1, w_2, \dots, w_n)$  with  $w_u = 4$ , and the  $u$ th element of  $\xi$  is  $-\frac{z_{uu}}{2\phi\sqrt{\lambda_u}}$ , where  $z_{uu}$  is the  $u$ th diagonal element of  $\mathbf{Z} = \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'$ ,  $u = 1, 2, \dots, n$ . It should be noted that the expression of  $\text{MSE}[\hat{\mu}(\mathbf{x})]$  changes with the choice of the link function. It also depends on the unknown  $\beta$ .

### 4.3. A Numerical Example

In this section, we present an example that illustrates the implementation of the QDGs for the log-linear model (that is, using a logarithmic link function) with a Poisson distribution. We also assess the robustness of the QDGs to the form of the link function and the distribution of the response variable.

The data considered in this example were taken from Piegorsch, Weinberg, and Margolin<sup>16</sup>. In a biomedical study of the immuno-activating ability of two agents, TNF (tumor necrosis factor) and IFN (interferon), to induce cell differentiation, the number of cells that exhibited differentiation after exposure to TNF and/or IFN was recorded using a  $4 \times 4$  factorial design. At each of the 16 dose combinations of TNF/IFN, 200 cells were examined. The number  $y$  of cells differentiating in one trial, and the corresponding dose levels of the two factors are given in Table 1.

Model (28) was fitted to the data set in Table 1 using the following model for the linear predictor

$$\eta(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2. \quad (34)$$

Here,  $x_1$  and  $x_2$  denote the coded values  $x_i = (X_i - 50)/50$ , with  $X_1$  and  $X_2$  being the actual values of agents TNF and IFN, respectively. The dispersion parameter  $\phi$  was taken to be 1.

Table 1. Experimental design and response values.

$y$	$X_1$	$X_2$	$x_1$	$x_2$
11	0	0	-1	-1
18	0	4	-1	-0.92
20	0	20	-1	-0.6
39	0	100	-1	1
22	1	0	-0.98	-1
38	1	4	-0.98	-0.92
52	1	20	-0.98	-0.6
69	1	100	-0.98	1
31	10	0	-0.8	-1
68	10	4	-0.8	-0.92
69	10	20	-0.8	-0.6
128	10	100	-0.8	1
102	100	0	1	-1
171	100	4	1	-0.92
180	100	20	1	-0.6
193	100	100	1	1

Table 2. Parameter estimates and model analysis using log link.

Parameter	Estimate	Std. error	P-value
$\beta_0$	6.7330	0.2316	< 0.0001
$\beta_1$	0.7939	0.0386	< 0.0001
$\beta_2$	0.3668	0.0364	< 0.0001
$\beta_{11}$	-1.9718	0.2344	< 0.0001
$\beta_{22}$	-0.5075	0.1153	< 0.0001
$\beta_{12}$	-0.1522	0.0350	< 0.0001

Note : Scaled deviance = 54.1883; DF = 10.

The experimental region,  $R$ , of the agent combinations is a square with  $-1 \leq x_1 \leq 1$ ,  $-1 \leq x_2 \leq 1$ . The parameter maximum likelihood estimates and their standard errors for model (34) are shown in Table 2.

We refer to the design in Table 1 as  $D_1$ . We also consider another design  $D_2$ , namely, a face-centered cube with replicated center and face-center points given in Table 3. Designs  $D_1$  and  $D_2$  have different design settings.

For each of the two designs, we consider the distribution of  $\tau_D(\mathbf{x}, \boldsymbol{\beta})$  on each of the several concentric squares,  $R_\nu$ , which are obtained by a reduction of the boundary of  $R$  using a shrinkage factor  $\nu$ ,  $0.5 < \nu \leq 1$ . Thus,  $R_\nu$  is determined by the inequalities

$$a_i + (1 - \nu)(b_i - a_i) \leq x_i \leq b_i - (1 - \nu)(b_i - a_i), i = 1, 2,$$

where  $a_i$  and  $b_i$  are the bounds on  $x_i$  in  $R$  ( $i=1,2$ ), that is,  $a_1 = -1$ ,  $a_2 = -1$ ,  $b_1 = 1$ ,  $b_2 = 1$  (see Figure 1).

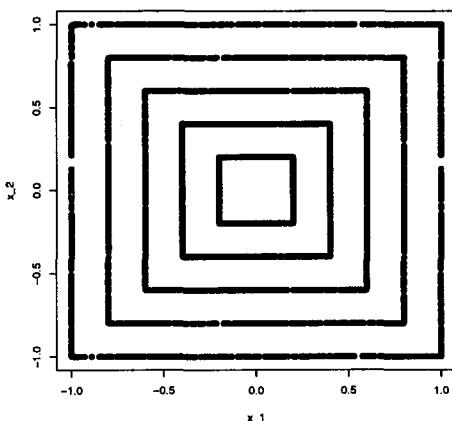


Fig. 1. Concentric squares within the region  $R$ .

To investigate the dependence of  $\tau_D(\mathbf{x}, \boldsymbol{\beta})$  on  $\boldsymbol{\beta}$ , a parameter space was established. For each parameter, a range consisting of the parameter's point estimate plus/minus four standard errors (from Table 2) was considered. A subset, C, of this parameter space was obtained by selecting three points within each parameter range, namely, the point estimate and the two end points. Thus, the number of points in C is  $3^6 = 729$ . The same parameter space was used for the two designs.

Table 3. Design  $D_2$ .

$X_1$	$X_2$	$x_1$	$x_2$
0	0	-1	-1
0	100	-1	1
100	100	1	1
100	0	1	-1
50	0	0	-1
100	50	1	0
50	100	0	1
0	50	-1	0
50	0	0	-1
100	50	1	0
50	100	0	1
0	50	-1	0
50	50	0	0
50	50	0	0
50	50	0	0
50	50	0	0

For each design and a selected value of  $\boldsymbol{\beta}$  in C, quantiles of the distribution of  $\tau_D(\mathbf{x}, \boldsymbol{\beta})$  are obtained for  $\mathbf{x} \in R_\nu$ , where  $\nu$  is one of several values chosen from the interval  $(0.5, 1]$ . The number of points chosen on each  $R_\nu$  was 1000, consisting of 250 points on each side. The quantiles are calculated for  $p = 0(0.05)1$ . The procedure is repeated for other values of  $\boldsymbol{\beta}$  in the subset C. Then,  $Q_D^{max}(p, \nu)$  and  $Q_D^{min}(p, \nu)$  were calculated using formulas (25) and (26). PROC GENMOD in SAS was used to calculate the parameter estimates. All other computations were done using the R software (version 1.6). For better representation, we take the natural logarithm (log) of the quantiles of the MSEP when plotting the QDGs.

To compare the two designs, we examine the corresponding QDGs shown in Figure 2. We note that the maximum quantiles of  $D_1$  are above those of  $D_2$  for most values of  $p$ , indicating that  $D_2$  has better prediction capability than  $D_1$ . The values of  $Q_D^{max}(p, \nu)$  and  $Q_D^{min}(p, \nu)$  for both designs are

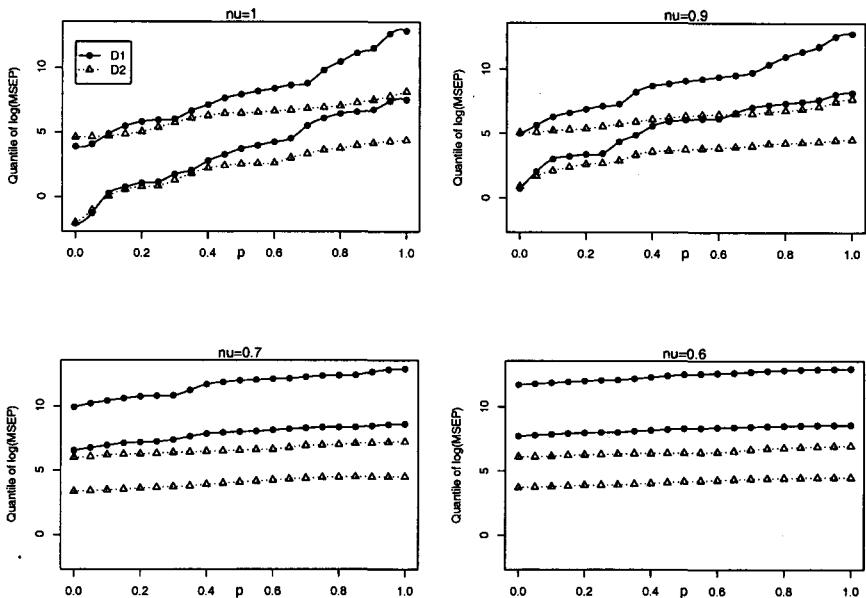


Fig. 2. QDGs for designs  $D_1$  and  $D_2$  using Poisson distribution ( $\phi = 1$ ) with a log link.

at some distance from each other, indicating sensitivity to the parameter values. As  $\nu$  decreases (that is, as we get closer to the center of  $R$ ), there is a slight decrease in the distance between the  $Q_D^{max}(p, \nu)$  and  $Q_D^{min}(p, \nu)$  for  $D_2$ , while the quantiles of  $D_1$  do not change much. Thus,  $D_2$  becomes slightly more robust to the parameter values near the design center, but the quantiles for  $D_1$  remain the same.

To assess the sensitivity of the QDGs to the form of the link function, we use the same data and assume the same Poisson distribution, as before, but employ a different link, namely, the square root.

We start with fitting model (31) to the data set in Table 1 using model (34) for the linear predictor. The dispersion parameter  $\phi$  was taken to be 1. The same two designs  $D_1$  and  $D_2$  are used as before. Our region of interest,  $R$ , and the several concentric squares,  $R_\nu$ , remain unchanged. The parameter estimates and their standard errors in Table 4 using the data set based on design  $D_1$ , however, change. The change occurs due to the different link function used. The corresponding quantiles were computed in a similar fashion.

Table 4. Parameter estimates and model analysis using square root link.

Parameter	Estimate	Std. error	P-value
$\beta_0$	17.7924	0.9231	< 0.0001
$\beta_1$	3.5137	0.1722	< 0.0001
$\beta_2$	1.5413	0.1807	< 0.0001
$\beta_{11}$	-6.5263	0.8782	< 0.0001
$\beta_{22}$	-2.0301	0.5035	< 0.0001
$\beta_{12}$	-0.2352	0.1852	0.2042

Note : Scaled deviance = 58.1596; DF = 10.

The QDGs for both designs are compared using the two different links. Figure 3 gives a comparison of the QDGs for  $D_1$  using the Poisson distribution with a log link and a square root link, and Figure 4 shows the same for  $D_2$ . The canonical link for the Poisson distribution is the log link. Mis-specifying the link as square root changes the robustness of the designs to the parameter values. The distance between the maximum and minimum quantiles for design  $D_1$  increases slightly as we change the link from log to square root. This indicates a slight increase in sensitivity to the parameter values for the square root link. The maximum and minimum quantiles for

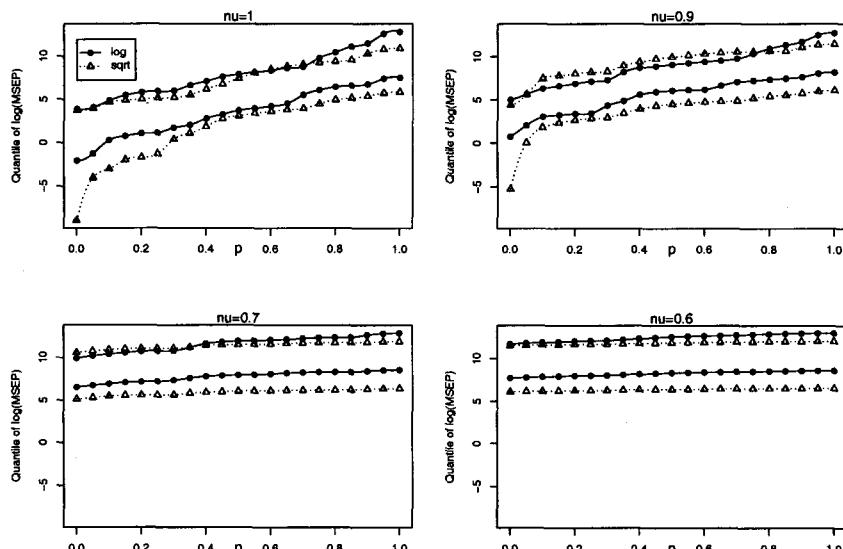


Fig. 3. Comparison of the QDGs for design  $D_1$  using Poisson distribution ( $\phi = 1$ ) with log and square root links.

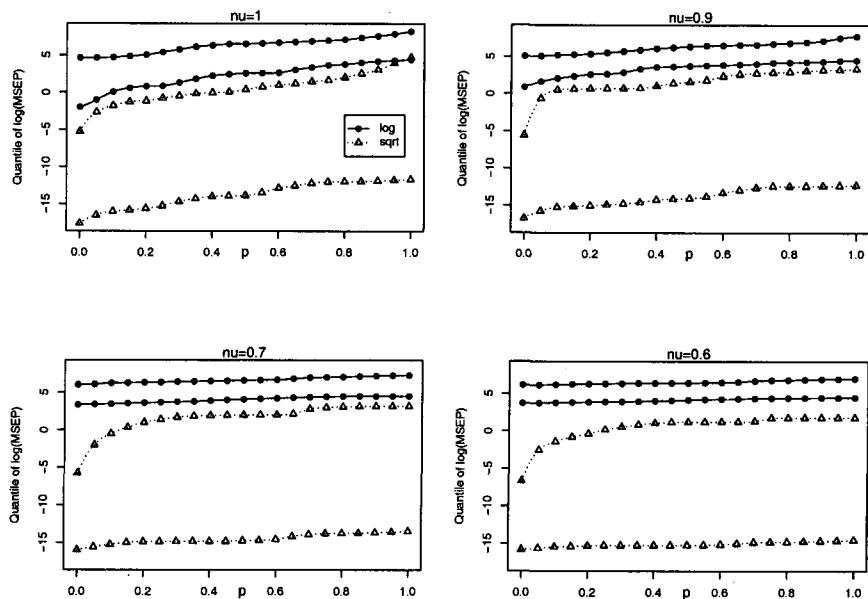


Fig. 4. Comparison of the QDGs for design  $D_2$  using Poisson distribution ( $\phi = 1$ ) with log and square root links.

$D_2$  for the square root link are further apart from each other than the log link case. Thus,  $D_2$  loses its robustness to the parameter values as we alter the link.

Next we try to study the robustness of the QDGs to different distributions. We use four distributions, namely, Poisson distribution with  $\phi = 1$ , Poisson distribution with  $\phi = 0.1824$  ( $\phi$  is estimated by using PROC GENMOD in SAS [options: scale = Pearson]), negative binomial distribution, and gamma distribution, each with log link. The model for the linear predictor remains the same and the experimental region  $R$  also remains unchanged. The parameter estimates and their respective standard errors, however, change.

Figure 5 gives the comparison of the QDGs with these four distributions for  $\nu = 1$ . The distance between the maximum and the minimum quantiles for both designs increases a little for the Poisson distribution with an estimated  $\phi$ , indicating a slight loss in robustness to the parameter values. For the negative binomial distribution, the designs become more sensitive to the parameters. In the case of the gamma distribution, the quantile plots for both  $D_1$  and  $D_2$  are very close to one another. We also notice that the

distance between the maximum and minimum is smaller for the gamma distribution by comparison to the estimated Poisson or the negative binomial distributions. Both designs show similar prediction capabilities for all four distributions. Overall, the shape of the QDGs has been affected by the distribution, but the superiority of  $D_2$  over  $D_1$ , in this example, appears to be unaffected by the distribution.

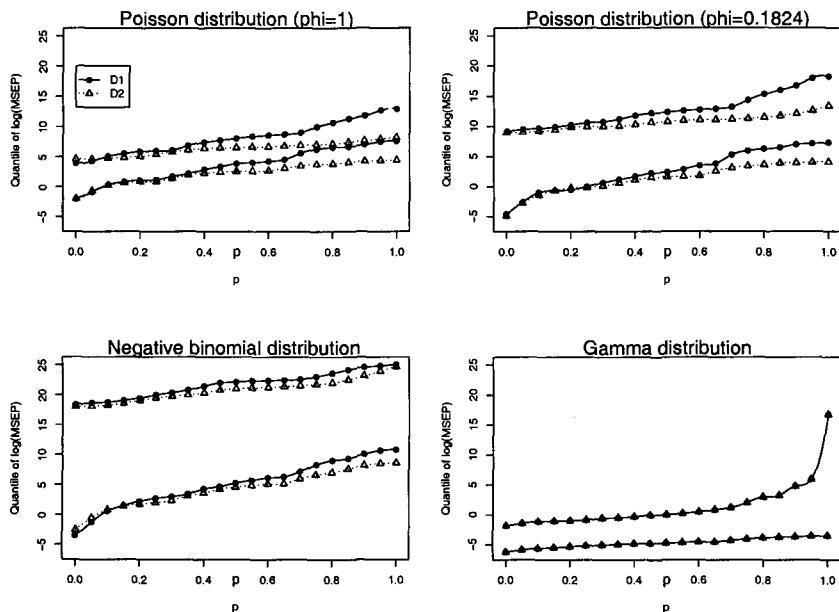


Fig. 5. Comparison of the QDGs for designs  $D_1$  and  $D_2$  using four different distributions, all with log link ( $\nu = 1$ ).

## 5. Conclusion

The numerical example demonstrates that the QDGs provide a convenient technique for evaluating and comparing designs for GLMs. They also provide information on the prediction capability of a design throughout the experimental region, and on its dependence on the model's parameters. Estimation bias is taken into account in the comparison of the designs. The example also shows that the choice of the link function can influence the shape of the QDGs for a given distribution. For a given link function, the example shows that the QDGs may also be influenced by the distribution of the response variable.

## Appendix A. Derivation of Formula (30)

For log-linear Poisson regression,  $\mu(\mathbf{x}) = \lambda(\mathbf{x})$ ,  $\eta = \log \lambda$ , and

$$\theta_u = \log \lambda_u, u = 1, 2, \dots, n, \quad (\text{A.1})$$

where  $\theta_u$  is the parameter shown in (1). Hence, from (A.1),

$$\lambda_u = \exp(\theta_u), u = 1, 2, \dots, n. \quad (\text{A.2})$$

Also from (13),  $\mathbf{W} = \text{Diag}(w_1, w_2, \dots, w_n)$ , where from (14),

$$w_u = \frac{1}{v_u} \left( \frac{d\mu_u}{d\eta_u} \right)^2 = \frac{[\exp \eta_u]^2}{\lambda_u} = \lambda_u, \quad (\text{A.3})$$

since from (28)

$$\begin{aligned} \frac{d\mu_u}{d\eta_u} &= \exp(\eta_u), \\ &= \lambda_u, u = 1, 2, \dots, n, \end{aligned} \quad (\text{A.4})$$

and

$$v_u = \frac{d\mu_u}{d\theta_u} = \lambda_u, u = 1, 2, \dots, n. \quad (\text{A.5})$$

Furthermore,  $\mathbf{Z}_d = \text{Diag}(z_{11}, \dots, z_{nn})$ , where  $z_{uu}$  is the  $u$ th diagonal element of  $\mathbf{Z} = \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'$ , and  $\mathbf{F} = \text{Diag}(f_{11}, \dots, f_{nn})$ , where

$$f_{uu} = \frac{1}{v_u} \left[ \frac{d^2 \lambda_u}{d\eta_u^2} \right] \left[ \frac{d\lambda_u}{d\eta_u} \right], u = 1, \dots, n. \quad (\text{A.6})$$

Here,

$$\begin{aligned} \frac{d^2 \lambda_u}{d\eta_u^2} &= \exp(\eta_u), \\ &= \lambda_u, u = 1, \dots, n. \end{aligned} \quad (\text{A.7})$$

Hence, from (A.4), (A.5) and (A.7),

$$f_{uu} = \lambda_u, u = 1, \dots, n. \quad (\text{A.8})$$

It follows that the  $u$ th element of  $\boldsymbol{\xi}$  in (19) is of the form

$$\begin{aligned} \xi_u &= (-1/2\phi) \frac{z_{uu} \lambda_u}{\lambda_u} \\ &= -\frac{z_{uu}}{2\phi}, u = 1, \dots, n. \end{aligned} \quad (\text{A.9})$$

Now, the mean-squared error of prediction at  $\mathbf{x}$  is

$$\text{MSE}[\hat{\lambda}(\mathbf{x})] = \text{Var}[\hat{\lambda}(\mathbf{x})] + \{\text{Bias}[\hat{\lambda}(\mathbf{x})]\}^2. \quad (\text{A.10})$$

From, (16) and (A.4), we have the approximation

$$\begin{aligned} \text{Var}[\hat{\lambda}(\mathbf{x})] &\approx \left[ \frac{d\lambda(\mathbf{x})}{d\eta(\mathbf{x})} \right]^2 \mathbf{f}'(\mathbf{x}) \frac{(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}}{\phi} \mathbf{f}(\mathbf{x}) \\ &= [\lambda(\mathbf{x})]^2 \mathbf{f}'(\mathbf{x}) \frac{(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}}{\phi} \mathbf{f}(\mathbf{x}). \end{aligned} \quad (\text{A.11})$$

Also from (15), (21), (A.4) and (A.7),

$$\begin{aligned} \text{Bias}[\hat{\lambda}(\mathbf{x})] &\approx \text{Bias}[\hat{\eta}(\mathbf{x})] \frac{d\lambda(\mathbf{x})}{d\eta(\mathbf{x})} + \frac{1}{2} \text{Var}[\hat{\eta}(\mathbf{x})] \frac{d^2\lambda(\mathbf{x})}{d\eta^2(\mathbf{x})} \\ &= \text{Bias}[\hat{\eta}(\mathbf{x})] \lambda(\mathbf{x}) + \frac{1}{2} \mathbf{f}'(\mathbf{x}) \frac{(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}}{\phi} \mathbf{f}(\mathbf{x}) \lambda(\mathbf{x}), \end{aligned} \quad (\text{A.12})$$

since from (15),  $\text{Var}[\hat{\eta}(\mathbf{x})]$  is approximately equal to  $\frac{\mathbf{f}'(\mathbf{x})(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{f}(\mathbf{x})}{\phi}$ . Note that from formula (20), the bias of  $\hat{\eta}(\mathbf{x})$  in (A.12) is approximated by

$$\text{Bias}[\hat{\eta}(\mathbf{x})] \approx \mathbf{f}'(\mathbf{x})(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\xi.$$

Using (A.11) and (A.12) in (A.10), we obtain formula (30).

## Appendix B. Derivation of Formula (33)

For Poisson regression with a square root link,  $\mu(\mathbf{x}) = \lambda(\mathbf{x})$ ,  $\eta = \sqrt{\lambda}$ , and

$$\theta_u = \log \lambda_u, u = 1, 2, \dots, n, \quad (\text{B.1})$$

where  $\theta_u$  is the parameter shown in (1). Hence, from (B.1),

$$\lambda_u = \exp(\theta_u), u = 1, 2, \dots, n. \quad (\text{B.2})$$

Also from (13),  $\mathbf{W} = \text{Diag}(w_1, w_2, \dots, w_n)$ , where from (14),

$$w_u = \frac{1}{v_u} \left( \frac{d\mu_u}{d\eta_u} \right)^2 = \frac{4\lambda_u}{\lambda_u} = 4, \quad (\text{B.3})$$

since from (31)

$$\begin{aligned} \frac{d\mu_u}{d\eta_u} &= 2\eta_u, \\ &= 2\sqrt{\lambda_u}, u = 1, 2, \dots, n, \end{aligned} \quad (\text{B.4})$$

and

$$v_u = \frac{d\mu_u}{d\theta_u} = \lambda_u, \quad u = 1, 2, \dots, n. \quad (\text{B.5})$$

Furthermore,  $\mathbf{Z}_d = \text{Diag}(z_{11}, \dots, z_{nn})$ , where  $z_{uu}$  is the  $u$ th diagonal element of  $\mathbf{Z} = \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'$ , and  $\mathbf{F} = \text{Diag}(f_{11}, \dots, f_{nn})$ , where

$$f_{uu} = \frac{1}{v_u} \left[ \frac{d^2\lambda_u}{d\eta_u^2} \right] \left[ \frac{d\lambda_u}{d\eta_u} \right], \quad u = 1, \dots, n. \quad (\text{B.6})$$

Here,

$$\frac{d^2\lambda_u}{d\eta_u^2} = 2, \quad u = 1, \dots, n. \quad (\text{B.7})$$

Hence, from (B.4), (B.5) and (B.7),

$$f_{uu} = \frac{4}{\sqrt{\lambda_u}}, \quad u = 1, \dots, n. \quad (\text{B.8})$$

It follows that the  $u$ th element of  $\xi$  in (19) is of the form

$$\begin{aligned} \xi_u &= (-1/2\phi) \frac{1}{4} z_{uu} \frac{4}{\sqrt{\lambda_u}} \\ &= -\frac{z_{uu}}{2\phi\sqrt{\lambda_u}}, \quad u = 1, \dots, n. \end{aligned} \quad (\text{B.9})$$

Now, the mean-squared error of prediction at  $\mathbf{x}$  is

$$\text{MSE}[\hat{\lambda}(\mathbf{x})] = \text{Var}[\hat{\lambda}(\mathbf{x})] + \{\text{Bias}[\hat{\lambda}(\mathbf{x})]\}^2. \quad (\text{B.10})$$

From, (16) and (B.4), we have

$$\begin{aligned} \text{Var}[\hat{\lambda}(\mathbf{x})] &\approx \left[ \frac{d\lambda(\mathbf{x})}{d\eta(\mathbf{x})} \right]^2 \mathbf{f}'(\mathbf{x}) \frac{(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}}{\phi} \mathbf{f}(\mathbf{x}) \\ &= [4\lambda(\mathbf{x})] \mathbf{f}'(\mathbf{x}) \frac{(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}}{\phi} \mathbf{f}(\mathbf{x}). \end{aligned} \quad (\text{B.11})$$

Also from (15), (21), (B.4) and (B.7),

$$\begin{aligned} \text{Bias}[\hat{\lambda}(\mathbf{x})] &\approx \text{Bias}[\hat{\eta}(\mathbf{x})] \frac{d\lambda(\mathbf{x})}{d\eta(\mathbf{x})} + \frac{1}{2} \text{Var}[\hat{\eta}(\mathbf{x})] \frac{d^2\lambda(\mathbf{x})}{d\eta^2(\mathbf{x})} \\ &= 2 \text{Bias}[\hat{\eta}(\mathbf{x})] \sqrt{\lambda(\mathbf{x})} + \mathbf{f}'(\mathbf{x}) \frac{(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}}{\phi} \mathbf{f}(\mathbf{x}), \end{aligned} \quad (\text{B.12})$$

since from (15),  $\text{Var}[\hat{\eta}(\mathbf{x})]$  is approximated by  $\frac{\mathbf{f}'(\mathbf{x})(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{f}(\mathbf{x})}{\phi}$ . Note that from (20), the bias of  $\hat{\eta}(\mathbf{x})$  in (B.12) is approximately given by

$$\text{Bias}[\hat{\eta}(\mathbf{x})] \approx \mathbf{f}'(\mathbf{x})(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\boldsymbol{\xi}.$$

Using (B.11) and (B.12) in (B.10), we obtain formula (33).

## References

1. G.E.P. Box and N.R. Draper, *Empirical Model-Building and Response Surfaces* (Wiley, New York, 1987).
2. A.I. Khuri and J.A. Cornell, *Response Surfaces*, 2nd Edition (Dekker, New York, 1996).
3. R.H. Myers and D.C. Montgomery, *Response Surface Methodology*, 2nd Edition (Wiley, New York, 2002).
4. W.G. Cochran, Experiments for nonlinear functions, *J. Amer. Statist. Assoc.* **68**, 771-781 (1973).
5. K.S. Robinson and A.I. Khuri, Quantile dispersion graphs for evaluating and comparing designs for logistic regression models, *Comput. Statist. Data. Anal.* **43**, 47-62 (2003).
6. A.I. Khuri and J. Lee, A graphical approach for evaluating and comparing designs for nonlinear models, *Comput. Statist. Data. Anal.* **27**, 433-443 (1998).
7. J.A. Nelder and R.W.M. Wedderburn, Generalized linear models, *J. Roy. Statist. Soc. A* **135**, 370-384 (1972).
8. P. McCullagh and J.A. Nelder, *Generalized Linear Models*, 2nd Edition (Chapman and Hall, London, 1989).
9. A.J. Dobson, *An Introduction to Generalized Linear Models*, 2nd Edition (Chapman and Hall, Boca Raton, 2002).
10. J.K. Lindsey, *Applying Generalized Linear Models* (Springer, New York, 1997).
11. C.E. McCulloch and S.R. Searle, *Generalized, Linear, and Mixed Models* (Wiley, New York, 2001).
12. R.H. Myers, D.C. Montgomery and G.G. Vining, *Generalized Linear Models with Applications in Engineering and the Sciences* (Wiley, New York, 2002).
13. G.M. Cordeiro and P. McCullagh, Bias correction in generalized linear models, *J. Roy. Statist. Soc. B* **53**, 629-643 (1991).
14. N.G. Cadigan, Bias approximation for maximum likelihood estimates, *J. Statist. Comput. Simul.* **51**, 89-95 (1994).
15. A. Agresti, *Categorical Data Analysis*, 2nd Edition (Wiley, New York, 2002).
16. W.W. Piegorsch, C.R. Weinberg and B.H. Margolin, Exploring simple independent action in multifactor tables of proportions, *Biometrics* **44**, 595-603 (1988).

This page is intentionally left blank

## CHAPTER 10

### DESIGN FOR A TRINOMIAL RESPONSE TO DOSE

Shenghua Kelly Fan<sup>1</sup> and Kathryn Chaloner<sup>2</sup>

<sup>1</sup>*Department of Statistics, California State University,  
East Bay, 25800 Carlos Bee Boulevard, Hayward, CA 94542-3087*

<sup>2</sup>*Department of Biostatistics, Department of Statistics & Actuarial Science,  
The University of Iowa, Iowa City, IA 52242  
E-mail: kathryn-chaloner@uiowa.edu*

This chapter discusses the design problem for a dose response experiment where the response is a trinomial. The literature is briefly reviewed and the proportional odds model and the continuation-ratio model are described. Results are presented for the continuation-ratio model. A special case of the continuation-ratio model with a constant slope coefficient is examined in detail. The robustness of the assumption of a constant slope coefficient when the slopes are not constant, but are almost equal, is explored. The designs are found to be reasonably robust to the assumption. In addition, because the criteria for design are based on large sample properties of the maximum likelihood estimator, the properties of an estimator for a special case with a small sample size are also presented. The special case is that of estimating the dose with the largest probability of a response in one category. In this simulation the c-optimal designs and the maximum likelihood estimator for the constant slope model are found to perform well, even when the data is generated from a model with different slopes.

#### 1. Introduction

Although not traditionally considered to be response surface methodology, dose response models and design for a binomial or multinomial response come under the very broad definition of response surface methods used by Mead and Pike.<sup>1</sup> For the trinomial response model used in this chapter, the second category of response corresponds to success (efficacy with no toxicity), and the probability of success is close to zero for very low doses,

then increases with dose to a maximum probability, and then decreases with dose as the probability of toxicity increases. Estimating the dose at which the maximum probability of success occurs is the example used in the simulations which parallels the more usual response surface problem of estimating the value of the explanatory variables at which the expected response is a maximum.

In dose response studies, the response of a subject to dose is often modeled as a binary response: with the response being either a binary toxicity (yes/no) or a binary efficacy (yes/no). This dose response design has been extensively studied and there is an extensive literature on this topic, see for example Refs. 2, 3, 4, 5. A bivariate binary response is often more appropriate and can incorporate both toxicity and efficacy; a bivariate binary response can however also be expressed as a single multinomial response.<sup>6</sup> Multinomial responses can be characterized as nominal or ordered. Nominal scales are devoid of structure; ordered scales can be ordinal with no numerical score attached, or interval with numerical scores attached.<sup>7</sup> Categories may also be considered nested, which is also referred to as being hierarchical<sup>7</sup> or contingent.<sup>8,9</sup> Glonek and McCullagh<sup>10</sup> and Glonek<sup>11</sup> provide a general class of models, denoted multivariate logistic models, for multinomial responses, with both continuous and categorical explanatory variables.

Zocchi and Atkinson<sup>12</sup> study optimal design for a special case of a multivariate logistic model. Perevozskaya, Rosenberger and Haines<sup>13</sup> provide locally D-optimal designs for the proportional odds model. They also provide designs which are optimal for estimating multiple quantiles and explore changes in the design as the parameters change. Fan and Chaloner<sup>14,15</sup> provide optimal designs for the trinomial continuation-ratio model. Design for a related model of bivariate binary response for toxicity and efficacy is described by Heise and Myers.<sup>16</sup>

Thall and Russell<sup>6</sup> develop an algorithm for conducting a dose finding clinical trial (a phase I/II trial) sequentially where the response follows the trinomial proportional odds model and the goal is to estimate the dose at which the probability of "success" (efficacy and no toxicity) is maximized. This probability initially increases as dose increases but then decreases because the probability of toxicity increases. Related sequential design strategies are also available based on either bandit strategies<sup>17</sup> or up-and-down strategies<sup>18</sup> which maximize the probability of a binary response which is assumed to be unimodal as a function of dose. Paul, Rosenberger and Flournoy<sup>19</sup> consider optimal sequential design for an ordinal toxicity

response with several response categories. Their motivating example has 5 response categories corresponding to toxicity severity ratings from 0 to 4 and their criterion is to find a set of dose levels corresponding to specified probabilities of different response categories. They compare their design strategies to strategies from the literature on phase I designs.

Design for binomial and multinomial responses with a single explanatory variable provides insight into other response surface problems for binomial or multinomial responses using mixtures of more than one drug, for example in chemotherapy. This is clearly an important application of response surface methodology which has yet to be extensively explored. An extensive overview of binary responses for drug combinations is given in Chapter 7 of the monograph by Hewlett and Plackett.<sup>20</sup> They present different models, and different approaches to modeling the combined action of 2 drugs with a binary response, and, as they note, the extension to 3 or more drugs is essentially no different. They also discuss the history of such modeling in pharmacology and discuss several ways in which drugs can interact. More recent references and developments in the statistics literature are given by Abdelbasit and Plackett;<sup>21</sup> Carter *et al.*<sup>22</sup> Unkelbach and Wolf<sup>23</sup> and Laska *et al.*<sup>24</sup> Kupchack<sup>25</sup> provides some intriguing recent results on design for this problem which focus on detecting and modeling drug synergy and drug interactions with a binomial response. He uses optimal design criteria such as D-optimality to derive designs. Some results for general binary responses with two or more explanatory variables<sup>26,27</sup> are also available. These publications all consider binary responses to mixtures of drugs. Modeling and design issues of a binomial or multinomial responses to mixtures of drugs with different doses is a response surface problem with additional methods and applications to be investigated. Additional design problems waiting for investigation are for the correlated multinomial responses that occur in developmental toxicity studies where correlations between and within litters are present.<sup>28</sup>

### 1.1. Notation and Definitions

An experimental design will be regarded as a probability measure on the dose domain  $\mathcal{X}$ . That is a design  $\eta$  puts weight  $m_i$  at dose  $x_i$  for  $i = 1, 2, \dots, k$ , where the  $m_i$ 's are non-negative and sum to one. For a sample size  $n$ , the values  $nm_i$  can be rounded to integers.<sup>29</sup>

For any particular model with unknown parameters  $\theta$ , the inverse of the Fisher information matrix gives the asymptotic variance-covariance matrix

for the maximum likelihood estimate (MLE) of the unknown parameters. The Fisher information matrix for a sample size of  $n = 1$  is denoted  $M(\theta, \eta)$  and usually depends on the value of  $\theta$  and the design  $\eta$ . Designs which optimize concave criteria based on the Fisher information matrix are often considered<sup>4</sup> and will be described here.

A locally D-optimal design is that which maximizes  $\log\det M(\theta, \eta)$  for a specified value  $\theta$ . If the specified value of  $\theta$  is close to the true value then maximizing this criterion should make the asymptotic variance-covariance matrix of the MLE of  $\theta$  "small." The same criterion averaged over a "prior" distribution,  $\pi$ , on the parameters gives "Bayesian" D-optimality: maximizing  $\int \log\det M(\theta, \eta) d\pi(\theta)$ . This criterion is typically a more robust criterion than local D-optimality in that it is typically more efficient for over  $\theta$  values close to the best guess and includes local D-optimality as a special case.<sup>2,4</sup> A locally c-optimal design minimizes the asymptotic variance of the MLE of a function of interest at a specified value of  $\theta$  and a corresponding Bayesian c-optimal design<sup>30</sup> minimizes the asymptotic variance averaged over a prior distribution:  $\int c(\theta)^T M(\theta, \eta)^{-1} c(\theta) d\pi(\theta)$  where  $c(\theta)$  is the gradient vector of the function to be estimated and  $\pi(\theta)$  is the prior distribution on the parameter  $\theta$ .

## 1.2. Overview

Section 2 presents two popular models for a trinomial response: the proportional odds model and the continuation-ratio model.<sup>7</sup> Section 3 provides examples of optimal designs and also designs called "limiting optimal" designs<sup>14,15</sup> for the continuation-ratio model. A sequence of limiting optimal designs is optimal in a limiting sense for a sequence of prior distributions. A member of the sequence can be used as an approximately optimal design for a prior distribution which is a member of the corresponding sequence of prior distributions. This concept is useful when algebraic forms for optimal designs are not available, but an algebraic form is available for a sequence of limiting optimal designs.

The continuation-ratio model has 4 parameters in general and a special case, leading to a 3 parameter model, is where the 2 slope parameters are equal. It may be convenient to use the 3 parameter model if the slopes are close, but not quite equal. Section 4 presents a simulation study exploring the sensitivity of the performance of the optimal designs in this situation. In addition, since optimal design criteria are based on large sample properties, and the sample sizes for early phase clinical trials are often small,

a simulation study of the performance of optimal design for a small sample size is also presented. Fan and Chaloner<sup>15</sup> briefly describe an overview of the numerical investigation but do not present specific results. These and additional numerical results are also described in the thesis of the first author, available on request. A concluding discussion is given in Section 5.

## 2. The Trinomial Response Model

In a clinical trial, efficacy and toxicity can be considered separately as bivariate outcomes. The risk-benefit can then be assessed, as described in the literature<sup>31,32</sup> for situations with a fixed dose. When the toxicity outcome is very severe, such as death or permanent kidney damage requiring lifelong dialysis, the concept of efficacy in the presence of toxicity becomes meaningless. For such cases, instead of defining the response of a subject as a bivariate binary response (efficacy: yes/no; toxicity: yes/no), the response of the subject may be better considered as a nested trinomial response: first “no reaction” (no toxicity, no efficacy), second “success” (no toxicity, efficacy) and third “adverse reaction” (toxicity, irrespective of efficacy).

The response when  $n_i$  subjects are given a dose  $x_i$  (often in log units) is therefore trinomial,  $(y_{1i}, y_{2i}, y_{3i})$ ,  $y_{1i} + y_{2i} + y_{3i} = n_i$  and the corresponding cell probabilities are  $(p_1(\theta, x_i), p_2(\theta, x_i), p_3(\theta, x_i))$  respectively, where  $\theta$  denotes the parameters of the model and  $p_k(\theta, x)$  is the probability of a single subject given a dose  $x$  responding in the  $k^{th}$  response category,  $k = 1, 2, 3$ . For any  $x$  and  $\theta$ ,  $p_1(\theta, x) + p_2(\theta, x) + p_3(\theta, x) = 1$ . The values of  $x$  can be chosen from some set  $\mathcal{X}$ .

Thall and Russell<sup>6</sup> considered this response as ordered because of the nature of the dose-response relationship: the response is most likely to be “no reaction” for low doses, “success” for moderate doses, and “adverse reaction” for high doses. They therefore used the proportional odds model<sup>7,33,13</sup> to model the dose-response relationship. The proportional odds model for a trinomial response is:

$$\begin{aligned}\log[p_3(\theta, x)/(1 - p_3(\theta, x))] &= a + bx \\ \log[(p_2(\theta, x) + p_3(\theta, x))/p_1(\theta, x)] &= a + c + bx\end{aligned}$$

where  $c \geq 0, b > 0$ . There are only three parameters. The model is easily extended to more than 3 response categories under the assumption that the effect of dose is the same across the cumulative logits. The proportional odds model is popular for small experiments with a small number of subjects. The assumption of proportional odds might, however, be invalid, especially when

the number of response categories is greater than 3. Even for a trinomial response the assumption might sometimes fail. Agresti<sup>33</sup> gives a data set showing a poor fit of the proportional odds model.

The response probabilities in the trinomial proportional odds model can be easily shown to be:

$$\begin{aligned} p_1(\theta, x) &= \frac{1}{1 + e^{a+c+bx}} \\ p_2(\theta, x) &= \frac{e^{a+c+bx}}{(1 + e^{a+bx})(1 + e^{a+c+bx})} \\ p_3(\theta, x) &= \frac{e^{a+bx}}{1 + e^{a+bx}}. \end{aligned}$$

The continuation-ratio model is an alternative which is more natural for nested responses.<sup>7</sup> It has been used for phase I/II clinical trials.<sup>34</sup> Assuming  $b_1, b_2 > 0$  the continuation-ratio model for a trinomial response is:

$$\log[p_3(\theta, x)/(1 - p_3(\theta, x))] = a_1 + b_1 x \quad (1)$$

$$\log[p_2(\theta, x)/p_1(\theta, x)] = a_1 + c + b_2 x. \quad (2)$$

The number of parameters is 4 and the response probabilities are given by:

$$\begin{aligned} p_1(\theta, x) &= \frac{1}{(1 + e^{a_1+b_1x})(1 + e^{a_1+c+b_2x})} \\ p_2(\theta, x) &= \frac{e^{a_1+c+b_2x}}{(1 + e^{a_1+b_1x})(1 + e^{a_1+c+b_2x})} \\ p_3(\theta, x) &= \frac{e^{a_1+b_1x}}{1 + e^{a_1+b_1x}}. \end{aligned}$$

A special case, with 3 parameters, assumes  $b_1 = b_2$  and will be referred to as the “constant slope continuation-ratio model.” Note that this special case is not the same as the proportional odds model. In many typical applications of this model, for a low dose, ( $x = 0$  say in log units), the probability of adverse reaction,  $e^{a_1}/(1 + e^{a_1})$ , is expected to be lower or at least not much higher than the conditional probability of efficacy conditional on no adverse reaction,  $e^{a_1+c}/(1 + e^{a_1+c})$ . The value of  $c$ , therefore, is positive, or, if negative,  $|c|$  is small.

### 3. Optimal Designs

#### 3.1. Constant Slope Continuation-ratio Model

Figure 1 gives plots of  $(p_1(\theta, x), p_2(\theta, x), p_3(\theta, x))$  against  $x$  for 3 values of  $\theta$  for the constant slope continuation-ratio model, ( $b_1 = b_2$ ).

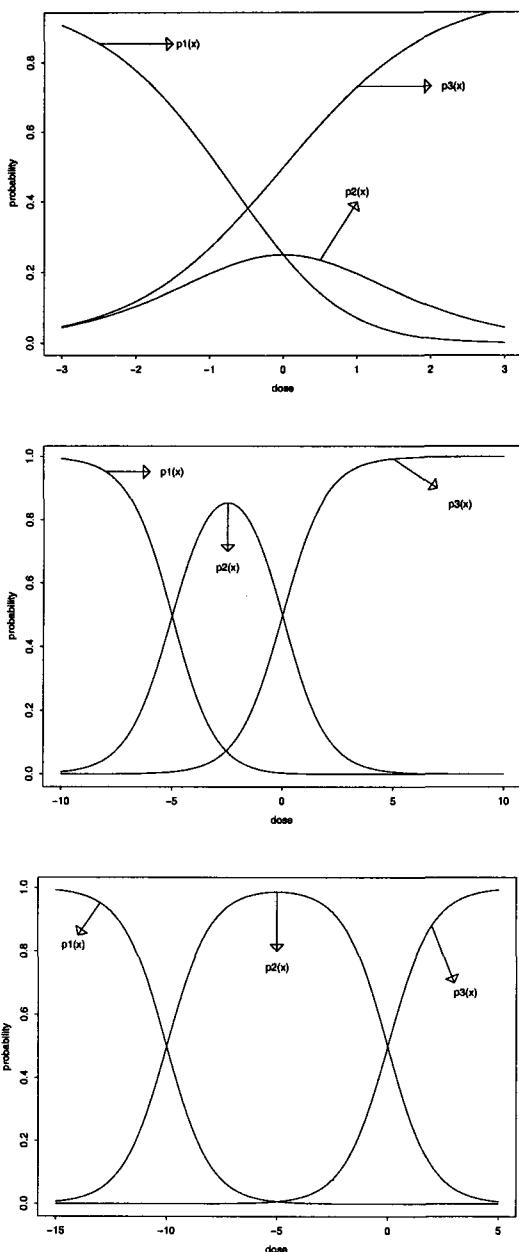


Fig. 1. Probability plot, probability vs. dose, for  $a_1 = 0$ ,  $b_1 = 1$ , and  $c = 0, 5$ , and  $10$ , from top to bottom.

The plots in Fig. 1 indicate that a larger value of  $c$  gives a wider range of dose levels where the probability of adverse reaction is low and the probability of success is high. In the following discussion,  $\theta$  will be rewritten as  $(c, a_1, b_1)$  for convenience. The parameter  $c$  plays an important role in finding optimal designs: it determines the shape of probability plot while other parameters,  $a_1$  and  $b_1$ , are like location and scale parameters. The value of  $c$  drives the number of design points and the weights of the locally D-optimal designs. Locally D-optimal designs are found numerically to have two, three, or four design points, the number of design points increasing with  $c$ . Specifically, for  $c = 0, 5$ , and  $10$  and  $a_1$  fixed at  $0$  and  $b_1$  fixed at  $1$ , corresponding to the 3 distributions shown in Fig. 1, the locally D-optimal designs have 2, 3 and 4 design points respectively. This can be explained heuristically by examining Fig. 1 and Eqs. (1) and (2): the value of  $x$  at which the probability of adverse reaction is  $\frac{1}{2}$  is  $x = -a_1/b_1$  and the value of  $x$  at which the conditional probability of success is  $\frac{1}{2}$ , given no adverse reaction, is  $x = -(a_1 + c)/b_1$ . As  $c$  increases and  $a_1$  and  $b_1$  remain fixed, these two  $x$  values become further apart and information is maximized around  $x$  values close to these two points and each region provides different information. Note that the response is multivariate and even though the model has 3 parameters, all 3 are estimable with a 2-point design. A 2-point design leads to an information matrix of full rank.

No closed form expressions have been found for the locally optimal designs, although Fan and Chaloner<sup>15</sup> derived limiting locally optimal designs which can be expressed in closed form. Consider a sequence of parameter values with  $a_1$  and  $b_1$  fixed and  $c$  changing:  $\theta_i = (c_i, a_1, b_1)$ , for  $i = 1, 2, \dots$  and  $c_i \rightarrow \infty$ . For this sequence, a sequence of limiting locally D-optimal designs puts equal numbers of observations at each of  $x = -(a_1 \pm 1.223)/b_1$  and  $x = -(a_1 + c_i \pm 1.223)/b_1$ . This sequence of limiting locally D-optimal designs is motivated by considering that as  $c$  increases the value of  $p_1$  at  $x = -(a_1 + c)/b_1$  converges to  $\frac{1}{2}$  and the value of  $p_3$  at  $x = -a_1/b_1$  is  $\frac{1}{2}$ . These limiting locally D-optimal designs are approximately optimal for large  $c$ .

For Bayesian D-optimality, a simple prior distribution is used for illustration. Let the values of  $a_1$  and  $b_1$  be fixed and the value of  $c$  be equally likely to be  $c = 0$  or  $c = c_0$ . Let the sequence of prior distributions be a sequence of distributions with  $c_0 = c_i$  for  $i = 1, 2, \dots$  and  $c_i \rightarrow \infty$ . Bayesian D-optimal designs for these simple prior distributions can be found numerically and typically have three dose levels. Fan and Chaloner<sup>15</sup> derived a sequence of limiting optimal designs for this sequence of prior distributions

which can be expressed in closed form. These limiting Bayesian D-optimal designs are approximately optimal for large  $c_0$ . For fixed  $a_1, b_1$  and a sequence of  $c_i$ ,  $i = 1, 2, \dots$  with  $c_i \rightarrow \infty$ , the sequence of designs putting weight 0.167, 0.448, and 0.385 at  $x = -(a_1 + c_i)/b_1$ ,  $-(a_1 + 1.47)/b_1$ , and  $-(a_1 - 1.14)/b_1$ , respectively, is a sequence of limiting Bayesian D-optimal designs.

Suppose now that the goal of the experiment is to find the dose,  $x_{max}$ , where  $p_2(\theta, x)$ , the probability of success (efficacy with no toxicity), is maximized:  $x_{max} = -(2a_1 + c)/2b_1$ . Then  $c(\theta)$  is the gradient vector of  $x_{max}$ ,  $c(\theta) = \nabla x_{max}$ . Locally c-optimal designs found numerically are typically 2-point designs. In addition, the sequence of designs putting weight 1/2 at each of  $x = -a_1/b_1$  and  $x = -(a_1 + c)/b_1$  is limiting c-optimal for a sequence of  $\theta$  where the value of  $c$  becomes large:  $c \rightarrow \infty$ . Note that the middle point of the two design points is the target dose  $x_{max}$ . In practice, the values of parameters are unknown but can be estimated with a two or more point design. When  $c$  is 0, the locally c-optimal design becomes a singular one-point design at  $x_{max}$ . Singular designs are of limited direct practical use as the parameters are not all estimable. They are however useful in sequential strategies and as benchmarks. (A definition of c-optimality extended to singular information matrices is given, for example, in Refs. 35 and 36).

### 3.2. General Continuation-ratio Model

For a more general model where  $b_1$  is not necessarily equal to  $b_2$  let  $r = b_1/b_2$ ,  $a_2 = a_1 + c$  and  $d = a_1 - ra_2$ . Because  $c$  is typically large and positive, the parameter  $d$  is correspondingly large and negative: that is  $d < 0$  and  $|d|$  is large. Similarly to the parameter  $c$  in the constant slope model, as  $-d/r$  gets large, the plot of the probability curves for  $p_1(\theta, x)$  and  $p_3(\theta, x)$  look something like two separate single logistic regressions in two different regions of dose, as illustrated in Fig. 2.

The parameters  $d$  and  $r$  determine the form of the optimal designs while the values of  $a_2$  and  $b_2$  are similar to location and scale parameters. For convenience, another parameterization for the general model is therefore used:  $\theta = (a_2, b_2, d, r)$ . Similarly to the model with  $b_1 = b_2$  locally D-optimal designs have no closed form expression. They have 2, 3, or 4 design points. Table 1 shows the locally D-optimal designs for  $a_2 = 0, b_2 = 1$  and different values of  $d$  and  $r$ . This can be heuristically motivated using a similar argument to the one for the constant slope model. For the trinomial

continuation-ratio model the value of  $x$  at which the probability of adverse reaction  $p_3(\theta, x)$  is  $\frac{1}{2}$  is  $x = -a_1/b_1 = -(a_2 + d/r)/b_2$  and the value of  $x$  at which the conditional probability of efficacy given no adverse reaction is  $\frac{1}{2}$  is  $x = -a_2/b_2$ . As observed in Fig. 2, when the ratio  $-d/r$  is large, these two related response models center in different regions of the design space. The design putting equal weight at each of  $-(a_2 \pm 1.543)/b_2$  and  $-(a_2 + d/r \pm 1.543/r)/b_2$  is then a good candidate for being almost optimal under local D-optimality and has been shown<sup>14</sup> to be limiting locally D-optimal. A closed form expression for limiting Bayesian D-optimal designs for the simple prior distributions described in Sec. 3.1 is provided in Ref. 14.

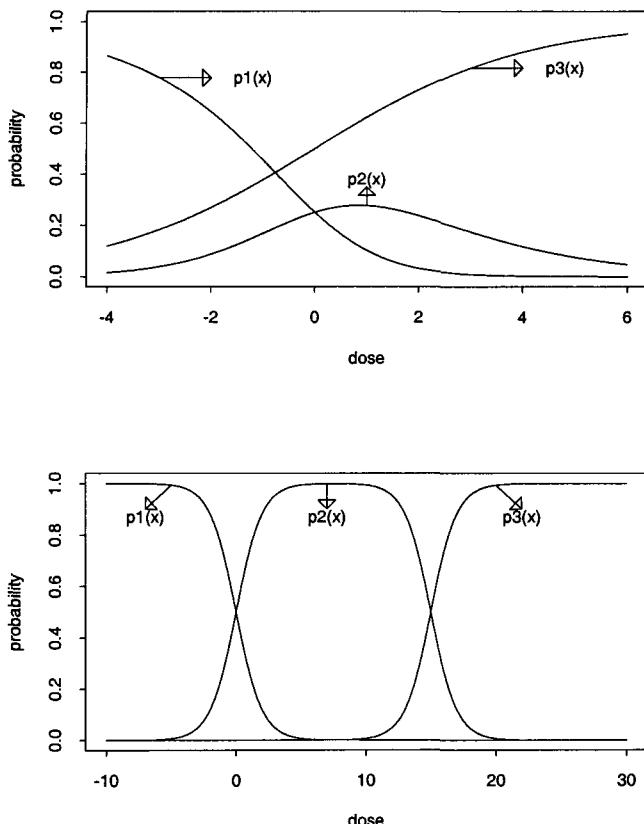


Fig. 2. Plots of the probabilities  $(p_1(\theta, x), p_2(\theta, x), p_3(\theta, x))$  for fixed  $\theta$ , versus  $x$  ( $x$  = dose): top  $(a_2, b_2, d, r) = (0, 1, 0, 0.5)$ , and bottom  $(a_2, b_2, d, r) = (0, 1, -15, 1)$ .

Table 1. Locally D-optimal designs for different  $d$  and  $r$  values.

	$d = 0$		$d = 1$		$d = 3$	
	dose	weight	dose	weight	dose	weight
$r = 0.5$	-2.59	0.50	-3.54	0.50	-8.34	0.25
	1.26	0.50	0.60	0.50	-3.57	0.38
$r = 1$	-2.04	0.50	-2.67	0.50	-4.14	0.50
	0.77	0.50	0.19	0.50	0.91	0.50
$r = 2$	-1.40	0.50	-1.78	0.50	-2.62	0.50
	0.37	0.50	-0.03	0.50	-0.86	0.50
$r = 3$	-2.48	0.21	-2.65	0.19	-3.03	0.14
	-0.51	0.42	-0.93	0.41	-1.68	0.41
	0.37	0.37	0.03	0.40	-0.64	0.45

The criterion of c-optimality for estimating  $x_{max}$  is not straightforward to implement in this more general model as there is no closed form expression for  $x_{max}$ . The value of  $x_{max}$  is the solution to a nonlinear equation. The probability  $p_2(\theta, x)$  is maximized at dose  $x$  if, and only if

$$g(\theta, x) = b_2(1 + e^{-a_1 - b_1 x}) - b_1(1 + e^{a_2 + b_2 x}) = 0.$$

The solution exists and is unique because  $b_2(1 + e^{-a_1 - b_1 x})$  is strictly decreasing and  $b_1(1 + e^{a_2 + b_2 x})$  is strictly increasing. Atkinson and Haines<sup>37</sup> provided a method which could be applied to any function defined implicitly as a solution to an equation. Suppose the function of interest,  $x(\theta)$ , is the root of  $g(\theta, x)$ . If the function  $g(\theta, x)$  has continuous first derivatives, and  $x(\theta)$  is continuous, then an application of the implicit function theorem gives:

$$\nabla x(\theta) = \frac{\partial x}{\partial \theta} \Big|_{\theta} = \left[ \frac{\partial g}{\partial x} \Big|_{(\theta, x(\theta))} \right]^{-1} \frac{\partial g}{\partial \theta} \Big|_{(\theta, x(\theta))}.$$

This expression of  $\nabla x(\theta)$  does not require an explicit  $x(\theta)$  to be calculated. An appropriate vector  $c(\theta)$  for the calculation of c-optimal designs for estimating  $x_{max}$  is therefore

$$\nabla x_{max} = \begin{pmatrix} e^{-a_1 - b_1 x_{max}} / [b_1(e^{-a_1 - b_1 x_{max}} + e^{a_2 + b_2 x_{max}})] \\ x_{max} e^{-a_1 - b_1 x_{max}} / [b_1(e^{-a_1 - b_1 x_{max}} + e^{a_2 + b_2 x_{max}})] \\ e^{a_2 + b_2 x_{max}} / [b_2(e^{-a_1 - b_1 x_{max}} + e^{a_2 + b_2 x_{max}})] \\ x_{max} e^{a_2 + b_2 x_{max}} / [b_2(e^{-a_1 - b_1 x_{max}} + e^{a_2 + b_2 x_{max}})] \end{pmatrix}$$

where  $\theta = (a_1, b_1, a_2, b_2)$ ,  $x(\theta) = x_{max}$ ,  $g(\theta, x)$  is as above. For example, if the value of  $\theta$  is  $(-3.3, 0.5, 3.4, 1)$ , the corresponding  $x_{max}$  is 0.4104, and

the c-optimal design puts mass (0.0012, 0.8003, 0.1985) at (-5.67, -0.64, 4.84) respectively.

### 3.3. The Efficiency of Limiting Optimal Designs

The efficiency of a design  $\eta$  is defined to be the percentage of sample size required for an experiment using the optimal design to reach the same value of the criterion as an experiment using the design  $\eta$  with sample size one. The limiting D-optimal designs in Section 3.1 for the constant slope model were shown<sup>15</sup> to be very efficient, even for designs early in the sequence. All efficiencies are higher than 95%, even for  $c = 0$ , or close to 0. The limiting locally c-optimal designs given in Section 3.1, are quite efficient for moderate to large  $c$  (for  $c > 2$  the efficiency is at least 75%, and at least 95% for  $c > 4$ ). The limiting locally D-optimal designs for the general continuation-ratio model given in Section 3.2, with  $b_1$  not necessarily equal to  $b_2$  are, however, not very efficient. Other limiting c-optimal or D-optimal designs that have been found for this model and for similar sequences of prior distributions are also inefficient. In some cases, the efficiency is as little as 60% even for quite large values of  $-d/r$ .<sup>39</sup>

## 4. When Can Constant Slopes be Assumed?

The robustness of the constant slope assumption is explored with an objective of determining whether the simpler 3 parameter model can be used if the slopes are not equal, but “close”. A guideline for how close is close enough is determined. Designs are investigated by:

- (1) when the slopes  $b_1$  and  $b_2$  are not equal, examining the efficiency of the optimal design for the model where they are equal.
- (2) when the slopes  $b_1$  and  $b_2$  are not equal, calculating the values to which the MLEs under the model where they are equal converge, and examining how much the fitted probability curves differ from the true probability curves (using the result from Huber<sup>38</sup>). In addition, when the slopes are not equal, calculating the MLE under the constant slope assumption for the target dose  $x_{max}$ . The difference between the asymptotic limit of the estimator and its true value is examined graphically.
- (3) when the slopes  $b_1$  and  $b_2$  are not equal, and the sample size is  $n = 20$ , the sampling distribution of the MLE of  $x_{max}$  under the assumption that the slopes are equal is examined by simulation. That is data were generated assuming unequal slopes, but  $x_{max}$  was estimated assuming that the slopes were equal. The error in estimating  $x_{max}$  was examined.

#### 4.1. Efficiency of Optimal Designs

Let  $\tilde{\eta}^*$  be the locally D-optimal design for the constant slope model with a parameter vector  $\theta = (c, a_1, b_1)$  and  $\eta^*$  be the locally D-optimal design of the general continuation-ratio model with a parameter vector having the same values of  $a_1, b_1$ , and  $c$ , but  $b_2$  not necessarily equal to  $b_1$ . It can be shown algebraically that the efficiency of  $\tilde{\eta}^*$  to  $\eta^*$  under the general model only depends on the value of  $b_2$  through  $r$ .<sup>39</sup> Hence, without loss of generality,  $b_2 = 1$  is assumed for the rest of this subsection and the efficiency for different values of  $r$  are examined.

Efficiency plots of the efficiency of  $\tilde{\eta}^*$  against  $d$  were drawn for different values of  $a_2$  and  $r$ . Recall that the value of  $c$  is typically large and positive in practice. The lower bound of  $c$  is taken to be 3 for these plots and thus the upper bound of  $d$ , ( $d = (1 - r)a_2 - c$ ) in these plots was  $(1 - r)a_2 - 3$ . These plots were all quite similar and the plot for  $a_2 = -1$  is shown in Fig. 3 for illustration. The efficiencies are all quite high when the ratio of slopes,  $r$ , is close to 1. Numerical investigation indicates that, over this range of parameter values, if  $0.9 \leq r \leq 1.1$ , that is  $b_1$  within 10% of  $b_2$ , the efficiency is at least 88%. As seen in Fig. 3, the efficiency is monotone when  $d$  is negative with a large magnitude but is not monotone elsewhere.

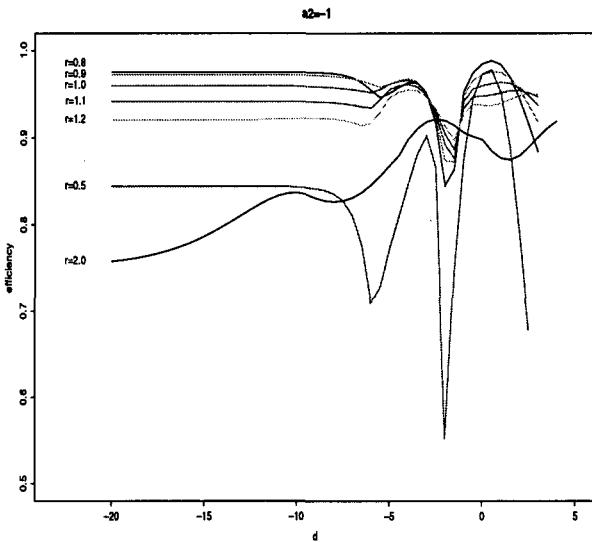


Fig. 3. Efficiency plot for  $a_2 = -1$ : efficiency vs.  $d$ .

#### 4.2. Convergence of MLE's

Are the fitted probability curves, using the constant slope model, close to the true curves asymptotically when the slopes are actually not equal? Huber<sup>38</sup> showed that if a distribution does not lie within a specified parametric class but the MLE is calculated assuming it does, then, as the sample size increases, the MLE approaches the parameter value corresponding to the distribution in the parametric family closest to the true distribution in terms of the directed Kullback divergence. In the case considered here the parametric class assumes that  $b_1 = b_2$ , but  $b_1 \neq b_2$  holds.

**Definition.** Given any two distributions  $Q$  and  $P$  with densities  $q$  and  $p$  with respect to some measure  $\mu$ , the Kullback divergence between  $Q$  and  $P$  is  $K(Q, P) = \int q \log(q/p)d\mu$ .

In this case, the true distribution of the observations depends on the design, the values to which the MLE's, under the parametric class which does not include the true model, approach will also depend on the design. Given a design  $\eta$ , putting weight  $m_i$  at  $x_i$ ,  $i = 1, 2, \dots, k$ , let  $n_i, i = 1, 2, \dots, k$  be the smallest positive integers such that  $n_1 : n_2 : \dots : n_k = m_1 : m_2 : \dots : m_k$ , let  $P_i$  be the true probability vector at  $x_i$ :  $(p_1(\theta, x_i), p_2(\theta, x_i), p_3(\theta, x_i))^T$  and let  $\tilde{P}_i$  be the modeled probability vector at  $x_i$  which assumes  $b_1 = b_2$ .

Consider an observation

$$Y_\eta = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{pmatrix}$$

where  $y_i = (y_{i1}, y_{i2}, y_{i3})^T$  follows a trinomial distribution with parameters  $n_i$  and  $P_i$ , denoted as  $Q_{x_i, n_i}$ , and  $y_1, y_2, \dots, y_k$  are independent. Hence  $Y_\eta$  is distributed as the product distribution  $Q = Q_{x_1, n_1} \times Q_{x_2, n_2} \times \dots \times Q_{x_k, n_k}$ . This true distribution  $Q$ , however, need not lie in the modeled distribution class  $\tilde{P}$ , which is  $\{P_\theta = P_{x_1, n_1} \times P_{x_2, n_2} \times \dots \times P_{x_k, n_k}\}$ , where  $P_{x_i, n_i}$  is the trinomial distribution with parameters  $n_i$  and  $\tilde{P}_i$ ,  $i = 1, 2, \dots, k$ . The question becomes, if  $b_1$  is actually not equal to  $b_2$ , and  $Q$  does therefore not lie in the class  $\tilde{P}$ , then, as the sample size goes to  $\infty$ , to what value does  $\hat{b}_1$ , the MLE of  $b_1$  approach? Similarly, to what values do  $\hat{a}_1$  and  $\hat{a}_2$  approach?

Let  $\mu_{n_i}$  be the measure assigning measure one at every  $(t_1, t_2, t_3)^T$  where  $t_1, t_2, t_3$  are non-negative integers and sum to  $n_i$ , and measure zero otherwise. Define the measure  $\mu$  as the product measure  $\mu_{n_1} \times \mu_{n_2} \times \dots \times \mu_{n_k}$ .

For convenience, write  $\tilde{P}_i$  as  $(\tilde{p}_{i1}, \tilde{p}_{i2}, \tilde{p}_{i3})^T$  and  $P_i$  as  $(p_{i1}, p_{i2}, p_{i3})^T$ . Then with respect to this measure  $\mu$  the density of  $Q$ ,  $q$ , and the density of  $P$ ,  $p$ , are:

$$q(Y_\eta) = \prod_{i=1}^k \prod_{j=1}^3 \frac{n_i!}{y_{i1}! y_{i2}! y_{i3}!} p_{ij}^{y_{ij}}$$

$$p(Y_\eta) = \prod_{i=1}^k \prod_{j=1}^3 \frac{n_i!}{y_{i1}! y_{i2}! y_{i3}!} \tilde{p}_{ij}^{y_{ij}}.$$

Hence

$$\begin{aligned} K(Q, P) &= \int \log \frac{q}{p} q d\mu \\ &= \int \left[ \sum_{i=1}^k \sum_{j=1}^3 y_{ij} (\log p_{ij} - \log \tilde{p}_{ij}) \right] q d\mu \\ &= \sum_{i=1}^k \sum_{j=1}^3 n_i p_{ij} \log p_{ij} - \sum_{i=1}^k \sum_{j=1}^3 n_i p_{ij} \log \tilde{p}_{ij}. \end{aligned}$$

For every given true  $Q$ ,  $\sum_{i=1}^k \sum_{j=1}^3 n_i p_{ij} \log p_{ij}$  is a constant, so minimizing  $K(Q, P)$  is equivalent to maximizing  $\sum_{i=1}^k \sum_{j=1}^3 n_i p_{ij} \log \tilde{p}_{ij}$ , denoted as  $f_\eta$ . Denote the values of  $a_1, a_2$ , and  $b_1$  which maximize  $f_\eta$  as  $a_1^*, a_2^*$ , and  $b_1^*$ . Then using the design  $\eta$  and the constant slope model, the MLE  $\hat{\theta} = (\hat{a}_1, \hat{b}_1, \hat{a}_2)$  will go to  $(a_1^*, b_1^*, a_2^*)$  as the sample size  $n$  goes to  $\infty$  by Huber's Theorem. Values of  $(a_1^*, b_1^*, a_2^*)$  can be found numerically.

Suppose  $b_1 = b_2$  is assumed and the D-optimal (or c-optimal) designs for the constant slope model are used. Let  $a_1, a_2$ , and  $r$  be fixed. It can be shown that, as  $n$  increases, in the constant slope model with  $b_1 = 1$ , if the MLE  $\hat{\theta} = (\hat{a}_1, \hat{b}_1, \hat{a}_2)$  goes to  $\theta^* = (a_1^*, b_1^*, a_2^*)$  then the MLE with  $b_1 = b$  converges to  $(a_1^*, b_1^*, a_2^*)$ . Without loss of generality, therefore,  $b_1 = 1$  is assumed in later parts of this subsection.

First suppose the D-optimal design for the constant slope model is used. The true probability curves and the asymptotic fitted probability curves were drawn on the same plot for 208 parameter values. The 208 values were all combinations of:  $c = -5, 0, 5, 10$ ,  $a_1 = -1, -0.5, 0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5$ ,  $b_1 = 1$ , and  $b_2 = 0.9, 0.95, 1.05, 1.1$ . In the plots where  $b_1$  is less than 10% different from  $b_2$ , the asymptotic fitted probability curves can barely be distinguished from the true probability curves. Figure 4 is an example. It is therefore reasonable to conclude that if  $b_1$  is less than 10%

different from  $b_2$  and the D-optimal designs of the constant slope model are used then the fitted probability curves are very close to the true probability curves asymptotically.

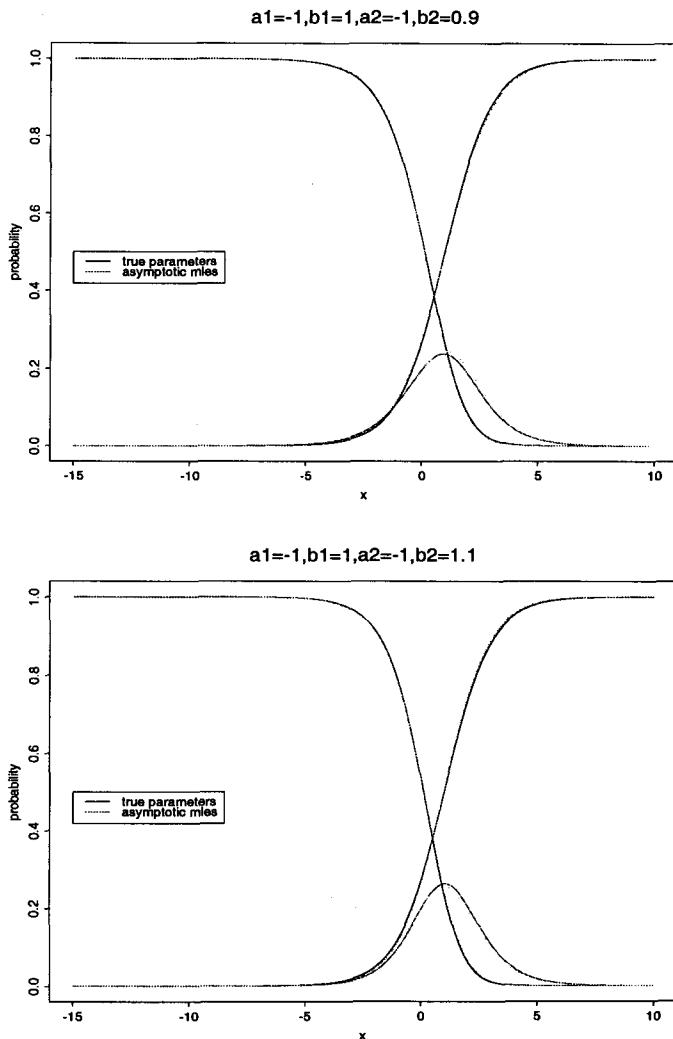


Fig. 4. Probability plots:  $a_1 = -1, b_1 = 1, a_2 = -1, b_2 = 0.9, 1.1$ . The solid lines are from the true parameter values and the dashed lines are from the asymptotic limit of the MLEs. The lines are almost indistinguishable.

Now consider the case where the dose  $x_{max}$  is of interest and the c-optimal design for the constant slope model is used. Denote the MLE of  $x_{max}$  in the constant slope model as  $\hat{x}_{max}$ , that is  $\hat{x}_{max} = -(\hat{a}_1 + \hat{a}_2)/(2\hat{b}_1)$ , and the asymptotic limit of  $\hat{x}_{max}$  is  $x_{max}^* = -(a_1^* + a_2^*)/(2b_1^*)$ . The asymptotic error of the estimator  $\hat{x}_{max}$  is defined as  $asyerror = x_{max}^* - x_{max}$ . Figures 5, 6, and 7 illustrate how  $asyerror$  changes as  $b_2$  changes from 0.5, 0.7, 0.9, 0.95, 1.05, 1.1, 1.5, and 2 for different  $c$  and  $a_1$  values.

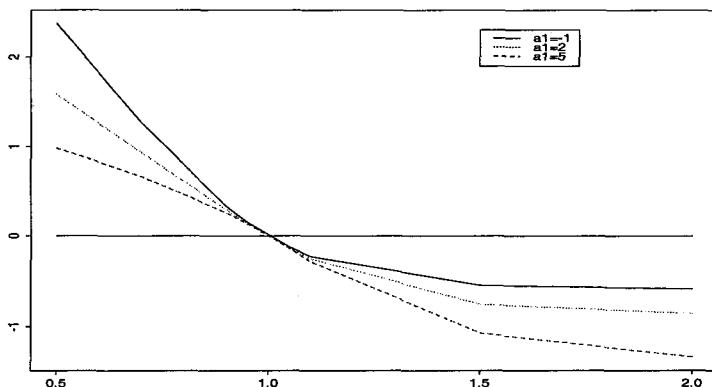


Fig. 5. Plots of  $asyerror$  vs.  $b_2$  for  $c = -3$ . The solid lines are for  $a_1 = -1$ , the dashed lines are for  $a_1 = 2$ , and the double dashed lines are for  $a_1 = 5$ .

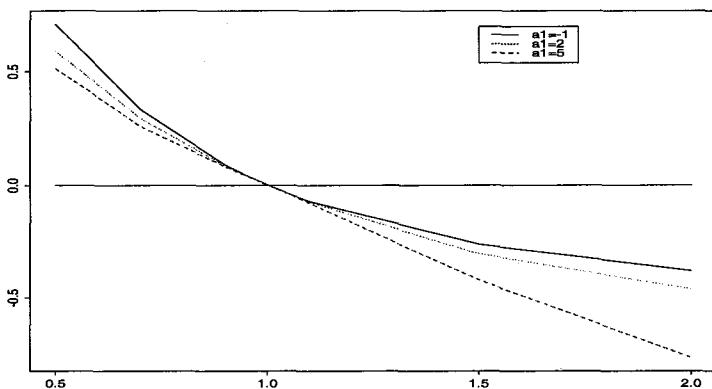


Fig. 6. Plots of  $asyerror$  vs.  $b_2$  for  $c = 1$ . The solid lines are for  $a_1 = -1$ , the dashed lines are for  $a_1 = 2$ , and the double dashed lines are for  $a_1 = 5$ .

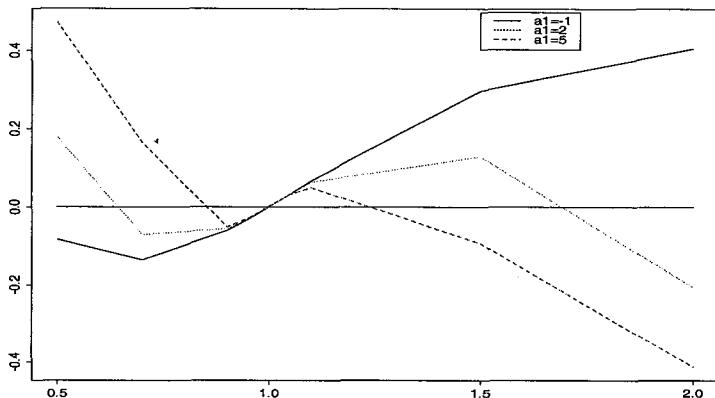


Fig. 7. Plots of  $|asyerror|$  vs.  $b_2$  for  $c = 5$ . The solid lines are for  $a_1 = -1$ , the dashed lines are for  $a_1 = 2$ , and the double dashed lines are for  $a_1 = 5$ .

Table 2. The target dose  $x_{max}$  at different  $c$  values.

$(a_1, b_1, a_2, b_2)$	$x_{max}$	$(a_1, b_1, a_2, b_2)$	$x_{max}$	$(a_1, b_1, a_2, b_2)$	$x_{max}$
$(-1, 1, -4, 1.1)$	2.61	$(-1, 1, 0, 1.1)$	0.55	$(-1, 1, 4, 1.1)$	-1.37
$(-1, 1, -4, 0.9)$	2.32	$(-1, 1, 0, 0.9)$	0.44	$(-1, 1, 4, 0.9)$	-1.63
$(2, 1, -1, 1.1)$	-0.20	$(2, 1, 3, 1.1)$	-2.28	$(2, 1, 7, 1.1)$	-4.21
$(2, 1, -1, 0.9)$	-0.79	$(2, 1, 3, 0.9)$	-2.70	$(2, 1, 7, 0.9)$	-4.77
$(5, 1, 2, 1.1)$	-3.01	$(5, 1, 6, 1.1)$	-5.12	$(5, 1, 10, 1.1)$	-7.05
$(5, 1, 2, 0.9)$	-3.90	$(5, 1, 6, 0.9)$	-5.84	$(5, 1, 10, 0.9)$	-7.91
$c = -3$		$c = 1$		$c = 5$	

Table 2 lists values of  $x_{max}$  for  $b_2 = 0.9$  and  $b_2 = 1.1$  in the plots. From Figs. 5–7, it can be seen that  $|asyerror|$  decreases as  $c$  increases. In addition,  $|asyerror|$  is small in the range  $0.9 \leq b_2 \leq 1.1$ ; it is less than 0.3 for  $c = -3$ , less than 0.1 for  $c = 1$ , and less than 0.07 for  $c = 5$ . Small errors might, however, be clinically significant because dose  $x$  is typically in log units. It therefore may be appropriate to transform  $x$  back to regular units for interpreting the error. An example for  $c = -3$  illustrates this point. Suppose that  $(a_1, b_1, a_2, b_2) = (-1, 1, -4, 1.1)$ , then  $x_{max}$  is 2.61 (Table 2). Because from Fig. 5,  $-0.3 \leq asyerror \leq 0.3$ , the (asymptotic) estimate of  $x_{max}$ ,  $x_{max}^*$ , is between 2.31 and 2.91 in log units; in regular units,  $x_{max}$  is 13.60 and  $x_{max}^*$  is between 10.07 and 18.36. In contrast, if  $(a_1, b_1, a_2, b_2) = (-1, 1, 4, 1.1)$ , then  $x_{max} = -1.37$ . Figure 7 gives  $|asyerror| \leq 0.07$  which means that  $x_{max}^*$  is between -1.44 and -1.30 in log units and  $x_{max} = 0.25$ .

and  $x_{max}^*$  is between 0.24 and 0.27 in regular units. Because of the concavity of the log transformation and the decreasing property of the magnitude of *asyerror* (as  $c$  increases), it is reasonable to conclude that these c-optimal designs for the constant slope model provide a good estimate of  $x_{max}$  when  $b_2$  is less than 10% different from  $b_1$ , and especially when  $c$  is positive.

#### 4.3. The Performance of Optimal Designs for a Small Sample

The c-optimal designs use asymptotic variance approximations which are not always accurate, especially in small samples. The variance of the normal distribution which approximates the distribution of the MLE is not necessarily related to the actual variance of the MLE. In addition Sun *et al*<sup>40</sup> provide an illuminating example of a logistic regression model where there is considerable variability in the posterior variance. One important question, is whether the estimates of  $x_{max}$  from optimal designs are close to the true value. In addition, in small samples, it can also be examined how well the estimates using c-optimal or D-optimal designs for the constant slope model perform when the slopes are not equal. This section will address these questions and examine estimates of  $x_{max}$  from both c-optimal designs and D-optimal designs for the constant slope model, for a sample size of 20, when the data are not necessarily from a model with constant slopes.

It can be shown<sup>39</sup> that the distribution of  $error = \hat{x}_{max} - x_{max}$  depends on the value of  $b_1$  only through  $r$  and thus  $b_1 = 1$  is assumed. Since a large value of  $a_1$  is not typically expected, only  $a_1 = -1, 0$ , and 2 were explored. If  $r = b_1/b_2$  is not close to one, the constant slope model should not be used, so only  $r = 0.9, 0.95, 1, 1.05$ , and 1.1 were investigated. Recall that if  $r$  is close to 1, the value of  $c = a_2 - a_1$  determines approximately the width of the region of  $x$  values for which the probability of success is close to optimal (see Figs. 1, 2), and so  $c$  values  $-1, 1$ , and 5 were chosen. In the simulations several extreme outlying points were found and so “box only” boxplots of *error* are shown: that is only the boxes, not the whiskers. Figures 8 and 9 show these boxplots for  $c = -1$  using the c-optimal design and D-optimal design respectively. Figures 10 and 11 are for  $c = 1$  and Figures 12 and 13 are for  $c = 5$ . For  $c = -1$  and 1, the medians in the boxplots increase with  $r$ . This can be explained: let  $x'_{max}$  denote  $-(2a_1 + c)/2b_1$ , the target dose based on the constant slope model. Because  $p_2(\theta, x)$  has only one local maximum for a given  $\theta$  and  $\frac{\partial p_2(\theta, x)}{\partial x}|_{x=x'_{max}}$  is approximately proportional

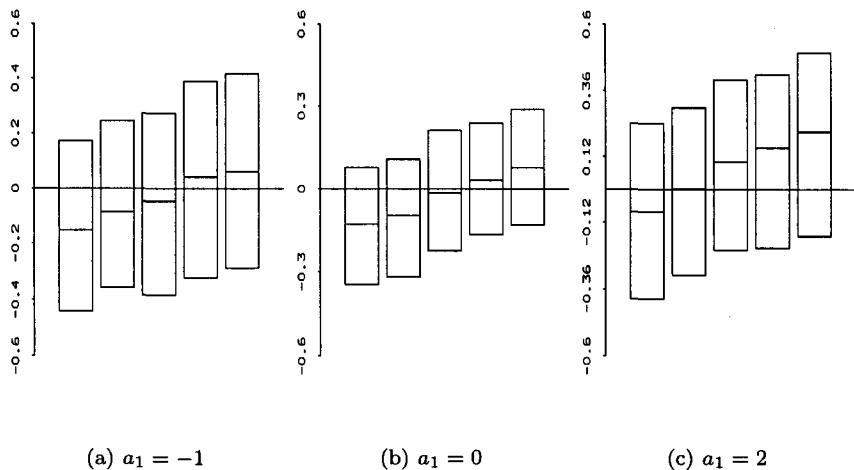


Fig. 8. "Box only" boxplots of error for  $c = -1$  and  $r = 0.9, 0.95, 1, 1.05$ , and  $1.1$  from left to right using c-optimal designs.

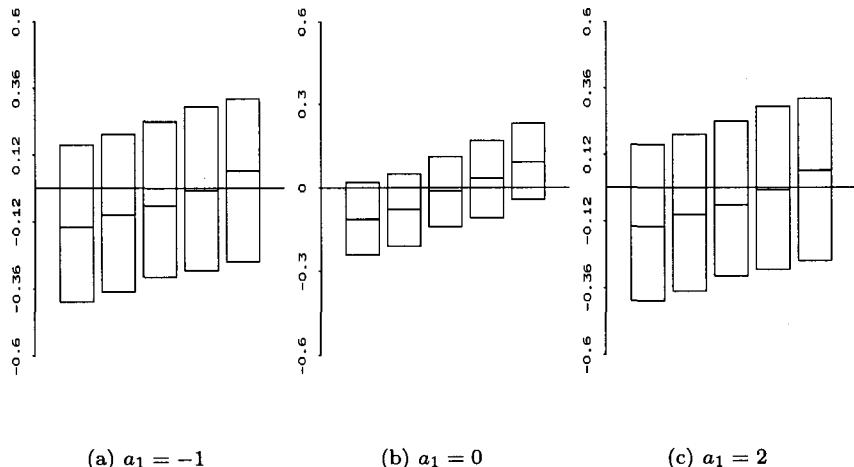


Fig. 9. "Box only" boxplots of error for  $c = -1$  and  $r = 0.9, 0.95, 1, 1.05$ , and  $1.1$  from left to right using D-optimal designs.

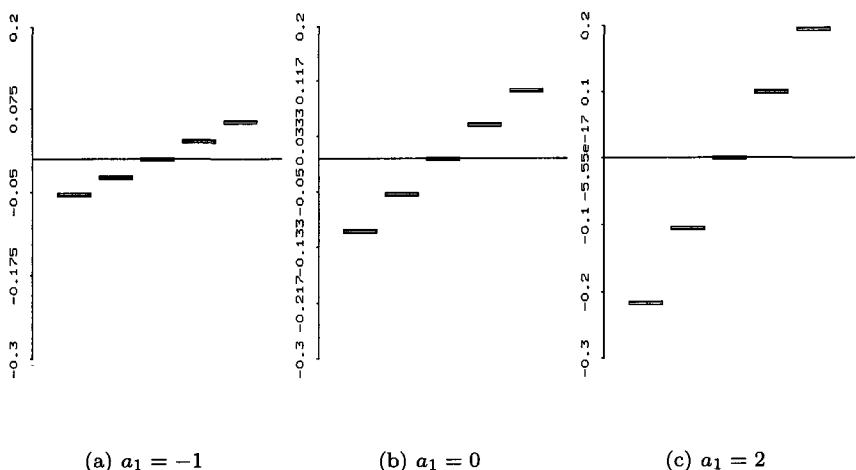


Fig. 10. "Box only" boxplots of error for  $c = 1$  and  $r = 0.9, 0.95, 1, 1.05$ , and  $1.1$  from left to right using  $c$ -optimal designs.

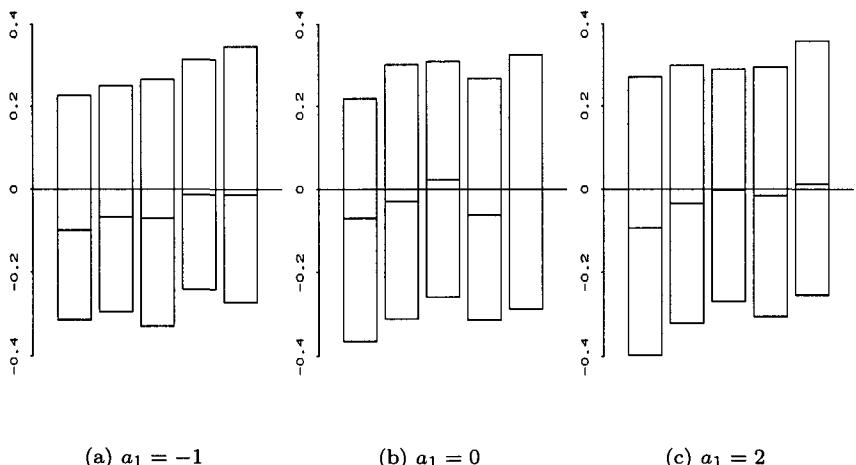


Fig. 11. "Box only" boxplots of error for  $c = 1$  and  $r = 0.9, 0.95, 1, 1.05$ , and  $1.1$  from left to right using D-optimal designs.

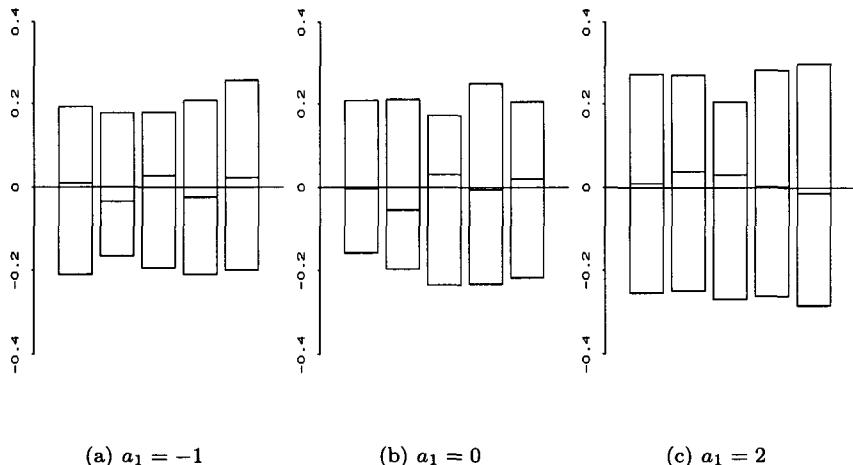


Fig. 12. "Box only" boxplots of error for  $c = 5$  and  $r = 0.9, 0.95, 1, 1.05$ , and  $1.1$  from left to right using c-optimal designs.

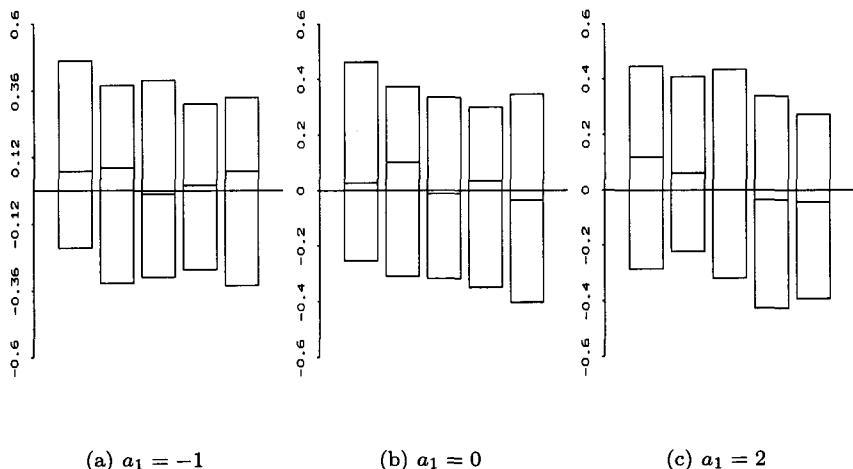


Fig. 13. "Box only" boxplots of error for  $c = 5$  and  $r = 0.9, 0.95, 1, 1.05$ , and  $1.1$  from left to right using D-optimal designs.

to  $(1 - r)$  for small  $|c|$  values and  $r$  close to 1,  $x'_{max}$  is smaller than  $x_{max}$  for  $r < 1$  and larger than  $x_{max}$  for  $r > 1$ . That is,  $x'_{max} - x_{max}$  will increase from negative to positive when  $r$  increases from 0.9 to 1.1 for small  $|c|$ .

The interquartile range (IQR) is quite small and so *error* is, in general, small, although the tails are very long. Note that the range of the IQR of *error* is larger than that of *asyerror*. It is also typically in log units of dose. Hence, to interpret the error, *error* must be transferred back to regular units as discussed in Section 4.2.

In most cases, the IQR of *error* using D-optimal designs was larger than that using c-optimal designs. For  $c = -1$ ,  $a_1 = 0$ , and  $r = 0.9$ , however, the IQR of *error* using D-optimal designs was found to be slightly smaller than that using c-optimal designs.

Overall, results indicate that c-optimal designs perform reasonably for positive  $c$  values, even for a sample size of 20, if  $b_1$  is less than 10% different from  $b_2$ . In some cases the c-optimal designs perform dramatically better than the D-optimal designs: this is seen, for example Figs. 10 and 11 for the case when  $c = 1$  (and note that the scale on the y-axes in Fig. 10 is different to that in Fig. 11). The c-optimal designs are, of course, based on estimating  $x_{max}$ .

#### 4.4. Summary

The results in this section confirm that assuming that the slopes are equal leads to reasonable designs and reasonable estimators if the slopes are not equal but close, and  $c \geq 0$ . The c-optimal designs also perform reasonably for a small sample size of  $n = 20$  if the goal is to estimate  $x_{max}$ . Moreover, a useful guideline for how close is “close” has been found: if  $b_1$  is within 10% of  $b_2$ , this is close enough.

### 5. Discussion

This chapter has briefly reviewed design for multinomial responses to dose. Details of designs for the continuation-ratio model with a trinomial response have been presented. Algebraic expressions for optimal designs are intractable. A concept of a sequence of limiting optimal designs which are optimal in the limit, can be used as approximately optimal designs. Sequences of limiting optimal designs are available in algebraic form for the constant slopes model and have been shown to be efficient.

The robustness of the constant slope assumption has been examined and guidelines provided on when it can be used. Small sample simulations

are also presented in details examining the behaviour of estimators, using optimal designs based on asymptotic approximations. The constant slope assumption is shown to be robust here if the slopes are close and  $a_2 \geq a_1$ , and a guideline has been found for "close". The c-optimal designs for the constant slope model are also shown here to perform reasonably well for a small sample,  $n = 20$ , if the slopes are not equal but close.

One of the advantages of the continuation-ratio model is that the MLE of the parameters can be found easily using statistical software such as SAS and SPSS, regardless of the number of levels of response (for details, see Ref. 33). The continuation-ratio model is practically appealing for nested responses.

Although designs are explored here only for trinomial responses, some properties appear to hold for more general multinomial responses. In numerical examples, for the continuation-ratio model with  $k$  categories of hierarchical response, the locally D-optimal designs have at most  $2(k-1)$  design points. This and other interesting design problems for the continuation-ratio model are waiting to be explored further.

The problem of mixtures of several drugs, and determining an optimal mixture, for either a binary or multinomial response is also a problem waiting to be explored further. There is some guidance from the literature, as described in the Sec. 1. Many diseases require combination drug therapy, but drugs rarely have additive effects and drug interactions are important to understand.

### Acknowledgments

This work was supported in part by a grant from the National Security Agency. We thank the referee for a very thoughtful and thorough review.

### References

1. R. Mead and D. J. Pike, *Biometrics* **31**, 803–851 (1975).
2. K. Chaloner and K. Larntz, *Journal of Statistical Planning and Inference* **21**, 191–208 (1989).
3. B. J. T. Morgan, *Analysis of Quantal Response Data*. Chapman and Hall, London (1992).
4. K. Chaloner and I. Verdinelli, *Statistical Science* **10**, 273–304 (1995).
5. D. M. Smith and M. S. Ridout, *Journal of Statistical Planning and Inference*, to appear (2005).
6. P. F. Thall and K. E. Russell, *Biometrics* **54**, 251–264 (1998).
7. P. McCullagh and J. A. Nelder, *Generalized Linear Models*. 2nd Edition. Chapman & Hall: London (1989).

8. Z. Li, S.D. Durham, and N. Flournoy, in *Adaptive Design*, eds. N. Flournoy and W.F. Rosenberger, IMS Lecture Notes, **25**, 179–196 (1995).
9. H. Rabie and N. Flournoy, in *Moda 7, Advances in Model-oriented Design and Analysis*, eds. A. D. Buccianico, H.L. Suter, H.P. Wynn, Springer-Verlag, 133–142 (2004).
10. G. F. V. Golnek, and P. McCullagh, *J. Roy. Stat. Soc. Ser. B.* **57**, 533–546 (1995).
11. G. F. V. Glonek, *Biometrika*, **83**, 15–28 (1996).
12. S. S. Zocchi and A. C. Atkinson, *Biometrics*, **55**, 437–444.
13. I. Perevozskaya, W. F. Rosenberger, and L. M. Haines, *Canadian Journal of Statistics*, **31**, 225–235 (2003).
14. S. K. Fan, and K. Chaloner, in *mODa6-Advances in Model-Oriented Design and Analysis*, eds. A. C. Atkinson, P. Hackl, and W. G. Muller. Physica-Verlag, 77–86 (2001).
15. S. K. Fan, and K. Chaloner, *Journal of Statistical Planning and Inference*, **126**, 347–360 (2004).
16. M. A. Heise and R. H. Myers, *Biometrics* **52**, 613–624 (1996).
17. J. Hardwick and Q.F. Stout, in *Optimal Design 2000*, eds. A.C. Atkinson, B. Bogacka and A. Zhigljavsky, 195–208 (2001).
18. E. E. Kpamegan and N. Flournoy, in *Optimal Design 2000*, eds. A.C. Atkinson, B. Bogacka and A. Zhigljavsky, 211–223 (2001).
19. R. K. Paul, W. F. Rosenberger, and N. Flournoy, *Statistics in Medicine* **23**, 2483–2495 (2004).
20. P. S. Hewlett and R. L. Plackett, *The Interpretation of Quantal Responses in Biology*, University Park Press, Baltimore, (1979).
21. K. M. Abdelbasit and R. L. Plackett, *Biometrics* **38**, 171–179 (1982).
22. W. H. Carter, Jr. et al, *The American Statistician* **40**, 124–128 (1986).
23. H. D. Unkelbach and T. Wolf, *Statistics in Medicine* **4**, 77–85 (1985).
24. E. M. Laska, M. Meisner and C. Siegel, *Biometrics* **50**, 834–841 (1994).
25. P. I. Kupchak, *Optimal Designs for the Detection of Drug Interaction*. PhD Thesis, University of Toronto, Department of Statistics (2000).
26. R.R. Sitter and B. Torsney, *Statistica Sinica* **5**, 405–419 (1995).
27. B. Torsney and N. Gunduz, in *Optimal Design 2000*, eds. A.C. Atkinson, B. Bogacka and A. Zhigljavsky, Kluwer Academic, 275–285 (2001).
28. Y. Zhu, D. Krewski, and W. H. Ross, *Applied Statistics* **43**, 583–598 (1994).
29. F. Pukelsheim, *Optimal Design of Experiments*. John Wiley & Sons Inc, New York (1993).
30. C. Han and K. Chaloner, *Journal of Statistical Planning and Inference* **115**, 585–601 (2003).
31. J. Bryant and R. Day, *Biometrics* **51**, 1372–1383 (1995).
32. M. R. Conaway and G. R. Petroni, *Biometrics* **51**, 656–664 (1995).
33. A. Agresti, *Categorical Data Analysis*. John Wiley & Sons Inc., New York (1990).
34. P. F. Thall and J. D. Cook, *Biometrics* in press.
35. S. D. Silvey, *Optimal Design*. Chapman and Hall (1980).

36. S.K. Fan and K. Chaloner, *Journal of Statistical Planning and Inference* **113**, 249–257 (2003).
37. A. C. Atkinson and L. M. Haines, *Handbook of Statistics* **13**, 437–475 (1996).
38. P. J. Huber, in *proceedings of the Fifth Berkeley Symposium*. Berkeley: University of California Press, 221–233 (1967).
39. S. K. Fan, *Multivariate Optimal Designs*. PhD Thesis, School of Statistics, University of Minnesota (1999).
40. D. Sun, R. K. Tsutakawa, and W. S. Lu, *Journal of Statistical Planning and Inference* **52**, 289–306 (1996).

## CHAPTER 11

### EVALUATING THE PERFORMANCE OF NON-STANDARD DESIGNS: THE SAN CRISTOBAL DESIGN

Linda M. Haines

*Department of Statistical Sciences,  
University of Cape Town, Rondebosch 7700,  
South Africa  
E-mail: lhaines@stats.uct.ac.za*

The San Cristobal is a little known design that was introduced by Rojas in 1962 for use in experiments in sugar farming. The design was constructed to accommodate the fitting of a quadratic response surface to  $k$  factors in situations in which the levels of the factors, as for example levels of fertilizer, are restricted in real terms to be positive or zero and comprises  $2^k$  factorial points, centre points and axial points all contained within the positive orthant. In this article methods for evaluating the performance of non-standard response surface designs such as the San Cristobal design are reviewed and critically appraised. Criteria based on the precision of the parameter estimates, on the prediction variances and on measures of model misspecification are described and recent developments and new techniques are highlighted. Many of the key features presented in the review are illustrated by means of the San Cristobal design and in addition some new ideas and pointers for future research are given.

#### 1. Introduction

Response surface methodology (RSM) has enjoyed immense development and wide-ranging application since the seminal papers of Box and Wilson<sup>1</sup> and Box and Draper<sup>2</sup> appeared in the 1950's. The fundamentals of the methodology are set out in the books by Myers<sup>3</sup>, Box and Draper<sup>4</sup>, Khuri and Cornell<sup>5</sup> and Myers and Montgomery<sup>6</sup> and in three key review articles, namely those of Hill and Hunter<sup>7</sup>, Myers, Khuri and Carter<sup>8</sup> and Myers, Montgomery, Vining, Borror and Kowalski<sup>9</sup>. A vital component of RSM is the fitting of an appropriate empirical model, usually a first- or a second-

order polynomial model, to observed responses which depend on several explanatory variables in order to examine the nature of the response surface in the vicinity of the optimum operating conditions. There is an enormous array of flexible and robust designs for such response surface modelling which are well-documented in the literature and which are immediately available to the practitioner. These include, *inter alia*, orthogonal designs for first-order polynomial models and central composite, Box-Behnken and small composite designs for second-order polynomials. At the same time there is considerable and continuing interest in devising new designs and new classes of designs to accommodate the specific problems and needs that arise within the RSM framework. For example, Morris<sup>10</sup> has introduced a novel class of designs termed augmented pairs designs for fitting second-order polynomial response surfaces which rival the more traditional  $3^k$  factorial and central composite designs, Mee<sup>11</sup> has presented an interesting class of non-central composite designs suitable for augmenting  $2^k$  factorials and Zahran, Anderson-Cook, Myers and Smith<sup>12</sup> have constructed designs which accommodate certain non-standard restrictions on the explanatory variables. In addition the experimenter may well generate new and useful designs over non-standard regions which are optimal with respect to particular criteria by using statistical packages and other specialist software<sup>6</sup>.

The aim of the present article is to review methods for evaluating designs which are new and which are considered to be, in some sense, appropriate within the response surface modelling context. Attention is restricted to experiments involving single response variables, to explanatory variables which act independently of each other or in other words which do not relate to the components of a mixture, and to empirical models which are first- or second-order polynomials in the explanatory variables. It is natural in evaluating a new design to pose questions as to the basic properties of the design and how these can be most effectively calculated and, more importantly, as to which criteria should be invoked in order to appraise the performance of the design and which established designs should be used to benchmark this performance. For a particular design the answers to such questions lie very broadly in the form of the model matrix, in an appraisal of the precision of the parameter estimates, in the nature of the prediction variances and in the manner in which model misspecification is accommodated. These issues are dealt with in the standard references on RSM, as for example those cited earlier. However while it is recognized that designs which behave well over a range of criteria should be sought<sup>13,14,15</sup>, attention has nevertheless focussed to a greater extent on individual criteria

and on the optimization of those criteria. The emphasis in the present article is on the development of criteria suitable for the evaluation and comparison of designs and optimal and near-optimal designs are introduced strictly within that context.

The chapter is structured as follows. Some preliminary ideas are introduced in Section 2 and in addition a composite design, termed the San Cristobal design, which is not well-known in the Statistics literature is presented and used for illustrative purposes in the remainder of the article. In Section 3 the basic properties of designs suitable for response surface modelling are reviewed and in Section 4 the appropriateness or otherwise of using the precision of parameter estimates in evaluating designs is considered. Section 5 is devoted to examining the variances of the predictions of the responses over the region of experimentation using measures of rotatability and “single number” criteria such as the integrated or average variance and, in addition, using techniques such as variance dispersion graphs which summarize the prediction variances over the design region more comprehensively. Model misspecification is a key issue in RSM and is examined within the context of design comparison in Section 6, specifically in relation to prediction bias, the detection of lack of fit, model checking and model robustness. Further robustness issues are also considered briefly in that section. Finally some broad conclusions and pointers for future research are given in Section 7.

## 2. Preliminaries and an Example

### 2.1. Problem Setting

Suppose that an experimenter is interested in investigating the response of a single variable to  $k$  explanatory variables and elects to approximate the associated response surface with a first-order or a second-order polynomial model and to use a particular design comprising  $N$  settings of the  $k$  explanatory variables in fitting the chosen model to the data. Then the design of interest can be summarized by a design matrix of the form  $D = [x_1, \dots, x_k]$ , where the column vector  $x_i$  has elements  $x_{ui}$  corresponding to the  $u$ th setting of the  $i$ th explanatory variable for  $u = 1, \dots, N$ ,  $i = 1, \dots, k$ . Note that the explanatory variables are taken to be appropriately coded and to belong to a specified region of experimentation  $R$ . Note also that for convenience the region  $R$ , also termed the design space, is taken to coincide with the region of interest and of operability. The observed responses can

then be modelled in terms of a first-order polynomial as

$$y_u = \beta_0 + \sum_{i=1}^k \beta_i x_{ui} + \epsilon_u, \quad u = 1, \dots, N \quad (1)$$

or in terms of a second-order polynomial as

$$y_u = \beta_0 + \sum_{i=1}^k \beta_i x_{ui} + \sum_{i=1}^k \beta_{ii} x_{ui}^2 + \sum_{i=1}^k \sum_{j=1, i < j}^k \beta_{ij} x_{ui} x_{uj} + \epsilon_u, \quad u = 1, \dots, N \quad (2)$$

where  $y_u$  is the response corresponding to the  $u$ th setting,  $\beta_0, \beta_i, \beta_{ii}$  and  $\beta_{ij}$  are unknown parameters corresponding to the intercept, linear, quadratic and interaction or bilinear terms respectively, and  $\epsilon_u$  is an error term assumed to be independently distributed with mean zero and variance  $\sigma^2$ . The models (1) and (2) can be represented succinctly in matrix form as

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (3)$$

where  $\mathbf{y}$  corresponds to the vector of  $N$  responses,  $\boldsymbol{\epsilon}$  to the vector of  $N$  error terms and  $X$  is the model matrix. Note that for the first-order polynomial  $X$  is an  $N \times (k+1)$  matrix equal to  $[1 \ D]$ , where  $1$  represents a vector of ones, and that for the second-order polynomial,  $X$  is an  $N \times \frac{(k+1)(k+2)}{2}$  matrix with the order of the columns mirroring the order of terms specified in (2).

Suppose now that the experimenter wishes to evaluate the performance of the design of interest over a range of criteria. Then it is necessary to select designs that are appropriate for benchmarking this performance. The choice of such designs is wide and the following guidelines may prove useful. First it is important, if not essential, to include well-established designs such as the  $2^k$  factorial designs for first-order polynomial models and the  $3^k$  designs, the central composite designs and Box-Behnken designs for second-order polynomials in any comparison. Second there may be special designs that are appropriate to the needs of the experimenter, such as small composite, Notz and Hoke designs within the context of near-saturated designs<sup>6</sup>, and these should be considered. Third the design itself may belong to a class of designs for which certain parameters, such as the number of centre points, can be varied. In this case a comparison within the class can be made and designs from the class which are optimal for a particular criterion can be constructed and compared. Fourth, it is often possible to obtain designs which are based on exactly  $N$  points and which optimize specific criteria.

Such designs have, almost invariably, to be constructed using computer-intensive methods and for that reason are referred to in the RSM literature as computer-generated designs. There is a wealth of computer packages, including SAS, JMP, Minitab, Design-Ease and ACED, which provide options for finding exact  $N$ -point designs for various criteria and which are readily available to the practitioner. Such exact designs play a particularly valuable role in comparative studies in that they can be used as the basis for efficiency calculations. However the designs are only optimal for a specific criterion and may well be less useful as designs robust to a range of different criteria<sup>9,15</sup>.

## 2.2. An Example: The San Cristobal Design

The dependence of the yield of sugar cane on varying amounts of fertilizers containing nutrients such as nitrogen, phosphorus and potassium can be modelled empirically using a second-order polynomial<sup>16,17</sup>. A design appropriate to this model setting was formulated and developed in a series of conference presentations and notes by Rojas<sup>16,18,19</sup> and is termed the San Cristobal design after the old sugar milling company of that name in Mexico. The design is a variant of the central composite design but with the important difference that it includes a control, that is a setting corresponding to no application of fertilizer. Specifically the San Cristobal design in  $k$  explanatory variables comprises

- (i) a full  $2^k$  factorial or a  $2^{k-p}$  fractional factorial design repeated  $r$  times with the explanatory variables taking the coded values 0 or 2 giving a total of  $f = r 2^k$  or  $r 2^{k-p}$  such points,
- (ii)  $n_0$  center points of the form  $(1, 1, \dots, 1)$ ,
- (iii)  $c$  axial points of the form  $(1, \dots, 1-\alpha, \dots, 1)$  for each of the  $k$  variables, where  $0 < \alpha \leq 1$ , giving a total of  $c k$  such points,
- (iv)  $k$  star or axial points of the form  $(1, \dots, c\alpha + 1, \dots, 1)$ , one for each of the variables.

Note that the design has a total of  $N = f + n_0 + k(c + 1)$  points, that the constraint  $\alpha \leq 1$  must hold in order to ensure non-negative applications of fertilizer and that otherwise if  $\alpha > 1$  with  $c = 1$  the design reduces to a central composite design. Note also that this specification is the most general for the San Cristobal design and that Rojas<sup>16</sup> initially proposed a somewhat simpler form. In examining the class of San Cristobal designs, Rojas<sup>18</sup> focussed on choosing values for the parameters  $r, n_0, c$  and  $\alpha$  in

order to achieve orthogonality in the sense of Box and Hunter<sup>20</sup>, while Mapham<sup>17</sup> introduced regions of interest defined by the intersection of a  $k$ -dimensional cube or hypersphere with the positive orthant, namely  $[0, \infty)^k$ . In addition certain properties of the San Cristobal design, in particular those relating to the prediction variances, were examined by Rojas<sup>19</sup> and in more detail by Mapham<sup>17</sup> and their results are summarized briefly in Dicks<sup>21</sup>.

A set of designs from the class of San Cristobal designs for two explanatory variables is introduced here for illustrative purposes. The selected designs are constrained to lie in a region of experimentation in 2-dimensional space defined by the intersection of the disk of radius  $\sqrt{8}$  with the positive quadrant, denoted  $R_0$ , and have the form specified above with a single replication of the  $2^2$  factorial points,  $n_0$  centre points, and  $\alpha$  chosen so that the star points  $(c\alpha + 1, 1)$  and  $(1, c\alpha + 1)$  lie on the boundary of the region of experimentation, i.e.  $\alpha = \frac{\sqrt{7} - 1}{c}$ . In addition, the design with one centre point,  $c = 2$  and  $\alpha = 0.8229$  and thus with axial points  $(0.1771, 1)$  and  $(0.1771, 1)$  repeated twice and with the single star points  $(2.6458, 1)$  and  $(1, 2.6458)$ , is of particular interest. This design, denoted  $SC2$ , is summarized together with the design space  $R_0$  in Figure 1(a) and its performance over a range of criteria is compared at appropriate points in the text with that of the  $3^2$  factorial design and of related San Cristobal and optimal designs.

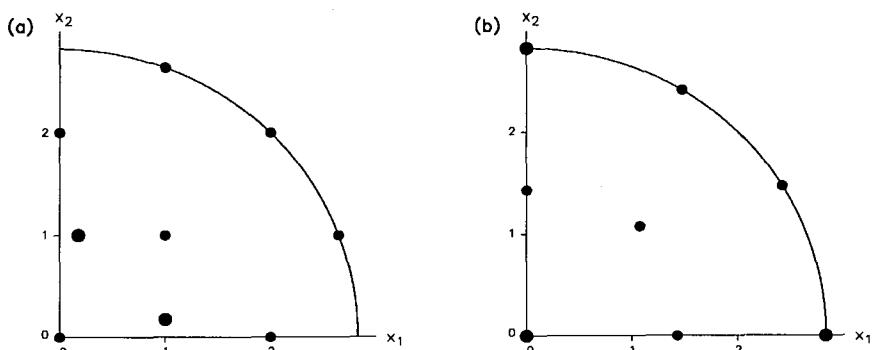


Fig. 1. Designs: (a) The San Cristobal design  $SC2$  and (b) the 11-point  $D$ -optimal design in the region of experimentation  $R_0$ .

### 3. Basic Properties

In evaluating a given design it is important to first identify certain of its basic properties, in particular the number of runs, the number of degrees of freedom available for the estimation of pure error and a skeleton analysis of variance table. Furthermore it is sensible to specify the form of the moment matrix,  $\frac{1}{N} X^T X$ , since this matrix is a basic building block in the computation of a wide range of information-based criteria. Note that the elements of the moment matrix are the moments of the design, written in standard notation as

$$[1^{\delta_1} 2^{\delta_2} \dots k^{\delta_k}] = \frac{1}{N} \sum_{u=1}^N x_{u1}^{\delta_1} x_{u2}^{\delta_2} \dots x_{uk}^{\delta_k}$$

where  $\delta = \sum_{i=1}^k \delta_i$  defines the order. Note also that it is helpful in later calculations to evaluate the moments, and hence the moment matrix and its inverse, in an algebraically explicit form.

It is of some interest to identify orthogonal designs. Such designs have diagonal moment matrices and, as a consequence, possess particularly attractive properties. For first-order polynomial models, orthogonal designs are those designs for which the first-order and mixed second-order moments are equal to zero. For second-order polynomials, the orthogonality or otherwise of a design is less clear. Thus when the model is expressed in the form (2), the moment matrix of any associated design cannot be diagonalized. However it is possible, following Box and Hunter<sup>20</sup> and more recently DuMouchel and Jones<sup>22</sup>, to construct a moment matrix for a design for a second-order polynomial model which is diagonal by coding the explanatory variables from  $x_i$  to  $\tilde{x}_i$  so that the first- and second-order moments satisfy  $[i] = 0$  and  $[i^2] = [ii] = 1$ , by replacing the quadratic terms  $x_i^2$  in the model formulation (2) with the new variables

$$x_i^{(2)} = \tilde{x}_i^2 - [iii]\tilde{x}_i - 1, \quad i = 1, \dots, k$$

and by setting certain of the moments and design parameters to zero. There are some advantages to be gained by achieving orthogonality in this way, mostly computational and related to model-fitting. However in redefining the explanatory variables, the sense of the second-order polynomial is lost and, in any case, the redefinition depends on the specific design through the third-order moments  $[i^3] = [iii]$  and should therefore be treated with some caution in comparative studies.

### 3.1. Example: The San Cristobal Design

The total number of runs for the class of San Cristobal designs of interest is given by  $N = n_0 + 2c + 6$  and the degrees of freedom for error by  $N - 6$ , with the latter partitioned into 3 degrees of freedom for lack of fit and  $n_0 + 2c - 3$  for pure error. The form of the matrix  $X^T X$  is readily derived as

$$\begin{bmatrix} N & N1^T & (N+a)1^T & N \\ N1 & aI + NJ & (2a+b)I + (N+a)J & (N+a)1 \\ (N+a)1 & (2a+b)I + (N+a)J & gI + hJ & (N+3a+b)1 \\ N & (N+a)1^T & (N+3a+b)1^T & (N+2a+4) \end{bmatrix}$$

where  $I$  is the identity matrix of order 2,  $J$  represents a  $2 \times 2$  matrix with all entries equal to 1 and  $a = (f + c(c+1)\alpha^2)$ ,  $b = c(c^2-1)\alpha^3$ ,  $d = c(c^3+1)\alpha^4$ ,  $g = 4(a+b)+d$  and  $h = N+2a+f$ . The inverse of the matrix  $X^T X$  can also be derived straightforwardly but, for the sake of brevity, is not presented here. Orthogonality can be achieved by an appropriate redefinition of the explanatory variables, by taking  $\alpha = \sqrt{\frac{2(\sqrt{N}-2)}{c(c+1)}}$  as in Rojas<sup>18</sup> and by imposing the constraint that the resultant designs lie in the region of experimentation and thus that  $n_0$  does not exceed the largest integer less than or equal to

$$\frac{(c+1)(23 - 8\sqrt{7} + 39c - 12\sqrt{7}c - 2c^2)}{c^2}.$$

However, as observed earlier, these orthogonal designs are not especially useful.

## 4. Parameter Estimates

An obvious approach to evaluating a design of interest for a particular model is to examine the precision of the parameter estimates. Criteria formulated specifically for this purpose, such as  $D$ - and  $A$ -optimality, are immediately available in the optimal design literature. However an appraisal of the properties of the parameter estimates for an empirical model in the response surface setting, either a first- or a second-order polynomial, should be treated with some caution. First the parameter estimates are not the focus of interest but rather the predictions. Second for the parameter estimates to have physical meaning the model should be couched in terms of the original units of the explanatory variables and not in coded units. As a consequence problems with respect to the dependence of a chosen criterion on scale and location can arise. At the same time the precision of

the parameter estimates is important when performing tests of hypothesis to ascertain whether or not the associated terms should be retained in the model.

A criterion used extensively for assessing the precision of the parameter estimates in a given model is that of generalized variance. This criterion is defined as the determinant of the inverse of the moment matrix and designs for which it is a minimum are termed *D*-optimal. The generalized variance criterion is particularly appealing in that it is invariant to linear transformations of the explanatory variables. For a given design within the RSM setting it is straightforward to evaluate the determinant  $\det\{N(X^T X)^{-1}\}$ , very often explicitly. Furthermore it is usually possible to construct a *D*-optimal or a near *D*-optimal design based on  $N$  runs and with points of support constrained to lie within the region of experimentation. In fact such optimal designs can be obtained from most packages with design construction capabilities such as SAS and Minitab. The *D*-efficiency of the design of interest can then be computed as

$$D\text{-efficiency} = \left\{ \frac{\det(X^T X)}{\det(X_D^T X_D)} \right\}^{\frac{1}{p}}$$

where  $X_D$  is the model matrix for the associated *D*-optimal design and  $p$  is the number of parameters in the model<sup>23</sup> and this efficiency can be used as a basis for design comparison. More recently Morris<sup>10</sup> introduced an attractive variant of *D*-efficiency, termed the generalized standard deviation and defined as

$$GSD = \sqrt{N \det\{(X^T X)^{-\frac{1}{p}}\}},$$

which is well-suited to comparative studies. This criterion is straightforward to compute since the construction of the attendant *D*-optimal design is not required. In addition, values of the *GSD* associated with different designs can be compared graphically following the ideas of Morris<sup>10</sup>.

It is not uncommon for a subset of parameters, such as the quadratic or bilinear terms in a second-order polynomial model, to be of specific interest. In this case criteria based on the notion of generalized variance and related to *D*-efficiency and to the *GSD* criterion can readily be formulated. Specifically, suppose that model (3) is written as

$$y = X_1\beta_1 + X_2\beta_2 + \epsilon$$

where  $X_1$  and  $X_2$  are model matrices relating to the parameter vectors  $\beta_1$  and  $\beta_2$  respectively and suppose that the parameters  $\beta_2$  are of interest.

The generalized variance associated with the estimates of  $\beta_2$  is given by the determinant  $\det\{N[X_2^T X_2 - X_2^T X_1 (X_1^T X_1)^{-1} X_1^T X_2]^{-1}\}$  and designs minimizing this criterion are termed  $D_s$ -optimal. Measures of  $D_s$ -efficiency and subset  $GSD$  can then be defined accordingly. Note however that  $D_s$ -optimal designs are not routinely available from standard packages and indeed must invariably be calculated directly by the practitioner. As a consequence the  $GSD$  criterion associated with a subset of parameters which does not require the computation of the corresponding  $D_s$ -optimal design offers a more attractive measure of the precision of those parameter estimates than does  $D_s$ -efficiency itself.

The average variance of the parameter estimates which is proportional to the trace of the inverse of the moment matrix, and thus to  $\text{tr}\{N(X^T X)^{-1}\}$ , can also be used to assess the precision of those estimates. This criterion is termed  $A$ -optimality and the associated optimal designs are routinely available in many packages. However the criterion is not invariant to linear transformations of the explanatory variables and must therefore be treated with considerable caution. Specifically a design which performs well under  $A$ -optimality for a particular setting of the explanatory variables may perform less well when the scale and location of those variables are changed.

Finally it is tempting to compare designs on the basis of the standard errors associated with individual parameter estimates. Note immediately however that while the standard errors for the linear terms in a first-order and the quadratic and bilinear terms in a second-order polynomial are affected only by a change in scale of the explanatory variables, those of the linear parameters in a second-order polynomial are affected by changes in both scale and location. Any comparison of designs involving the latter should therefore be made with some circumspection. The efficiency of a given design in terms of the precision of an individual parameter estimate can be formulated as the ratio of the minimum variance attainable for that estimate to the variance provided by the design. There are however problems in deriving the lower bound which forms the numerator of this ratio. Specifically designs minimizing the variance of a single parameter estimate for a particular experimental setting fall into the class of  $c$ -optimal designs and are very often singular. Thus the associated  $c$ -efficiency is not always computable. Recently Mee<sup>11</sup> suggested an efficiency measure, termed  $b_{ii}$ -efficiency, which incorporates an idealized lower bound on the variance of a parameter estimate. However the identification of this lower bound is not altogether straightforward and, in addition, the bound itself is not meaningful in practice. Designs can also be evaluated on the basis of the rel-

ative efficiencies of individual parameter estimates as discussed by Mee<sup>11</sup> but more direct measures of comparison are arguably to be preferred. For example the efficiencies of the bilinear in relation to the pure quadratic effects for a second-order polynomial model could well be appraised from the scale-invariant ratio

$$\frac{Var(b_{ij})}{\sqrt{Var(b_{ii})Var(b_{jj})}} \quad i, j = 1, \dots, k, \quad i \neq j$$

where  $b_{ij}$ ,  $b_{ii}$  and  $b_{jj}$  are estimates of the parameters  $\beta_{ij}$ ,  $\beta_{ii}$  and  $\beta_{jj}$  respectively.

A summary of the criteria discussed above and associated with the precision of the parameters estimates is presented in Table 1.

Table 1. Criteria for assessing the precision of the parameter estimates.

Criterion	Formulation
$D$ -optimal	$\det\{N(X^T X)^{-1}\}$
$GSD$	$\sqrt{N\det\{(X^T X)^{-\frac{1}{p}}\}}$
$D_s$ -optimal	$\det\{N[X_2^T X_2 - X_2^T X_1 (X_1^T X_1)^{-1} X_1^T X_2]^{-1}\}$
$A$ -optimal	$\text{tr}\{N(X^T X)^{-1}\}$
$se(b_{ij})$	$\sqrt{N[(X^T X)^{-1}]_{ij}}$

#### 4.1. Example: The San Cristobal Design

The determinant and the trace of the inverse of the moment matrix and values of the  $D$ -efficiency and the  $GSD$  criterion associated with the second-order polynomial model parameters for the  $3^2$  factorial design and for the San Cristobal design  $SC2$  are summarized in Table 2. The  $D_s$ -efficiency and the  $GSD$  relating to the pure quadratic parameters  $\beta_{11}$  and  $\beta_{22}$  for these model-design settings are presented in Table 3. The exact  $D$ -optimal designs based on 9 and 11 points, which are required in the efficiency calculations,

Table 2. Precision of the parameter estimates for the second-order polynomial model.

Design	$\det\{N(X^T X)^{-1}\}$	$\text{tr}\{N(X^T X)^{-1}\}$	$D$ -efficiency	$GSD$
$3^2$	102.5156	62.0000	0.5530	1.4708
$SC2$	29.6001	56.4115	0.6986	1.3262

Table 3. Precision of the parameter estimates of  $\beta_{11}$  and  $\beta_{22}$  for the second-order polynomial model.

Design	$D_s$ -efficiency	GSD
$3^2$	0.3667	2.1213
SC2	0.8122	1.5507

were obtained by using SAS to construct optimal designs on a finite grid of points in the design region and by then refining the resultant designs using a nonlinear optimization routine available in the programming language GAUSS. As an aside the 11-point  $D$ -optimal design has support at the points  $(0, 0)$ ,  $(0, 0)$ ,  $(0, 1.496)$ ,  $(1.496, 0)$ ,  $(0, 2.828)$ ,  $(0, 2.828)$ ,  $(2.828, 0)$ ,  $(2.828, 0)$ ,  $(1.078, 1.078)$ ,  $(1.366, 2.477)$  and  $(2.477, 1.366)$  and these points form an interesting radial pattern in the design space  $R_0$  as shown in Figure 1(b). The 9-point  $D$ -optimal design is similar to the 11-point design but with the points at  $(2.828, 0)$  and  $(0, 2.828)$  occurring only once. The requisite  $D_s$ -optimal designs based on 9 and 11 points of support were constructed directly using appropriate nonlinear optimization routines in GAUSS and have a similar form to their  $D$ -optimal counterparts. It is clear from the results presented in Tables 2 and 3 that the San Cristóbal design SC2 outperforms the traditional  $3^2$  factorial design with respect to the selected criteria relating to the precision of the parameter estimates. In fact this design is  $D$ -optimal over the class of all San Cristóbal designs involving two explanatory variables and constrained to lie in the design region  $R_0$ .

## 5. Prediction Variances

In response surface methodology interest necessarily centres on the predicted values of the responses and on their associated variances. As a consequence criteria for evaluating designs based on prediction variances are of particular importance and are indeed widely used in practice. Specifically, the prediction at a setting  $x$  of the explanatory variables can be written as  $\hat{y}(x) = f(x)^T b$ , where  $f(x) = (1, x_1, \dots, x_k)$  for a first-order and  $f(x) = (1, x_1, \dots, x_k, x_1^2, \dots, x_k^2, x_1 x_2, \dots, x_{k-1} x_k)$  for a second-order polynomial model and  $b$  is the least squares estimator of the parameter vector  $\beta$ . Then, for a design with model matrix  $X$ , the scaled prediction variance at  $x$  is given by

$$\frac{NVar\{\hat{y}(x)\}}{\sigma^2} = Nf(x)^T(X^T X)^{-1}f(x)$$

and provides the basic building block for a broad range of design criteria.

### 5.1. Rotatability

Rotatable designs are those designs for which the prediction variances at coded explanatory variable settings a fixed distance from the origin of the region of experimentation are constant<sup>20</sup>. Such designs are easily characterized. In particular, a design for a first-order polynomial model is rotatable if and only if it is orthogonal and the pure second-order moments are equal. Furthermore a design for a second-order polynomial is rotatable if and only if all odd moments up to order 4 are zero and in addition the even fourth-order moments satisfy the condition  $[iiii]/[iijj] = 3$ . Not all designs are rotatable however and a number of indices of rotatability which assess the closeness or otherwise of a given design to rotatability have been devised. Specifically, Draper and Guttman<sup>24</sup> introduced a measure of rotatability based on approximating contours of constant prediction variance with functions of the form  $\sum_{i=1}^k |x_i|^m$ , where  $m$  is an integer, but the attendant calculations are a little cumbersome. Arguably the most attractive indices of rotatability are those developed by Draper and Pukelsheim<sup>25</sup>, based on the earlier work of Khuri<sup>26</sup>. These indices are derived by regressing the moments of a given design against the idealized moments for a rotatable design to give a rotatable component of the moment matrix and by quantifying the discrepancy between this rotatable component and the moment matrix itself. Importantly the resultant indices are invariant to rotations of the design space. More recently Park, Lim and Baba<sup>27</sup> suggested a measure of rotatability based on the sum of the squared differences between the prediction variance at a setting  $x$  in the design region and the average of the prediction variances over the sphere of radius  $\sqrt{x^T x}$ . Their measure is appealing in that it relates to the ideas underpinning variance dispersion graphs which are discussed later and in that it can be found explicitly in certain cases.

### 5.2. Single Number Criteria

The average scaled prediction variance over the region of interest,  $R$ , provides an immediate overall measure of the precision associated with the predicted responses. Specifically, if the region  $R$  comprises a finite set of  $K$  points, commonly a grid of points, the average scaled prediction variance is given by

$$V = \frac{1}{K} \sum_{k=1}^K N f(x_k)^T (X^T X)^{-1} f(x_k)$$

and designs minimizing this criterion are termed  $V$ -optimal. Alternatively, if the region  $R$  is a subset of  $k$ -dimensional space, the average scaled prediction variance is then

$$IV = \frac{N \int_R f(x)^T (X^T X)^{-1} f(x) dx}{\int_R dx} = N \text{tr}\{(X^T X)^{-1} M\}$$

where  $M$  is the region moment matrix with elements of the form

$$\frac{\int_R x_1^{\delta_1} \dots x_k^{\delta_k} dx}{\int_R dx}$$

which mirror the design moments introduced earlier. Designs minimizing this  $IV$ -criterion are termed  $IV$ -optimal and also, and equivalently,  $Q$ -optimal<sup>28</sup>.

The calculation of the  $V$ -optimality criterion for a given design defined on a region of experimentation comprising a finite set of points is straightforward, provided the number of points is not so large that complete enumeration becomes expensive in terms of computer-time. In contrast, the calculation of the  $IV$ -criterion for continuous design spaces is more involved and essentially hinges on the evaluation of the region moment matrix  $M$ . For hyperspherical design spaces the region moments can be obtained directly using the formulae provided by Box and Draper<sup>2</sup> and by Myers<sup>3</sup>. For cuboidal regions the integrals defining the moments separate out and can thus be found routinely, and for non-standard design regions it is usually possible, with a little algebraic ingenuity, to obtain explicit expressions for the region moments<sup>29</sup>. However the tools for performing such symbolic calculations are not always readily available. An alternative approach is thus to approximate the  $IV$ -criterion by evaluating the average scaled prediction variance over a specified set of points in the design region, as is done in *SAS*, or over a set of randomly generated points in that region, as in *RS/Discover*. Borkowski<sup>29</sup> has shown however that such approximations tend to produce inflated values of the  $IV$ -criterion and thus explicit formulae are generally to be preferred.

It is also helpful to introduce the notions of  $V$ - and  $IV$ -efficiency. Specifically, the  $V$ -efficiency of an  $N$ -point design defined on a finite design space

and with a model matrix  $X$  is given by

$$V\text{-efficiency} = \frac{\text{tr}\{(X_V^T X_V)^{-1} M\}}{\text{tr}\{(X^T X)^{-1} M\}},$$

where  $X_V$  is the model matrix for the corresponding  $N$ -point  $V$ -optimal design. The  $IV$ -efficiency of a design in a continuous design space can be similarly defined. The  $V$ - and  $IV$ -optimal designs required in order to calculate these efficiencies are not always easy to obtain however. Only selected packages, such as ACED, provide routines for calculating  $V$ -optimal designs and, for the reasons discussed above, none generate  $IV$ -optimal designs. Thus, in constructing  $V$ - and  $IV$ -optimal designs, recourse must invariably be made to user-based programmes and routines.

The maximum scaled prediction variance over the region of interest, defined by

$$N \max_{x \in R} f(x)^T (X^T X)^{-1} f(x),$$

is a worst scenario measure of the precision of the predicted responses and provides a valuable counterpoint to average variance criteria. Designs minimizing this maximum variance are termed  $G$ -optimal and the criterion itself the  $G$ -criterion. For a design defined on a region of interest comprising a finite set of points the maximum prediction variance can be evaluated by complete enumeration. For continuous design spaces the situation is less straightforward. For a first-order polynomial with coded variables satisfying  $\sum_{u=1}^n x_{ui} = 0$  the scaled prediction variance can be written as

$$1 + Nx^T (D^T D)^{-1} x$$

where  $D$  is the design matrix and the maximum can, in many cases, be readily obtained. For example if the region of interest is a hypersphere of radius  $r$ , the maximum scaled prediction variance is given in Giovannitti-Jensen and Myers<sup>30</sup> by  $1 + N\lambda_{\max}r^2$  where  $\lambda_{\max}$  is the largest eigenvalue of  $(D^T D)^{-1}$ . Also if the design region is cuboidal, or is defined by a set of linear constraints on the explanatory variables, the maximization of the prediction variance can be formulated as a quadratic programming problem and solved accordingly.

For second-order polynomial models defined on continuous design spaces, evaluating the  $G$ -criterion for a design of interest can be particularly problematic. Borkowski<sup>31</sup> derived an explicit expression for the prediction variances associated with a central composite design and showed that in the case of a face-centred design and a cuboidal region of experimentation the maximum can be obtained by searching over the set of barycentric points.

However this would seem to be an isolated example of an analytic solution. More generally, the easiest approach to evaluating the maximum prediction variance is by approximation over a finite set of specified or simulated points in the design region, as discussed earlier within the context of the *IV*-criterion, but the accuracy associated with the resultant maximum is limited by the number of points used in the enumeration. Otherwise, if accuracy is important, constrained nonlinear optimization routines must be invoked in order to maximize the prediction variance and there is no guarantee that the maximum so found is global. Specialist software implementing this approach for particular design settings is available. For example Vining<sup>32</sup> has written a highly robust FORTRAN program that evaluates the maximum prediction variance over the surfaces of concentric hyperspheres centred at the design origin and thus provides a maximum over the associated hyperspherical design region, while Borror<sup>33</sup> has developed a similar program for cuboidal regions. However for non-standard regions of interest it is in general necessary for the user to provide the appropriate optimization routines.

The *G*-efficiency of a given design, defined as the ratio of the maximum prediction variance for the *G*-optimal design to that for the design itself, is commonly computed and used in design evaluation studies. From the classical theory of optimal design it follows immediately that the lower bound on the maximum prediction variance associated with the approximate *G*-optimal design is given by  $p$ , the number of parameters in the model, and thus that the associated *G*-efficiency is simply the ratio

$$\frac{p}{N \max_{x \in R} f(x)^T (X^T X)^{-1} f(x)}.$$

Strictly *G*-optimal designs based on  $N$  runs and defined on the region of experimentation should be used in computing *G*-efficiencies but the construction of such designs is not entirely straightforward. Welch<sup>34</sup> developed an elegant exchange algorithm for finding *G*-optimal designs over design spaces comprising a finite set of points and incorporated this routine into the software package ACED. More generally, minimizing the maximum value of the *G*-criterion for continuous design spaces is a minimax problem and must be handled using derivative-free routines such as simulated annealing<sup>35</sup> or genetic algorithms<sup>36,37</sup>. In practice therefore *G*-efficiencies based on the lower bound  $p$  for the maximum scaled prediction variance are commonly invoked and, following Giovannitti-Jensen and Myers<sup>30</sup> and Borkowski<sup>31</sup>, efficiency values of at least 50% are considered acceptable.

Table 4. Single number criteria for the prediction variances.

Criterion	Formulation
$V$ -optimal	$V = \frac{1}{K} \sum_{k=1}^K N f(x_k)^T (X^T X)^{-1} f(x_k)$
$IV = Q$ -optimal	$\frac{N \int_R f(x)^T (X^T X)^{-1} f(x) dx}{\int_R dx} = N \text{tr}\{(X^T X)^{-1} M\}$
$G$ -optimal	$N \max_{x \in R} f(x)^T (X^T X)^{-1} f(x)$

A summary of the criteria associated with the prediction variances and discussed above is presented in Table 4. Other measures relating to the  $G$ -criterion, in particular the minimum and the range of the scaled prediction variances over the design space of interest, can also be invoked. Indeed the minimum prediction variance can be computed using methods similar to those developed for the maximum. These criteria have received little attention in the literature however, most probably because they do not fall within the ambit of classical optimal design theory.

### 5.3. Snapshot Criteria

The measures relating to prediction variances discussed thus far are single number criteria. It is clearly of interest however to provide more comprehensive summaries of the prediction variances and in particular to capture some of the key features and patterns of those variances over the region of experimentation. An obvious approach to this problem is to plot contours of constant prediction variance over the design region but this is only feasible for cases involving two and possibly three explanatory variables<sup>6</sup>. As a consequence a number of strategies which involve "snapshot" graphical summaries of the prediction variances and which emanate from the notion of rotatability have been devised.

Box and Behnken<sup>38</sup> introduced the concept of a variance profile as a plot of the maximum and minimum prediction variances over the surface  $S_\rho$  of a hypersphere centred at the origin of the design region and of radius  $\rho$  against the radius  $\rho$  itself. More recently Giovannitti-Jensen and Myers<sup>30</sup> and Myers, Vining, Giovannitti-Jensen and Myers<sup>39</sup> formalized this approach and extended it to include the spherical prediction variance

defined as

$$v_\rho = \frac{N \int_{S_\rho} f(x)^T (X^T X)^{-1} f(x) dx}{\int_{S_\rho} dx} = N \text{tr}\{(X^T X)^{-1} M_\rho\}$$

where  $M_\rho$  is the region moment matrix over the surface  $S_\rho$ . Plots of  $v_\rho$  and the maximum and minimum variances over  $S_\rho$ , denoted  $\text{VMIN}_\rho$  and  $\text{VMAX}_\rho$  respectively, against the radius  $\rho$  are termed variance dispersion graphs (VDG's). These graphs summarize the near-rotatability or otherwise of a particular design defined on a spherical region of experimentation and, at the same time, capture much of the nature of the prediction variances associated with that design. It is clearly preferable to obtain explicit expressions for the prediction variances  $\text{VMIN}_\rho$ ,  $\text{VMAX}_\rho$  and  $v_\rho$ , as for example those derived by Borkowski<sup>40,41</sup> for the Box-Behnken and central composite designs. Such analytic results are however elusive and in general recourse must be made to obtaining these variances numerically using, for example, the program of Vining<sup>32</sup>.

Some interesting variants on the basic VDG's have been introduced into the literature more recently. Thus Khuri, Kim and Um<sup>42</sup> suggested finding quantiles of the scaled prediction variances on the surface  $S_\rho$  of a hypersphere of radius  $\rho$  by means of simulation and plotting the resultant quantiles for a range of  $\rho$  values. This extension to the basic VDG can be readily implemented using the S-Plus routines provided by the authors and is particularly helpful in detailed design comparisons. In addition Trinca and Gilmour, in two separate papers, provide valuable guidelines on the use of VDG's for blocked experiments<sup>43</sup> and for situations in which the variance of the difference between the predictions and a specified benchmark prediction are of interest<sup>44</sup>. The use of VDG's to compare designs defined on regions other than the hypersphere has received surprisingly limited attention. Myers and Montgomery<sup>6</sup> describe the nature of VDG's for cuboidal design regions and Borror<sup>33</sup> provides a program for computing such graphs. For non-standard regions of interest it is possible to obtain VDG's by introducing a series of nested subregions, following the guidelines given in the context of mixture experiments by Khuri, Harrison and Cornell<sup>45</sup>, but this approach does not seem to have been pursued.

A simple but attractive approach to summarizing prediction variances over the entire design region is to simulate points within that region and to present a quantile plot of the attendant scaled prediction variances. This

idea was introduced within the context of mixture experiments by Khuri, Harrison and Cornell<sup>45</sup> and the plots referred to as scaled prediction variance quantile (SPVQ) plots. More recently Zahran, Anderson-Cook and Myers<sup>46</sup> observed that VDG's do not properly reflect volumes in the design space and addressed this issue by introducing fraction of design space plots which are essentially the same as SPVQ plots but for polynomial models in a non-mixture situation. One of the problems associated with obtaining SPVQ plots, and also quantile VDG's, is the efficient simulation of points in the design region and, in particular, in a design region which is non-standard. In fact there is a wealth of information on algorithms suitable for this purpose available in, for example, the book by Fishman<sup>47</sup>.

#### 5.4. Example: The San Cristobal Design

The integrated, maximum and minimum scaled prediction variances, denoted VAVG, VMAX and VMIN respectively, and the *IV*-efficiency for the  $3^2$  factorial and the *SC2* designs taken over the region of experimentation  $R_0$ , that is over the intersection of the disk of radius  $\sqrt{8}$  centered at the origin  $(0,0)$  with the positive quadrant, and associated with the fitting of a second-order polynomial model are presented in Table 5. Note that the integrated variance criterion was calculated explicitly using the result

$$\int_{R_0} x_1^{\delta_1} x_2^{\delta_2} dx_1 dx_2 = \frac{8^{\frac{\delta_1+\delta_2+2}{2}}}{(\delta_1 + \delta_2 + 2)} \frac{\Gamma(\frac{\delta_1+1}{2}) \Gamma(\frac{\delta_2+1}{2})}{2 \Gamma(\frac{\delta_1+\delta_2+2}{2})}$$

and that the maximum and minimum prediction variances were obtained using appropriate optimization routines. The *IV*-optimal designs required for the efficiency calculations were constructed by implementing similar procedures to those used for the corresponding *D*-optimal designs and have the same form as those designs. In addition the VDG's obtained for points in  $R_0$  equidistant from the origin and the SPVQ or fraction of design space plots for the  $3^2$  factorial and the *SC2* design are presented in Figures 2 and 3 respectively. It is clear from these results that the *SC2* design outperforms the  $3^2$  factorial design in terms of prediction variance. It is also interesting to observe that the scaled prediction variances for points farthest from the origin of the design space  $R_0$  are particularly high for both designs.

Table 5. Criteria for the prediction variances.

Design	VAVG	VMAX	VMIN	<i>IV</i> -efficiency
$3^2$	7.7231	47.6566	3.2000	0.5490
<i>SC2</i>	5.3228	28.7508	2.3710	0.8849

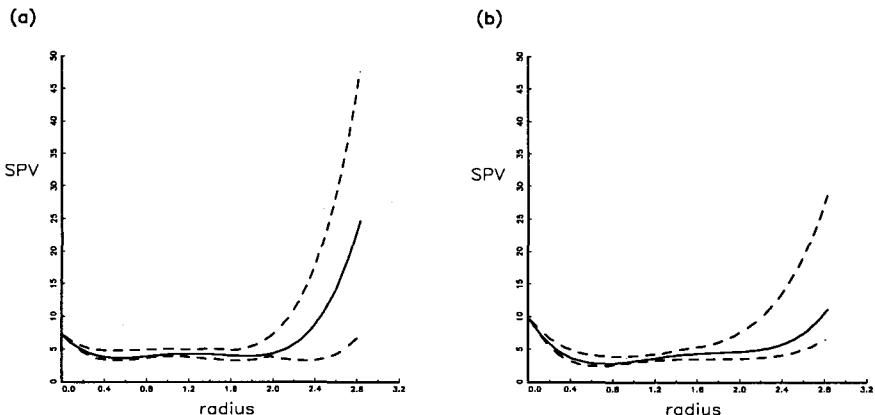


Fig. 2. Variance dispersion graphs: (a) The  $3^2$  factorial design and (b) the San Cristobal design  $SC2$  in the region of experimentation  $R_0$ . The solid lines correspond to  $V_\rho$  and the dotted lines to  $VMIN_\rho$  and  $VMAX_\rho$  in an obvious way.

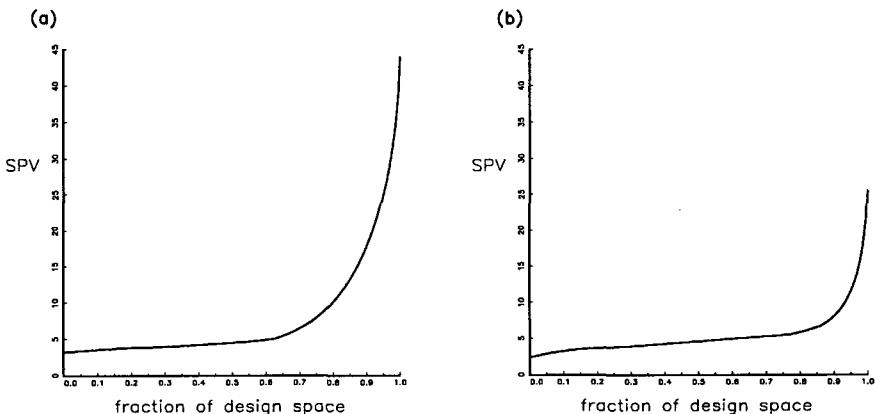


Fig. 3. Fraction of design space plots: (a) The  $3^2$  factorial design and (b) the San Cristobal design  $SC2$  in the region of experimentation  $R_0$ .

## 6. Robustness Issues

### 6.1. Model Checking

Suppose that a polynomial model of the form  $y = X_1\beta_1 + \epsilon$  is fitted to response surface data but that the experimenter is concerned that additional polynomial terms have been neglected and that in fact the model

$y = X_1\beta_1 + X_2\beta_2 + \epsilon$ , where  $\beta_2$  represents the vector of extra parameters with associated model matrix  $X_2$ , is more appropriate. There is a large literature on devising suitable design criteria relating to the protection against and the detection of such departures from the fitted model. However the emphasis in these studies is almost exclusively on the construction of designs which are in some sense optimal for these criteria and surprisingly little attention has been given to using the criteria to compare designs. The aim of this section therefore is to provide a brief review of optimality criteria relating to departures from the fitted model and to suggest how these, and possibly other criteria, may be sensibly used in comparative studies.

### 6.1.1. Bias

For the fitted model, the prediction at the setting  $x$  is given by  $\hat{y}(x) = f_1(x)^T b_1$  where  $b_1$  is the least squares estimator of  $\beta_1$  and the bias in this predictor in relation to the model with expected response  $f_1(x)^T \beta_1 + f_2(x)^T \beta_2$  is then given by  $\{f_1(x)^T A - f_2(x)^T\} \beta_2$  where  $A = (X_1^T X_1)^{-1} X_1^T X_2$  is the alias matrix<sup>3</sup>. A suitable measure of prediction bias over the design region  $R$  can be formulated by appropriately squaring and averaging or integrating the biases over the experimental settings to yield the criterion

$$B = \beta_2^T (A^T \mu_{11} A - \mu_{21} A - A^T \mu_{12} + \mu_{22}) \beta_2 = \beta_2^T C \beta_2$$

where following Myers<sup>3</sup> the terms  $\mu_{ij}$  represent the region moment matrices

$$\mu_{ij} = \frac{\int_R f_i(x) f_j(x)^T dx}{\int_R dx}, \quad i, j = 1, 2.$$

This bias criterion is in turn incorporated into the mean squared error of prediction integrated over the design region  $R$  and, specifically, the integrated mean squared error is given by

$$IMSE = \frac{N}{\sigma^2} \left\{ \text{tr}\{(X_1^T X_1)^{-1} \mu_{11}\} + \beta_2^T C \beta_2 \right\}.$$

Note that the criteria  $B$  and  $IMSE$  depend crucially on the unknown parameters  $\beta_2$  and that it is therefore necessary to introduce assumptions to counter this dependence.

Designs minimizing the bias and the integrated mean squared error criteria provide protection against departures from the fitted model and numerous studies on the construction of such designs have been reported. In contrast the paper by Vining and Myers<sup>48</sup>, which considers comparing designs by using graphs analogous to VDG's but with mean squared error replacing the scaled prediction variance, would seem to be an isolated one. In fact when comparing designs it is not unreasonable to decouple bias and prediction variance and to allow the practitioner to decide on the balance between them. For this reason, criteria relating to bias in the predictions are considered here in their own right.

The integrated squared bias  $B = \beta_2^T C \beta_2$  can be used directly to compare designs provided that the parameter  $\beta_2$  is a scalar or that a good guess for that parameter is available. Otherwise assumptions about the nature of  $\beta_2$  must be introduced in order to formulate usable criteria and those discussed here are gleaned from studies on designs which minimize the IMSE criterion. Thus Allen, Yu and Schmitz<sup>49</sup> suggested placing a prior distribution on  $\beta_2$  with mean zero and variance matrix  $\tau^2 I$  and adopting the expectation

$$E[\beta_2^T C \beta_2] = E[tr\{C \beta_2 \beta_2^T\}] = \tau^2 tr(C)$$

as an appropriate criterion. This approach is attractive in that it is compatible with the one formulated earlier by Welch<sup>50</sup> and, in addition, in that it coincides with that developed by Vining and Myers<sup>48</sup> and based on averaging  $\beta_2^T C \beta_2$  over all parameters  $\beta_2$  lying on a sphere of radius  $\tau$ . At the same time the trace criterion  $tr(C)$  is not invariant to changes in scale and location of the explanatory variables and should therefore be used with caution. An alternative approach, following Vining and Myers<sup>48</sup>, is to adopt a worst case scenario and to invoke the maximum of  $\beta_2^T C \beta_2$  over the sphere  $\beta_2^T \beta_2 = 1$ , namely the largest eigenvalue of  $C$ , as a measure of bias. In fact arguably the most attractive criterion for comparing designs on the basis of bias alone is the determinant of the matrix  $C$ , which corresponds to the volume of the ellipsoid defined by  $\beta_2^T C \beta_2$  for arbitrary  $\beta_2$ . This criterion does not emerge naturally from a consideration of the IMSE criterion and merits further attention. Finally note that the minimum integrated squared bias is given explicitly by  $\beta_2^T (\mu_{22} - \mu_{21}\mu_{11}^{-1}\mu_{12})\beta_2$  and can be used to benchmark values of the criteria suggested here. A summary of the criteria associated with bias, together with the appropriate references, is given in Table 6.

Table 6. Robustness criteria involving bias.

Criterion	Formulation	References
$\beta_2$ scalar or known	$\beta_2^T C \beta_2$	
$\beta_2 \sim N(0, \tau^2 I)$	$\tau^2 \text{tr}(C)$	48,49,50
worst case	$\lambda_{\max}(C) = \max_{\beta_2^T \beta_2 = 1} \beta_2^T C \beta_2$	48
volume of ellipsoid	$\det(C)$	

### 6.1.2. Detecting Lack of Fit

The practitioner may well be interested in comparing designs within the context of detecting departures from the fitted model  $y = X_1\beta_1 + \epsilon$  in the direction of the extended model  $y = X_1\beta_1 + X_2\beta_2 + \epsilon$ . To this end it is sensible to adopt the approach introduced by Atkinson<sup>51</sup> in which the non-centrality parameter of the  $F$  test for the hypothesis  $\beta_2 = 0$ , namely  $N\Delta$  where

$$\Delta = \frac{1}{N} \beta_2^T (X_2^T X_2 - X_2^T X_1 (X_1^T X_1)^{-1} X_1^T X_2) \beta_2 = \beta_2^T L \beta_2,$$

is used to quantify the power of a particular design to detect lack of fit. Designs maximizing the criterion  $\beta_2^T L \beta_2$  belong to the class of  $T$ -optimal designs<sup>23</sup> and are clearly attractive. However the  $T$ -criterion can only be evaluated provided the parameter vector  $\beta_2$  is either a scalar or is known and this is rarely the case<sup>51</sup>. Indeed the emphasis in the work of Atkinson<sup>51</sup> is on accommodating the fact that the parameter  $\beta_2$  is unknown and on constructing designs which maximize the criteria so obtained. There would seem to be few, if any, later studies devoted specifically to using Atkinson's criteria for design comparison and some tentative suggestions are therefore provided here.

The criterion most widely used in constructing designs optimal for detecting departures from the fitted model is the determinant of the matrix  $L$ , written  $\det(L)$ , and this is eminently suitable for comparative studies. The criterion has particular appeal in that it is essentially a  $D_s$ -optimal or  $GSD$  criterion and all the advantages of such criteria immediately accrue. Other measures, such as the trace and the smallest eigenvalue of  $L$ , which

are derived on the basis of assumptions about the parameter  $\beta_2$  described earlier for bias, can also be used to compare designs but are less tractable than the determinant criterion. In addition DeFeo and Myers<sup>52</sup>, following the work of Jones and Mitchell<sup>53</sup>, considered the average and the minimum of  $\beta_2^T L \beta_2$  subject to a linear constraint on the squared bias  $\beta_2^T C \beta_2$ , that is the trace and the smallest eigenvalue of the matrix  $C^{-1} L$ , as appropriate criteria for constructing optimal designs but these measures have limited appeal for design comparison.

The determinant criterion  $\det(L)$  can in fact be regarded as a measure of the precision with which the parameters  $\beta_2$  in the extended model are estimated. However these parameters are assumed to be negligible in the original model formulation and thus a criterion which combines efficiency of estimation for the parameters  $\beta_1$  with that for the additional unknown parameters  $\beta_2$  would be more appealing. To this end, the composite criterion

$$\frac{\alpha}{p_1} \ln\{\det(X_1^T X_1)\} + \frac{1-\alpha}{p_2} \ln\{\det(X_2^T X_2 - X_2^T X_1 (X_1^T X_1)^{-1} X_1^T X_2)\}$$

where  $p_1$  and  $p_2$  are the number of parameters associated with  $\beta_1$  and  $\beta_2$  respectively and  $\alpha \in [0, 1]$  represents a weight selected by the experimenter, is appropriate<sup>23</sup> and is based on the earlier work of Stigler<sup>54</sup> and of Cook and Wong<sup>55</sup>. Alternatively the fully Bayesian criterion developed by DuMouchel and Jones<sup>22</sup>, which involves redefining the terms associated with the parameters  $\beta_2$  so that the resultant model matrix has the form  $X^* = [X_1 \ X_2^*]$  with  $X_1^T X_2^* = 0$  and placing a prior distribution with mean zero and variance  $\tau^2 I$  on the induced additional parameters, can be invoked. This criterion can be simply formulated as  $\det(X^{*T} X^* + \frac{1}{\tau^2} E)$  where  $E$  is a diagonal matrix with the first  $p_1$  entries equal to 0 and the last  $p_2$  entries equal to 1 and is particularly appealing in that it allows for parsimonious model checking in the sense that the number of runs in a design of interest can be less than the total number of parameters  $p_1 + p_2$ <sup>22,56</sup>. Both the composite and the Bayesian criteria for detecting lack of fit have been used almost exclusively in the context of optimal design construction but, with some adaptation, could well prove valuable in design comparison.

A summary of the criteria associated with lack of fit introduced here, together with the appropriate references, is presented in Table 7.

Table 7. Lack of fit criteria.

Criterion	Formulation	References
$T$ -optimal	$\beta_2^T L \beta_2$ with $\beta_2$ scalar or known	23,51
$D_s$ -optimal	$\det(L)$	51
average	$\text{tr}(T^{-1}L)$ where $T = I$ or $C$	51,52,53
worst case	$\lambda_{\min}(T^{-1}L)$ where $T = I$ or $C$	51,52,53
composite	$\frac{\alpha}{p_1} \ln\{\det(X_1^T X_1)\} + \frac{1-\alpha}{p_2} \ln\{\det(NL)\}$	23,54,55
Bayesian	$\det\{X^{*T} X^* + \frac{1}{\tau^2} E\}$	22,56

### 6.1.3. Model Robustness

The practitioner may suspect that the model to be fitted to a particular response surface is overparametrized and may therefore wish to select designs which perform well not only for the fitted model but also for selected submodels of that model. Borkowski and Valeroso<sup>57</sup> addressed this issue and recommended the simple but elegant expedient of comparing a range of designs on the basis of the values of the  $D$ -,  $A$ -,  $G$ - and  $IV$ -criteria for various submodels of the fitted model. An alternative approach is to use the model robust criteria introduced by Läuter<sup>58</sup> and developed by Cook and Nachtsheim<sup>59</sup> which comprise weighted averages of appropriate design efficiencies for various models, with the weights reflecting the interest of the experimenter in those models. This idea has not been pursued however.

A related problem involves settings in which a number of explanatory variables will be ultimately dropped from the fitted model, as for example in screening experiments. In such cases the projective properties of the design points in the original design space of  $k$  variables onto points in the space of  $k^*$  of those variables, with  $k^*$  strictly less than  $k$ , are important. Morris<sup>10</sup>, following Lin and Draper<sup>60</sup>, has provided some sensible guidelines as to how these projective properties, and in particular the distributions of the projected points, can be summarized and compared over a range of designs.

### 6.2. Other Robustness Measures

Box and Draper<sup>13</sup> advocated the use of a broad range of criteria for comparing designs and in particular stressed the importance of introducing

measures of comparison other than those associated with the variances of the parameter estimates and of the predictions. Specifically, they devised a measure of the insensitivity of a design to the presence of outliers as the sum of squares of the diagonal entries in the "hat" matrix  $X(X^T X)^{-1}X^T$  which can in turn be related to other diagnostic criteria<sup>61</sup>. Furthermore Andrews and Herzberg<sup>62</sup> investigated criteria based on a consideration of missing values while Vuchkov and Boyadjieva<sup>63</sup> proposed two criteria for appraising designs in terms of robustness to errors in the explanatory variables. However the studies of Box and Draper<sup>13</sup>, Andrews and Herzberg<sup>62</sup> and Vuchkov and Boyadjieva<sup>63</sup> appear to be isolated ones and the criteria so developed would seem to have been rarely used.

### 6.3. Example: The San Cristobal Design

Suppose that the mixed cubic terms with associated parameters  $\beta_{112}$  and  $\beta_{122}$  should not have been neglected in the choice of fitted model. Then the induced bias in the parameters can be measured using the criteria  $\det(B)$  and  $\text{tr}(B)$  and the lack of fit of the second-order polynomial model using  $\det(L)$  and  $\text{tr}(L)$ . Values of these criteria for the  $3^2$  factorial design and for the *SC2* design are presented in Table 8. From these results it is clear that the  $3^2$  factorial design provides better protection against and detection of the mixed cubic terms than does the San Cristobal design. Note however that the San Cristobal design comprises five levels of each factor and can thus provide information about the pure cubic terms whereas the  $3^2$  factorial design cannot. The measure of robustness to outliers developed by Box and Draper<sup>13</sup>, denoted  $v^2$ , is also included in Table 8 and it is clear that the  $3^2$  factorial design is again more robust than the San Cristobal design in this regard.

Table 8. Measures of prediction bias, lack of fit and robustness to outliers.

Design	Bias		Lack of fit		Outliers $v^2$
	$\det(C)$	$\text{tr}(C)$	$\det(L)$	$\text{tr}(L)$	
$3^2$	0.0472	0.4363	0.0219	0.2963	335.2500
<i>SC2</i>	0.0507	0.4508	0.0149	0.2480	463.3947

## 7. Conclusions

The main aim of the present chapter is to provide a comprehensive review of methods for evaluating and comparing designs used in the fitting of a response surface and specifically in fitting first- and second-order polynomial models to that surface. In particular three broad but distinct approaches to this problem, namely those based on the precision of the parameter estimates for the fitted model, on the variances of the predictions and on measures of robustness to model misspecification, are identified and critically appraised. The key features to emerge from the review are now highlighted and some pointers for future research are given.

Summary measures of the precision associated with the parameter estimates of the fitted models such as  $D$ -efficiency and generalized standardized deviation are introduced. However it is emphasized that, since the parameters themselves are not meaningful, these criteria have a limited appeal, most specifically within the context of performing tests of hypothesis to ascertain whether or not the associated terms should be retained in the model. As a counterpoint to this, prediction variances play a key role in comparative studies and measures of rotatability and single number and snapshot criteria are described in some detail. From the discussion of these measures it is clear that there is still scope for further research, particularly in relation to snapshot criteria. For example variance dispersion graphs for hyperspherical design regions are well established but there is a need to document and formalize the use of such graphs for cuboidal and, more particularly, for non-standard regions. Also the appropriateness of VDG's in summarizing prediction variances is arguably unclear in that the graphs are based on criteria evaluated over surfaces and not volumes of the design space<sup>46</sup>. Fraction of design space or equivalently SPVQ plots address this issue but other approaches, such as cumulative plots of suitably evaluated criteria, merit attention. Furthermore there are some computational challenges associated with the evaluation of prediction variance criteria. Specifically, obtaining maximum and minimum prediction variances over a chosen design region is computationally non-trivial and multi-resolution approaches similar to that introduced by Vining<sup>32</sup> and in a different context by Noubiap and Seidel<sup>64</sup> are worth further investigation. In addition the efficient simulation of points in a given design region, which is required in for example the construction of quantile VDG's and SPVQ plots, is not straightforward, particularly if the number of explanatory variables is large. Established approaches to this problem require systematic investi-

gation and newer techniques in simulation, such as the quasi-Monte Carlo method introduced by Fang and Wang<sup>65</sup> and based on space-filling designs, could well be explored to some advantage. Criteria associated with bias and variance in the predictions and with model checking and robustness are well established and widely used in the context of optimal design. However these criteria have received, and continue to receive, surprisingly little attention in comparative studies and the need to redress this balance is emphasized. In particular the fact that certain measures of bias and lack of fit depend on unknown parameters is problematic and needs to be more fully resolved, while the role of composite criteria in design comparison requires careful investigation. Finally measures of robustness to outliers, to errors-in-variables and to missing observations have only been considered in a few isolated studies and it is clear that these and other measures pertaining, for example, to extrapolation merit further attention.

A number of interesting and more general issues relating to response surface methodology emerge from this review. In particular it is clear that computer-generated designs are valuable benchmarks in comparative studies even though their use as designs robust to a range of criteria may well be questionable. This leads to the problem of constructing such designs. The branch-and-bound approach introduced by Welch<sup>66</sup> for obtaining *D*-optimal designs, although expensive of computer-time, guarantees global optimality and it would be appealing to extend the methodology to other criteria. More generally heuristics such as exchange algorithms, simulated annealing and genetic algorithms are widely used in the context of response surface methodology but it is nevertheless tempting to consider invoking other optimization procedures for design construction such as the Tabu search<sup>67</sup>.

The San Cristobal design is of interest in its own right as a composite design suitable for agricultural experiments and a more detailed account of its properties and use is presented elsewhere<sup>68</sup>. In a broader context it is clear that response surface methodology has enjoyed limited use in agriculture settings and this may well be attributed to the fact that agricultural experiments are long term and therefore not amenable to sequential study<sup>69</sup>. However glasshouse experiments such as those described by Edmondson<sup>70</sup> are medium term and dose-response studies on, for example, insecticides are of short duration and techniques used in the study of response surfaces could thus be used to advantage in these situations. In the more traditional experimental settings, designs such as the San Cristobal offer the agriculturalist interesting alternatives to factorial and fractional factorial designs in

that they can accommodate unusual design spaces and often require fewer runs.

Finally the guidelines for comparing and evaluating designs within the context of response surface methodology provided here should prove helpful in related settings such as in the modelling of binary responses which depend on several explanatory variables<sup>9,15</sup> and in the use of empirical models in computer experiments<sup>71</sup>.

### Acknowledgments

Much of the work for this paper was carried out while the author was in the School of Mathematics, Statistics and Information Technology at the University of KwaZulu-Natal, Pietermaritzburg, South Africa.

The study was supported by the University of KwaZulu-Natal, The University of Cape Town, and the National Research Foundation, South Africa. The author would like to thank Bill Notz for his helpful and insightful comments and the referee for some particularly valuable observations.

### References

1. G.E.P. Box and K.B. Wilson, *Journal of the Royal Statistical Society, Series B* **13**, 1–45 (1951).
2. G.E.P. Box and N.R. Draper, *Journal of the American Statistical Association* **54**, 622–654 (1959).
3. R.H. Myers, *Response Surface Methodology*. (Allyn and Bacon, Boston, 1971).
4. G.E.P. Box and N.R. Draper, *Empirical Model-Building and Response Surfaces* (Wiley, New York, 1987).
5. A.I. Khuri and J.A. Cornell, *Response Surfaces: Designs and Analyses*. Second Edition. (Marcel Dekker, New York, 1996).
6. R.H. Myers and D.C. Montgomery, *Response Surface Methodology*. (Wiley, New York, 2002).
7. W.J. Hill and W.G. Hunter, *Technometrics* **8**, 571–590 (1966).
8. R.H. Myers, A.I. Khuri and W.H. Carter, *Technometrics*, **31**, 137–157 (1989).
9. R.H. Myers, D.C. Montgomery, G.G. Vining, C.M. Borror and S.M. Kowalski, *Journal of Quality Technology* **36**, 53–77 (2004).
10. M.D. Morris, *Technometrics* **42**, 111–121 (2000).
11. R. Mee, *Technometrics* **43**, 34–43 (2001).
12. A. Zahran, C.M. Anderson-Cook, R.H. Myers and E.P. Smith, *Journal of Quality Technology* **35**, 387–392 (2003).
13. G.E.P. Box and N.R. Draper, *Biometrika* **62**, 347–352 (1975).
14. J. Kiefer, *Biometrika* **62**, 277–288 (1975).
15. R.H. Myers, *Journal of Quality Technology* **31**, 30–44 (1999).
16. B.A. Rojas, *Proceedings of the 11th Congress of the International Society for Sugar Cane Technology*, 197–203 (1962).

17. W.R. Mapham, *Some Biometrical Aspects of Soil Calibration*. (M.Sc. thesis, University of Natal, South Africa, 1975).
18. B.A. Rojas, *Proceedings of the 14th Congress of the International Society for Sugar Cane Technology*, 1085–1093 (1971).
19. B.A. Rojas, Unpublished note (1973).
20. G.E.P. Box and J.S. Hunter, *Annals of Mathematical Statistics* **28**, 195–241 (1957).
21. H.M. Dicks, *Journal of Tropical Forest Resources* **15**, 43–52 (1999).
22. W. DuMouchel and B. Jones, *Technometrics* **36**, 37–47 (1994).
23. A.C. Atkinson and A.N. Donev, *Optimum Experimental Designs* (Clarendon Press, Oxford, 1992).
24. N.R. Draper and I. Guttman, *Technometrics* **30**, 105–111 (1988).
25. N.R. Draper and F. Pukelsheim, *Technometrics* **32**, 195–202 (1990).
26. A.I. Khuri, *Technometrics* **30**, 95–104 (1988).
27. S.H. Park, J.H. Lim and Y. Baba, *Annals of the Institute of Statistical Mathematics* **45**, 655–664 (1993).
28. W.R. Myers, R.H. Myers and W.H. Carter, *Journal of Statistical Planning and Inference* **42**, 57–77 (1994).
29. J.J. Borkowski, *Journal of Quality Technology* **35**, 70–77 (2003).
30. A. Giovannitti-Jensen and R.H. Myers, *Technometrics* **31**, 159–171 (1989).
31. J.J. Borkowski, *Communications in Statistics - Theory and Methods* **24**, 2041–2058 (1995).
32. G.G. Vining, *Journal of Quality Technology* **25**, 45–58 (1993).
33. C.M. Borror, *Response surface methods for experiments involving noisy variables* (Ph.D. dissertation, Arizona State University, Tempe, AZ, 1998).
34. W.J. Welch, *Technometrics* **26**, 217–224 (1984).
35. L.M. Haines, *Technometrics* **29**, 439–447 (1987).
36. J.J. Borkowski, *Journal of Probability and Statistical Science* **1**, 65–88 (2003).
37. A. Heredia-Langner, W.M. Carlyle, D.C. Montgomery, C.M. Borror and G.C. Runger, *Journal of Quality Technology* **35**, 28–46 (2003).
38. G.E.P. Box and D.W. Behnken, *Technometrics* **2**, 455–475 (1960).
39. R.H. Myers, G.G. Vining, A. Giovannitti-Jensen and S.L. Myers, *Journal of Quality Technology* **24**, 1–11 (1992).
40. J.J. Borkowski, *Communications in Statistics - Theory and Methods* **24**, 2581–2600 (1995).
41. J.J. Borkowski, *Technometrics* **37**, 399–410 (1995).
42. A.I. Khuri, H.J. Kim and Y. Um, *Computational Statistics and Data Analysis* **22**, 395–407 (1996).
43. L.A. Trinca and S.G. Gilmour, *Journal of Quality Technology* **30**, 314–327 (1998).
44. L.A. Trinca and S.G. Gilmour, *Applied Statistics* **48**, 441–455 (1999).
45. A.I. Khuri, J.M. Harrison and J.A. Cornell, *Applied Statistics* **48**, 521–532 (1999).
46. A. Zahran, C.M. Anderson-Cook and R.H. Myers, *Journal of Quality Technology* **35**, 377–386 (2003).

47. G.S. Fishman, *Monte Carlo: Concepts, Algorithms and Applications* (Springer, New York, 1996).
48. G.G. Vining and R.H. Myers, *Technometrics* **33**, 315–326 (1991).
49. T.T. Allen, L. Yu and J. Schmitz, *Applied Statistics* **52**, 103–117 (2003).
50. W.J. Welch, *Biometrika* **70**, 205–213 (1983).
51. A.C. Atkinson, *Biometrika* **59**, 275–293 (1972).
52. P. DeFeo and R.H. Myers, *Biometrika* **79**, 375–380 (1992).
53. E.R. Jones and T.J. Mitchell, *Biometrika* **65**, 541–551 (1978).
54. S.M. Stigler, *Journal of the American Statistical Association* **66**, 311–318 (1971).
55. R.D. Cook and W.K. Wong, *Journal of the American Statistical Association* **89**, 687–692 (1994).
56. A.C. Atkinson, B. Bogacka and S.S. Zocchi, *Biometrical Letters* **37**, 67–95 (2000).
57. J.J. Borkowski and E.S. Valero, *Technometrics* **43**, 468–477 (2001).
58. E. Läuter, *Mathematische Operationsforschung und Statistik* **5**, 379–398 (1974).
59. R.D. Cook and C.J. Nachtsheim, *Technometrics* **24**, 49–54 (1982).
60. D.K.J. Lin and N.R. Draper, *Technometrics* **34**, 423–428 (1992).
61. R.D. Cook, *Technometrics* **42**, 65–68 (2000).
62. D.F. Andrews and A.M. Herzberg, *Journal of Statistical Planning and Inference* **3**, 249–258 (1979).
63. I.N. Vuchkov and L.N. Boyadjieva, *Journal of Statistical Computation and Simulation* **17**, 31–41 (1983).
64. R.F. Noubiap and W. Seidel, *Journal of Statistical Planning and Inference* **91**, 151–168 (2000).
65. K.T. Fang and Y. Wang, *Number-Theoretic Methods in Statistics*. (Chapman and Hall, London, 1994).
66. W.J. Welch, *Technometrics* **24**, 41–48 (1982).
67. F. Glover, *Interfaces* **20**, 74-94 (1990).
68. L.M. Haines, Unpublished report (2004).
69. R. Mead and D.J. Pike, *Biometrics* **31**, 803–851 (1975).
70. R.N. Edmondson, *Biometrics* **47**, 1435–1448 (1991).
71. T.J. Santner, B.J. Williams and W.I. Notz, *The Design and Analysis of Computer Experiments* (Springer, New York, 2003).

This page is intentionally left blank

## **CHAPTER 12**

### **50 YEARS OF MIXTURE EXPERIMENT RESEARCH: 1955–2004**

Greg F. Piepel

*Statistical Sciences, Battelle–Pacific Northwest Division,  
P.O. Box 999, Richland, WA 99352  
E-mail: greg.piepel@pnl.gov*

The year 2004 marked the 50<sup>th</sup> anniversary of published statistical research on mixture experiments, counting from a 1955 article by Claringbold. Subsequent seminal articles by Scheffé in 1958 and 1963 provided a firm foundation for mixture experiment research over the ensuing 50 years. This article summarizes mixture experiment research from 1955–2004 published in the Statistics, Quality Engineering, and Chemometrics literature (journals, proceedings, books, and technical reports). The large number of mixture experiment publications in the literature of other disciplines precluded their inclusion in this summary. Publications per year are tallied, and are listed chronologically under several topic areas: (1) designs, (2) methods for assessing and comparing designs, (3) models, (4) other data analysis topics, and (5) reviews, summaries, and tutorials. The publications in each topic area are briefly discussed. Finally, ideas for future mixture experiment research are presented.

#### **1. Introduction**

This article summarizes statistical research related to experiments with mixtures published during the 50-year period from 1955 to 2004. In 1955, Claringbold<sup>50</sup> published the first statistics journal article that described a mixture experiment. Three years later in 1958, Scheffé<sup>311</sup> introduced a class of mixture designs and canonical polynomial models that still are widely used. Scheffé's paper laid the foundation for much of the development of mixture experiment methods that followed. Three editions of a book by Cornell<sup>58</sup> played a key role in the documentation and application of mixture experiment methods.

Mixture experiment research (theoretical, methods, and applications) has appeared in statistics-related and other literature. A bibliography of mixture experiment publications maintained by the author contains over 700 entries, with roughly half of them appearing in non-statistics literature. It would have required too many pages to list all of these publications in this article. Hence, this summary of mixture experiment research is limited to publications in the Statistics, Quality Engineering, and Chemometrics literature. Types of publications include: journals, statistical proceedings, textbooks on statistics, and other statistical research publications (such as technical reports or bulletins).

Although the majority of the publications appearing in other literature utilize mixture experiment methods standard at the time they appeared, many present new methods. For space reasons, such publications after 1980 were excluded with a few exceptions. Publications from other literature prior to 1980 were included because of the important roles such publications had in the early years of mixture experiment research and practice. Technical reports subsequently published as journal articles were also excluded, unless the technical reports contain substantive content not included in the journal articles. Finally, only publications in English were included.

The remainder of this article is divided into several sections. Section 2 summarizes the mixture experiment problem. Section 3 discusses the chronology and authors of mixture experiment research publications from 1955 to 2004. Sections 4 to 8 list in table form and discuss mixture experiment research in the following topic areas: (1) designs, (2) methods for assessing and comparing designs, (3) models, (4) other data analysis topics, and (5) reviews, summaries, and tutorials. The tables in Sections 4 to 8 are designed to assist the reader in easily locating publications on specific topics. Finally, Section 9 discusses possible topics for future mixture experiment research.

## **2. The Mixture Experiment Problem**

In mixture experiments, the variables controlled by the experimenter are nonnegative proportional amounts of the ingredients (components) in a mixture. The proportions may be expressed as mass, mole, volume or

other fractions of the total amount of the components. When expressed as fractions of a mixture, the proportions sum to 1. Thus, for a  $q$ -component mixture in which  $x_i$  represents the proportion of the  $i^{\text{th}}$  component, the applicable mixture constraints are

$$0 \leq x_i \leq 1, i = 1, 2, \dots, q \text{ and } \sum_{i=1}^q x_i = 1. \quad (1)$$

The composition space of the  $q$  components defined by the restrictions in Eq. (1) takes the form of a  $(q-1)$ -dimensional simplex. Additional constraints may be imposed on the component proportions in the form of lower and upper bounds on single components

$$0 < L_i \leq x_i \leq U_i < 1, i = 1, 2, \dots, q \quad (2)$$

or on linear multicomponent expressions

$$C_j \leq A_{1j}x_1 + A_{2j}x_2 + \dots + A_{qj}x_q \leq D_j, j = 1, 2, \dots, h, \quad (3)$$

where  $L_i$ ,  $U_i$ ,  $A_{ij}$ ,  $C_j$ , and  $D_j$  are constants. Constraints of the forms in Eqs. (2) and/or (3) may alter the shape of the experimental region from a simplex to an irregularly-shaped convex polyhedron inside the simplex. Thus, the experimental region for a mixture experiment is the whole simplex, or a smaller simplex or a convex polyhedron inside the whole simplex. In any case, the region has dimensionality  $q-1$ . Because the component proportions must sum to 1, mixture experiments are different from experiments in which the variables can be changed independently.

In a classical mixture experiment, the total amount of the mixture is held fixed and only the relative proportions of the components are varied. However, in a *mixture-amount (MA)* experiment, the responses can depend on the amount of the mixture as well as the relative proportions of the components. In a *mixture-process variable (MPV)* experiment, the responses can depend on the levels of non-mixture factors (denoted *process variables*, PVs) as well as the mixture component proportions. Non-mixture PVs can be quantitative or qualitative.

In mixture, MA, and MPV experiments: (1) data are collected according to an experimental design, (2) models are developed to predict response variables as functions of the mixture composition, and (3) the resulting models are used to understand how composition affects the responses and to develop optimum mixture compositions. Subsequent

sections list and discuss publications associated with designs, models, and data analysis methods for experiments with mixtures.

### **3. Mixture Experiment Research Chronology**

Table 1 lists chronologically the authors of mixture experiment publications (as discussed in Section 1) that appeared from 1955 to 1980. Within each year the authors are listed in alphabetical order. For all 50 years from 1955 to 2004, the number of publications in each year is also provided in Table 1. Space limitations did not allow including the listing of authors for all 50 years of publications, but such a version of Table 1 is available from the author upon an email request.

In the Introduction, the published statistical literature on mixture experiments was stated to have begun with Claringbold's 1955 article. However, in commenting on the mixture experiment models proposed by Scheffé<sup>311</sup>, Quenouille<sup>303</sup> noted earlier work on models for mixtures (pp. 260-262)

in his 1953 book<sup>302</sup>. This elicited a reply from Scheffé<sup>312</sup>, who acknowledged that not everyone will agree on which model forms are best for describing the shapes of mixture response surfaces. This mild disagreement between Quenouille and Scheffé seemed to have served as fuel for later research on model forms. Because the work of Quenouille<sup>302</sup> was not about mixture experiments in the traditional sense, it is not counted as the first published research on the topic.

Authors who were especially active researchers with four or more publications in the early period of 1955-1980 were: Snee (10), Cornell (9), Nigam (7), Gorman (6), Draper (6), Lambrakis (5), Kiefer (5), Becker (4), and St. John (4). Key publications by McLean and Anderson (1966) and Cox (1971) have also had lasting impact over the years.

From 1955 to 2004, the researchers with 10 or more mixture experiment publications listed in the Reference section are: Cornell (53), Piepel (51), Draper (18), Redgate (14), Snee (13), Montgomery (12), Chan (11), Gorman (10), and Prescott (10).

Many of the publications summarized in Table 1 and listed in the References were authored by applied statisticians and researchers

from other disciplines. This suggests that research on many mixture experiment topics was motivated by the need to solve real problems.

Table 1. A Chronological Listing of the Statistical Research Literature on Mixture Experiments from 1955 to 2004.

Year	Authors of Publications	#
1955	Claringbold	1
1958	Scheffé	1
1959	Quenouille	1
1961	Scheffé; Kiefer	2
1962	Gorman & Hinman; Wagner & Gorman	2
1963	Kenworthy; Scheffé; Wagner & Gorman	3
1964	Uranisi	1
1965	Bownds, Kurotori & Cruise; Draper & Lawrence (1965a,b)	3
1966	Cruise; Box & Gardner; Gorman; Kurotori; McLean & Anderson; Murty; Nelder	7
1967	Diamond; Drew; Farrell, Kiefer & Walbran	3
1968	Becker; Lambrakis (1968a,b,c); Murty & Das; Thompson & Myers	6
1969	Atwood; Becker; Kennard & Stone; Lambrakis (1969a,b); Watson	6
1970	Becker; Cornell & Good; Gorman; Nigam	4
1971	Cornell; Cox; Li; Paku, Manson & Nelson; Snee; Van Schalkwyk	6
1972	Narcy & Renaud;	1
1973	Cornell; Nigam; Saxena & Nigam; Snee	4
1974	Anderson & McLean; Marquardt & Snee; Nigam; Snee & Marquardt	4
1975	Cornell; Cornell & Ott; Draper & St. John (1975a,b); Kiefer; Laake; Mendieta <i>et al.</i> ; Rusin; Snee (1975a, 1975b)	10
1976	Nigam; Snee & Marquardt; Sobolev & Chemleva	3
1977	Cornell; Draper & St. John (1977a,b); Galil & Kiefer (1977a,b); Hare & Brown; Saxena & Nigam	7
1978	Becker; Cornell & Gorman; Kiefer; Park; Vuchkov <i>et al.</i>	5
1979	Cornell; Cornell & Khuri; Goel & Nigam; Hare; Morris & Snee; Snee (1979a,b)	7
1980	Daniel & Wood; Goel (1980a,b); Piepel	4

#### Number of Mixture Experiment Research Publications by Year from 1981 to 2004

Year	#	Year	#	Year	#	Year	#
1981	5	1987	10	1993	14	1999	11
1982	8	1988	16	1994	9	2000	14
1983	10	1989	11	1995	14	2001	7
1984	8	1990	13	1996	14	2002	13
1985	11	1991	10	1997	16	2003	11
1986	6	1992	11	1998	16	2004	15

#### 4. Research on Mixture Experiment Design Topics

Table 2 lists publications associated with mixture experiment design. Mixture design topics are listed in the table in chronological order of the first publication in each topic area. Following each underlined topic are the authors of the publications(s) that addressed the topic. Within each topic, the authors' names are listed in chronological order of the publication. The number in parentheses next to the author(s) name is the year the publication appeared in print. Several of the more important design topics listed in Table 2 are now briefly discussed.

Simplex Designs: Simplex designs and the years they were introduced include: simplex-lattice (1958), simplex-centroid (1963), augmented simplex-centroid (1965), axial (1977), and radial (1978). As early as 1961, research occurred to show that some of these designs were optimal for given mathematical optimality criteria (*e.g.*, D-optimality) and Scheffé model forms. Subsequent research over the years developed optimal simplex designs for many different mixture model forms (see Section 6) and optimality criteria (*e.g.*, D-, G-, V-, and A-optimality).

MIV Designs: Several approaches for constructing mixture designs from mathematically independent variable (MIV) designs have been discussed. All approaches involve transforming the  $q$  mixture components to  $q-1$  MIVs, and choosing a factorial or response-surface type design. These design approaches include: ratio variables (1963), projection (1966), other matrix transformation (1968), categorized components with PVs (1971), and restricted region simplex (1982). The MIV approach to constructing mixture and MPV designs became less popular as software for generating designs directly in terms of mixture components and PVs became more available.

Constrained Region Designs: From 1966-1983, various approaches were proposed for extreme vertices designs on polyhedral regions defined by lower and upper bounds on the component proportions. Beginning in 1975, publications began appearing on the use of optimal design software to generate mixture designs on constrained regions. In 1979, Snee<sup>326</sup> discussed constructing mixture designs for regions defined by multi- and single-component constraints. In 1988, Piepel<sup>258</sup> provided algorithms and software to generate vertices and centroids when there are single- and/or multicomponent constraints on mixture variables and/or non-mixture variables (*e.g.*, process variables and/or the total amount of the mixture).

Table 2. Chronological Listing of Mixture Experiment Designs and Publications.

Year	Type of Design: Publications by Author (year)
1955	<u>Two-part simplex design at 3 amounts</u> : Claringbold (1955)
1958	<u>Simplex-lattice and modifications</u> : Scheffé (1958); Gorman & Hinman (1962); Wagner & Gorman (1962); Lambrakis (1969a); Chen & Zhu (1984); Cornell (1986); Chasalow & Brand (1995)
1961	<u>Alphabetic-optimal for various models on simplex regions</u> : Kiefer (1961, 1978); Uranisi (1964); Farrell <i>et al.</i> (1967); Atwood (1969); Laake (1975); Galil & Kiefer (1977b); Guan (1983); Mikaeili (1987b, 1988, 1989, 1993); Donev (1988); Lim (1990); Liu & Neudecker (1995, 1997); Meng <i>et al.</i> (1995); Hilgers (2000); Chan & Guan (2001)
1963	<u>Simplex-centroid</u> : Scheffé (1963); Bownds <i>et al.</i> (1965); Klein (2002a,b) <u>Mixture and process or categorical variables</u> : Scheffé (1958, 1963); Murty & Das (1968); Cornell (1971); Li (1971); Hare (1979); Vuchkov <i>et al.</i> (1981); Vuchkov (1982); Cornell & Gorman (1984); John (1984); Cornell (1988); Czitrom (1988, 1989, 1992); Duineveld <i>et al.</i> (1993a,b); DuMouchel & Lane (1994); Murthy & Manga (1996); Anderson & Whitcomb (1998, 2000); Campisis <i>et al.</i> (1999); Kowalski <i>et al.</i> (2000, 2002); Goos & Donev (2003); Prescott (2004)
	<u>Factorial in ratios (<math>x_i/x_j</math>)</u> : Kenworthy (1963)
1965	<u>Augmented simplex-centroid</u> : Bownds <i>et al.</i> (1965); Kurotori (1966); Cornell (1986) <u>Minimum bias &amp; MSE, simplex</u> : Draper & Lawrence (1965a, 1965b); Becker (1970); Paku <i>et al.</i> (1971); Chakrabarti & Mandal (1995)
1966	<u>Extreme vertices</u> : McLean & Anderson (1966); Gorman (1966); Diamond (1967); Snee & Marquardt (1974, 1976); Nigam <i>et al.</i> (1983) <u>Block mixture &amp; MPV</u> : Murty (1966); Nigam (1970, 1976); Saxena & Nigam (1973); John (1984); Atkinson & Donev (1988); Czitrom (1989); Murthy & Murty (1992); Dean <i>et al.</i> (1992); Draper <i>et al.</i> (1993); Prescott <i>et al.</i> (1993, 1997); Lewis <i>et al.</i> (1994); Atkinson & Donev (1996, Section 14); Prescott (1998, 2000); Prescott & Draper (1998); Chan & Sandhu (1999); Ghosh & Liu (1999); Aggarwal, Sarin & Singh (2002); Goos & Donev (2003) <u>Projection</u> : Box & Gardner (1966); Hau & Box (1990a,b,c); Hau (1990); Duineveld <i>et al.</i> (1993a,b); Bjerke, Naes & Ellekjaer (2000); Prescott (2000); Prescott & Draper (2004)
1968	<u>Response surface in <math>q-1</math> independent variables</u> : Thompson & Myers (1968); Becker (1969); Cornell & Good (1970); Murthy & Murty (1982); Piepel & Cornell (1992, 1994); Lepeniotis & Vigezzi (1995) <u>Design with all proportions <math>&gt; 0</math></u> : Lambrakis (1968c)

Table 2. (Cont'd)  
Chronological Listing of Mixture Experiment Designs and Publications.

Year	Type of Design: Publications by Author (year)
	<u>Categorized component mixture, mixture-of-mixtures, multifactor mixture:</u> Lambrakis (1968b); Cornell & Good (1970); Nigam (1973); Kumari & Mittal (1986); Murthy & Murty (1989); Drava <i>et al.</i> (1996); Cornell & Ramsey (1997, 1998); Piepel (1999); Dingstad <i>et al.</i> (2003)
	<u>Symmetric-simplex &amp; adaptation to constrained regions:</u> Murty & Das (1968); Saxena & Nigam (1973, 1977)
1969	<u>Distance-based &amp; space-filling:</u> Kennard & Stone (1969); Snee & Marquardt (1974); Piepel <i>et al.</i> (1993b)
1971	<u>Categorized components &amp; process variables:</u> Cornell (1971) <u>Restricted simplex:</u> Narcy & Renaud (1971)
1973	<u>Sequential mixture &amp; MPV designs:</u> Nigam (1974); Vuchkov <i>et al.</i> (1978, 1981); Vuchkov (1982); Goel & Nigam (1979)
1975	<u>Alphabetic-optimal on constrained regions:</u> Mendieta <i>et al.</i> (1975); Snee (1975a, 1979a, 1981, 1985); Donev (1988); Yonchev (1988); Atkinson & Donev (1988); Meyer & Nachtsheim (1989); Piepel <i>et al.</i> (1993b); Atkinson & Donev (1996, Section 16.3); Martin <i>et al.</i> (2000); Piepel <i>et al.</i> (2002) <u>For Cox's models:</u> Cornell (1975) <u>For Scheffé models with inverse terms:</u> Draper & St. John (1975b, 1977b)
1976	<u>Screening:</u> Snee & Marquardt (1976); Piepel (1990, 1991); Piepel <i>et al.</i> (1993b) <u>Axial:</u> Cornell (1977); Chen <i>et al.</i> (1989); Meng <i>et al.</i> (1995)
1978	<u>Radial:</u> Becker (1978); Cornell & Gorman (1978)
1979	<u>Mixture experiments with multicomponent constraints:</u> Snee (1979a); Piepel (1988a, 1989a, 1989b, 2003); Koons (1989)
1980	<u>Systematic design for mixture data subject to a time trend:</u> Goel (1980a) <u>Uniform exploration of constrained regions:</u> Goel (1980b)
1982	<u>Restricted region simplex &amp; others constructed from factorials:</u> Murthy & Murty (1982, 1983, 1989); Aggarwal & Sarin (1996); Murthy & Manga (1996) <u>Slope-rotatable:</u> Park & Kim (1982, 1988) <u>Check points for model lack-of-fit:</u> Shelton (1982); Shelton <i>et al.</i> (1983)
1983	<u>Alphabetic optimal for MPV experiments:</u> Vuchkov <i>et al.</i> (1983); Welch (1984) <u>Centroids for designs:</u> Piepel (1983b)
1985	<u>For Scheffé models with log terms:</u> Chen <i>et al.</i> (1985); Zhu <i>et al.</i> (1987a,b)

**Table 2. (Cont'd)**  
**Chronological Listing of Mixture Experiment Designs and Publications.**

Year	Type of Design: Publications by Author (year)
	<u>Mixture-amount</u> : Piepel (1985); Cornell & Piepel (1986); Piepel & Cornell (1986, 1987); Cornell (1990c); Hilgers & Bauer (1995); Atkinson & Donev (1996, Section 12.4); Tian & Fang (1999); Heiligers & Hilgers (2003);
	<u>Component-amount</u> : Piepel (1985); Piepel & Cornell (1986, 1987); Cornell (1990c); Hilgers (1999); Heiligers & Hilgers (2003); Prescott & Draper (2004)
1987	<u>For log-contrast models</u> : Lim (1987); Chan (1988, 1992); Chan & Guan (1998)
1988	<u>Paired-comparison experiments</u> : Charnet & Beaver (1988); Ghani (1992)
	<u>Single- or multi-variable constraints on mixture and/or non-mixture variables</u> : Piepel (1988a, 1989a, 1989b, 2003)
	<u>D-optimal for mixture &amp; qualitative factors or blocks</u> : Donev (1989); Atkinson & Donev (1996, Section 13.5); Goos & Donev (2003)
1990	<u>Symmetric constrained</u> : Crosier (1990, 1991)
1992	Candidate-point-free optimal design: Hardin & Sloane (1992); Piepel <i>et al.</i> (2004)
1993	<u>Central composite analogue</u> : Piepel <i>et al.</i> (1993b)
	<u>Layered</u> : Piepel <i>et al.</i> (1993b); Chan & Guan (1998); Piepel <i>et al.</i> (1999); Piepel <i>et al.</i> (2002); Cooley <i>et al.</i> (2003); Piepel, Cooley & Jones (2004)
	<u>Minimum bias &amp; MSE, constrained</u> : Piepel <i>et al.</i> (1993b)
1994	<u>Robust design, noise factors</u> : Tsubaki (1994); Goldfarb, Borror, <i>et al.</i> (2004b)
	<u>Bayesian D-optimal &amp; model robust</u> : DuMouchel & Jones (1994); Andere-Rendon <i>et al.</i> (1997); Heredia-Langner <i>et al.</i> (2004)
	<u>Augmentation with components and/or design points</u> : Hrma <i>et al.</i> (1994); Piepel <i>et al.</i> (1995); Piepel, Cooley, <i>et al.</i> (2002); Cooley, <i>et al.</i> (2003); Piepel <i>et al.</i> (2004)
	<u>Uniform</u> : Fang and Wang (1994); Wang & Fang (1996); Chan & Guan (1998); Tian & Fang (1999); Fang & Yang (2000); Tang <i>et al.</i> (2004)
1996	<u>Extended hexagon lattice</u> : Bruns <i>et al.</i> (1996)
1998	<u>For additive quadratic or cubic models</u> : Chan <i>et al.</i> (1998a, 1998b)
2000	<u>Bayesian two-stage optimal</u> : Lin, Myers & Ye (2000)
2003	<u>Using a genetic algorithm</u> : Heredia-Langner <i>et al.</i> (2003, 2004)
	<u>D-optimal split-plot MPV</u> : Goos & Donev (2003)

Screening Designs: From 1976 to 1993, several publications by industrial and national laboratory statisticians<sup>262,263,271,331</sup> focused on screening designs for identifying important mixture components.

MPV Designs: Scheffé<sup>311,313</sup> first mentioned mixture-process variable (MPV) designs, with substantial research occurring over successive years. Some topics were: constructing MPV designs<sup>90-92,155,174,192,205,227,234,347-349</sup>, fractionating MPV designs<sup>74</sup>, and applying split-plot designs to MPV experiments<sup>62,144,193</sup>.

MA and CA Designs: Mixture-amount (MA) and component-amount (CA) designs have been discussed and compared by several publications (see 1985 entry in Table 2). A MA design is expressed in the component proportions and total amount of the mixture. A CA design is expressed in the amounts of individual components, such that factorial or response-surface type designs can be used. MA designs have the advantage of separately quantifying mixture component and total amount effects.

Block Designs: Table 2 lists 19 publications from 1966-2003 that addressed designs for mixture or MPV experiments performed in blocks. Much of the work focused on designs with blocks orthogonal to mixture effects, which can also be applied to MPV designs so PV effects are orthogonal to mixture effects.

Optimal Design: Optimal design for mixture, MA, or MPV experiments involves selecting design points to minimize a mathematical optimality criterion. Most optimality criteria (e.g., D-, G-, V-, and A-optimality) are related to minimizing the variance of coefficients or predictions for an assumed model form. Some work has focused on: (i) minimizing bias or mean-squared-error (variance plus squared bias) properties of the design<sup>16,32,106,107,244,271</sup>, and (ii) model-robust design<sup>7,122,166</sup>. Some publications address space-filling<sup>178,271,330</sup> and uniform designs (see 1994 in Table 2) using distance-based criteria that do not depend on an assumed model form. Most optimal design algorithms for mixture experiments select design points from a pre-generated candidate set<sup>258,326</sup>. However, a few publications<sup>154,165,166,274</sup> discuss algorithms that generate optimal mixture design points directly without the need for a candidate set.

Multi-layer Designs: To combat the tendency of variance-optimal design approaches to place most or all points on or near the boundary of

the experimental region, *layered designs* and *central composite analogue designs* with points on the interior and boundary of an experimental region have been proposed (see the 1993 “Layered” entry in Table 2).

Designs for Categorized Components (CC), Mixture-of-Mixtures (MoM), and Multifactor Mixture (MFM) Experiments: Some mixture experiments involve mixing component categories (denoted *major components*), which are themselves mixtures of components (denoted *minor components*). This type of experiment is referred to as a CC or MoM experiment. The MFM class of experiments is similar, but involves conducting mixture experiments for two or more independent mixture factors. Research for CC/MoM and MFM experiments has been conducted in spurts (see 1968, 1986, and 1996 in Table 2).

Robust Design: Although much literature exists for robust design with control and noise factors for non-mixture situations, only two publications<sup>143,341</sup> have addressed this topic for mixture experiments.

## 5. Research on Methods for Assessing and Comparing Mixture Experiment Designs

Table 3 lists publications associated with methods for assessing or comparing mixture experiment designs. The organization of Table 3 is similar to that described for Table 2 at the start of Section 4.

Over the years, various methods have been proposed to assess and compare mixture, MA, and MPV designs. For three (and sometimes four) mixture components, and three or fewer non-mixture variables, graphical displays of designs have been used. With more mixture components and constrained designs, pairwise component correlations and scatterplot matrices have been used. Beginning in 1974 (see Table 3), various efficiency measures based on design optimality criteria have been used to assess and compare designs. In 1985, Snee<sup>329</sup> discussed using collinearity diagnostics, variance inflation factors (VIFs) and eigenanalysis results, while in 1994 Montgomery and Voth<sup>225</sup> discussed comparing constrained designs using collinearity and leverage measures.

Other methods for assessing and comparing mixture designs have been primarily graphical. Several are based on prediction variance, including: variance trace plots (see 1989 in Table 3), variance dispersion

graph (VDG) plots (1992), variance quantile plots (1999), and fraction-of-design-space (FDS) plots (2004). In 2004, Goldfarb *et al.*<sup>140,142,143</sup> suggested 3D versions of VDG and FDS plots for MPV situations.

Table 3. Chronological Listing of Methods and Publications for Assessing and Comparing Mixture Experiment Designs.

Year	Method for Assessing Designs: Publications by Author (Year)
1966	<u>Plotting mixtures on simplex coordinates</u> : Cruise (1966)
1974	<u>Comparing designs using efficiency measures</u> : Snee & Marquardt (1974); Kiefer (1975); Galil & Kiefer (1977a,b); Piepel <i>et al.</i> (1993b); Kamoun <i>et al.</i> (2002)
1983	<u>Consistent constraints</u> : Piepel (1983a); Crosier (1984); Goos & Vandebroek (2001)
1985	<u>Collinearity diagnostics, VIFs, eigenanalysis</u> : Snee (1985)
1989	<u>Prediction variance trace plots</u> : Vining <i>et al.</i> (1989, 1993); Jang & Yoon (1997) <u>Prediction bias trace plots</u> : Vining <i>et al.</i> (1989)
1992	<u>Variance dispersion graphs for designs on polyhedral regions</u> : Piepel & Anderson (1992); Piepel, Anderson & Redgate (1993a,b, 1995)
1993	<u>Bias dispersion graphs</u> : Piepel <i>et al.</i> (1993b) <u>Plots comparing design efficiencies for several models and optimality criteria</u> : Duineveld <i>et al.</i> (1993a,b)
1994	<u>Comparing collinearity &amp; leverage</u> : Montgomery & Voth (1994)
1995	<u>Scatterplot matrix</u> : Piepel <i>et al.</i> (1995a, 2002)
1996	<u>Component slope variance trace plots</u> : Jang & Na (1996)
1998	<u>Simulation to assess optimal mixture sensitivity to designs &amp; models</u> : Brandvik (1998) <u>Kiefer ordering</u> : Draper & Pukelsheim (1999); Draper <i>et al.</i> (1998, 1999)
1999	<u>Prediction variance quantile plots</u> : Khuri <i>et al.</i> (1999)
2004	<u>3D variance dispersion graphs for MPV experiments</u> : Goldfarb, Borror, <i>et al.</i> (2004a,b) <u>Fraction-of-design-space plots</u> : Goldfarb, Borror, <i>et al.</i> (2004b); Goldfarb, Anderson-Cook, <i>et al.</i> (2004); Godfrey <i>et al.</i> (2004)

## 6. Research on Mixture Experiment Models

Table 4 lists publications associated with models for mixture experiment data. The organization of Table 4 is similar to that described for Table 2 at the start of Section 4.

Several early publications discussed fitting mixture experiment data to standard polynomial models after expressing the mixtures in terms of  $q-1$  mathematically independent variables (MIVs). The MIVs were defined as ratios (see 1963 in Table 4) or other specified transformations (see 1955 in Table 4) of the component proportions. However, as Scheffé and other mixture models became better known and researchers learned how to fit them using existing regression software, the use of standard polynomial models in MIVs rapidly declined.

Scheffé and subsequent authors (see 1958 in Table 4) presented the canonical polynomial models for mixture experiments up to the fourth degree. The first-degree or linear blending model in  $q$  components is

$$Y = \eta_{q,1} + \varepsilon = \sum_{i=1}^q \beta_i x_i + \varepsilon \quad (4)$$

where  $Y$  represents the observed response,  $\eta_{q,1}$  represents the expected (true) response,  $\beta_i$  is the expected response for pure component  $i$  (*i.e.*, the response at  $x_i = 1$ ), and  $\varepsilon$  is the random error in the observed response value having expectation 0 and variance  $\sigma^2$ . The second-degree or quadratic nonlinear blending model is

$$Y = \eta_{q,2} + \varepsilon = \sum_{i=1}^q \beta_i x_i + \sum_{i < j}^{q-1} \sum_{j=1}^q \beta_{ij} x_i x_j + \varepsilon \quad (5)$$

where  $\beta_{ij}$  is a measure of the quadratic nonlinear blending of components  $i$  and  $j$ . The special-cubic model adds to Eq. (5) nonlinear blending terms of the form  $\beta_{ijk} x_i x_j x_k$ , while the full-cubic model adds the special-cubic terms plus full-cubic terms of the form  $\beta_{ij} x_i x_j (x_i - x_j)$ . The special-quartic and full-quartic models are presented by Cornell (2002, pp. 71 and 93).

Table 4. Chronological Listing of Mixture Experiment Models and Publications.

Year	Type of Model: Publications by Author (Year)
1955	<u>Standard polynomials using <math>q-1</math> independent variables</u> : Claringbold (1955); Murty (1966); Thompson & Myers (1968); Becker (1969)
1958	<u>Scheffé canonical polynomials</u> : Scheffé (1958, 1963); Gorman & Hinman (1962); Wagner & Gorman (1962)
	<u>Mixture-process variable</u> : Scheffé (1958, 1963); Murty & Das (1968); Nigam (1974); Cornell (1995); Kowalski <i>et al.</i> (2000, 2002); Prescott (2004)
1963	<u>Standard polynomials using ratios of component proportions</u> : Kenworthy (1963); Snee (1973)
1966	<u>Inverse polynomials for CA experiments</u> : Nelder (1966)
1968	<u>Homogeneous-of-degree-one (for additive components)</u> : Scheffé (1961); Becker (1968); Snee (1973); Becker (1978); Cornell & Gorman (1978)
	<u>Categorized components mixture, mixture-of-mixtures, multifactor mixture</u> : Lambrakis (1968b); Nigam (1973); Kumari & Mittal (1986); Cornell & Ramsey (1997, 1998); Piepel (1999); Dingstad <i>et al.</i> (2003)
1971	<u>Cox polynomial models</u> : Cox (1971); Smith & Beverly (1997)
1975	<u>Scheffé models with inverse terms</u> : Draper & St. John (1975b, 1977a)
	<u>Power model</u> : Rusin (1975)
1976	<u>Piecewise linear models</u> : Sobolev & Chernleva (1976)
1978	<u>Inactive components</u> : Cox (1971); Becker (1978); Aitchison & Bacon-Shone (1984)
1979	<u>Gasoline blending</u> : Morris & Snee (1979); Snee (1981)
1982	<u>Intercept mixture models</u> : Snee & Rayner (1982)
1984	<u>Log contrast (for inactive or additive components)</u> : Aitchison <i>et al.</i> (1984)
1985	<u>Mixture-amount</u> : Piepel (1985, 1987, 1988b); Cornell & Piepel (1986); Piepel & Cornell (1985, 1986); Atkinson & Donev (1996, Section 12.4)
	<u>Component-amount</u> : Piepel (1985) <sup>(a)</sup> ; Piepel & Cornell (1985, 1986); Hilgers (1999)
	<u>Component-wise mixture</u> : Piepel (1985)
	<u>Additive</u> : Darroch & Waller (1985); Chan <i>et al.</i> (1998a,b); Piepel <i>et al.</i> (2002)
	<u>Scheffé models with log terms</u> : Chen <i>et al.</i> (1985); Zhu & Chan (1987); Zhu <i>et al.</i> (1987a,b)
1989	<u>Response-additive model for 2-component MA</u> : Chen <i>et al.</i> (1989)
1992	<u>Loglinear canonical model</u> : Ghani (1992)
	<u>Slack variable</u> : Piepel & Cornell (1992, 1994) <sup>(a)</sup> ; Cornell (2000a)

Table 4. (Cont'd)  
Chronological Listing of Mixture Experiment Models and Publications.

Year	Type of Model: Publications by Author (Year)
1993	<u>Binary and multinomial response</u> : Dunsmore & Ward (1993); Redgate & Piepel (1996)
1994	<u>Mechanistic model for PV or amount with parameters expanded as mixture models</u> : Hrma, Piepel, <i>et al.</i> (1994); White <i>et al.</i> (2004) <u>Partial quadratic mixture</u> : Hrma, Piepel, <i>et al.</i> (1994); Piepel & Redgate (1996b, 1997); Piepel <i>et al.</i> (2002)
1996	<u>Logistic for quantal response data</u> : Chen <i>et al.</i> (1996)
1997	<u>Kronecker homogeneous polynomials (K-models)</u> : Draper & Pukelsheim (1998); Prescott <i>et al.</i> (2002) <u>Additive spline partial least squares</u> : Durand & Sabatier (1997)
2001	<u>Generalized linear models</u> : Hamada <i>et al.</i> (2001); Godfrey <i>et al.</i> (2004)
2003	<u>Constrained canonical mixture polynomials</u> : White <i>et al.</i> (2004) <u>Modified L-pseudocomponent &amp; centered-and-scaled intercept</u> : Cornell & Gorman (2003)
2004	<u>Generalized linear mixed models, non-normal responses</u> : Robinson <i>et al.</i> (2004) <u>Component slope linear</u> : Piepel (2004)

<sup>(a)</sup> Many publications in the literature of other disciplines prior to this date.

In 1968, Becker<sup>14</sup> presented three models homogeneous-of-degree-one, which are appropriate when one or more components may have additive or negligible effects:

$$H1: \eta = \eta_{q,1} + \sum_{i=1}^{q-1} \sum_{j=i+1}^q \beta_{ij} \min(x_i, x_j) + \cdots + \beta_{12\dots q} \min(x_1, x_2, \dots, x_q) \quad (6)$$

$$H2: \eta = \eta_{q,1} + \sum_{i=1}^{q-1} \sum_{j=i+1}^q \beta_{ij} \frac{x_i x_j}{x_i + x_j} + \cdots + \beta_{12\dots q} \frac{x_1 x_2 \cdots x_q}{(x_1 + x_2 + \cdots + x_q)^{q-1}} \quad (7)$$

$$H3: \eta = \eta_{q,1} + \sum_{i=1}^{q-1} \sum_{j=i+1}^q \beta_{ij} (x_i x_j)^{1/2} + \cdots + \beta_{12\dots q} (x_1 x_2 \cdots x_q)^{1/q} \quad (8)$$

See Section 6.5 of Cornell<sup>58</sup> for additional discussion of these models.

In 1971, Cox<sup>84</sup> proposed the use of standard first- and second-degree polynomial models, with constraints on the parameters to compensate for overparameterization. The intercept in both models is the expected response value at a specified reference mixture. The coefficients of linear, crossproduct, and squared terms have interpretations related to expected changes in the response as component proportions change from the reference mixture (see Cox<sup>84</sup> and Section 6.7 of Cornell<sup>58</sup> for further explanation). In 2004, Piepel<sup>268</sup> proposed the *component slope linear mixture model*, which is a standard first-degree polynomial with a different constraint on the parameters so that the coefficients are the expected slopes of the response surface along the Cox effect directions for the components.

Draper and St. John<sup>113,114</sup> proposed adding inverse terms  $\beta_{-i}x_i^{-1}$  to the Scheffé linear and quadratic models in Eqs. (4) and (5). Such inverse terms are useful when the value of the response increases rapidly as a component approaches a zero boundary. Modifications to the basic form of the inverse term are needed when  $x_i$  can equal zero, and when the value of the response increases rapidly as a component approaches a lower bound  $L_i$  or an upper bound  $U_i$ .

From 1985-1987, three papers<sup>48,359,360</sup> proposed adding logarithmic terms  $\gamma_i \log(x_i)$  to the linear and quadratic Scheffé models in Eqs. (4) and (5) for situations when the value of the response decreases rapidly as a component approaches a zero boundary. Similar modifications as discussed in the preceding paragraph are also needed for log terms.

Additive polynomial models of the forms

$$\eta = \sum_{i=1}^q \beta_i x_i + \sum_{i=1}^q \beta_{ii} x_i^2 \quad (9)$$

and

$$\eta = \sum_{i=1}^q \beta_i x_i + \sum_{i=1}^q \beta_{ii} x_i^2 + \sum_{i=1}^q \beta_{iii} x_i^3 \quad (10)$$

have been proposed<sup>39,40,94</sup> for responses where mixture component linear and nonlinear blending effects are additive.

Piepel *et al.*<sup>292</sup> proposed partial quadratic mixture (PQM) models consisting of the linear Scheffé model in Eq. (4) plus selected

crossproduct and/or squared terms. The complete Scheffé quadratic model in Eq. (5) contains only crossproduct terms and not squared terms, but both kinds of terms can be present in a PQM model. PQM models include as special cases reduced forms of the Scheffé quadratic model in Eq. (5) and the additive quadratic model in Eq. (9).

Other mixture model forms that have been proposed include: log contrast<sup>3</sup>, Kronecker homogeneous polynomials<sup>109,297</sup>, intercept<sup>332</sup>, modified L-pseudocomponent<sup>75</sup>, centered-and-scaled intercept<sup>75</sup>, piecewise linear<sup>333</sup>, power model<sup>307</sup>, and constrained canonical polynomial<sup>357</sup>. Models have also been discussed for responses taking binary<sup>124,305</sup>, multinomial<sup>305</sup>, proportion<sup>124</sup>, or count<sup>46</sup> values.

Models for mixture-process variable (MPV) experiments were first discussed by Scheffé<sup>311</sup> and subsequently by several authors (see 1958 in Table 4). Typically, MPV models are formed by crossing separate models for the mixture components and the process variables. However, other MPV model forms can be obtained by addition as well as crossing operations, such as

$$\eta = \left( \sum_{i=1}^q \beta_i x_i \right) \left( \alpha_0 + \sum_{k=1}^K \alpha_k z_k \right) + \sum_{i < j}^{q-1} \sum_j^q \beta_{ij} x_i x_j \quad (11)$$

The choice of addition and crossing operations to form MPV models depends on expectations regarding how the mixture linear and nonlinear blending properties are affected or not affected by the PVs.

Models for MA experiments have been discussed in several publications (see 1985 in Table 4). MA models are generally analogous to MPV models with one PV. However, mixture experiments when the total amount takes a zero value (MAZ) require different models<sup>257,259</sup>.

Finally, models for CC/MoM and MFM experiments (see Section 4) have been discussed by several authors (see 1968 in Table 4). However, only some of these papers contain examples of fitting the models to data.

## 7. Research on Other Data Analysis Topics

Table 5 lists publications associated with data analysis topics (other than models) for mixture experiment data. The organization of Table 5 is

similar to that described for Table 2 at the start of Section 4. Because a wide variety of data analysis topics, both analytical and graphical are listed in Table 5, we briefly mention only a few key topics.

Table 5. Chronological Listing of Data Analysis Topics for Experiments with Mixtures and Associated Publications.

Year	Data Analysis Topic: Publications by Author (Year)
1965	<u>Contour plots</u> : Bownds <i>et al.</i> (1965) <u>Pseudocomponent and other component transformations</u> : Bownds <i>et al.</i> (1965); Kurotori (1966); Gorman (1966, 1970); Snee & Rayner (1982); Prescott <i>et al.</i> (2002); Cornell & Gorman (2003)
1966	<u>Ill-conditioning/collinearity</u> : Gorman (1966, 1970); Snee & Rayner (1982); St. John (1984); Altekar (1991); Redgate <i>et al.</i> (1992); Montgomery & Voth (1994); Prescott <i>et al.</i> (2002); Cornell & Gorman (2003) <u>Analysis of projection designs</u> : Box & Gardner (1966); Hau & Box (1990a,b,c); Bjerke <i>et al.</i> (2000)
1968	<u>Model coefficients for simplex-lattice designs</u> : Lambrakis (1968a)
1969	<u>Adaptations of standard regression methods for mixtures</u> : Becker (1969)
1971	<u>Assessing component effects, additivity, and interactions</u> : Cox (1971); Snee & Marquardt (1976); Piepel (1980, 1982); Darroch & Waller (1985); Hare (1984, 1985); Mikaeili (1987a); Piepel & Redgate (1996a, 1996b, 1997, 1998); Piepel, Hicks, <i>et al.</i> (2002)
1973	<u>Screening components and model reduction</u> : Snee (1973); Snee & Marquardt (1976); Park (1978); Piepel & Redgate (1996a, 1996b, 1997, 1998)
1974	<u>Test statistics for mixture models</u> : Marquardt & Snee (1974); Park (1978)
1975	<u>Nonlinear blending model structure</u> : Rusin (1975) <u>Gradient plots</u> : Cornell & Ott (1975); Snee (1975b) <u>Response trace, multiple trace, and effects plots</u> : Snee (1975b); Snee & Marquardt (1976); Piepel (1980, 1982); Hare (1984, 1985); Piepel <i>et al.</i> (1993); Lepeniotis & Vigezzi (1995); Jang & Yoon (1997); Piepel & Redgate (1996a, 1996b, 1997, 1998); Piepel, Hicks, <i>et al.</i> (2002); Kalicin (2003)
1976	<u>3D and/or contour plots</u> : Hare & Brown (1976); Koons & Heasley (1981); Cornell <i>et al.</i> (1983); Hare (1984, 1985)
1977	<u>Weighted vs. unweighted fits of Scheffé models</u> : Cornell (1977)
1979	<u>Blocking for sequential exploration</u> : Goel & Nigam (1979)
1982	<u>Model lack of fit</u> : Shelton (1982); Shelton <i>et al.</i> (1983); Bruns <i>et al.</i> (1996) <u>MPV model reduction</u> : Gorman & Cornell (1982) <u>Analysis with block effects</u> : Singh <i>et al.</i> (1982)

Table 5. (Cont'd)

Chronological Listing of Data Analysis Topics for Experiments with Mixtures and Associated Publications.

Year	Data Analysis Topic: Publications by Author (Year)
1983	<u>Using SAS PROC REG to analyze mixture data</u> : Showers (1983)
1984	<u>Ridge regression</u> : St. John (1984); Ghani (1992); Jang & Yoon (1997)
1985	<u>Overlaid contour plots</u> : Koons & Wilt (1985); Heinsman & Montgomery (1995) <u>Modeling a mixture response surface with eutectics</u> : Gorman & Cornell (1985) <u>MA model reduction</u> : Piepel (1985) <u>Analysis of symmetric simplex designs</u> : Singh & Pratap (1985)
1986	<u>Selecting optimal formulations</u> : Cain & Price (1986)
1987	<u>Ridge analysis</u> : Hoerl (1987); Peterson (1993); Draper & Pukelsheim (2002) <u>Bayesian control of mixture processes</u> : Berliner (1987)
1988	<u>Multiresponse optimization, desirability functions and other approaches</u> : Bohl (1988); Chitra (1990); Del Castillo <i>et al.</i> (1993); Heinsman & Montgomery (1995); Peterson (2000); Gupta (2001); Peterson (2004) <u>Analysis of split-plot MPV</u> : Cornell (1988); Kowalski <i>et al.</i> (2002); Robinson <i>et al.</i> (2004)
1990	<u>Robustness to noise factors and/or noise in mixture compositions</u> : De Boer <i>et al.</i> (1990, 1991, 1992); Tsubaki (1994); Steiner & Hamada (1997); Naes <i>et al.</i> (1998); Goldfarb <i>et al.</i> (2003)
1991	<u>Comparing performance of Scheffé-type versus subject-matter models</u> : Bures <i>et al.</i> (1991); Piepel, Redgate & Masuga (1995, 1996)
1992	<u>Partial least squares and/or principal components analysis in mixture or MPV experiments</u> : Kettaheh-Wold (1992); Wegscheider & Walner (1993); Lepeniotis & Vigezzi (1995); Brandvik & Daling (1998a,b) <u>Second-order model selection</u> : Redgate <i>et al.</i> (1992)
1993	<u>Optimal mixtures with binary response models</u> : Dunsmore & Ward (1993) <u>Biplot display for multiple responses</u> : Smith & Cornell (1993)
1994	<u>Multicriteria steepest ascent via local designs &amp; models</u> : Duineveld <i>et al.</i> (1994); Duineveld & Coenegracht (1995)
1997	<u>Fuzzy set approach combining two overlapping models into one</u> : Setz <i>et al.</i> (1997)
2001	<u>Optimizing Z-scores</u> : Stanard (2001)
2001	<u>Sizing fixed effects &amp; computing power</u> : Oehlert & Whitcomb (2001) <u>Grobner basis methods</u> : Giglio <i>et al.</i> (2001)
2002	<u>Confidence region for optimum point</u> : Peterson <i>et al.</i> (2002); Cahya <i>et al.</i> (2004)
2003	<u>MPV experiments with noise factors</u> : Goldfarb <i>et al.</i> (2003); Goldfarb, Borror, <i>et al.</i> (2004) <u>Adjustment &amp; normalization of analyzed compositions</u> : Weier & Piepel (2003)

In 1965-1966, pseudocomponent transformations were proposed<sup>22,195</sup> for mixture components with lower bounds. Such transformations allowed the existing simplex designs and models to be applied. The value of pseudocomponent transformations in reducing collinearity for mixture regions defined by constraints of the forms in Eqs. (2) and (3) was later recognized<sup>146,147</sup>. Other component transformations and methods to address collinearity have also been proposed<sup>75,173,297,332,334</sup>.

Various graphical methods have been proposed for understanding the linear and nonlinear blending effects of mixture components. These include: contour and 3D surface plots (see 1965 and 1976 in Table 5), gradient plots<sup>79,325</sup>, response trace and effects plots (see 1975), biplot displays<sup>321</sup>, and component interaction plots<sup>281</sup>.

Methods for assessing component effects, additivity, and interactions have been discussed in publications spanning 1971-2002 (see Table 5). The related topics of screening components and model reduction have been discussed in publications from 1973-1998 (see Table 5).

In 1974, a key paper by Marquardt and Snee<sup>211</sup> discussed test statistics for mixture models. They noted that some portions of standard analysis of variance (ANOVA) table and related statistics (*e.g.*,  $R^2$ ) for no-intercept models are incorrect for mixture experiment models and presented the proper formulas. Unfortunately, some publications still appear containing incorrect results.

Various methods for multiresponse optimization have been discussed, including: overlaid contour plots, various types of desirability functions, constrained nonlinear optimization, and a Bayesian posterior predictive approach (see 1985 and 1988 in Table 5).

Several publications have addressed the analysis of MPV experiments, including: MPV model reduction<sup>148</sup>, split-plot MPV experiments<sup>62,193,306</sup>, and MPV experiments with noise factors<sup>141,143</sup>.

Seven publications from 1990-2003 (see Table 5) address robustness to noise factors in mixture experiments and accounting for noise in mixture compositions. A 2002 publication<sup>355</sup> discusses methods to adjust and normalize analyzed component proportions so they sum to 1.

The application of partial least squares and/or principal components analysis to mixture and MPV experiments has been discussed by five publications beginning in 1992 (see Table 5).

## 8. Review, Summary, and Tutorial Publications

Table 6 lists review, summary, and tutorial publications. Foremost among these are the three editions of *Experiments with Mixtures* by Cornell (1981, 1990a, 2002)<sup>58</sup>.

Table 6. Chronological Listing of Review, Summary, and Tutorial Publications for Experiments with Mixtures.

Year	Nature of Review, Summary, or Tutorial	Publication
1967	Mixture methods (non-statistics journal)	Drew (1967)
1971	Mixture designs and models	Snee (1971)
1973	Mixture literature from 1955 to 1971	Cornell (1973)
1974	Review of simplex and extreme vertices designs	Anderson & McLean (1974, Section 13.1.4)
1975	Mixture designs, models & methods	Draper & St. John (1975a)
1979	Mixture literature from 1973 to 1979	Cornell (1979)
	Mixture experiments with examples	Snee (1979b)
1981	<i>Experiments with Mixtures</i> book	Cornell (1981, 1990a, 2002)
1983	Mixture "How To" booklet summarizing the design & analysis of mixture experiments	Cornell (1983, 1990b)
1985	Article in statistics encyclopedia	Cornell (1985)
1987	Chapter in response surface book giving overview of mixture designs and analyses	Khuri & Cornell (1987, 1996), Chapter 9
1990	Book chapter on mixture experiment methods	Cornell (1990d)
1991	Simplex designs and mixture models	Cornell & Linda (1991)
	Summary & review of mixture methods	Cornell (1991a)
	Catalog of mixture experiment examples	Piepel & Cornell (1991, 2004)
1992	Five approaches for mixture experiments with examples and recommendations	Piepel & Cornell (1992, 1994)
1995	Review of mixture and MA models and optimal designs for these models	Chan (1995, 2000)
1997	Survey of software with mixture capabilities	Piepel (1997, 1998)
	Mixture, MA, and MPV designs & models	Cornell & Harrison (1997)
	Tutorial using margarita example	Bowles & Montgomery (1987)
1998	Tutorial with 10-step process illustrated using two examples, Chemometrics aspects	Eriksson <i>et al.</i> (1998)
	Review of mixture designs & models, discussion of needed future research	Cornell (1998, 2000b)
1999	Review of optimal design algorithms and software, summary of designs for different shapes of 3-component constrained regions	Martin <i>et al.</i> (1999)

## **9. Topics for Future Research and Software Implementation**

Following are some possible topics for future mixture experiment research. Ideas for extending the mixture experiment capabilities of statistical software are also mentioned.

### ***9.1 Future Research Topics for Experimental Design***

1. Develop methods for generating optimal designs over constrained regions for mixture, mixture-amount (MA), mixture-process variable (MPV), categorized components/mixture-of-mixtures (CC/MoM), and multifactor mixture (MFM) experiments that do not first require generating candidate design points.
2. Develop methods for generating space-filling designs for constrained mixture, MA, MPV, CC/MoM, and MFM experiments that locate sufficient points on the interior of the experimental region as well as the boundary, regardless of the dimensionality of the problem.
3. Develop methods and software for generating optimal designs for mixture, MA, and MPV experiments (simplex and constrained regions) to minimize bias or MSE criteria rather than only variance-based criteria (*e.g.*, D- and G-optimality).
4. Develop model-free or model-robust techniques for generating and/or fractionating mixture and crossed MA & MPV designs.
5. Develop concepts similar to resolution, minimum aberration, etc. for fractional mixture, MA, and MPV designs.
6. Extend recent work in optimizing designs with hard-to-change and easy-to-change factors (and other randomization restrictions) for non-mixture experiments to MPV, CC/MoM, and MFM experiments.

### ***9.2 Future Research Topics for Mixture Experiment Models***

1. Publish case-study papers where Becker, inverse term, log contrast and other non-Scheffé models outperform Scheffé models.
2. Develop models nonlinear in their parameters and publish papers with examples where such models outperform linear-in-parameters mixture

models. Adapting *inverse polynomial models* (Nelder<sup>236</sup>) for mixture experiments is one option.

3. Publish case-study papers where models for binary, multinomial, or count response variables are used.
4. Develop and illustrate methods for stepwise regression fitting of Scheffé and other mixture models. Such procedures would drop terms consisting of components that have zero effects or combine terms involving components that have similar effects.

### 9.3 Suggestions for Statistical Software

1. Implement designs for CC/MoM and MFM experiments.
2. Implement designs for experimental regions defined by linear equality constraints (of the form in Eq. (3), but with equal signs) in addition to linear inequality constraints.
3. Provide for blocking mixture, MA, and MPV designs.
4. Provide for fitting and comparing the fits of mixture model forms other than Scheffé canonical polynomials (see Section 6).
5. Provide for fitting user-specified mixture models and obtain the correct ANOVA table and summary statistics.
6. Provide automated methods for reducing the number of components in linear and other forms of mixture experiment models.
7. Implement variable selection methods (such as stepwise regression) for nonlinear blending terms of various forms, including crossproduct terms, additive quadratic and cubic terms, inverse terms, logarithmic terms, and Becket H1, H2, or H3 terms (see Section 6).

## References

1. Aggarwal, M.L. and V. Sarin, "Use of Asymmetrical Factorials for Generating Designs for Quadratic Mixture Model", *Statistics and Probability Letters*, 27, 275-279 (1996).
2. Aggarwal, M.L., V. Sarin and P. Singh, "Optimal Designs in Two Blocks for Becker's Mixture Models in Three and Four Components", *Statistics & Probability Letters*, 59, 385-396 (2002).
3. Aitchison, J. and J. Bacon-Shone, "Log Contrast Models for Experiments with Mixtures", *Biometrika*, 71, 323-330 (1984).

4. Altekar, M., "Ill-Conditioning in Mixture Experiments", *1991 Proceedings of the Section on Quality and Productivity*, 28-31 (American Statistical Association, Alexandria, VA, 1991).
5. Altekar, M., "Use of Binders in Tablets—Developing A Formulating Guide Through Statistically Designed Experiments", *1997 Proceedings of the Section on Quality and Productivity*, 57-60 (American Statistical Association, Alexandria, VA, 1997).
6. Altekar, M. and A.N. Scarlatti, "Resin Vehicle Characterization Using Statistically Designed Experiments", *Chemometrics and Intelligent Laboratory Systems*, 36, 207-211 (1997).
7. Andere-Rendon, J., D.C. Montgomery and D.A. Rollier, "Design of Mixture Experiments Using Bayesian D-Optimality", *Journal of Quality Technology*, 29, 451-463 (1997).
8. Anderson, V.L. and R.A. McLean, *Design of Experiments: A Realistic Approach*, (Marcel Dekker, New York, 1974) pp. 335-348.
9. Anderson, M.J. and P.J. Whitcomb, "DOE on Paraffin Blend for Therapeutic Bath", *1998 Proceedings of the Section on Physical and Engineering Sciences*, pp. 100-103 (American Statistical Association, Alexandria, VA, 1998).
10. Anderson, M.J. and P.J. Whitcomb, "How to Design and Analyze Mixture Designs that Include Process Factors and/or Categorical Variables", *Proceedings of the Industrial Statistics in Action International Conference*, Vol. 1, 38-49, Edited by S. Coleman, D. Stewardson and L. Fairbairn, University of Newcastle, UK (2000).
11. Atkinson, A.C. and A.N. Donev, "Algorithms, Exact Designs, and Blocking in Response Surface and Mixture Designs", from *Optimal Design and Analysis of Experiments*, pp. 61-69, Y. Dodge, V.V. Federov and H.P. Wynn (Editors), (North Holland, 1988).
12. Atkinson, A.C. and A.N. Donev, *Optimum Experimental Designs*, (Clarendon Press, Oxford, UK, 1996).
13. Atwood, C.L., "Optimal and Efficient Designs of Experiments", *Annals of Mathematical Statistics*, 40, 1570-1602 (1969).
14. Becker, N.G., "Models for the Response of a Mixture", *Journal of the Royal Statistical Society, B*, 30, 349-358 (1968).
15. Becker, N.G., "Regression Problems When the Predictor Variables are Proportions", *Journal of the Royal Statistical Society, B*, 31, 107-112 (1969).
16. Becker, N.G., "Mixture Designs for a Model Linear in the Proportions", *Biometrika*, 57, 329-338 (1970).
17. Becker, N.G., "Models and Designs for Experiments with Mixtures", *Australian Journal of Statistics*, 3, 195-208 (1978).
18. Berliner, L.M., "Bayesian Control in Mixture Models", *Technometrics*, 29, 455-460 (1987).
19. Bjerke, F., T. Naes and M.R. Ellekjaer, "An Application of Projection Design in Product Development", *Chemometrics and Intelligent Laboratory Systems*, 51, 23-36 (2000).
20. Bohl, A.H., "A Formulation Tool", *Chemtech*, May, 284-289 (1988).

21. Bowles, M.L. and D.C. Montgomery, "How to Formulate the Ultimate Margarita: A Tutorial on Experiments with Mixtures", *Quality Engineering*, 10(2), 239-253 (1997).
22. Bownds, J.M., I.S. Kurotori and D.R. Cruise, *Notes on Simplex Models in the Study of Multi-component Mixtures*, NAVWEPS Report 8670, NOTS TP 3719, Copy 197, (U.S. Naval Ordnance Test Station, China Lake, CA, 1965).
23. Box, G.E.P. and C.J. Gardner, *Constrained Designs—Part 1, First Order Designs*, Technical Report No. 89, Department of Statistics, University of Wisconsin, Madison, WI (1966).
24. Brandvik, P.J., "Statistical Simulation as an Effective Tool to Evaluate and Illustrate the Advantage of Experimental Designs and Response Surface Methods", *Chemometrics and Intelligent Laboratory Systems*, 42, 51-61 (1998).
25. Brandvik, P.J. and P.S. Daling, "Optimisation of Oil Spill Dispersant Composition by Mixture Design and Response Surface Methods", *Chemometrics and Intelligent Laboratory Systems*, 42, 63-72 (1988a).
26. Brandvik, P.J. and P.S. Daling, "Optimising Oil Spill Dispersants as a Function of Oil Type and Weathering Degree: A Multivariate Approach Using Partial Least Squares (PLS)", *Chemometrics and Intelligent Laboratory Systems*, 42, 73-91 (1988b).
27. Bruns, R.E., J.C. De Andrade, C. Reis and D. Nakai, "Is Statistical Lack of Fit a Reliable Criterion for Chemical Complexity?", *Chemometrics and Intelligent Laboratory Systems*, 33, 159-166 (1996).
28. Bures, S., F.A. Pokorny and G.O. Ware, "Shrinkage in Ternary Mixes of Container Media", *Proceedings of the 1991 Kansas State University Conference on Applied Statistics in Agriculture*, 43-53 (1991).
29. Cain, M. and M.L.R. Price, "Optimal Mixture Choice", *Applied Statistics*, 35, 1-7 (1986).
30. Campisi, B., D. Vojnovic, D. Chicco and R. Phan-Tan-Luu , "Melt Granulation in a High Shear Mixer: Optimization of Mixture and Process Variables using a Combined Experimental Design", *Chemometrics and Intelligent Laboratory Systems*, 48, 59-70 (1999).
31. Cahya, S., E. Del Castillo and J.J. Peterson, "Computation of Confidence Regions for Optimal Factor Levels in Constrained Response Surface Problems", *Journal of Computational & Graphical Statistics*, 13, 499-518 (2004).
32. Chakrabarti, H. and N.K. Mandal, "Mixture Experiments: All-Variance and Minimum-Bias Designs", *Calcutta Statistical Association Bulletin*, 45(179-180), 219-234 (1995).
33. Chan, L.Y., "Optimal Design for Linear Log Contrast Model for Experiments with Mixtures", *Journal of Statistical Planning and Inference*, 20, 105-113 and erratum from Vol. 21, pg. 129 (1988).
34. Chan, L.Y., "D-Optimal Design for a Quadratic Log Contrast Model for Experiments with Mixtures", *Communications in Statistics—Theory and Methods*, 21, 2909-2930 (1992).
35. Chan, L.Y., "A Review on Optimal Design for Mixture Models", In Chan, K.Y., Liu, M.C. (Eds.), *Five Decades as a Mathematician and Educator-On the 80th*

- Birthday of Professor Yung-Chow Wong*, pp. 45-88 (World Scientific, Singapore, 1995).
36. Chan, L.Y., "Optimal Designs for Experiments with Mixtures: A Survey", *Communications in Statistics—Theory and Methods*, 29, 2281-2312 (2000).
  37. Chan, L.Y. and Y.N. Guan, "Optimal and Uniform Optimal Designs for a Log Contrast Model for Experiments with Mixtures", Department of Industrial and Manufacturing Systems Engineering, The University of Hong Kong, Hong Kong (1998).
  38. Chan, L.Y. and Y.N. Guan, "A- and D-Optimal Designs for a Log Contrast Model for Experiments with Mixtures", *Journal of Applied Statistics*, 28, 537-546 (2001).
  39. Chan, L.Y., Y.N. Guan and C.Q. Zhang, "A-Optimal Designs for an Additive Quadratic Mixture Model", *Statistica Sinica*, 8, 979-990 (1998).
  40. Chan, L.Y., J.H. Meng, Y.C. Jiang and Y.N. Guan, "D-Optimal Axial Designs for Quadratic and Cubic Additive Mixture Models", *Australian and New Zealand Journal of Statistics*, 40, 359-371 (1998).
  41. Chan, L.Y. and M.K. Sandhu, "Optimal Orthogonal Block Designs for a Quadratic Mixture Model for Three Components", *Journal of Applied Statistics*, 26, 19-34 (1999).
  42. Chardon, J., J. Nony, M. Sergent, D. Mathieu and R. Phan-Tan-Luu, "Experimental Research Methodology Applied to the Development of a Formulation for Use with Textiles", *Chemometrics and Intelligent Laboratory Systems*, 6, 313-321 (1989).
  43. Charnet, R. and R.J. Beaver, "Design and Analysis of Paired Comparison Experiments Involving Mixtures", *Journal of Statistical Planning and Inference*, 20, 91-103 (1988).
  44. Chasalow, S.D. and R.J. Brand, "Generation of Simplex Lattice Points", *Applied Statistics*, 44, 534-545 (1995).
  45. Chen, J.J., R.H. Heflich and B.S. Hass, "A Response-Additive Model for Analyzing Mixtures of Two Chemicals in the Salmonella Reversion Assay", *Biometrics Journal*, 31, 495-503 (1989).
  46. Chen, J.J., L.A. Li and C.D. Jackson, "Analysis of Quantal Response Data from Mixture Experiments", *Environmetrics*, 7, 503-512 (1996).
  47. Chen, W. and W. Zhu, "An Algorithm for Construction of Simplex Lattice Points: In Mixture Design", *Proceedings of the Statistical Computing Section*, American Statistical Association, 224-228 (1984).
  48. Chen, W.S., W.Y. Zhu and C.J. Hu, "A New Mixture Model with Three or Four Components and Their Quasi D-Optimality", Northern Illinois University, DeKalb, IL (1985).
  49. Chitra, S.P., "Multi-Response Optimization for Designed Experiments", Hercules Research Center, Wilmington, DE (1990).
  50. Claringbold, P.J., "Use of the Simplex Design in the Study of Joint Action of Related Hormones", *Biometrics*, 11, 174-185 (1955).
  51. Clayton, C.A., M.M. Goldberg and B.B. Potter, "Design and Analysis of an Experiment for Assessing Cyanide in Gold Mining Wastes", *Chemometrics and Intelligent Laboratory Systems*, 36, 181-193 (1997).

52. Cooley, S.K., G.F. Piepel, H. Gan, W.K. Kot and I.L. Pegg, "A Two-Stage Layered Mixture Experiment Design for a Nuclear Waste Glass Application", Parts 1 and 2, *2003 Proceedings of the American Statistical Association*, pp. 1036-1051 (American Statistical Association, Alexandria, VA, 2003).
53. Cornell, J.A., "Process Variables in the Mixture Problem for Categorized Components", *Journal of the American Statistical Association*, 66, 42-48 (1971).
54. Cornell, J.A., "Experiments with Mixtures: A Review", *Technometrics*, 15, 437-455 (1973).
55. Cornell, J.A., "Some Comments on Designs for Cox's Mixture Polynomial", *Technometrics*, 17, 25-35 (1975).
56. Cornell, J.A., "Weighted Versus Unweighted Estimates Using Scheffé's Mixture Model for Symmetrical Error Variance Patterns", *Technometrics*, 19, 237-247 (1977).
57. Cornell, J.A., "Experiments with Mixtures: An Update and Bibliography", *Technometrics*, 21, 95-106 (1979).
58. Cornell, J.A., *Experiments With Mixtures: Designs, Models, and the Analysis of Mixture Data*, First Edition (1981), Second Edition (1990), Third Edition (2002), John Wiley & Sons, New York (1981, 1990a, 2002).
59. Cornell, J.A., *How to Run Mixture Experiments for Product Quality*, Vol. 5, First Edition (1983), Second Edition (1990b), The ASQC Basic References in Quality Control: Statistical Techniques, (American Society for Quality Control, Milwaukee, WI, 1983, 1990b).
60. Cornell J.A., "Mixture Experiments", *Encyclopedia of Statistical Sciences*, edited by Kotz and Johnson, Vol. 5, 569-579 (1985).
61. Cornell, J.A., "A Comparison Between Two Ten-point Designs for Studying Three-component Mixture Systems", *Journal of Quality Technology*, 18, 1-15 (1986).
62. Cornell, J.A., "Analyzing Data from Mixture Experiments Containing Process Variables: A Split-Plot Approach", *Journal of Quality Technology*, 20, 2-23 (1988).
63. Cornell, J.A., "Embedding Mixture Experiments Inside Factorial Experiments", *Journal of Quality Technology*, 22, 265-276 (1990c).
64. Cornell, J.A., "Mixture Experiments" in *Statistical Design and Analysis of Industrial Experiments*, edited by Subir Ghosh, Chapter 7, pp. 175-209 (1990d).
65. Cornell, J.A., "Mixture Designs for Product Improvement Studies", *Communications in Statistics—Theory and Methods*, 20(2), 391-416 (1991a).
66. Cornell, J.A., "The Fitting of Scheffé-Type Models for Estimating Solubilities of Multisolvent Systems", *Journal of Biopharmaceutical Statistics*, 1, 303-329 (1991b).
67. Cornell, J.A., "Saving Money With a Mixture Experiment", *ASQC Statistics Division Newsletter*, 14, 11-12 (1993).
68. Cornell, J.A., "Fitting Models to Data from Mixture Experiments Containing Other Factors", *Journal of Quality Technology*, 27, 13-33 (1995).
69. Cornell, J.A., "Mixture Experiment Research, Are We Done?", *1998 Proceedings of the Section on Physical and Engineering Sciences*, pp. 94-99, (American Statistical Association, Alexandria, VA, 1998).

70. Cornell, J.A., "Fitting a Slack-Variable Model to Mixture Data: Some Questions Raised", *Journal of Quality Technology*, 32, 133-147 (2000a).
71. Cornell, J.A., "Developing Mixture Models, Are We Done?", *Journal of Statistical Computing and Simulation*, 66, 127-144 (2000b).
72. Cornell, J.A. and I.J. Good, "The Mixture Problem for Categorized Components", *Journal of the American Statistical Association*, 65, 339-355 (1970).
73. Cornell, J.A. and J.W. Gorman, "On the Detection of an Additive Blending Component in Multicomponent Mixtures", *Biometrics*, 34, 251-263 (1978).
74. Cornell, J.A. and J.W. Gorman, "Fractional Design Plans for Process Variables in Mixture Experiments", *Journal of Quality Technology*, 16, 20-38 (1984).
75. Cornell, J.A. and J.W. Gorman, "Two New Mixture Models: Living with Collinearity but Removing Its Influence", *Journal of Quality Technology*, 35, 78-88 (2003).
76. Cornell, J.A. and J.M. Harrison, *Models and Designs for Experiments with Mixtures, Part II: Exploring a Subregion of the Simplex and the Inclusion of Other Factors in Mixture Experiments*, Agriculture Experiment Station Technical Bulletin 899, Institute of Food and Agricultural Sciences, University of Florida, Gainesville, FL (1997).
77. Cornell, J.A. and A.I. Khuri, "Obtaining Constant Prediction Variance on Concentric Triangles for Ternary Mixture Systems", *Technometrics*, 21, 147-157 (1979).
78. Cornell, J.A. and S.B. Linda, *Models and Designs for Experiments with Mixtures, Part I: Exploring the Whole Simplex Region*, Agricultural Experiment Station Bulletin 879, Institute of Food and Agricultural Sciences, University of Florida, Gainesville, FL (1991).
79. Cornell, J.A. and L. Ott, "The Use of Gradients to Aid in the Interpretation of Mixture Response Surfaces", *Technometrics*, 17, 409-424 (1975).
80. Cornell, J.A. and G.F. Piepel, "Experimental Designs for Mixture-Amount Experiments", *Proceedings of the 4th Conference on Statistical Methods in Experimental Research and Quality Control*, pp. 54-74, Varna, Bulgaria (1986).
81. Cornell, J.A. and P.J. Ramsey, "Modeling the Component Linear and Nonlinear Blending Properties in a Two-Stage Mixture Experiment", *Nonlinear Analysis, Theory, Methods & Applications*, 30, 4041-4050 (1997).
82. Cornell, J.A. and P.J. Ramsey, "A Generalized Mixture Model for Categorized-Components Problems With an Application to a Photoresist-Coating Experiment", *Technometrics*, 40, 48-61 (1998).
83. Cornell, J.A., J.T. Shelton, R. Lynch and G.F. Piepel, *Plotting Three Dimensional Response Surfaces for Three-Component Mixtures or Two-Factor Systems*, Institute of Food and Agricultural Sciences Bulletin No. 836, University of Florida, Gainesville, FL (1983).
84. Cox, D.R., "A Note on Polynomial Response Functions for Mixtures", *Biometrika*, 58, 155-159 (1971).
85. Crosier, R.B., "Mixture Experiments: Geometry and Pseudocomponents", *Technometrics*, 26, 209-216 (1984).

86. Crosier, R.B., "The Geometry of Constrained Mixture Experiments", *Technometrics*, 28, 95-102 (1986).
87. Crosier, R.B., *Symmetry and Design in Mixture Experiments*, CRDEC-TR-218, (U.S. Army Armament Munitions Chemical Command, Aberdeen Proving Ground, MD, 1990).
88. Crosier, R.B., "Symmetry in Mixture Experiments", *Communications in Statistics—Theory and Methods*, 20, 1911-1935 (1991).
89. Cruise, D.R., "Plotting the Composition of Mixtures on Simplex Coordinates", *Journal of Chemical Education*, 43, 30-33 (1966).
90. Czitrom, V., "Mixture Experiments With Process Variables: D-Optimal Orthogonal Experimental Designs", *Communications in Statistics—Theory and Methods*, 17(1), 105-121 (1988). Erratum from Vol. 19(3), pg. 1169 (1990).
91. Czitrom, V., "Experimental Design for Four Mixture Components With Process Variables", *Communications in Statistics—Theory and Methods*, 18, 4561-4581 (1989).
92. Czitrom, V., "Note on a Mixture Experiment With Process Variables", *Communications in Statistics, Part B: Simulation and Computation*, 21(2), 493-498 (1992).
93. Daniel, C. and F.S. Wood, *Fitting Equations to Data, Chapter 9*, (John Wiley and Sons, New York, 1980).
94. Darroch, J.N. and J. Waller, "Additivity and Interaction in Three-Component Experiments with Mixtures", *Biometrika*, 72, 153-163 (1985).
95. De Boer, J.H., A.K. Smilde and D.A. Doornbos, "Introduction of a Robustness Coefficient in Optimization Procedures: Implementation in Mixture Design Problems, Part I: Theory", *Chemometrics and Intelligent Laboratory Systems*, 7, 223-236 (1990).
96. De Boer, J.H., A.K. Smilde and D.A. Doornbos, "Introduction of a Robustness Coefficient in Optimization Procedures: Implementation in Mixture Design Problems, Part II: Some Practical Considerations", *Chemometrics and Intelligent Laboratory Systems*, 10, 325-336 (1991).
97. De Boer, J.H., A.K. Smilde and D.A. Doornbos, "Introduction of a Robustness Coefficient in Optimization Procedures: Implementation in Mixture Design Problems. Part III: Validation and Comparison With Competing Criteria", *Chemometrics and Intelligent Laboratory Systems*, 15, 13-28 (1992).
98. Dean, A.M., S.M. Lewis, P. Prescott and N.R. Draper, "Use of a Symbolic Algebra Computer System to Investigate the Properties of Mixture Designs in Orthogonal Blocks", *Computational Statistics: Proceedings of 10th Symposium on Computational Statistics*, Vol. 2, 215-220 (1992).
99. Del Castillo, E. and D.C. Montgomery, "A Nonlinear Programming Solution to the Dual Response Problem", *Journal of Quality Technology*, 25, 199-204 (1993).
100. Diamond, W.J., "Three Dimensional Models of Extreme Vertices Designs for Four Component Mixtures", *Technometrics*, 9, 472-475 (1967).
101. Dingstad, G., B. Egelandsdal and T. Naes, "Modeling Methods for Crossed Mixture Experiments A Case Study from Sausage Production", *Chemometrics and Intelligent Laboratory Systems*, 66, 175-190 (2003).

102. Donev, A.N., *The Construction of Exact D-Optimum Designs*, Ph.D. Thesis, University of London (1988).
103. Donev, A.N., "Design of Experiments with Both Mixture and Qualitative Factors", *Journal of the Royal Statistical Society, B*, 51, 297-302 (1989).
104. Draper, N.R., B. Heiligers and F. Pukelsheim, *Kiefer Ordering of Simplex Designs for Second-Order Mixture Models with Four or More Ingredients*, Report 403, Institut für Mathematik, Universität Augsburg, Germany (1998).
105. Draper, N.R., B. Heiligers and F. Pukelsheim, "Kiefer Ordering of Second-Degree Mixture Designs for Four Ingredients", *1999 Proceedings of the Section on Physical and Engineering Sciences*, pp. 1-9, (American Statistical Association, Alexandria, VA, 1999).
106. Draper, N.R. and W.E. Lawrence, "Mixture Designs for Three Factors", *Journal of the Royal Statistical Society, B*, 27, 450-465 (1965a).
107. Draper, N.R. and W.E. Lawrence, "Mixture Designs for Four Factors", *Journal of the Royal Statistical Society, B*, 27, 473-478 (1965b).
108. Draper, N.R., P. Prescott, S.M. Lewis, A.M. Dean, P.W.M. John and M.G. Tuck, "Mixture Designs for Four Components in Orthogonal Blocks", *Technometrics*, 35, 268-276 (1993). Correction in *Technometrics*, 1994, 36, 234 (1994).
109. Draper, N.R. and F. Pukelsheim, "Mixture Models Based on Homogeneous Polynomials", *Journal of Statistical Planning and Inference*, 71, 303-311 (1998).
110. Draper, N.R. and F. Pukelsheim, "Kiefer Ordering of Simplex Designs for First- and Second-Degree Mixture Models", *Journal of Statistical Planning and Inference*, 79, 325-348 (1999).
111. Draper, N.R. and F. Pukelsheim, "Generalized Ridge Analysis Under Linear Restrictions, With Particular Applications to Mixture Experiments Problems", *Technometrics*, 44, 250-259 (2002).
112. Draper, N.R. and R.C. St. John, *Models and Designs for Experiments With Mixtures: I, Background Material*, Department of Statistics Technical Report# 360, University of Wisconsin, Madison, WI (1975a).
113. Draper, N.R. and R.C. St. John, *Models and Designs for Experiments With Mixtures: II, A New Mixtures Model*, Department of Statistics Technical Report# 361, University of Wisconsin, Madison, WI (1975b).
114. Draper, N.R. and R.C. St. John, "A Mixtures Model with Inverse Terms", *Technometrics*, 19, 37-46 (1977a).
115. Draper, N.R. and R.C. St. John, "Designs in Three and Four Components for Mixtures Models with Inverse Terms", *Technometrics*, 19, 117-130 (1977b).
116. Drava, G., R. Leardi, A. Portesani and E. Sales, "Application of Chemometrics to the Production of Friction Materials: Analysis of Previous Data and Search of New Formulations", *Chemometrics and Intelligent Laboratory Systems*, 32, 245-255 (1996).
117. Drew, B.D., "Experiments with Mixtures", *The Minnesota Chemist*, 19, 4-9 (1967).
118. Duineveld, C.A.A., C.H.P. Bruins, A.K. Smilde, G.K. Bolhuis, K. Zuurman and D.A. Doornbos, "Multicriteria Steepest Ascent", *Chemometrics and Intelligent Laboratory Systems*, 25, 183-201 (1994).

119. Duineveld, C.A.A. and P.M.J. Coenegracht, "Multicriteria Steepest Ascent in a Design Space Consisting of Both Mixture and Process Variables", *Chemometrics and Intelligent Laboratory Systems*, 30, 23-36 (1995).
120. Duineveld, C.A.A., A.K. Smilde and D.A. Doornbos, "Comparison of Experimental Designs Combining Process and Mixture Variables. Part I. Design Construction and Theoretical Evaluation", *Chemometrics and Intelligent Laboratory Systems*, 19, 295-308 (1993a).
121. Duineveld, C.A.A., A.K. Smilde and D.A. Doornbos, "Comparison of Experimental Designs Combining Process and Mixture Variables. Part II. Design Evaluation on Measured Data", *Chemometrics and Intelligent Laboratory Systems*, 19, 309-318 (1993b).
122. Dumouchel, W. and B. Jones, "A Simple Bayesian Modification of D-Optimal Designs to Reduce Dependence on an Assumed Model", *Technometrics*, 36, 37-47 (1994).
123. Dumouchel, W.H. and T.P. Lane, "Construction of Mixture-Process Designs Via the Franklin-Bailey Algorithm", *1994 Proceedings of the Section on Quality and Productivity*, pp. 7-16 (American Statistical Association, Alexandria, VA, 1994).
124. Dunsmore, I.R. and N.J. Ward, "Optimal Mixture Choice for Binary Response Models", *The Statistician* (Royal Statistical Society), 42, 145-152 (1993).
125. Durand, J.F. and R. Sabatier, "Additive Splines for Partial Least Squares Regression", *Journal of the American Statistical Association*, 92, 1546-1554 (1997).
126. Eriksson, L., E. Johansson and C. Wikström, "Mixture Design—Design Generation, PLS Analysis, and Model Usage", *Chemometrics and Intelligent Laboratory Systems*, 43, 1-24 (1998).
127. Fang, K.T. and Y. Wang, "Uniform Designs of Experiments with Mixtures", Section 5.5 in *Number-Theoretic Methods in Statistics*, (Chapman & Hall, London, 1994).
128. Fang, K.T. and Z.H. Yang, "On Uniform Design of Experiments with Restricted Mixtures and Generation of Uniform Distribution on Some Domains", *Statistics and Probability Letters*, 46, 113-120 (2000).
129. Farrell, R.H., J.C. Kiefer and A. Walbran, "Optimum Multivariate Designs", In Le Cam, L.M., Neyman, J., (Eds.), *Proceedings of the 5<sup>th</sup> Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, CA, Vol 1., 113-138 (1967). Also in Kiefer, J.C., Brown, L.D., Olkin, I., Sacks, J. and Wynn, H.P., *Jack Carl Kiefer Collected Papers III. Design of Experiments*, pp. 247-272 (Springer, New York, 1985).
130. Galil, Z. and J. Kiefer, "Comparison of Box-Draper and D-Optimum Designs for Experiments with Mixtures", *Technometrics*, 19, 441-444 (1977a).
131. Galil, Z. and J. Kiefer, "Comparison of Simplex Designs for Quadratic Mixture Models", *Technometrics*, 19, 445-453 (1977b).
132. Gardenier, T.K., "Mixture Experiments and Fractional Factorials Used to Tailor Computer Simulations", *Proceedings of the 20th Symposium on the Interface*, 20, 593-598 (1988).

133. Ghani, I.A., "Ridge Regression Analysis of Paired Comparison Loglinear Canonical Model for Experiment With Mixtures", *Proceedings of the Third Islamic Countries Conference on Statistical Sciences*, Munir Ahmad editor, Islamic Society of Statistical Sciences, Lahore, Pakistan, 131-142 (1992).
134. Ghosh, S. and T. Liu, "Optimal Mixture Designs for Four Components in Two Orthogonal Blocks", *Journal of Statistical Planning and Inference*, 78, 219-228 (1999).
135. Giglio, B., H.P. Wynn and E. Riccomagno, "Groebner Basis Methods in Mixture Experiments and Generalisations", *Nonconvex Optimization and Its Applications*, 51, 33-44 (Kluwer Academic Publications, Dordrecht, 2001).
136. Godfrey, A.O., C.M. Anderson-Cook and R.H. Myers, "Fraction of Design Space Plots for Generalized Linear Models", Technical Report Number 04-3, Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, VA (2004).
137. Goel, B.S., "Systematic Designs For Experiments With Mixtures", *Biometrical Journal*, 22, 345-350 (1980a).
138. Goel, B.S., "Designs for Restricted Exploration in Mixture Experiments", *Biometrical Journal*, 22, 351-358 (1980b).
139. Goel, B.S. and A.K. Nigam, "Sequential Exploration in Mixture Experiments", *Biometrical Journal*, 21, 277-285 (1979).
140. Goldfarb, H.B., C.M. Anderson-Cook, C.M. Borror and D.C. Montgomery, "Fraction of Design Space Plots for Assessing Mixture and Mixture-Process Designs", *Journal of Quality Technology*, 36, 169-179 (2004).
141. Goldfarb, H.B., C.M. Borror, D.C. Montgomery, "Mixture-Process Variable Experiments with Noise Variables", *Journal of Quality Technology*, 35, 393-405 (2003).
142. Goldfarb, H.B., C.M. Borror, D.C. Montgomery and C.M. Anderson-Cook, "Three-Dimensional Variance Dispersion Graphs for Mixture-Process Experiments", *Journal of Quality Technology*, 36, 109-124 (2004a).
143. Goldfarb, H.B., C.M. Borror, D.C. Montgomery and C.M. Anderson-Cook, "Evaluating Mixture-Process Designs with Control and Noise Variables", *Journal of Quality Technology*, 36, 245-262 (2004b).
144. Goos, P. and A.N. Donev, *The D-Optimal Design of Blocked and Split-Plot Experiments with Mixture Components*, Research Report 0303, Departament Toegepaste Economische Wetenschappen, Katholieke Universiteit Leuven, Leuven, Belgium (2003).
145. Goos, P. and M. Vandebroek, "How to Relax Inconsistent Constraints in a Mixture Experiment", *Chemometrics and Intelligent Laboratory Systems*, 55, 147-149 (2001).
146. Gorman, J.W., "Discussion of 'Extreme Vertices Designs of Mixture Experiments' by R.A. McLean and V.L. Anderson", *Technometrics*, 8, 455-456 (1966).
147. Gorman, J.W., "Fitting Equations to Mixture Data with Restraints on Compositions", *Journal of Quality Technology*, 2, 186-194 (1970).

148. Gorman, J.W. and J.A. Cornell, "A Note on Model Reduction for Experiments With Both Mixture Components and Process Variables", *Technometrics*, 24, 243-247 (1982).
149. Gorman, J.W. and J.A. Cornell, "A Note on Fitting Equations to Freezing-Point Data Exhibiting Eutectics for Binary and Ternary Mixture Systems", *Technometrics*, 27, 229-239 (1985).
150. Gorman, J.W. and J.E. Hinman, "Simplex Lattice Designs for Multicomponent Experiments", *Technometrics*, 4, 463-487 (1962).
151. Guan, Y., *On Some Optimality of Simplex-Centroid Design*, Technical Report# 5, Department of Management Science and Engineering, Tokyo Institute of Technology, Tokyo, Japan (1983).
152. Gupta, A., "Optimization of Product Performance of a Paint Formulation Using a Mixture Experiment", *Journal of Applied Statistics*, 28, 199-213 (2001).
153. Hamada, M., V. Kutschy and S.H. Steiner, *Finding a Robust Bubble Solution Via Mixture Experiments*, LA-UR-01-1193, Los Alamos National Laboratory, Los Alamos, NM (2001).
154. Hardin, R.H. and N.J.A. Sloane, *Operating Manual for Gosset: A General Purpose Program for Constructing Experimental Designs* (Second Edition), Mathematical Sciences Research Center, AT&T Bell Laboratories, Murray Hill, NJ (1992).
155. Hare, L.B., "Designs for Mixture Experiments Involving Process Variables", *Technometrics*, 21, 159-173 (1979).
156. Hare, L.B., "Graphical Display of Mixture Responses", *1984 ASQC Quality Congress Transactions*, 315-324, American Society for Quality Control, Milwaukee, WI (1984).
157. Hare, L.B., "Graphical Display of the Results of Mixture Experiments", *Experiments in Industry-Design, Analysis, and Interpretation of Results*, pp. 99-109, edited by R.D. Snee, L.B. Hare and J.R. Trout, American Society for Quality Control, Milwaukee, WI (1985).
158. Hare, L.B. and P.L. Brown, "Plotting Response Surface Contours for Three Component Mixtures", *Journal of Quality Technology*, 9, 193-197 (1977).
159. Hau, I., *Constrained Experimental Designs*, Ph.D. thesis, University of Wisconsin, Madison, WI (1990).
160. Hau, I. and G. Box, *Constrained Experimental Designs, Part I: Construction of Projection Designs*, Report No. 53, Center for Quality and Productivity Improvement, University of Wisconsin, Madison, WI (1990a).
161. Hau, I. and G. Box, *Constrained Experimental Designs, Part II: Analysis of Projection Designs*, Report No. 54, Center for Quality and Productivity Improvement, University of Wisconsin, Madison, WI (1990b).
162. Hau, I. and G. Box, *Constrained Experimental Designs, Part III: Steepest Ascent and Properties of Projection Designs*, Report No. 55, Center for Quality and Productivity Improvement, University of Wisconsin, Madison, WI (1990c).
163. Heiligers, B. and R.D. Hilgers, "A Note on Optimal Mixture and Mixture Amount Designs", *Statistica Sinica*, 13, 709-725 (2003).

164. Heinsman, J.A. and D.C. Montgomery, "Optimization of a Household Product Formulation Using a Mixture Experiment", *Quality Engineering*, 7, 583-599 (1995).
165. Heredia-Langner, A., W.M. Carlyle, D.C. Montgomery, C.M. Borror and G.C. Runger, "Genetic Algorithms for the Construction of D-Optimal Designs", *Journal of Quality Technology*, 35, 28-46 (2003).
166. Heredia-Langner, A., D.C. Montgomery, W.M. Carlyle and C.M. Borror, "Model-Robust Optimal Designs: A Genetic Algorithm Approach", *Journal of Quality Technology*, 36, 263-279 (2004).
167. Hilgers, R.D., "Design Efficiency and Estimation for Component Amount Models", *Biometrical Journal*, 41, 783-798 (1999).
168. Hilgers, R.D., "D-Optimal Design for Becker's Minimum Polynomial", *Statistics & Probability Letters*, 49, 175-179 (2000).
169. Hilgers, R.D. and P. Bauer, "Optimal Designs for Mixture Amount Experiments", *Journal of Statistical Planning and Inference*, 48, 241-246 (1995).
170. Hoerl, R.W., "The Application of Ridge Techniques to Mixture Data: Ridge Analysis", *Technometrics*, 29, 161-172 (1987).
171. Hrma, P.R., G.F. Piepel, M.J. Schweiger, D.E. Smith, D.S. Kim, P.E. Redgate, J.D. Vienna, C.A. Lopresti, D.B. Simpson, D.K. Peeler and M.H. Langowski, *Property Composition Relationships for Hanford High-Level Waste Glasses Melting at 1150EC*, PNL-10359, Volumes 1 and 2, Pacific Northwest Laboratory, Richland, WA (1994).
172. Jang, D.H. and H.J. Na, "A Graphical Method for Evaluating Mixture Designs with Respect to the Slope", *Communications in Statistics—Theory and Methods*, 25(5), 1043-1058 (1996).
173. Jang, D.H. and M. Yoon, "Graphical Methods for Evaluating Ridge Regression Estimator in Mixture Experiments", *Communications in Statistics—Part B: Simulation and Computation*, 26(3), 1049-1061 (1997).
174. John, P.W.M., *Experiments With Mixtures Involving Process Variables*, Statistics Department Technical Report No. 8, University of Texas, Austin, TX (1984).
175. Kacker, R.N., E.S. Lagergren, M.D. Hill, W. Wong-Na, C.K. Chiang and E.R. Fuller Jr. (1991), "An Efficient Experiment to Study Superconducting Ceramics", *Communications in Statistics—Theory and Methods*, 20(2), 441-456 (1991).
176. Kalicin, S., "Trace Plots: Do You Really Know Your Mixture?", *Joint Newsletter for the Section on Physical and Engineering Sciences and the Quality and Productivity Section of the American Statistical Association*, 11, 6-9 (2003).
177. Kamoun, A., M. Chaabouni, M. Sergent and R. Phan-Tan-Luu, "Mixture Design Applied to the Formulation of Hydrotropes for Liquid Detergents", *Chemometrics and Intelligent Laboratory Systems*, 63 (1), 69-79 (2002).
178. Kennard, R.W. and L. Stone, "Computer Aided Design of Experiments", *Technometrics*, 11, 137-148 (1969).
179. Kenworthy, O.O., "Factorial Experiments with Mixtures Using Ratios", *Industrial Quality Control*, 19, 24-26 (1963).
180. Kettaneh-Wold, N., "Analysis of Mixture Data With Partial Least Squares", *Chemometrics and Intelligent Laboratory Systems*, 14, 57-69 (1992).

181. Khuri, A.I., "Invariance of Prediction from a Mixture Model Under a Nonsingular Linear Transformation", *Journal of Propagations in Probability and Statistics*, 2, 145-148 (2002).
182. Khuri, A.I. and J.A. Cornell, *Response Surfaces: Designs and Analyses*, First Edition (1987), Second Edition (1996), (Marcel Dekker, Inc., New York, 1987, 1996).
183. Khuri, A.I., J.M. Harrison and J.A. Cornell, "Using Quantile Plots of the Prediction Variance for Comparing Designs for a Constrained Mixture Region: An Application Involving a Fertilizer Experiment", *Applied Statistics*, 48, 521-532 (1999).
184. Kiefer, J.C., "Optimum Designs in Regression Problems, II", *Annals of Mathematical Statistics*, 32, 298-325 (1961).
185. Kiefer, J., "Optimal Design: Variation in Structure and Performance Under Change of Criterion", *Biometrika*, 62, 277-288 (1975).
186. Kiefer, J.C., "Asymptotic Approach to Families of Design Problems", *Communications in Statistics, Theory and Methods A*, 7, 1347-1362 (1978).
187. Klein, T., *Invariant Symmetric Block Matrices for the Design of Mixture Experiments*, Technical Report No. 443, Institut für Mathematik, Universität Augsburg, Augsburg, Germany (2002a).
188. Klein, T., *Optimal Designs for Second-Degree Kronecker Model Mixture Experiments*, Technical Report No. 444, Institut für Mathematik, Universität Augsburg, Augsburg, Germany (2002b).
189. Koons, G.F., "Effect of Sinter Composition on Emissions: A Multi-Component, Highly-Constrained Mixture Experiment", *Journal of Quality Technology*, 21, 261-267 (1989).
190. Koons, G.F. and R.H. Heasley, "Response Surface Contour Plots for Mixture Problems", *Journal of Quality Technology*, 13, 207-214 (1981).
191. Koons, G.F. and M.H. Wilt, "Design and Analysis of an ABS Pipe Compound Experiment", *Experiments in Industry—Design, Analysis, and Interpretation of Results*, pp. 111-117, edited by R.D. Snee, L.B. Hare and J.R. Trout, (American Society for Quality Control, Milwaukee, WI, 1985).
192. Kowalski, S., J.A. Cornell and G.G. Vining, "A New Model and Class of Designs for Mixture Experiments with Process Variables", *Communications in Statistics—Theory and Methods*, 29, 2255-2280 (2000).
193. Kowalski, S.M., J.A. Cornell and G.G. Vining, "Split-Plot Designs and Estimation Methods for Mixture Experiments With Process Variables", *Technometrics*, 44, 72-79 (2002).
194. Kumari, R. and S.P. Mittal, "A Note on Two Factor Mixture Experiments", *Journal of the Indian Society of Agricultural Statistics*, 38, 141-147 (1986).
195. Kurotori, I.S., "Experiments With Mixtures of Components Having Lower Bounds", *Industrial Quality Control*, 22, 592-596 (1966).
196. Laake, P., "On the Optimal Allocation of Observations in Experiments with Mixtures", *Scandinavian Journal of Statistics*, 2, 153-157 (1975).
197. Lambrakis, D.P., "Estimated Regression Function of the {q,m} Simplex-Lattice Design", *Bulletin of the Hellenic Mathematical Society*, 9, 13-19 (1968a).

198. Lambrakis, D.P., "Experiments With Mixtures: A Generalization of the Simplex-Lattice Design", *Journal of the Royal Statistical Society, B*, 30, 123-136 (1968b).
199. Lambrakis, D.P., "Experiments With p-Component Mixtures", *Journal of the Royal Statistical Society, B*, 30, 137-144 (1968c).
200. Lambrakis, D.P., "Experiments With Mixtures: An Alternative to the Simplex-Lattice Design", *Journal of the Royal Statistical Society, B*, 31, 234-245 (1969a).
201. Lambrakis, D.P., "Experiments With Mixtures: Estimated Regression Function of the Multiple-Lattice Design", *Journal of the Royal Statistical Society, B*, 31, 276-284 (1969b).
202. Lawson, J.S., "A Case Study of Effective Use of Statistical Experimental Design in a Smoke Stack Industry", *Journal of Quality Technology*, 20, 51-62 (1988).
203. Lepeniotis, S.S. and M.J. Vigezzi, "Lowering Manufacturing Cost of Material by Formulating It Through Statistical Modeling and Design", *Chemometrics and Intelligent Laboratory Systems*, 29, 133-139 (1995).
204. Lewis, S.M., A.M. Dean, N.R. Draper and P. Prescott, "Mixture Designs for q Components in Orthogonal Blocks", *Journal of the Royal Statistical Society, B*, 56, 457-467 (1994).
205. Li, J., *Designs of Experiments with Mixtures and Independent Variable Factors*, Ph.D. thesis, Rutgers, The State University, New Brunswick, NJ (1971).
206. Lim, Y.B., "Symmetric D-Optimal Designs for Log Contrast Models With Mixtures", *Journal of the Korean Statistical Society*, 16, 71-79 (1987).
207. Lim, Y.B., "D-Optimal Design for Cubic Polynomial Regression on the q-Simplex", *Journal of Statistical Planning and Inference*, 25, 141-152 (1990).
208. Lin, H., R.H. Myers and K. Ye, "Bayesian Two-Stage Optimal Design for Mixture Models", *Journal of Statistical Computation and Simulation*, 66, 209-231 (2000).
209. Liu, S. and H. Neudecker, "A V-Optimal Design for Scheffé's Polynomial Model", *Statistics and Probability Letters*, 23, 253-258 (1995).
210. Liu, S. and H. Neudecker, "Experiments with Mixtures: Optimal Allocations for Becker's Models", *Metrika*, 45, 53-66 (1997).
211. Marquardt, D.W. and R.D. Snee, "Test Statistics for Mixture Models", *Technometrics*, 16, 533-537 (1974).
212. Martin, R.J., M.C. Bursnall and E.C. Stillman, "Efficient Designs for Constrained Mixture Experiments", *Statistics and Computing*, 9, 229-237 (1999).
213. Martin, R.J., M.C. Bursnall and E.C. Stillman, "Further Results on Optimal and Efficient Designs for Constrained Mixture Experiments" (with discussion), in *Optimum Design 2000*, edited by A. Atkinson, B. Bogacka and A. Zhigljavsky, pp. 225-239, Kluwer Academic Publishers, Dordrecht/Boston/London (2000).
214. Martin, R.J., L.M. Platts, A.B. Seddon and E.C. Stillman, "The Design and Analysis of a Mixture Experiment on Glass Durability", *Proceedings of the Industrial Statistics in Action International Conference*, Vol. 1, 265-275, Edited by S. Coleman, D. Stewardson and L. Fairbairn, University of Newcastle, UK (2000).

215. Martin, R.J., L.M. Platts, A.B. Seddon and E.C. Stillman, "The Design and Analysis of a Mixture Experiment on Glass Durability", *Australian and New Zealand Journal of Statistics*, 45, 19-27 (2003).
216. Mclean, R.A. and V.L. Anderson, "Extreme Vertices Design of Mixture Experiments", *Technometrics*, 8, 447-456 (1966).
217. Mendieta, E.J., H.N. Linssen and R. Doornbos, "Optimal Designs for Linear Mixture Models", *Statistica Neerlandica*, 29, 145-150 (1975).
218. Meng, J.H., L.Y. Chan and K.W. Ng, "D-Optimal Axial Design for an Additive Cubic Mixture Model", *Proceedings of the Workshop on Quasi-Monte Carlo Methods and Their Applications*, Hong Kong Baptist University, Hong Kong (1995).
219. Meyer, R.K. and C.J. Nachtsheim, "Computer-Aided Construction of Mixture Designs for Constrained Regions of Higher Dimensions", *Proceedings of the 21st Symposium on the Interface*, pp. 491-495, American Statistical Association, Alexandria, VA (1989).
220. Mikaeili, F., "On Measuring Component Effect in Experiments with Mixtures", *Proceedings of the 9th International Conference on Quality Control*, Tokyo, Japan, pp. 559-562 (1987a).
221. Mikaeili, F., "Simplex and D-Optimum Design", Contributed Papers of the 46th Session of the International Statistical Institute, 291-292 (1987b).
222. Mikaeili, F., "Allocation of Measurements in Experiments with Mixtures", *Keio Science and Technology Reports*, 41, 25-37 (1988).
223. Mikaeili, F., "D-Optimum Design for Cubic Without 3-Way Effect on the Simplex", *Journal of Statistical Planning and Inference*, 21, 107-115 (1989).
224. Mikaeili, F., "D-optimum Design for Full Cubic on Q-Simplex", *Journal of Statistical Planning and Inference*, 35, 121-130 (1993).
225. Montgomery, D.C. and S.M. Voth, "Multicollinearity and Leverage in Mixture Experiments", *Journal of Quality Technology*, 26, 96-108 (1994).
226. Morris, W.E. and R.D. Snee, "Blending Relationships Among Gasoline Component Mixtures", *Automotive Technologist*, 3, 56-61 (1979).
227. Murthy, M.S.R. and P.L. Manga, "Restricted Region Simplex Designs for Mixture Experiments in the Presence of Process Variables", *Sankhya*, B, Indian Journal of Statistics, 58, 231-239 (1996).
228. Murthy, M.S.R. and J.S. Murty, "A Note on Mixture Designs Derived from Factorials", *Journal of the Indian Society of Agricultural Statistics*, 34, 87-93 (1982).
229. Murthy, M.S.R. and J.S. Murty, "Restricted Region Simplex Design for Mixture Experiments", *Communications in Statistics-Theory and Methods*, 12(22), 2605-2615 (1983).
230. Murthy, M.S. and J.S. Murty, "Restricted Region Designs for Multifactor Mixture Experiments", *Communications in Statistics-Theory and Methods*, 18, 1279-1295 (1989).
231. Murthy, M.S.R. and J.S. Murty, "Block Designs for Mixture Experiments", *Journal of the Indian Society of Agricultural Statistics*, 44, 55-71 (1992).

232. Murty, J.S., "Problems of Construction and Analysis of Designs of Experiments", Ph.D. thesis, Delhi University, India (1966).
233. Murty, J.S. and M.N. Das, "Design and Analysis of Experiments with Mixtures", *Annals of Mathematical Statistics*, 39, 1517-1539 (1968).
234. Naes, T., E.M. Faergestad and J. Cornell, "A Comparison of Methods for Analyzing Data from a Three Component Mixture Experiment in the Presence of Variation Created by Two Process Variables", *Chemometrics and Intelligent Laboratory Systems*, 41, 221-235 (1998).
235. Narcy, J.P. and J. Renaud, "Use of Simplex Experimental Designs in Detergent Formulation", *Journal of the American Oil Chemists' Society*, 49, 598-608 (1972).
236. Nelder, J.A., "Inverse Polynomials, A Useful Group of Multi-Factor Response Functions", *Biometrics*, 22, 128-141 (1966).
237. Nigam, A.K., "Block Designs for Mixture Experiments", *Annals of Mathematical Statistics*, 41, 1861-1869 (1970).
238. Nigam, A.K., "Multifactor Mixture Experiments", *Journal of the Royal Statistical Society*, 35, 51-56 (1973).
239. Nigam, A.K., "Some Designs and Models for Mixture Experiments for the Sequential Exploration of Response Surfaces", *Journal of the Indian Society of Agricultural Statistics*, 26, 120-124 (1974).
240. Nigam, A.K., "Correction to 'Block Designs for Mixture Experiments'", *Annals of Statistics*, 4, 1294-1295 (1976).
241. Nigam, A.K., S.C. Gupta and S. Gupta, "A New Algorithm for Extreme Vertices Designs for Linear Mixture Models", *Technometrics*, 25, 367-371 (1983).
242. Obermiller, D.J., "Tips on JMPing into Mixture Experimentation," *Proceedings of the Twenty-First Annual SAS User's Group International Conference*, pp. 799-804, SAS Institute, Inc., Cary, NC (1996).
243. Oehlert, G.W. and P. Whitcomb, "Sizing Fixed Effects for Computing Power in Experimental Designs", *Quality and Reliability Engineering International*, 17, 291-306 (2001).
244. Paku, G.A., A.R. Manson and L.A. Nelson, *Minimum Bias Estimation in the Mixture Problem*, North Carolina State University Institute of Statistics Mimeo, Series No. 757, Raleigh, NC (1971).
245. Park, S.H., "Selecting Contrasts Among Parameters in Scheffé's Mixture Models: Screening Components and Model Reduction", *Technometrics*, 20, 273-279 (1978).
246. Park, S.H. and J.H. Kim, "Axis-Slope-Rotatable Designs for Experiments with Mixtures", *Journal of the Korean Statistical Society*, 11, 36-44 (1982).
247. Park, S.H. and J.I. Kim, "Slope-Rotatable Designs for Estimating the Slope of Response Surfaces in Experiments with Mixtures", *Journal of the Korean Statistical Society*, 17, 121-133 (1988).
248. Peterson, J.J., "A General Approach to Ridge Analysis With Confidence Intervals", *Technometrics*, 35, 204-214 (1993).
249. Peterson, J.J., "A Probability-Based Desirability Function for Multiresponse Optimization", *2000 Proceedings of the Section on Quality and Productivity*, pp. 64-69 (American Statistical Association, Alexandria, VA, 2000).

250. Peterson, J.J., "A Posterior Predictive Approach to Multiple Response Surface Optimization", *Journal of Quality Technology*, 36, 139-153 (2004).
251. Peterson, J.J., S. Cahya and E. Del Castillo, "A General Approach to Confidence Regions for Optimal Factor Levels of Response Surfaces", *Biometrics*, 58, 422-431 (2002).
252. Piepel, G.F., "Component Effects in Mixture Experiments", *Proceedings of the 1979 DOE Statistical Symposium*, CONF-791016, 7-19, Oak Ridge National Laboratory, Oak Ridge, TN (1980).
253. Piepel, G.F., "Measuring Component Effects in Constrained Mixture Experiments", *Technometrics*, 24, 29-39 (1982).
254. Piepel, G.F., "Defining Consistent Constraint Regions in Mixture Experiments", *Technometrics*, 25, 97-101 (1983a).
255. Piepel, G.F., "Calculating Centroids in Constrained Mixture Experiments", *Technometrics*, 25, 279-283 (1983b).
256. Piepel, G.F., *Models and Designs for Generalizations of Mixture Experiments Where the Response Depends on the Total Amount*, Ph.D. thesis, University of Florida (University Microfilms International, Ann Arbor, MI, 1985).
257. Piepel, G.F., "Models for Mixture-Amount Experiments When the Total Amount Can Equal Zero", BN-SA-2479, Battelle, Pacific Northwest Laboratories, Richland, WA (1987).
258. Piepel G.F., "Programs for Generating Extreme Vertices and Centroids of Linearly Constrained Experimental Regions", *Journal of Quality Technology*, 20, 125-139 (1988a).
259. Piepel, G.F., "A Note on Models for Mixture-Amount Experiments When the Total Amount Takes a Zero Value", *Technometrics*, 30, 449-450 (1988b).
260. Piepel, G.F., "MIXSOFT"—A Collection of FORTRAN Routines for the Design and Analysis of Mixture Experiments", *Proceedings of the 21st Symposium on the Interface*, pp. 481-490 (American Statistical Association, Alexandria, VA, 1989a).
261. Piepel, G.F., MIXSOFT—Software for the Design and Analysis of Mixture Experiments, User's Guide, Version 1, MIXSOFT—Mixture Experiment Software, Richland, WA (1989b).
262. Piepel, G.F., "Screening Designs for Constrained Mixture Experiments Derived From Classical Screening Designs", *Journal of Quality Technology*, 22, 23-33 (1990).
263. Piepel, G.F., "Screening Designs for Constrained Mixture Experiments Derived From Classical Screening Designs—An Addendum", *Journal of Quality Technology*, 23, 96-101 (1991).
264. Piepel, G.F., "Survey of Software With Mixture Experiment Capabilities", *Journal of Quality Technology*, 29, 76-85 (1997).
265. Piepel, G.F., "Survey of Software With Mixture Experiment Capabilities", PNWD-SA-4348, Rev. 1, Battelle, Pacific Northwest Laboratories, Richland, WA (1998).
266. Piepel, G.F., "Modeling Methods for Mixture-of-Mixtures Experiments Applied to a Tablet Formulation Problem", *Pharmaceutical Development and Technology*, 4, 593-606 (1999).

267. Piepel, G.F., MIXSOFT—Software for the Design and Analysis of Mixture and Other Constrained Region Experiments, User's Guide, Version 2.4.1, MIXSOFT—Mixture Experiment Software, Richland, WA (2003).
268. Piepel, G.F., "A Component Slope Linear Model for Mixture Experiments", PNWD-SA-6313, Rev. 1, Battelle—Pacific Northwest Division, Richland, WA (2004).
269. Piepel, G.F. and C.M. Anderson, "Variance Dispersion Graphs for Designs on Polyhedral Regions", *1992 Proceedings of the Section on Physical and Engineering Sciences*, 111-117 (American Statistical Association, Alexandria, VA, 1992).
270. Piepel, G.F., C.M. Anderson and P.E. Redgate, "Variance Dispersion Graphs for Designs on Polyhedral Regions—Revisited", *1993 Proceedings of the Section on Physical and Engineering Sciences*, 102-107 (American Statistical Association, Alexandria, VA, 1993a).
271. Piepel, G.F., C.M. Anderson and P.E. Redgate, "Response Surface Designs for Irregularly-Shaped Regions" (Parts 1, 2, and 3), *1993 Proceedings of the Section on Physical and Engineering Sciences*, 205-227 (American Statistical Association, Alexandria, VA, 1993b).
272. Piepel, G.F., C.M. Anderson and P.E. Redgate, "Variance Dispersion Graphs for Evaluating and Comparing Experimental Designs on Polyhedral Regions", PNWD-SA-4218, Rev. 0, Pacific Northwest Laboratory, Richland, WA (1995).
273. Piepel, G.F., S.K. Cooley, D.K. Peeler, J.D. Vienna and T.B. Edwards, "Augmenting a Waste Glass Mixture Experiment Study with Additional Glass Components and Experimental Runs", *Quality Engineering*, 15, 91-111 (2002).
274. Piepel, G.F., S.K. Cooley and B. Jones, "Construction of a 21-Component Layered Mixture Experiment Design Using a New Mixture Coordinate-Exchange Algorithm", Rev. 4, PNNL-SA-37340, Pacific Northwest National Laboratory, Richland, WA (2004), to appear in *Quality Engineering* in 2005.
275. Piepel, G.F. and J.A. Cornell, "Models for Mixture Experiments When the Response Depends on the Total Amount", *Technometrics*, 27, 219-227 (1985).
276. Piepel, G.F. and J.A. Cornell, "A Comparison of Mixture-Amount and Component Amount Experiments", BN-SA-2179, Rev. 1, Battelle, Pacific Northwest Laboratories, Richland, WA (1986).
277. Piepel, G.F. and J.A. Cornell, "Designs for Mixture-Amount Experiments", *Journal of Quality Technology*, 19, 11-28 (1987).
278. Piepel, G.F. and J.A. Cornell, "A Catalog of Mixture Experiment Examples", BN-SA-3298, Rev. 0 (1991), Rev. 14.7 (2004), Battelle, Pacific Northwest Laboratories, Richland, WA (1991, 2004).
279. Piepel, G.F. and J.A. Cornell, "Mixture Experiment Approaches—Options, Examples, and Recommendations (shortened version)", *1992 Proceedings of the Section on Physical and Engineering Sciences*, 98-103, American Statistical Association, Alexandria, VA (1992).
280. Piepel, G.F. and J.A. Cornell, "Mixture Experiment Approaches: Examples, Discussion, and Recommendations", *Journal of Quality Technology*, 26, 177-196 (1994).

281. Piepel, G.F., R.D. Hicks, J.M. Szychowski and J.L. Loeppky, "Methods for Assessing Curvature and Interaction in Mixture Experiments", *Technometrics*, 44, 161-172 (2002).
282. Piepel, M.G. and G.F. Piepel, "How Soil Composition Affects Water Capacity—A Science Fair Project", *1998 Proceedings of the Section on Physical and Engineering Sciences*, pp. 104-109 (American Statistical Association, Alexandria, VA, 1998).
283. Piepel, M.G. and G.F. Piepel, "How Soil Composition Affects Density and Water Capacity—A Science Fair Project Using Mixture Experiment Methods", *STATS*, No. 28, 14-20 (2000).
284. Piepel, G. and T. Redgate, "A Mixture Experiment Analysis of the Hald Cement Data", *1996 Proceedings of the Section on Physical and Engineering Sciences*, 93-98 (American Statistical Association, Alexandria, VA, 1996a).
285. Piepel, G.F. and T. Redgate, "Mixture Experiment Techniques for Reducing the Number of Components Applied to Modeling Waste Glass Sodium Release", *1996 Proceedings of the Section on Physical and Engineering Sciences*, 308-317 (American Statistical Association, Alexandria, VA, 1996b).
286. Piepel, G.F. and T. Redgate, "Mixture Experiment Techniques for Reducing the Number of Components Applied to Modeling Waste Glass Sodium Release", *Journal of the American Ceramic Society*, 80, 3038-3044 (1997).
287. Piepel, G. and T. Redgate, "A Mixture Experiment Analysis of the Hald Cement Data", *The American Statistician*, 52, 23-30 (1998).
288. Piepel, G., T. Redgate and P. Hrma, "Phased Mixture Experiment Design of a Waste Glass Study", *1995 Proceedings of the Section on Physical and Engineering Sciences*, 220-228 (American Statistical Association, Alexandria, VA, 1995a).
289. Piepel, G., T. Redgate, P. Hrma and S. Hartley, "Mixture Experiment Design and Property Modeling in a Multi-Year Nuclear Waste Glass Study", *1995 Proceedings of the Section on Physical and Engineering Sciences*, 173-178 (American Statistical Association, Alexandria, VA, 1995b).
290. Piepel, G., T. Redgate and P. Masuga, "A Comparison of Mixture and Free Energy of Hydration Models for Waste Glass Durability", *1995 Proceedings of the Section on Physical and Engineering Sciences*, 161-166 (American Statistical Association, Alexandria, VA, 1995c).
291. Piepel, G., T. Redgate and P. Masuga, *Mixture Models Versus Free Energy of Hydration Models for Waste Glass Durability*, PNL-10823, Pacific Northwest National Laboratory, Richland, WA (1996).
292. Piepel, G.F., J.M. Szychowski and J.L. Loeppky, "Augmenting Scheffé Linear Mixture Models with Squared and/or Crossproduct Terms", *Journal of Quality Technology*, 34, 297-314 (2002).
293. Piepel, G.F., J.D. Vienna and P. Hrma, "Phase 1 Experimental Design for the INEEL HLW Glass Composition Variation Study", PNNL-SA-29594, Rev. 2, Pacific Northwest National Laboratory, Richland, WA (1999).
294. Prescott, P., "Nearly Optimal Orthogonally Blocked Designs for a Quadratic Mixture Model with q Components", *Communications in Statistics—Theory and Methods*, 27(10), 2559-2580 (1998).

295. Prescott, P., "Projection Designs for Mixture Experiments in Orthogonal Blocks", *Communications in Statistics, Part A, Theory and Methods*, 29, 2227-2253 (2000).
296. Prescott, P., "Modelling in Mixture Experiments Including Interactions with Process Variables", *Quality Technology & Quantitative Management*, 1, 87-103 (2004).
297. Prescott, P., A.M. Dean, N.R. Draper and S.M. Lewis, "Mixture Experiments: Ill-Conditioning and Quadratic Model Specification", *Technometrics*, 44, 260-268 (2002).
298. Prescott, P. and N.R. Draper, "Mixture Designs for Constrained Components in Orthogonal Blocks", *Journal of Applied Statistics*, 25, 613-638 (1998).
299. Prescott, P. and N.R. Draper, "Mixture Component-Amount Designs Via Projections, Including Orthogonally Blocked Designs", *Journal of Quality Technology*, 36, 413-431 (2004).
300. Prescott, P., N.R. Draper, A.M. Dean and S.M. Lewis, "Mixture Designs for Five Components in Orthogonal Blocks", *Journal of Applied Statistics*, 20, 105-117 (1993).
301. Prescott, P., N.R. Draper, S.M. Lewis and A.M. Dean, "Further Properties of Mixture Designs for Five Components in Orthogonal Blocks", *Journal of Applied Statistics*, 24, 147-156 (1997).
302. Quenouille, M.H., *The Design and Analysis of Experiments*, (Hafner Publishing Company, New York, 1953).
303. Quenouille, M.H., "Experiments With Mixtures", *Journal of the Royal Statistical Society, B*, 21, 201-202 (1959).
304. Redgate, P.E., G.F. Piepel and P.R. Hrma, "Second-Order Model Selection in Mixture Experiments", *1992 Proceedings of the Section on Physical and Engineering Sciences*, 104-110 (American Statistical Association, Alexandria, VA, 1992).
305. Redgate, P.E. and G.F. Piepel, *Predicting Primary Crystalline Phase and Liquidus Temperature Above or Below 1050°C as Functions of Glass Composition*, PNNL-10998, Pacific Northwest National Laboratory, Richland, WA (1996).
306. Robinson, T.J., R.H. Myers and D.C. Montgomery, "Analysis Considerations in Industrial Split-Plot Experiments with Non-Normal Responses", *Journal of Quality Technology*, 36, 180-192 (2004).
307. Rusin, M., "The Structure of Nonlinear Blending Models", *Chemical Engineering Science*, 30, 937-944 (1975).
308. Sahrmann, H.F., G.F. Piepel and J.A. Cornell, "In Search of the Optimum Harvey Wallbanger Recipe Via Mixture Experiment Techniques", *The American Statistician*, 41, 190-194 (1987).
309. Saxena, S.K. and A.K. Nigam, "Symmetric-Simplex Designs for Mixtures", *Journal of the Royal Statistical Society, B*, 35, 466-472 (1973).
310. Saxena, S.K. and A.K. Nigam, "Restricted Exploration of Mixtures by Symmetric-Simplex Design", *Technometrics*, 19, 47-52 (1977).
311. Scheffé, H., "Experiments with Mixtures", *Journal of the Royal Statistical Society, B*, 20, 344-360 (1958).

312. Scheffé, H., "Reply to Mr. Quenouille's Comments About My Paper on Mixtures", *Journal of the Royal Statistical Society*, B, 23, 171-172 (1961).
313. Scheffé, H., "The Simplex-Centroid Design for Experiments with Mixtures", *Journal of the Royal Statistical Society*, B, 25, 235-251 (1963).
314. Setz, S., M. Semling and R. Mühlaupt, "Fuzzy Set Approach for Fitting a Continuous Response Surface in Adhesion Formulation", *Journal of Chemometrics*, 11, 403-418 (1997).
315. Shelton, J.T., *Testing Lack of Fit in a Mixture Model*, Ph.D. thesis, University of Florida (University Microfilms International, Ann Arbor, MI, 1982).
316. Shelton, J.T., A.I. Khuri and J.A. Cornell, "Selecting Check Points for Testing Lack of Fit in Response Surface Models", *Technometrics*, 25, 357-365 (1983).
317. Showers, M.E., "Using PROC REG to Analyze Mixture Experiments", *Proceedings of the 1983 SAS User's Group Meeting*, SAS, Cary, NC, corrected version from author (1983).
318. Singh, S.P. and M. Pratap, "Analysis of Symmetric-Simplex Designs in Mixture Experiments", *Calcutta Statistical Association Bulletin*, 34, 65-73 (1985).
319. Singh, S.P., M. Pratap and M.N. Das, "Analysis of Mixture Experiments in Presence of Block Effects", *Sankhya*, 44, 270-277 (1982).
320. Smith, W.F. and T.A. Beverly, "Generating Linear and Quadratic Cox Mixture Models", *Journal of Quality Technology*, 29, 211-224 (1997).
321. Smith Jr., W.F. and J.A. Cornell, "Biplot Displays for Looking at Multiple Response Data in Mixture Experiments", *Technometrics*, 35, 337-350 (1993).
322. Snee, R.D., "Design and Analysis of Mixture Experiments", *Journal of Quality Technology*, 3, 159-169 (1971).
323. Snee, R.D., "Techniques for the Analysis of Mixture Data", *Technometrics*, 15, 517-528 (1973).
324. Snee, R.D., "Experimental Designs for Quadratic Models in Constrained Mixture Spaces", *Technometrics*, 17, 149-159 (1975a).
325. Snee, R.D., "Discussion of 'The Use of Gradients to Aid in the Interpretation of Mixture Response Surfaces'", *Technometrics*, 17, 425-430 (1975b).
326. Snee, R.D., "Experimental Designs for Mixture Systems with Multicomponent Constraints", *Communications in Statistics—Theory and Methods*, A8(4), 303-326 (1979a).
327. Snee, R.D., "Experimenting With Mixtures", *Chemtech*, 9, 702-710 (1979b).
328. Snee, R.D., "Developing Blending Models for Gasoline and Other Mixtures", *Technometrics*, 23, 119-130 (1981).
329. Snee, R.D., "Computer-Aided Design of Experiments: Some Practical Experiences", *Journal of Quality Technology*, 17, 222-236 (1985).
330. Snee, R.D. and D.W. Marquardt, "Extreme Vertices Designs for Linear Mixture Models", *Technometrics*, 16, 399-408 (1974).
331. Snee, R.D. and D.W. Marquardt, "Screening Concepts and Designs for Experiments with Mixtures", *Technometrics*, 18, 19-29 (1976).
332. Snee, R.D. and A.A. Rayner, "Assessing the Accuracy of Mixture Model Regression Calculations", *Journal of Quality Technology*, 14, 67-79 (1982).

333. Sobolev, N.N. and T.A. Chemleva, "Construction of Simplicial-Linear Models of Composition-Property Diagrams", *Industrial Laboratory*, 42 103-108, English translation of Russian language *Zavodskaya Laboratoriya*, 42, 70-76 (1976).
334. St. John, R.C., "Experiments with Mixtures, Ill-Conditioning, and Ridge Regression", *Journal of Quality Technology*, 16, 81-96 (1984).
335. Stanard, C., "Optimizing Z Scores in Mixture DOE's", Technical Information Series 2001CRD095, General Electric Research & Development Center, Schenectady, NY (2001).
336. Steiner, S.H. and M. Hamada, "Making Mixtures Robust to Noise and Mixing Measurement Errors", *Journal of Quality Technology*, 29, 441-450 (1997).
337. Tang, M., J. Li,, L.Y. Chan and D.K.J Lin, "Application of Uniform Design in the Formation of Cement Mixtures", *Quality Engineering*, 16, 461-474 (2004).
338. Thompson, W.O. and R.M. Myers, "Response Surface Designs for Experiments with Mixtures", *Technometrics*, 10, 739-756 (1968).
339. Tian, G.L. and K.T. Fang, "Uniform Designs for Mixture-Amount Experiments and for Mixture Experiments Under Order Restrictions", *Science in China, Series A*, 42, 456-470, in English (1999).
340. Tiede, M.L., "How to Use SAS for ANOVA of Mixture Problems", *Proceedings of 1981 SAS User's Group Meeting*, 233-238, SAS Institute, Cary, NC (1981).
341. Tsubaki, M., "Robust Design and Analysis for Mixture Experiment", *Proceedings of International Conference on Statistics in Industry, Science, and Technology*, p. 316-321 (1994).
342. Uranisi, H., "Optimum Design for the Special Cubic Regression on the q-Simplex", *Mathematical Reports*, 1, 7-12, General Education Department, Kyushu University, Japan (1964).
343. Van Schalkwyk, D.J., "On the Design of Mixture Experiments", Ph.D. thesis, University of London, England (1971).
344. Vining, G.G., J.A. Cornell, R.H. Myers and A. Giovannitti-Jensen, "A Graphical Approach for Evaluating Mixture Designs", *Department of Statistics Technical Report No. 360*, University of Florida, Gainesville, FL (1989).
345. Vining, G.G., J.A. Cornell and R.H. Myers, "A Graphical Approach for Evaluating Mixture Designs", *Applied Statistics*, 42, 127-138 (1993).
346. Vuchkov, I.N., "Sequentially Generated Designs", *Biometrical Journal*, 24, 751-763 (1982).
347. Vuchkov, I.N., D.L. Damgaliev and C.A. Yontchev, "Sequentially Generated Second Order Quasi D-Optimal Designs for Experiments With Mixture and Process Variables", *Technometrics*, 23, 233-238 (1981).
348. Vuchkov, I.N., C.A. Yontchev and D.L. Damgaliev, "Continuous D-Optimal Designs for Experiments with Mixture and Process Variables", *Mathematische Operationsforschung und Statistik, Series Statistics*, 14, 33-51 (1983).
349. Vuchkov, I.N., H.A. Yonchev, D.L. Damgaliev, V.K. Tsochev and T.D. Dikova, *Catalogue of Sequentially Generated Designs*, Department of Automation, Higher Institute of Chemical Technology, Sofia, Bulgaria, in Bulgarian and English (1978).

350. Wagner, T.O. and J.W. Gorman, "The Lattice Method for Design of Experiments With Fuels and Lubricants", *Application of Statistics and Computers to Fuel and Lubricant Research Problems*, pp. 123-145, Department of the Army, San Antonio, Texas (1962).
351. Wagner, T.O. and J.W. Gorman, "Fuels, Lubricants, and Experimental Designs", *Transactions of the Society of Automotive Engineers*, 196, 684-701 (1963).
352. Wang, Y. and K.T. Fang, "Uniform Design of Experiments with Mixtures", *Science in China, Series A*, 39, 264-275 (1996).
353. Watson, G.S., "Linear Regression on Proportions", *Biometrics*, 25, 585-587 (1969).
354. Wegscheider, W. and U. Walner, "Sparse Experimentation for Sparse Effects in a Mixed Variable Mixture Model", *Chemometrics and Intelligent Laboratory Systems*, 19, 169-174 (1993).
355. Weier, D.R. and G.F. Piepel, *Methodology for Adjusting and Normalizing Analyzed Glass Compositions*, PNWD-3260 (WTP-RPT-049), Battelle, Pacific Northwest Division, Richland, WA (2003).
356. Welch, W.J., "Computer-Aided Design of Experiments for Response Estimation", *Technometrics*, 26, 217-224 (1984).
357. White, D.B., H.M. Faessel, H.K. Slocum, L. Khinkis and W.R. Greco, "Nonlinear Response Surface and Mixture Experiment Methodologies Applied to the Study of Synergism", *Biometrical Journal*, 46, 56-71 (2004).
358. Yonchev, H.A., "New Computer Procedures for Generating Optimal Mixture Designs on Finite Design Spaces", *COMPSTAT, Proceedings of 8th Symposium on Computational Statistics*, edited by D. Edwards and N.E. Rauh, Physica-Verlag Heidelberg, 433-438 (1988).
359. Zhu, W., C. Hu and H. Chen, "D-Optimality and  $D_n$ -Optimal Designs for Mixtures Regression Models With Logarithmic Terms", *Acta Mathematicae Applicatae Sinica*, English Series, 3, 26-36 (1987a).
360. Zhu, W., C. Hu and H. Chen, "D-Optimal and  $D_n$ -Optimal Designs in Five and Six Components for Mixture Models With Logarithmic Terms", *Acta Mathematicae Applicatae Sinica*, English Series, 3, 317-329 (1987b).

This page is intentionally left blank

## **CHAPTER 13**

### **GRAPHICAL METHODS FOR COMPARING RESPONSE SURFACE DESIGNS FOR EXPERIMENTS WITH MIXTURE COMPONENTS**

Heidi B. Goldfarb

*The Dial Corporation  
Scottsdale, AZ 85254*

Douglas C. Montgomery

*Arizona State University  
Tempe, AZ 85287-5906*

Different types of experiments involving mixture components are studied in the response surface setting. These include the basic mixture experiment and mixture experiments with process variables. Designs, models, and analyses for these situations are considered. Graphical methods for design evaluation are presented, including variance dispersion graphs and fraction of design space plots. Examples illustrating these techniques are presented.

#### **1. Introduction**

##### ***1.1. Mixture and Mixture-Process Experiments***

Mixture and mixture-process (MP) experiments are commonly encountered in many fields including the chemical, food, pharmaceutical, and consumer products industries. For mixture experiments, the design factors are the proportions of the components that sum to a constant, and the response variables depend only on these component proportions. There may also be additional constraints on component proportions. The shape of the design space for these experiments depends on the number

of components and their constraints. For MP experiments, the experimenter is also interested in other variables that can be varied independently of one another and of the mixture components. The shape of the design space for these combined designs will depend on the number of process variables as well as the number of mixture components and the restrictions on them. See Cornell<sup>1</sup> for a comprehensive treatment of mixture and mixture-process experiments.

Both mixture and MP experiments provide design challenges. Mixture designs often involve component constraints that lead to irregularly-shaped regions. MP designs have the additional challenge of adding process variables and often requiring large numbers of runs to adequately cover the combined design space. Cornell<sup>1</sup> discusses many strategies for generating designs over restricted spaces as well as designs in the combined MP space. More details and additional approaches for MP experiments can be found in Cornell and Gorman<sup>2</sup>, Czitrom<sup>3,4</sup>, and Kowalski, Cornell and Vining<sup>5,6</sup>.

Many of these design strategies focus on the precise estimation of model coefficients. In this chapter we will use tools focused on scaled prediction variance (SPV) to evaluate various designs, a measure advocated by Box and Hunter<sup>7</sup>. They noted the advantages of examining SPVs over single-number summaries used by many of the alphabetic optimality criteria. Another useful measure is the unscaled prediction variance (PV), which differs from the SPV in that the prediction variance is not multiplied by the sample size. The idea of multiplying by the sample size is to penalize larger designs. A rule-of-thumb for good designs is to have a maximum PV less than or equal to 1. This corresponds to being able to predict the response at least as well as it can be measured. Both PV and SPV are important measures for evaluating designs on both absolute and relative/penalized scales. With the methodologies developed in this chapter, we will evaluate several mixture and MP designs with respect to the SPV over the entire design space.

## ***1.2. Variance Dispersion Graphs***

Giovannitti-Jensen and Myers<sup>8</sup> introduced the variance dispersion graph (VDG), a graphical technique that can be used to evaluate the prediction

variance properties of a design. They focused on classical response surface designs and considered the prediction variance properties moving along concentric spheres starting from the overall centroid of the design space. The VDGs plot the minimum, average, and maximum prediction variances versus the radii of the concentric spheres. Examples of the usefulness of these VDGs can be found in Myers, Vining, Giovannitti-Jensen, & Myers<sup>9</sup>, Vining<sup>10</sup>, Borkowski<sup>11</sup>, Trinca and Gilmour<sup>12</sup>, and Borror, Montgomery, and Myers<sup>13</sup>.

Rozum and Myers<sup>14</sup>, Myers and Montgomery<sup>15</sup>, and Borror, Montgomery, and Myers<sup>13</sup> discuss VDGs for designs on cuboidal regions. For these types of designs it is customary to plot the minimum, average, and maximum prediction variances on shrunken cubes.

VDGs have also been extended to designs involving mixture components and/or non-mixture variables with and without linear constraints by Piepel and Anderson<sup>16</sup>, and Piepel, Anderson, and Redgate<sup>17,18</sup>. They use a shrunken region approach, plotting the variances along constant shrinkage values of the original design polyhedron. 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<sup>19</sup> also develop 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. Khuri, Harrison, and Cornell<sup>20</sup> introduce a method to plot the entire distribution of the SPV for a given shrinkage level of the design space. Separate plots are used for each shrinkage level.

Three-dimensional VDGs for MP experiments were introduced in Goldfarb, Borror, Montgomery, and Anderson-Cook<sup>21</sup>. These VDGs plot the shrinkage in the mixture space along the x-axis, the shrinkage in the process-variable space along the y-axis, and the average or maximum SPV on the z-axis. The shapes of the VDG SPV surfaces show the imprecision of the prediction throughout the MP shrinkage plane. These VDGs offer an experimenter one way to visually evaluate the prediction variance properties of a design throughout the combined MP space. More details and examples of these plots will be given later in the chapter.

### **1.3. Fraction of Design Space Plots**

Zahran, Anderson-Cook, and Myers<sup>22</sup> introduce a complementary technique to the VDG called the Fraction of Design Space (FDS) plot. For this technique the SPV is calculated throughout the design space and then the fraction of the design space that is less than or equal to a given SPV value is determined. Plotting the cumulative fraction for each SPV value provides the experimenter with a profile of the SPV throughout the design space. A distinction between the VDG and FDS plot is that the VDG gives equal weight on the plot to each radii or shrinkage value regardless of the proportion of the design space represented. The FDS plot gives weight proportionate to the fraction of the design space, which results in more weight on the outer portions of the design where the volume is larger. The FDS graph also has the advantage of needing just a single line to represent each design, allowing an easy comparison of multiple designs on a single plot. Zahran, Anderson-Cook, and Myers<sup>22</sup> develop and demonstrate this technique for spherical and cuboidal designs, but do not address designs on non-regular regions. Later in the chapter we will discuss the extension of FDS plots to mixture and mixture-process problems as introduced in Goldfarb, Anderson-Cook, Borror, and Montgomery<sup>23</sup>.

The FDS technique and plots will be discussed for mixture designs in Section 2 and for MP designs in Section 3. Both will be illustrated with examples. In section 3, three-dimensional VDGs will be discussed for the MP design settings.

## **2. Mixture Experiments**

### **2.1. FDS Plots for Mixture Designs**

To calculate the FDS, points are sampled throughout the mixture design region as defined by the constraints on the components. See Goldfarb, Anderson-Cook, Borror, and Montgomery<sup>23</sup> for more details on the sampling procedure. After the points are generated, the SPV,  $v(x)$ , is calculated for each point,  $x_0$  according to

$$v(\mathbf{x}_0) = \frac{N \operatorname{var}(\hat{\mathbf{y}}_0)}{\sigma^2} = N \mathbf{x}'_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_0.$$

The PV is calculated as

$$\text{PV} = \frac{v(\mathbf{x}_0)}{N} = \frac{\operatorname{var}(\hat{\mathbf{y}}_0)}{\sigma^2} = \mathbf{x}'_0 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_0.$$

Here  $\operatorname{var}(\hat{\mathbf{y}}_0)$  is the variance of the estimated response at the point of interest in the design space,  $\mathbf{X}$  is the design matrix expanded to model form,  $\mathbf{x}_0$  is the point of interest expanded to model form, and  $N$  is the number of runs in the design. The empirical cumulative distribution function is then calculated for the SPV values and plotted on the FDS plot. The minimum value of the SPV is shown at an FDS of 0 and the maximum value is shown at the fraction 1. Along the FDS curve, the fraction of the design space at or below a particular SPV value can be determined. A good design is one that starts with small values and remains flat throughout. When comparing designs, those with lower and flatter lines are preferable from a scaled prediction variance standpoint. More details on this methodology can be found in Goldfarb, Anderson-Cook, Borror, and Montgomery<sup>23</sup>. The PV can also be plotted in the same manner.

### Example 1: The Flare Experiment

McLean and Anderson<sup>24</sup> introduced a four-component mixture experiment that examines the illumination produced by flares. The objective was to find the mixture that gave maximum illumination. The four components, magnesium ( $x_1$ ), sodium nitrate ( $x_2$ ), strontium nitrate ( $x_3$ ), and binder ( $x_4$ ), sum to 1 and engineering experience indicated that the individual components should have the following restrictions:

$$0.4 \leq x_1 \leq 0.6$$

$$0.1 \leq x_2 \leq 0.5$$

$$0.1 \leq x_3 \leq 0.5$$

$$0.03 \leq x_4 \leq 0.08.$$

McLean and Anderson<sup>24</sup> created a 15-point design (Design A) based on their extreme-vertices strategy to fit a 10-term quadratic mixture model as shown in Equation (1).

$$\begin{aligned} y = & \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{14} x_1 x_4 \\ & + \beta_{23} x_2 x_3 + \beta_{24} x_2 x_4 + \beta_{34} x_3 x_4 + \varepsilon \end{aligned} \quad (1)$$

Piepel, Anderson, and Redgate<sup>25</sup> considered three alternative 15-run designs based on those developed by Cornell<sup>1</sup> (Table 8.6). The first two (Designs B and C) were developed using the D-optimality criteria, with the second design also including an overall centroid. Designs formed using the D-optimality criteria seek to minimize confidence ellipses for the unknown parameters. The third design (Design D) was developed using the G-optimality criteria. Designs formed using the G-optimality criteria seek to minimize prediction variance. VDGs and FDS plots for the four designs are shown in Figures 1 and 2.

The VDGs are displayed on two plots to enhance readability, while the FDS plot shows all four designs. Since designs B and C have similar profiles and there is a tradeoff when comparing designs A and D, we have grouped the VDGs accordingly. The VDGs show that the designs have different values at the center of the design space but seem to converge for all three measures toward the edges of the design space. Piepel, Anderson, and Redgate<sup>25</sup> provide a comprehensive evaluation of the VDGs for the four designs. They conclude that design A is superior to design B and design D is superior to design C. Design A does better at the center and smaller shrinkage levels, while design D is superior at higher shrinkage levels on the outer regions of the design space.

The FDS plot shows, as do the VDGs, that designs B and C have inferior SPV properties compared to designs A and D for most of the design space. The FDS plot shows that designs A and D track closely to one another for the majority of the plot as do designs B and C to one another. The FDS curves for designs B and C are consistently higher and hence inferior to the other two designs. Examining designs A and D more closely we see that as the fraction of the design space approaches around 0.7, the SPV of design A begins to increase sharply. Design D

stays quite flat and is the superior design. This demonstrates the property of the FDS plot to give more weight to the outer regions of the design space, where there is more volume. Recall that the VDG indicated that design A had poor prediction at higher radii values, but it was less clear how much weight to associate with this information.

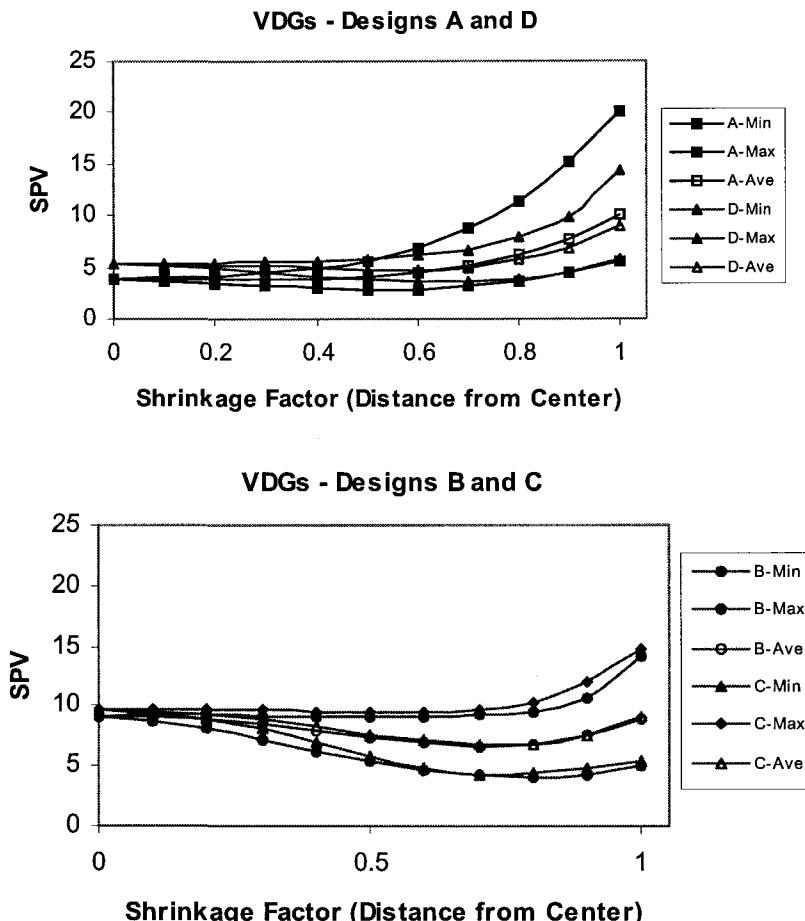


Figure 1. VDGs for the flare designs in Example 1.

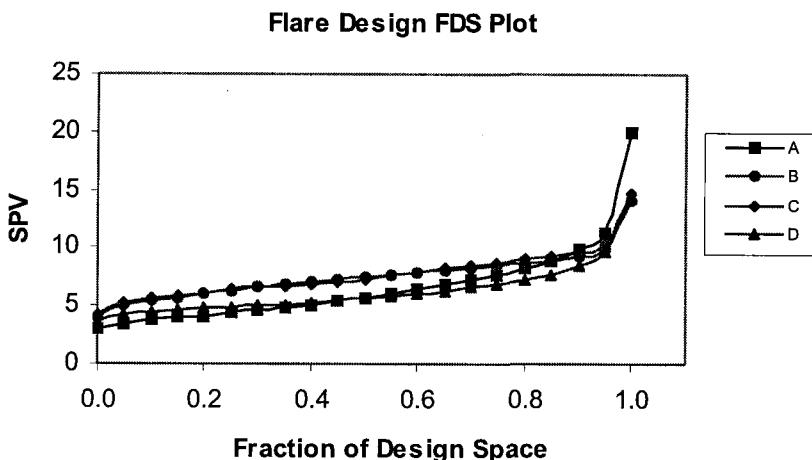


Figure 2. FDS plot for the flare designs in Example 1.

### 3. Mixture-Process Experiments

#### 3.1. Three-Dimensional VDGs for Mixture-Process Designs

For experiments with both mixture and process variables, we introduce VDGs that examine the prediction variance surface over the combined space. A standard VDG could be applied to this setting where the x-axis corresponds to the shrinkage factor for both design spaces, process and mixture. However this would only consider a small portion of the combined space. Prediction variances over shrinkage factors that are not equal are also of interest. By expanding the conventional VDGs to three dimensions, we can 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 values can be thought of as multipliers of the original design space. A shrinkage value of 1 would represent points on the edge of the original design space. A value of 0 would indicate that the point is the overall centroid of the regions. A shrinkage value of 0.7 would mean that the points fall along the outline of a shape that is 70% of the way between the overall centroid and the original outline.

When calculating the PV or SPV for the combined MP designs,  $\mathbf{X}$  and  $\mathbf{x}_0$  in the variance calculations contain both mixture and process variables. To characterize the variation over the shrunken regions it is useful to look at either the maximum or the average prediction variance. The maximum would allow us to see the worst case for each shrinkage value and could be used to apply G-optimal type criteria of minimizing the maximum prediction variance. The average SPV provides information that can be used for  $V$  or  $IV$  (also known as  $Q$ ) type criteria which seek to minimize the average prediction variance over a given set of points and over the entire region, respectively. The shapes of the prediction variance surfaces show the behavior of the prediction variance as the experimenter moves throughout the mixture-process plane. The minimum variance is the least critical of the three summaries traditionally included on VDGs and is therefore not usually included on three-dimensional VDGs.

Since many of the examples of VDGs have been of response surface designs on spherical regions, radii of the hyper-spheres are generally plotted for the process variables. Note that if the process factor design is on a sphere, then the shrinkage factor is equivalent to a radius. For cuboidal designs, the shrinkage factors represent shrunken cubes. For the mixture design, shrunken regions are used and the shape will vary depending on the constraints of the mixture components.

To generate the plots, a grid of shrinkage values ranging from 0 to 1 is generated in both spaces. For each combination of shrinkage values, equally spaced points are generated in both the mixture and process space and then the cartesian product of these points is taken. Then PV and SPV values are calculated for each of these generated points and then the average and maximum values are calculated. Standard statistics software can then be used to produce contour and surface plots, with the x-axis representing the mixture shrinkage, the y-axis representing the process shrinkage, and the z-axis showing the average or maximum prediction variance values. For more details on three-dimensional VDGs for MP designs, see Goldfarb, Borror, Montgomery, and Anderson-Cook<sup>21</sup>.

### ***3.2. FDS Plots for Mixture-Process Designs***

FDS plots can also be constructed for MP designs and used to aid in design evaluation. As the number of variables increases, the number of runs in these types of designs often gets quite large. It is important to have an effective way to compare designs as well as determine how to efficiently augment designs to improve prediction variance properties.

To construct FDS plots for MP experiments, an approach similar to that for mixture designs is used. Points are sampled in each of the design spaces, process and mixture, and then combined to yield the final points for which SPV values are calculated. In the mixture space, points are randomly sampled from the edges of the shrunken regions, where the outlines of the regions are determined by the constraints on the individual components. For the process variables, points are randomly selected around different shrinkage levels of either a cuboidal or a spherical region, depending on the type of design used. All of the points in the process space are then crossed with all of the points in the mixture space to form a grid of points throughout the combined design space.

After the points are chosen, the SPV for each is calculated. An empirical cumulative distribution function is then calculated for all of the SPV values and plotted as a global FDS. For each shrinkage level of the process variable space, FDS values are also calculated across all mixture shrinkage levels. These slices for different process space shrinkage values are shown as separate FDS lines for each design. The slices can be used to judge which of the two spaces, mixture or process, contributes more to changes in the SPV values. If the FDS values for the different process shrinkage slices are far apart, then changes in location throughout the process space have a larger effect. If the slices have rapidly increasing slopes, are close together, and follow closely the global FDS, then changes in location throughout the mixture space have the larger effect. If the slices are somewhere between horizontal and diagonal, then the spaces have more balanced contributions to changes in the SPV values.

This information can help an experimenter decide where to focus additional runs or how to move some runs to improve overall SPV properties. If the plots indicate that the location in the process variable

space has a greater effect on the SPV, additional points should be placed in areas of the process variable space where there are weaknesses. For example if the process slice corresponding to the center of the region shows high SPV values, then the additional points should be placed at the center of the process design region. If the process slice corresponding to the edges of the region shows high SPV values, then the additional points should be placed along the edges of the process variable design region. Likewise, if the plots indicate that the location in the mixture space has a greater effect on the SPV, additional points should be placed in areas of the mixture space where there are weaknesses.

Examining the order of the process shrinkage slices, the experimenter can see the shape of the SPV throughout the process design space. For example, if the FDS lines increase in value as the process shrinkage level increases, this indicates that the SPV increases as predictions are made further away from the center. If the SPV curves decrease and then increase again, it indicates that the SPV is higher in the center and along the edges of the design space and decreases in between the two extremes. As with pure mixture designs, the FDS plot can be used in conjunction with the VDG to evaluate designs.

### **Example 2: A Mixture-Process Experiment**

Kowalski, Cornell, and Vining<sup>5</sup> (KCV) presented a generic 3-mixture by 3-process variable example. They considered the following 21-term model, involving a quadratic mixture model, pure quadratic and two-factor interactions for the process variables, and interactions between the linear mixture and linear process variables.

$$\begin{aligned}
 y = & \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \\
 & \alpha_{11} z_1^2 + \alpha_{22} z_2^2 + \alpha_{33} z_3^2 \\
 & + \alpha_{12} z_1 z_2 + \alpha_{13} z_1 z_3 + \alpha_{23} z_2 z_3 \\
 & + \gamma_{11} x_1 z_1 + \gamma_{21} x_2 z_1 + \gamma_{31} x_3 z_1 + \gamma_{12} x_1 z_2 \\
 & + \gamma_{22} x_2 z_2 + \gamma_{32} x_3 z_2 + \gamma_{13} x_1 z_3 + \gamma_{23} x_2 z_3 + \gamma_{33} x_3 z_3 + \varepsilon
 \end{aligned}$$

The mixture variables are denoted by x's and the process variables by z's. KCV proposed 31 and 37-run designs focused on precise estimation

of model coefficients. The 37-run design is comprised of the 31-run design plus six additional points all at the center of the process design space. Both designs are shown in Table 1. The approach that KCV took in the generation of the designs was to start with central composite designs (CCD) in the process variables. Then simplex-centroid designs in the mixture components were placed at each of the CCD points, with only a fraction of the mixture blends being run at each point. The designs were balanced as much as possible to achieve symmetry. See Kowalski, Cornell, and Vining<sup>5</sup> for more details.

For comparison, we generated a 37-run design (DX637A) with the D-optimal design generator option of Design-Expert 6 (DX6) (Stat-Ease 2003), splitting the degrees of freedom beyond those needed to fit the model equally between lack-of-fit and replication. Splitting the extra runs equally between lack-of-fit and replication is often a good design approach. The resulting design is shown in Table 2. VDGs and FDS plots are shown for the three designs in Figures 3, 4, and 5.

The three-dimensional VDGs for the average SPVs for the two KCV designs, shown in Figure 3, confirm that they have very different prediction variance profiles. The SPV of the 31-run design increases significantly as the points move toward the outer parts of the mixture design space. For the 37-run design the SPV increases as the points move toward the outer edges of the process space. The VDG for the DX6 37-run design confirms that the average SPV values for this design are higher than those of the other two designs throughout the design space.

Figure 4 shows global FDS plots for both PV and SPV. As expected, the PV values of the larger designs are generally lower than those of the smaller design. However, the maximum PV for the DX6 design is greater than the maximum of the others. Note that the KCV designs have PV values less than or equal to 1 throughout the design space, which is a desirable quality.

The 37-run DX6 design is dominated by both of the KCV designs for SPV. For all fractions, the SPV of the DX6 design is greater than that of both of the KCV designs. In comparing the two KCV designs, we see that the 37-run design has more area with very small SPV values, but has slightly higher SPV values as the percentages get higher. From this plot

it appears that the additional six runs added in the 37-run design do not yield a significant advantage in prediction variance except over a very small portion of the design space.

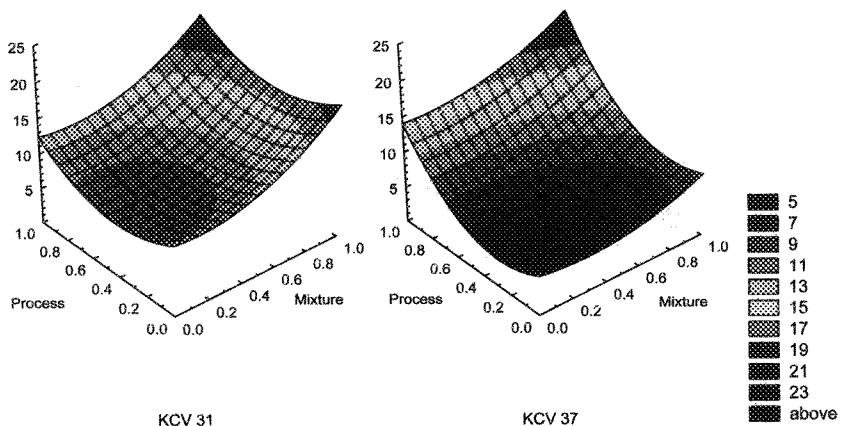
Table 1. KCV designs in example 1.

Run	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	Z <sub>1</sub>	Z <sub>2</sub>	Z <sub>3</sub>
1	1	0	0	-1	-1	1
2	1	0	0	-1	1	-1
3	1	0	0	1	1	1
4	1	0	0	1	-1	-1
5	0	1	0	-1	-1	1
6	0	1	0	-1	1	-1
7	0	1	0	1	1	1
8	0	1	0	1	-1	-1
9	0	0	1	-1	-1	1
10	0	0	1	-1	1	-1
11	0	0	1	1	1	1
12	0	0	1	1	-1	-1
13	0.5	0.5	0	-1	-1	-1
14	0.5	0.5	0	-1	1	1
15	0.5	0.5	0	1	-1	1
16	0.5	0.5	0	1	1	-1
17	0.5	0	0.5	-1	-1	-1
18	0.5	0	0.5	-1	1	1
19	0.5	0	0.5	1	-1	1
20	0.5	0	0.5	1	1	-1
21	0	0.5	0.5	-1	-1	-1
22	0	0.5	0.5	-1	1	1
23	0	0.5	0.5	1	-1	1
24	0	0.5	0.5	1	1	-1
25	0.33	0.33	0.33	-1	0	0
26	0.33	0.33	0.33	1	0	0
27	0.33	0.33	0.33	0	-1	0
28	0.33	0.33	0.33	0	1	0
29	0.33	0.33	0.33	0	0	-1
30	0.33	0.33	0.33	0	0	1
31	0.33	0.33	0.33	0	0	0
<b>Additional Points for the 37 Point Design</b>						
32	1	0	0	0	0	0
33	0	1	0	0	0	0
34	0	0	1	0	0	0
35	0.5	0.5	0	0	0	0
36	0.5	0	0.5	0	0	0
37	0	0.5	0.5	0	0	0

Table 2. Design DX6-37A from Example 2.

Run	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	Z <sub>1</sub>	Z <sub>2</sub>	Z <sub>3</sub>
1	1	0	0	-1	-1	-1
2	0	0	1	1	1	-1
3	0.5	0.5	0	-1	-1	1
4	0.5	0.5	0	1	1	1
5	0	0.5	0.5	0	1	0
6	0.5	0	0.5	-1	0	0
7	0	0	1	1	1	-1
8	0	0	1	-1	-1	-1
9	0	0	1	1	-1	0
10	0	1	0	1	-1	1
11	0	1	0	-1	-1	-1
12	1	0	0	-1	1	1
13	0.5	0.5	0	-1	1	-1
14	0	1	0	1	1	-1
15	0	1	0	-1	-1	-1
16	0.5	0.5	0	0	0	0
17	1	0	0	0	1	0
18	0.333	0.333	0.333	1	-1	-1
19	0	0.5	0.5	-1	0	1
20	0	0	1	-1	1	1
21	0	1	0	-1	1	1
22	1	0	0	0	-1	0
23	1	0	0	1	-1	1
24	1	0	0	1	0	0
25	0.333	0.333	0.333	1	-1	-1
26	0	1	0	1	-1	1
27	0.5	0.5	0	-1	1	-1
28	0	0	1	0	0	-1
29	0	0	1	-1	1	1
30	1	0	0	1	1	-1
31	0	1	0	-1	1	1
32	0	1	0	1	1	-1
33	0	1	0	-1	0	0
34	1	0	0	0	0	-1
35	1	0	0	0	0	1
36	0	0	1	1	0	1
37	0.5	0	0.5	0	-1	1

## Average Scaled Prediction Variance by Process and Mixture Shrinkage Values



## Average Scaled Prediction Variance by Process and Mixture Shrinkage Values

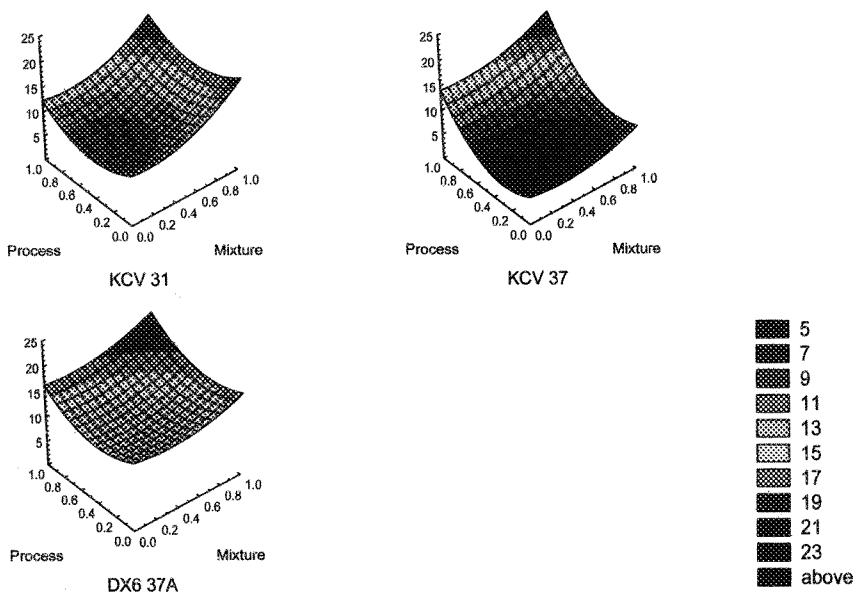


Figure 3. Average SPV VDGs for the designs in Example 2.

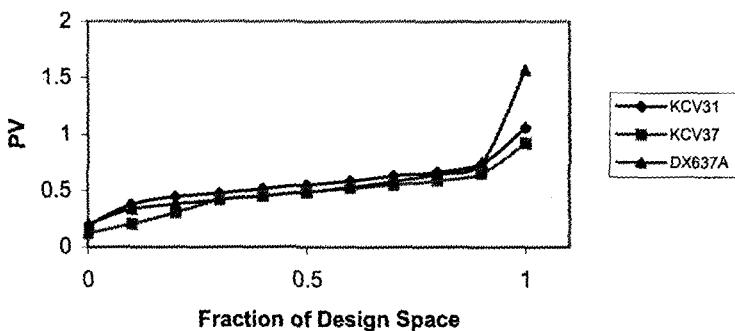
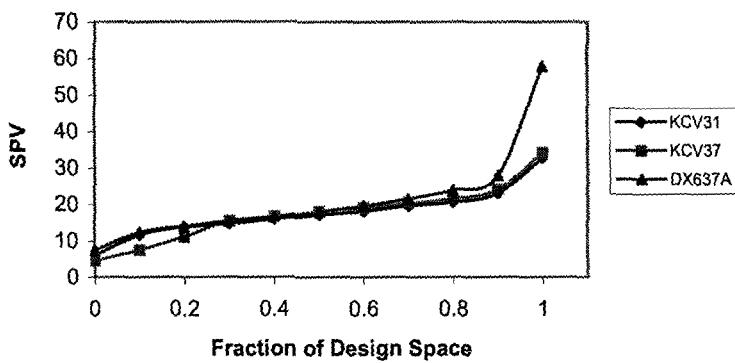
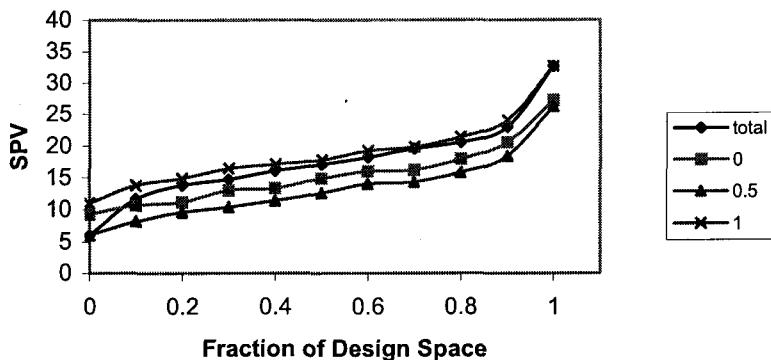
**Global FDS Plots for the PV of the Example 2 Designs****Global FDS Plots for the SPV of the Example 2 Designs**

Figure 4. Global FDS plots for the designs in Example 2.

The FDS plots showing slices across the process space, shown in Figure 5, reveal distinctly different patterns for the two KCV designs. For the 31-run design, all of the slices appear to follow a similar increasing pattern, meaning that for a fixed process shrinkage level, the SPV increases toward the edges of the mixture space and does so similarly for all levels. The fact that these lines are so close to one another indicates that the change in process shrinkage levels does not have as large of an effect as changes in the mixture levels. We also see that the line for the 0.5 slice has the lowest value which indicates that there is a bowl-shape to the SPV in the mixture space.

### FDS Plot for the KCV 31-Run Design



### FDS Plot for the KCV 37-Run Design

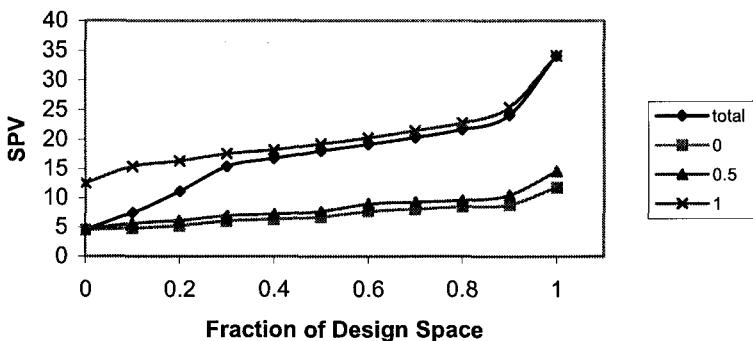


Figure 5. FDS plots with process slices for the designs in Example 2.

For the 37-run KCV design, the lines for the total process space and the process space with a shrinkage of 1, the outer part of the process design space, clearly have higher SPV values than those for the other shrinkage levels. The lines for shrinkage levels of 0 and 0.5 are low and flat. This means that at these shrinkage levels in the process space, there is little change in the SPV across the entire mixture space. This, along with the fact that the lines for some process shrinkage levels are far from others, indicate that the biggest changes in the SPV values are attributed

to changes in the process space. For this design, the levels of the shrinkage lines follow in order with the 0 shrinkage lowest, then the 0.5, and finally the 1. So unlike the 31-run design, we do not have a bowl-shaped SPV surface in the process space.

These patterns suggest that augmenting the 31-run design with the 6 additional runs has reduced the prediction variance toward the center of the process design space. This makes sense, since the six runs were all located at the center of the of the process design space. It also suggests that if additional runs can be added, they should be added to the outer portions of the process space.

The superior performance of the KCV designs over the DX6 design is not surprising. In Kowalski, Cornell, and Vining<sup>5</sup>, the KCV designs are shown to be superior to D-Optimal designs with Proc Optex in SAS.

This example shows how the FDS plots provide information complementary to the VDG. They also have the advantage that the minimum and maximum SPV values are easily read from them. We would need additional VDGs for the minimum and maximum values to get this information.

#### **4. Conclusions**

We have illustrated several types of response surface experiments involving mixture components. These important variations of response surface experiments occur in many industrial settings. Designs for these situations were considered, along with two graphical techniques, variance dispersion graphs (VDGs) and fraction of design space (FDS) plots, which are useful for design evaluation and comparison. These graphs can provide considerable insight about design alternatives. Because the design space for mixture and mixture-process experiments can be complicated, VDGs and FDS plots should be of considerable practical value to the experimenter.

#### **References**

1. Cornell, J. A. (2002). *Experiments with Mixtures: Designs, Models, and the Analysis of Mixture Data*, Third Edition. John Wiley & Sons, New York, NY.

2. Cornell, J. A., Gorman, J. W. (1984). "Fractional Design Plans for Process Variables in Mixture Experiments". *Journal of Quality Technology* 16, pp. 20-38.
3. Czitrom, V. (1988). "Mixture Experiments with Process Variables: D-Optimal Orthogonal Experimental Designs". *Communications in Statistics - Theory and Methods* 17, pp. 105-121.
4. Czitrom, V. (1990). "Experimental Designs for Four Mixture Components with Process Variables". *Communications in Statistics - Theory and Methods* 18, pp. 4561-4581.
5. Kowalski, S. M., Cornell, J. A., and Vining, G. G. (2000). "A New Model and Class of Designs for Mixture Experiments with Process Variables". *Communications in Statistics - Theory and Methods* 29, pp. 2255-2280.
6. Kowalski, S. M., Cornell, J. A., and Vining, G. G. (2002). "Split Plot Designs and Estimation Methods". *Technometrics* 44, pp. 72-79.
7. Box, G. E. P. and Hunter, J. S. (1957). "Multifactor Experimental Designs for Exploring Response Surfaces". *The Annals of Mathematical Statistics* 28, pp. 195-241.
8. Giovannitti-Jensen, A. and Myers, R. H. (1989). "Graphical Assessment of the Prediction Capability of Response Surface Designs". *Technometrics* 31, pp. 159-171.
9. Myers, R. H., Vining, G. 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.
10. Vining, G. (1993). "A Computer Program for Generating Variance Dispersion Graphs". *Journal of Quality Technology* 25, pp. 45-58.
11. Borkowski, J.J. (1995). "Spherical Prediction Variance Properties of Central Composite and Box-Behnken Designs". *Technometrics* 37, pp. 399-410.
12. 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.
13. 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, pp. 54-70.
14. 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.
15. Myers R. H. and Montgomery, D. C. (2002), *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, Second Edition. John Wiley & Sons, New York, NY.
16. 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, Alexandria, VA, pp. 111-117.
17. 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, VA, pp. 102-107.

18. 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, VA, pp. 205-227.
19. 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.
20. 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, Part 4, pp. 521-532.
21. Goldfarb, H. B., Borror, C. M., Montgomery, D. C., and Anderson-Cook, C. M. (2004). "Three-Dimensional Variance Dispersion Graphs for Mixture-Process Experiments". *Journal of Quality Technology* 36, pp.109-124.
22. 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.
23. Goldfarb, H. B., Anderson-Cook, C. M., Borror, C. M., and Montgomery, D. C. (2004). "Fraction of Design Space Plots for Assessing Mixture and Mixture-Process Designs". *Journal of Quality Technology* 36, pp.169-179.
24. McLean, R. A. and Anderson, V. L. (1966). "Extreme Vertices Design of Mixture Experiments". *Technometrics* 24, pp. 447-456.
25. Piepel, G., Anderson, C. M, and Redgate, P. E. (1995). "Variance Dispersion Graphs for Evaluating and Comparing Experimental Designs on Polyhedral Regions". Technical Report PNWD-SA-4218, Batelle, Pacific Northwest Laboratories, Richland, WA.

## CHAPTER 14

### GRAPHICAL METHODS FOR ASSESSING THE PREDICTION CAPABILITY OF RESPONSE SURFACE DESIGNS

John J. Borkowski

*Department of Mathematical Sciences  
Montana State University  
Bozeman, MT 59717  
E-mail: jobo@math.montana.edu*

In their 1960 *Technometrics* article “Some New Three Level Designs for the Study of Quantitative Variables”, Box and Behnken introduced plots called “variance profiles” to study the prediction capability of their proposed designs. These were graphs of a scaled prediction variance plotted as a function of  $\rho$ , the distance from the center of the design. For many years after this publication, plots for assessing the prediction capability of response surface designs were rarely used. Then, in the late 1980s and throughout the 1990s, the variance dispersion graph (VDG) which is a variation of Box and Behnken’s variance profile, was introduced into the response surface methodology literature.

VDGs were generated for many designs, design regions, and prediction variance functions. For example, VDGs have been studied for designs in spherical, cuboidal, and polyhedral regions, and they have been generated for the minimum, maximum, and average of prediction variance functions. The VDG approach has been extended to the prediction variance quantile (PVQ) plot for studying the distribution of prediction variance values, and, most recently, the fraction of design space (FDS) plot. PVQ plots and FDS plots provide additional information about a design’s prediction capability that complement VDGs. A summary of the major results and a review of these and other graphical methods for assessing the prediction capability of response surface designs will be presented. *Componentwise variance dispersion graphs* (CVDGs), a new graphical tool for mixture designs, will also be presented.

#### 1. The Prediction Variance Function

Consider the situation in which a researcher is planning a  $k$ -factor experiment having  $N$  experimental runs. Adopting the terminology of Box and

Hunter,<sup>1</sup> the *design matrix*  $\mathbf{D}$  is the  $N \times k$  matrix whose rows correspond to the  $N$  experimental runs. For any  $p$ -parameter polynomial response surface model, the *model matrix*  $\mathbf{X}$  associated with  $\mathbf{D}$  is the  $N \times p$  matrix whose columns correspond to the  $p$  terms in the model.  $N^{-1}\mathbf{X}'\mathbf{X}$  is the *moment matrix* and its inverse  $N(\mathbf{X}'\mathbf{X})^{-1}$  is the *precision matrix*. For any linear model, the  $N$  equations for the  $N$  experimental runs can be expressed in matrix notation as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}. \quad (1)$$

We will assume (1) is appropriate for the experiment where  $\mathbf{Y}$  is the  $N \times 1$  vector of responses,  $\boldsymbol{\beta}$  is the  $p \times 1$  vector of unknown coefficients, and  $\boldsymbol{\epsilon}$  is a  $N \times 1$  vector of random errors. We further assume  $E(\boldsymbol{\epsilon}) = \mathbf{0}$  and  $\text{var}(\boldsymbol{\epsilon}) = \sigma^2 \mathbf{I}_N$ .

Polynomials are typically used as empirical models for approximating the true model over the experimental region. For the  $k$ -factor case, the first-order model for response  $Y$  is

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

and second-order model is

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

The predicted value at  $\mathbf{x}$  is

$$\hat{Y}(\mathbf{x}) = \mathbf{f}'(\mathbf{x})\hat{\boldsymbol{\beta}} \quad (4)$$

where  $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$  is the ordinary least squares estimator of  $\boldsymbol{\beta}$  and  $\mathbf{f}'(\mathbf{x})$  is the vector corresponding to the model terms. For example,

$$\mathbf{f}'(\mathbf{x}) = [1 | x_1, \dots, x_k | x_1 x_2, \dots, x_{k-1} x_k | x_1^2, \dots, x_k^2]$$

is the (partitioned) vector corresponding to the model terms in (3).

The problem of choosing a “best” design  $\mathbf{D}$  for fitting (1) can be interpreted in more than one way. One interpretation is that  $\mathbf{D}$  satisfies the requirement that the coefficients defined by  $\boldsymbol{\beta}$  are estimated separately with smallest variance. For the first-order model in (2) with each of the  $x_i \in \{-1, +1\}$ , it is known that  $\text{var}(\beta_i)/\sigma^2$  is minimized for  $i = 1, 2, \dots, k$  if the  $N$ -point design  $\mathbf{D}$  satisfies  $N^{-1}\mathbf{X}'\mathbf{X} = \mathbf{I}_N$ . (e.g., see Myers and Montgomery p. 308).<sup>2</sup> For any best  $N$ -point first-order design, the moment matrix has a unique form and is realized by choosing  $\mathbf{D}$  to have orthogonal

columns subject to  $\mathbf{x}'_i \mathbf{1} = 0$  and  $\mathbf{x}'_i \mathbf{x}_i = N$  where  $\mathbf{x}_i$  is the column of  $\mathbf{X}$  corresponding to the  $\beta_i x_i$  term in the model and  $\mathbf{1}$  is a vector of ones.

Box and Hunter<sup>1</sup> noted that consideration of only the variances of the individual coefficient estimates does not, for the case of second or higher order models, lead to any unique class of "best" designs. Their conclusion was that the coefficient estimates should be studied simultaneously. Therefore, one desirable design property is to produce predicted values  $\hat{Y}(\mathbf{x})$  with low variance. The *prediction variance at point  $\mathbf{x}$*  is  $\text{var}(\hat{Y}(\mathbf{x})) = \sigma^2 V_1(\mathbf{x})$  where

$$V_1(\mathbf{x}) = \mathbf{f}'(\mathbf{x})(\mathbf{X}'\mathbf{X})^{-1}\mathbf{f}(\mathbf{x}). \quad (5)$$

The *scaled prediction variance* (SPV) function  $V(\mathbf{x})$  (and considered by Box and Hunter) is defined as

$$V(\mathbf{x}) = \frac{N}{\sigma^2} \text{var}(\hat{Y}(\mathbf{x})) = NV_1(\mathbf{x}) \quad (6)$$

Box and Hunter<sup>1</sup> stated that

For any experimental design  $V(\mathbf{x})$  provides a standardized measure of the precision of the estimated response at any point in the space of the variables. It is a function of  $x_1, x_2, \dots, x_k$  and the elements of the precision matrix alone and is uniquely defined for every  $k$  dimensional experimental design of order  $d$ .

Examples of graphical techniques involving the prediction variance function are found sporadically in the statistical literature (prior to Giovannitti-Jensen and Myers).<sup>3</sup> The first plots appeared in 1957 when Box and Hunter<sup>1</sup> introduced the concept of design rotatability. Instead of plotting  $V(\mathbf{x})$  they considered its reciprocal known as the *weight function*  $W(\mathbf{x}) = \{V(\mathbf{x})\}^{-1}$ . For various first and second-order rotatable designs, they plotted  $W(\rho)$  where  $\rho$  is the distance of  $\mathbf{x}$  from the design center.

For one and two factor response surface designs, it is easy to plot  $V(\mathbf{x})$  and have complete information about the prediction variance surface. For one factor, we can plot  $V(\mathbf{x})$  across the interval design region. For example, consider two 6-point designs consists of collecting data at  $\mathbf{x} = -1, -1, 0, 0, 1, 1$  (design D1) and at  $\mathbf{x} = -1, -.5, 0, 0, .5, 1$  (design D2). Then, for the first-order model:

$$\text{For D1 : } V(\mathbf{x}) = 1 + 1.5x^2 \quad \text{For D2 : } V(\mathbf{x}) = 1 + 2.4x^2$$

and for the second-order model:

$$\text{For D1 : } V(\mathbf{x}) = 3 - (9/2)x^2 + (9/2)x^4$$

$$\text{For D2 : } V(\mathbf{x}) = 51/26 - (144/65)x^2 + (72/13)x^4$$

The plots of  $V(\mathbf{x})$  for both designs and both models are given in Figure 1. For the first-order model depicted in Figure 1(a),  $V(\mathbf{x})$  is uniformly better for D1 than for D2. This is not true, however, for  $V(\mathbf{x})$  for the second-order model shown in Figure 1(b). Because D1 has replicated endpoints at  $\pm 1$ ,  $V(\mathbf{x})$  is smaller for D1 than D2 near  $\pm 1$ . The design points for D2, however, are equispaced across  $[-1, 1]$  with replicates at 0 leading to smaller  $V(\mathbf{x})$  in the center of the interval. Thus, each design has its strengths and weaknesses with respect to the prediction variance of this model.

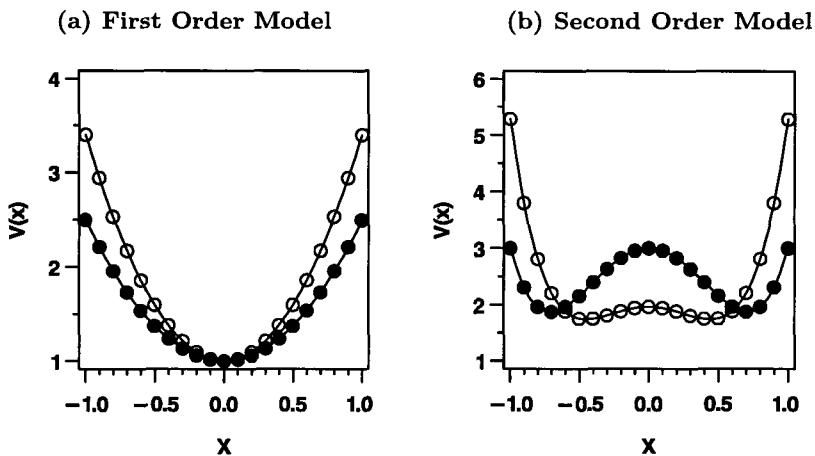


Fig. 1. Scaled prediction variance plot for two 6-point 1-factor designs. Solid dots correspond to design D1:  $\mathbf{x} = -1, -1, 0, 0, 1, 1$ . Open circles correspond to design D2:  $\mathbf{x} = -1, -0.5, 0, 0, 0.5, 1$ .

For the two-factor designs,  $V(\mathbf{x})$  can be depicted using either a contour plot or a three-dimensional surface plot. Box and Hunter<sup>1</sup> included contour plots of  $V(\mathbf{x})$  for three two-factor designs. One of the designs was the  $3^2$  design  $\{-1, 0, 1\} \times \{-1, 0, 1\}$  for which

$$V(\mathbf{x}) = 5 + 4.5(x_1^4 + x_2^4 - x_1^2 - x_2^2 + 0.5x_1^2x_2^2).$$

Figure 2(a) is a contour plot of  $V(\mathbf{x})$  describing the pattern of prediction variances in the square design region. A three-dimensional plot of  $V(\mathbf{x})$  is contained in Figure 2(b) which clearly depicts the prediction variance surface. Like many designs,  $V(\mathbf{x})$  is largest for the  $3^2$  design on the boundary of the design region. Two useful applications of prediction variance contour plots can be found in Hebble and Mitchell<sup>4</sup> and Khuri<sup>5</sup>. Hebble and

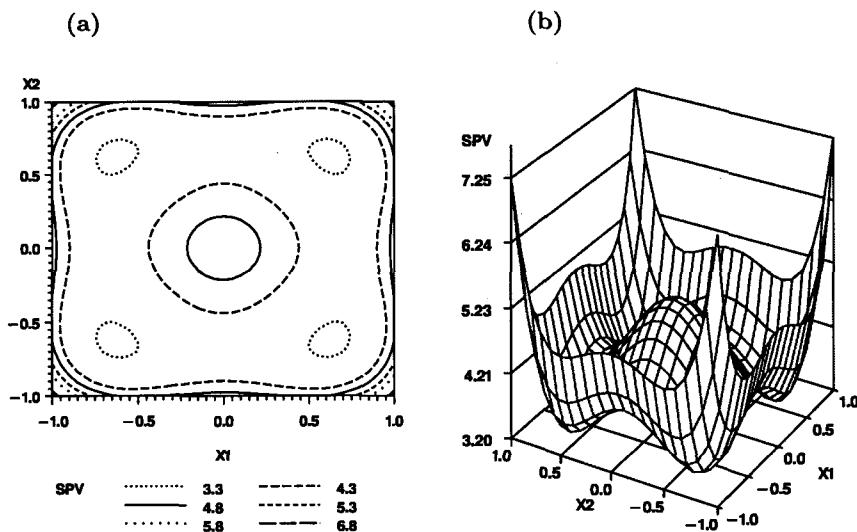


Fig. 2. Scaled prediction variance (SPV) function plots for  $3^2$  design. A contour plot of  $V(\mathbf{x})$  is shown in (a). A three-dimensional surface plot of  $V(\mathbf{x})$  is shown in (b).

Mitchell<sup>4</sup> used contour plots to compare the prediction variance before and after ‘repairing’ designs with augmentation points. Khuri<sup>5</sup> proposed a measure of rotatability and used contour plots to repair rotatability based on this measure.

## 2. The Prediction Variance and Optimality Criteria

Prior to running a proposed experimental design, the researcher may consider a design optimality criterion related to  $\text{var}(\hat{Y}(\mathbf{x}))$  as one of the criteria when choosing a design. Most design optimality criteria are single-value measures based on “optimal properties” of the  $\mathbf{X}'\mathbf{X}$  matrix. When several competing designs are proposed, their optimality properties can be compared to aid in the design choice. The assumption that the empirical model is adequate is essential when evaluating designs by properties of  $\mathbf{X}'\mathbf{X}$ , i.e., optimality criteria are highly model dependent.

Two quantities directly related to  $\text{var}(\hat{Y}(\mathbf{x}))$  are  $G$  and  $IV$  (which are, in turn, related to the  $G$  and  $IV$ -criteria). For a given design and a region of interest  $R$ , the  $G$ -criterion is tantamount to minimizing the maximum of  $V(\mathbf{x})$  over all design measures. For a given design, we define  $G$  as

$$G = \max_{\mathbf{x} \in R} V(\mathbf{x}). \quad (7)$$

Because “good” designs based on  $G$  tend to minimize the maximum prediction variance, the researcher then has variance information regarding the worst prediction scenario, or, in other words, has an upper bound for the prediction variance for a proposed design. Typically,  $G$  is approximated by taking the maximum of  $V(\mathbf{x})$  over a discrete set of experimenter-supplied points in the design region. Although the discrete set of points is often based on a set of lattice points, it can vary based on the software used. On the other hand, associated with the IV-criterion is the quantity

$$IV = \frac{N}{K\sigma^2} \int_R \text{var}(\hat{Y}(\mathbf{x})) d\mathbf{x} = \frac{1}{K} \int_R V(\mathbf{x}) d\mathbf{x}, \quad (8)$$

where  $K$  is the volume of region  $R$ . Thus,  $IV$  is the average of  $V(\mathbf{x})$  over  $R$ . Box and Draper<sup>6,7</sup> proposed designs that were robust to model misspecification based on  $IV$ . Borkowski<sup>8</sup> provided a critical assessment of statistical software and of the inconsistent output related to the average prediction variance. The inconsistencies arose because the integral in (8) was replaced by an average over a discrete set of points with the choice of points dependent on the software of choice. To address this problem, Borkowski<sup>9</sup> applied a genetic algorithm that explicitly used (8) to generate small exact response surface designs.

Three common design evaluation quantities are  $D$ ,  $G$ , and  $IV$  (where  $D$  is based on minimizing the generalized variance, i.e., minimizing  $|(\mathbf{X}'\mathbf{X})^{-1}|$ ). Although the single-value  $G$  and  $IV$  quantities provide useful information about the prediction variance, much information is lost when a design’s prediction variance properties are reduced to a single value. In particular, the values of  $G$  and  $IV$  (as well as  $D$ ) do not provide information regarding the distribution of  $\text{var}(\hat{Y}(\mathbf{x}))$  values in the design region. For example, for many response surface designs,  $\text{var}(\hat{Y}(\mathbf{x}))$  tends to be much larger on or near the boundary of the design region. Thus, when it is important to estimate the response near the boundary,  $IV$  does not provide the necessary information about  $\text{var}(\hat{Y}(\mathbf{x}))$ . Likewise,  $G$  does not provide information indicating regions where  $\text{var}(\hat{Y}(\mathbf{x}))$  tends to be large.

The experimenter should also be aware that a design superior to other designs by one optimality criterion may perform poorly when evaluated by another criterion (e.g., see Ref. 10). Therefore, the choice of design may be dependent upon the choice of an evaluation criterion. For example, consider  $V(\mathbf{x})$  for designs D1 and D2 in Figure 1. For the second-order model,  $G = 3$  and  $IV = 2.4$  for D1 while  $G \approx 5.3$  and  $IV \approx 2.33$  for D2. Thus, D1 is superior to D2 based on  $G$  while it is inferior to D2 based on  $IV$ .

In Borkowski and Valeroso,<sup>11</sup>  $G$ -efficiencies ( $= 100p/G$ ) and  $IV$  values were plotted against the number of parameters  $p$  for reduced models of (3). For example, Figure 3 contains an  $IV$  vs  $p$  plot for the 28-point 4-factor central composite design (CCD) in the hypercube across 224 types of reduced models. The plotting symbol is  $q =$  the number of  $\beta_i x_i^2$  terms in the reduced model. This plot highlights the strong dependence of the average prediction variance on the model, and in particular,  $q$ . For various designs, the dependence of both the  $G$  and  $IV$  values on the model choice was graphically shown in Borkowski and Valeroso.<sup>11</sup> Note, however, when the properties of  $V(\mathbf{x})$  are condensed to a single  $G$  or  $IV$  value, we lose all information regarding where  $V(\mathbf{x})$  is large or small in the design region.

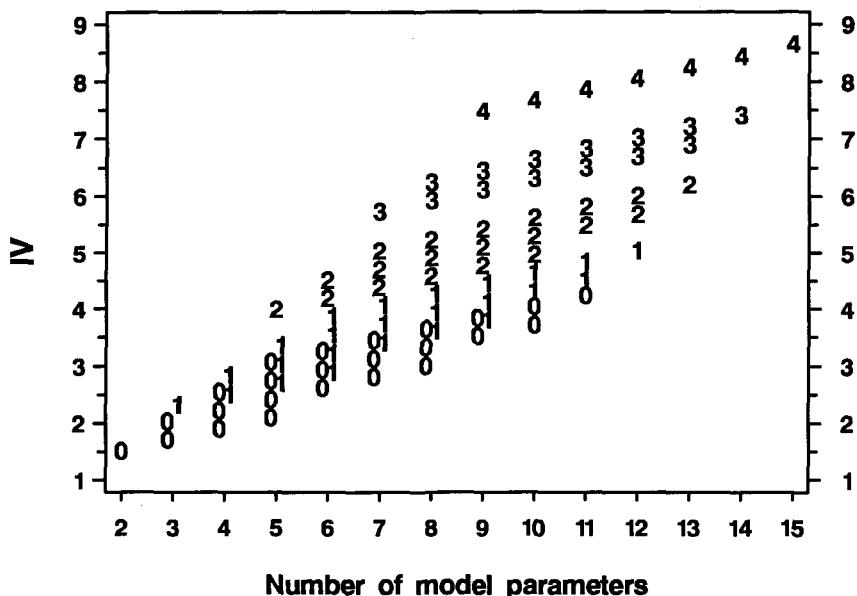


Fig. 3. A plot of  $IV$  values for the 4-factor, 28-point CCD in a hypercube design region across 224 types of reduced models. The plotting symbol is the number of  $\beta_i x_i^2$  terms in the reduced model.

A link relating optimality criteria and  $V(\mathbf{x})$  to graphical methods can be found in Haines.<sup>12</sup> In this paper, an annealing algorithm was used to construct exact optimal designs based on the  $G$ ,  $IV$ , and  $D$  optimality criteria. To compare the proposed designs, plots of  $V(\mathbf{x})$  against  $\mathbf{x}$  for 1 factor were presented. In Haines,<sup>13</sup> the prediction variance was studied for

optimal design for nonlinear regression models and examples of plots of the associated variance function against  $\mathbf{x}$  were included. In Hebble and Mitchell,<sup>4</sup> plots of  $G$  against varying  $N$  after a design is modified with additional points shows the impact of design augmentation on the maximum prediction variance.

For an overview of optimality criteria see Atkinson and Donev,<sup>14</sup> chapter 14 of Box and Draper,<sup>15</sup> and chapter 8 of Myers and Montgomery.<sup>2</sup> Other examples of plots relating optimality criteria and  $V(\mathbf{x})$  for one factor designs can be found in chapters 6, 9, and 11 of Atkinson and Donev.<sup>14</sup>

As an alternative to the single number approach to design selection and because plots like those presented in Figures 1 and 2 can only be made for  $k = 1$  and  $k = 2$  factors, graphical assessment techniques for evaluating prediction variance properties throughout the experimental region have been developed for studying  $k \geq 3$  factors. Three particular properties of interest are the maximum, minimum, and spherical prediction variances computed on the surfaces of several concentric spheres centered at the origin within the region of interest. Background on spherical prediction variance is presented in Section 3. The variance dispersion graph (VDG) is introduced in Section 4. Sections 5 and 6 include, respectively, discussion of the prediction variance quantile (PVQ) plot and the fraction of design space (FDS) plot. In Section 7, the graphical methods are extended to mixture experiments.

### 3. Spherical Prediction Variance (SPV)

For comparing designs, the maximum prediction variance  $G$  and the average prediction variance  $IV$  over the design region have often been used. Although  $G$  and  $IV$  provide useful information about the prediction variance, they do not provide information regarding its distribution in the design region.

A simple approach that provides substantially more information about the prediction variance throughout the design region utilizes various spherical prediction variance functions. The **average spherical prediction variance**  $V_\rho$  is the average of the scaled prediction variance function  $V(\mathbf{x})$  taken over  $S_\rho$ , the sphere of radius  $\rho$  centered at the origin. (See Hussey et al.<sup>16</sup>) This average is given by:

$$V_\rho = \frac{N}{\omega_\rho} \int_{S_\rho} \mathbf{f}'(\mathbf{x})(\mathbf{X}'\mathbf{X})^{-1}\mathbf{f}(\mathbf{x}) d\mathbf{x} = \text{tr } M(\mathbf{X}'\mathbf{X})^{-1} \quad (9)$$

where  $\omega_\rho$  is the surface area of  $S_\rho$  and  $M = \frac{N}{\omega_\rho} \int_{S_\rho} \mathbf{f}(\mathbf{x})\mathbf{f}'(\mathbf{x}) d\mathbf{x}$

is the matrix of region moments. For a  $k$ -dimensional sphere,  $\omega_\rho = 2\rho^{k-1}(\sqrt{\pi})^k/\Gamma(k/2)$  (where  $\Gamma(\cdot)$  is the gamma function).

Two other important spherical prediction variance functions are the **minimum** and **maximum prediction variances** and are defined as:

$$VMIN_\rho = \min_{\mathbf{x} \in S_\rho} V(\mathbf{x}) \quad (10)$$

$$VMAX_\rho = \max_{\mathbf{x} \in S_\rho} V(\mathbf{x}). \quad (11)$$

For the first-order response surface model in (2), Giovannitti-Jensen and Myers<sup>3</sup> showed

$$V_\rho = 1 + \frac{N\rho^2}{k} \sum_{i=1}^k \lambda_i$$

$$VMIN_\rho = 1 + N(\lambda_{min})\rho^2$$

$$VMAX_\rho = 1 + N(\lambda_{max})\rho^2$$

where  $\lambda_i$  are the eigenvalues of  $(\mathbf{X}'\mathbf{X})^{-1}$  with  $\lambda_{min}$  and  $\lambda_{max}$  being the smallest and largest eigenvalues. However, for most second-order response surface designs, analogous functions have not been found. On spherical and cuboidal design regions, Giovannitti-Jensen and Myers<sup>3</sup> employed a computer algorithm developed by Vining<sup>17,18</sup> for approximating these spherical prediction variance functions. This algorithm uses the trace form for  $V_\rho$  in (9). It was shown by Borkowski,<sup>19,20</sup> however, that for central composite designs (CCDs) and Box-Behnken designs (BBDs) that  $V_\rho$ ,  $VMIN_\rho$ , and  $VMAX_\rho$  can be determined analytically and from these functional forms exact values can be evaluated without the need of an optimization algorithm.

#### 4. The Variance Dispersion Graph (VDG)

In their important 1989 paper, Giovannitti-Jensen and Myers<sup>3</sup> aptly state the need for visualization tools for design assessment:

In recent years, more statisticians have recognized the value of graphical methods in data analysis. Since the performance of an experimental design (particularly in an RSM setting) so obviously presents a multidimensional problem, it would seem that creative graphical techniques in comparing and evaluating designs would be an obvious approach.

Their focus would be plotting various quantities against the radius of a sphere centered at the origin. Plotting predicted quantities against the radius of a sphere, however, was not a new technique. For example, in ridge analysis, Hoerl<sup>21,22</sup> plotted the predicted responses of the second order response surface model against the spherical radius subsequent to running an experiment. Likewise, Box and Behnken<sup>23</sup> plotted the standardized minimum and maximum prediction variances against the spherical radius for several of their proposed designs (a precursor of the VDG). What is common to these plotting techniques is their usefulness for studying the quantity of interest on a design region scale. For example, it may be important to make predictions at certain locations in the design region with high precision (e.g., near or on the design boundary). Whether or not a design's prediction variance values are reasonable at these locations cannot be ascertained from the  $G$  and  $IV$  values. This shortcoming of single-value optimality criteria will be addressed with the application of VDGs that will assist the experimenter when exploring the distribution of  $V(\mathbf{x})$  values in the design space.

To be able to compare the prediction variance properties of response surface designs,  $V_\rho$ ,  $VMIN_\rho$  and  $VMAX_\rho$  were plotted against  $\rho$ , the radius of the sphere with origin at the design's center.<sup>3,24</sup> These plots are called *variance dispersion graphs* or VDGs. By plotting  $VMIN_\rho$  and  $VMAX_\rho$ , the experimenter is provided with some insight regarding the relative stability of  $V(\mathbf{x})$  throughout a spherical design region. In addition, because the maximum value of  $V(\mathbf{x}) \geq p$  in the design region where  $p$  is the number of model parameters (Myers and Montgomery, p. 789).<sup>2</sup>, a horizontal reference line at  $V(\mathbf{x}) = p$  is typically included for a comparison to the optimal  $VMAX_\rho$  value.

To see the usefulness of VDGs, consider the plots in Figure 4 which contains plots of  $V_\rho$ ,  $VMIN_\rho$ , and  $VMAX_\rho$  for three five-factor designs: the 44-point Box-Behnken design (BBD), the 30-point rotatable central composite design (CCD) with  $\alpha = 2$ , and the 30-point CCD with  $\alpha = \sqrt{5}$  for the star-points. Assuming the second-order model in (3), we have the following equations for  $V_\rho$ :

$$V_\rho = \begin{cases} 11.0000 - 3.3000\rho^2 + 1.1210\rho^4 & \text{for the BBD (scaled)} \\ 7.0000 - 1.7500\rho^2 + 1.1250\rho^4 & \text{for the rotatable CCD} \\ 7.5000 - 1.8462\rho^2 + 1.0190\rho^4 & \text{for the CCD with } \alpha = \sqrt{5} \end{cases}$$

Analogous equations can be found for  $VMIN_\rho$  and  $VMAX_\rho$  using the closed-form expressions derived by Borkowski<sup>19,20</sup> for CCDs and BBDs. Note that when BBDs are compared to CCDs (as in Refs. 3, 19, 20, 24,

26), it is common to scale the design so that extreme design points are at a distance  $\rho = \sqrt{k}$ . For a BBD with  $k = 5$ , the scaling factor is  $\sqrt{5}/2$ .

From Figure 4 we see that the BBD is uniformly worse than either of the CCDs near the center of design region up to  $\rho \approx 1.5$ . The similarity of the  $\rho^4$  term coefficients for the BBD and the two CCDs is reflected in similar  $V_\rho$  graphs for large  $\rho$ . For  $\rho > 1.5$ ,  $V_\rho$  and  $VMAX_\rho$  are similar for all three designs while the performance of  $VMIN_\rho$  for the rotatable CCD is poorer in comparison to the other two designs. The variance reference line at  $V(x) = 21$  indicates that except for values of  $\rho$  near  $\sqrt{5}$ , all three design are performing well with respect to  $V(x)$  relative to the ideal.

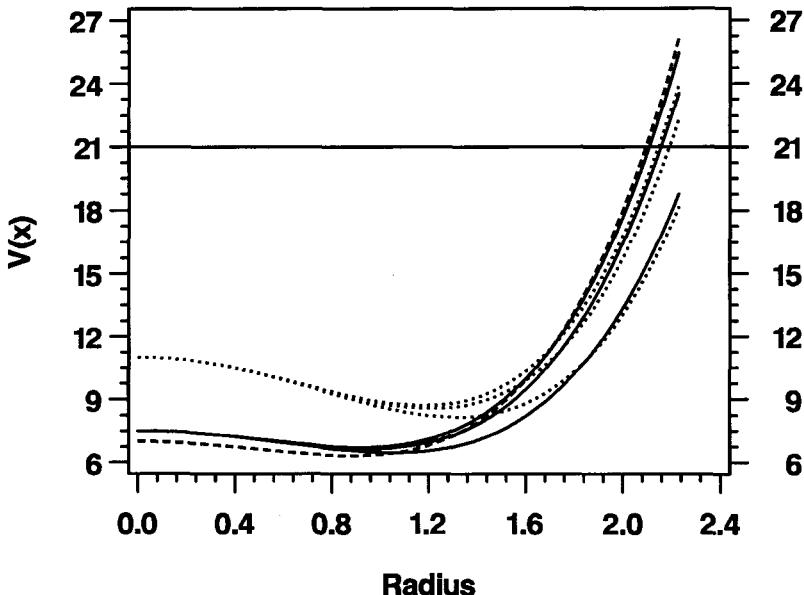


Fig. 4. VDGs for five-factor designs in a spherical design region: the 44-point BBD, the 30-point rotatable CCD ( $\alpha = 2$ ), and the 30-point CCD with  $\alpha = \sqrt{5}$ . Each design has four center points. Dotted lines are  $VMAX_\rho$ ,  $V_\rho$ , and  $VMIN_\rho$  for the BBD, solid lines are  $VMAX_\rho$ ,  $V_\rho$ , and  $VMIN_\rho$  for the CCD with  $\alpha = \sqrt{5}$ , and the dashed line is  $VMAX_\rho = V_\rho = VMIN_\rho$  for the rotatable CCD.

Figure 4 can also be viewed as a warning to the experimenter who is considering running a BBD. The  $V_\rho$  graphs within Figure 4 suggest that the BBD is comparable to the two CCDs for larger  $\rho$ . This comparison,

however, holds only as long as the experimenter is able to run the BBD at the scaled minimum and maximum variable levels and the CCD star-points at the extended  $\alpha$  level. For example, suppose a CCD and a BBD are considered for a five-factor experiment. Each factorial point of the CCD would be  $\sqrt{5}$  from the origin because each of the five coded factor levels is either  $\pm 1$ . For a non-center point in the BBD, however, the distance from the origin is  $\sqrt{2}$  because only two coded factor levels are  $\pm 1$  with the other three being 0. Thus, these BBD points must be scaled by a factor of  $\sqrt{5/2}$  to be comparable to the CCD's factorial points. Let temperature be a design variable. If the experimenter decides that the coded levels of  $-1$  and  $+1$  represent  $150^\circ$  and  $250^\circ$  for the CCDs, then the BBD having a comparable  $V_\rho$  graph must be run at extreme levels  $200 \pm 50\sqrt{5/2} (\approx 121^\circ$  and  $279^\circ)$ . Similarly, the uncoded star-point levels for the rotatable CCD with  $\alpha = 2$  and the CCD with  $\alpha = \sqrt{5}$  will be  $(-\alpha, +\alpha) \approx (100^\circ, 300^\circ)$  and  $(88^\circ, 312^\circ)$ , respectively. Only if these BBD scaled design variable levels and uncoded star-point levels are feasible for the experimenter, i.e., a spherical design region of radius  $\sqrt{5}$  is reasonable, then comparing graphs of spherical prediction variance properties for a BBD (with scaled design variables) to a CCD is appropriate.

The VDG is also a useful tool for assessing the impact of replicating design points. For example, consider the three-factor BBD. Figure 5(a) contains plots of  $V_\rho$  for 2, 3, 4, and 5 center point (CP) replicates (assuming a second-order model). Ideally we want a design with uniformly smaller  $V_\rho$  but that is not achieved through center point replication. In Figure 5(b), the range  $VMAX_\rho - VMIN_\rho$  is plotted for these four designs. We see that the prediction variances become increasingly unstable as  $\rho$  increases. If the experimenter believes it is more important to minimize the risk of a large  $V(\mathbf{x})$  value, a plot of  $VMAX_\rho$  would be a more suitable function to plot. By deleting one or more points from a design, the VDG can also provide useful information to assess the impact of missing experimental runs.

Examples of the use of VDGs to compare other response surface designs, including hybrid designs, small composite designs (SCDs), computer-generated designs, and Notz, Hoke, and Box-Draper designs were presented in Myers et al.<sup>24</sup>

The examples considered so far assume that the experimental design region is spherical. Often, however, the design region is cuboidal with coded extremes of each factor coded as  $\pm 1$ . Because BBDs do not include the vertices of the  $k$ -dimensional cube, the precision of prediction at or near these points is very poor. The face-centered cube (FCC) designs (i.e., CCDs with

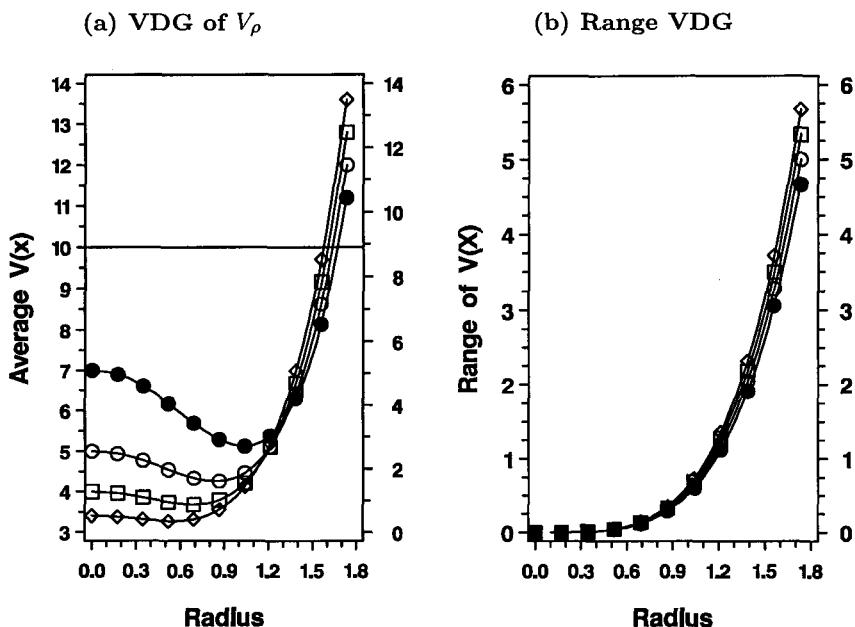


Fig. 5. VDG Plots for 3 factor Box-Behnken designs: solid dots for 2 CP, open circles for 3 CP, squares for 4 CP, and diamonds for 5 CP.

$\alpha = 1$ ), on the other hand, do include some or all of the  $2^k$  factorial hypercube vertices making CCDs popular designs for cuboidal regions. When considering VDGs for designs in a hypercube design region, there is no problem for  $\rho \leq 1$  because the hypersphere of radius  $\rho$  is contained in the hypercube with vertices  $(\pm 1, \pm 1, \dots, \pm 1)$ . However, a problem naturally arises when considering  $V(\mathbf{x})$  for  $\mathbf{x}$  with  $\rho > 1$  but still inside the  $k$ -dimensional hypercube. That is, for  $1 < \rho \leq \sqrt{k}$ , the points of distance  $\rho$  from the design center form subregions of the hypercube which will be referred to as hyperarcs. Let  $V_{H,\rho}$ ,  $VMIN_{H,\rho}$ , and  $VMAX_{H,\rho}$ , respectively, be the analogs of  $V_\rho$ ,  $VMIN_\rho$ , and  $VMAX_\rho$  for all points of radius  $\rho$  contained in the hypercube including the hyperarcs. Myers et al.<sup>24</sup> generated VDGs to assess the impact of center point replication for FCC designs and to compare a BBD to a FCC design.

Figure 6 contain examples of  $VMIN_{H,\rho}$  and  $VMAX_{H,\rho}$  for the six-factor face-centered cube designs having one or two sets of star-point replicates (assuming a second-order model). The cusps on the  $VMAX_{H,\rho}$  plot occur at  $\rho = \sqrt{t}$  for  $t = 1, 2, \dots, k$ . Borkowski<sup>19</sup> showed that when star-

points are replicated, there is no significant improvement in  $VMAX_{H,\rho}$  for  $k \leq 5$ , while there is a considerable reduction in the global performance of  $VMAX_{H,\rho}$  for  $k > 5$ . Only as  $\rho \rightarrow \sqrt{k}$  is the performance of  $VMAX_{H,\rho}$  poorer when star-points are replicated. Closed-form expressions for  $V_{H,\rho}$ ,  $VMIN_{H,\rho}$ , and  $VMAX_{H,\rho}$  for FCC designs and BBDs can also be found in Refs. 19, 20.

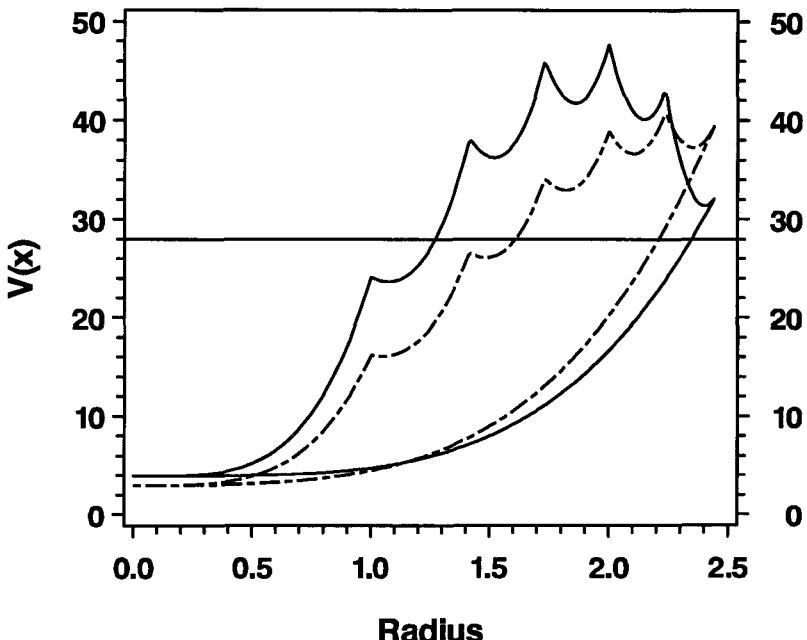


Fig. 6. Six-factor comparison in a cuboidal design space: VDGs of minimum and maximum  $V(x)$  for a  $n = 48$ -point FCC design 1 star-point replicate (solid line) and a for a  $n = 60$ -point FCC design with 2 star-point replicates (dashed line).

As an alternative to considering  $V(x)$  on hyperarcs for the cuboidal design space, calculation of the average, minimum, and maximum prediction variances can be restricted to the points on the surface of nested hypercubes with vertices at  $(\pm a, \pm a, \dots, \pm a)$  for  $0 \leq a \leq 1$ . These will be denoted  $V_{a,\rho}$ ,  $VMIN_{a,\rho}$ , and  $VMAX_{a,\rho}$ . Rozum and Myers<sup>25</sup> derived cuboidal region moments that were used to evaluate  $V_{a,\rho}$ ,  $VMIN_{a,\rho}$ , and  $VMAX_{a,\rho}$ . This

VDG approach was also discussed in Myers and Montgomery (pp. 409-412).<sup>2</sup> Figure 7 contains  $VMIN_{a,\rho}$  and  $VMAX_{a,\rho}$  for the six-factor designs previously studied in Figure 6. It shows that the FCC design with 2 star-point replicates is uniformly better than the FCC design with only 1 star-point replicate when studying  $V(\mathbf{x})$  on the surface of hypercubes. This was not the case, however, when using the surface of hyperspheres. Despite the differences, both figures provide useful information regarding the distribution and stability of  $V(\mathbf{x})$  in a cuboidal design region.

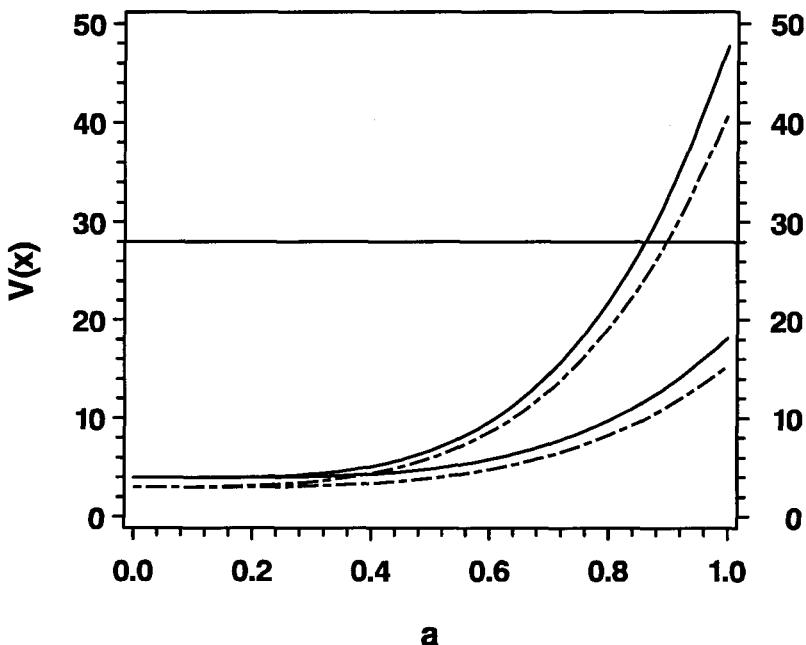


Fig. 7. Six-factor comparison in a cuboidal design space. The curves represent the minimum  $VMIN_{a,\rho}$  and maximum  $VMAX_{a,\rho}$  of  $V(\mathbf{x})$  over the set of points on nested hypercubes with vertices at  $(\pm a, \pm a, \dots, \pm a)$ . These VDGs correspond to the 48-point FCC design with 1 star-point replicate (solid line) and the 60-point FCC design with 2 star-point replicates (dashed line).

An extension of the use of VDGs to response surface experiments requiring blocking of experimental runs was explored by Trinca and Gilmour.<sup>26</sup> They concluded that designs can be run in small blocks and still retain good prediction variance properties, and that the choice of arrangement of

experimental runs into blocks can have a large impact on the performance of prediction.

Borror et al.<sup>27</sup> extended the use of VDGs to process robustness studies for which the goal is to simultaneously minimize the influence of process noise factors while determining optimal settings of the controllable process factors. Scaled prediction variance expressions were developed for both a mean model and for the slope. Another application was presented by Jang and Park<sup>28</sup> who introduced *slope variance dispersion graphs* (SVDGs). SVDGs are plots of the average, maximum, and minimum of the scaled average slope variance (SASV) on hyperspheres. These plots provide a way to compare the slope estimation properties of different designs.

## 5. The Prediction Variance Quantile (PVQ) Plot

Although VDGs provide useful information to the experimenter regarding the average and extremes for the prediction variance in the design region, they do not provide information regarding its distribution. A graphical technique was developed by Khuri et al.<sup>29</sup> to describe such a distribution using plots of its quantiles on each of several concentric spheres  $S_\rho$  inside a region of interest  $R$ . More specifically, if  $V(\mathbf{x}|\rho)$  denotes the value of  $V(\mathbf{x})$  at a point  $\mathbf{x}$  on the surface of  $S_\rho$ , then a plot of the quantiles of the distribution of  $V(\mathbf{x}|\rho)$  is obtained. These quantiles are estimated using a set of points randomly generated on the surface of  $S_\rho$ . Hyperspherical coordinates are used to represent the generated points. This process can be repeated for different values of  $\rho$  provided that  $S_\rho$  is contained inside the region  $R$ . This *prediction variance quantile plot* (PVQ plot) can then be used to study the properties of a design of interest or to compare competing designs at multiple radii. (Note, hyperspherical coordinates were also instrumental in Borkowski<sup>19,20</sup> when generating VDGs for BBDs and CCDs).

Figure 8 contains eight PVQ plots for comparison of three 3-factor designs: the 15-point BBD, the 17-point CCD with  $\alpha = \sqrt{3}$ , and the 13-point small composite design (SCD) with  $\alpha = \sqrt{3}$ . Each design has three center points. Figures 8(a)-(h) correspond respectively to  $\rho = 0.1, 0.3, \dots, 1.7$  while  $p$  on the horizontal axis, is the (approximate) probability of observing a value of  $V(\mathbf{x}|\rho)$  less than or equal to the corresponding quantile. From these plots we can see how stable  $V(\mathbf{x})$  is for the BBD and CCD relative to the SCD (especially as  $\rho \rightarrow \sqrt{3}$ ). The flatness of the PVQ plots for the CCD is expected because  $\alpha = \sqrt{3}$  is close to  $\sqrt[4]{8}$ , the value for the rotatable CCD. If a design was rotatable, it would yield perfectly horizontal

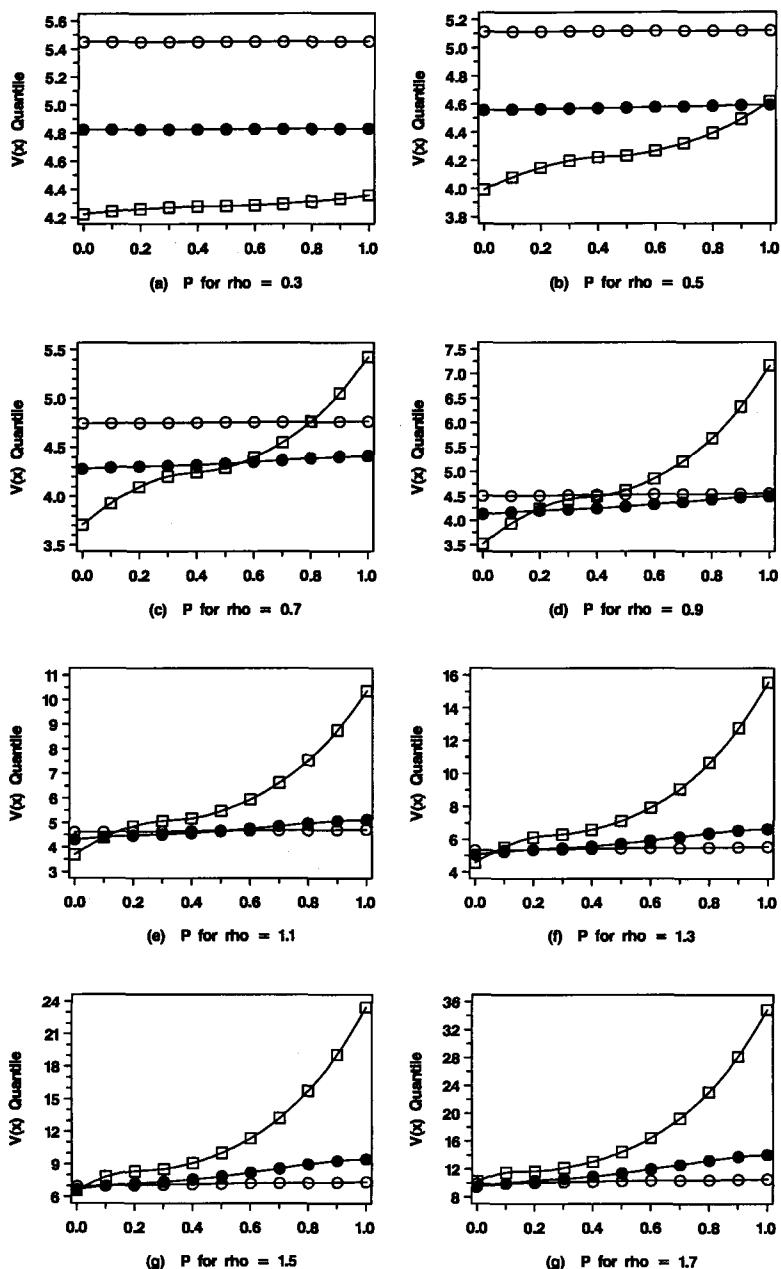


Fig. 8. PVQ plots for three 3-factor designs: the 15-point BBD (solid dots), the 17-point CCD (open circles), and the 13-point SCD (open squares).

PVQ plots for each  $\rho$ . For  $\rho \leq 0.5$ , the PVQ plot for the CCD is uniformly above the PVQ plots for the BBD and SCD. For  $\rho = 0.7$ , the PVQ plot for the CCD is again uniformly above the PVQ plot for the BBD while slightly lower in the right tail of the distribution for the SCD. However, as  $\rho \rightarrow \sqrt{3}$ , the distribution of  $V(\mathbf{x})$  values tends to be smaller for the CCD. Thus, if it is important to obtain high quality predictions near the boundary of the spherical design region, the CCD is the best design. For  $\rho$  close to zero, the SCD performs well because it has three center points like the other two designs but has fewer overall number of points. Thus, on a per point basis, the SCD is considered best for  $\rho$  near zero.

The direct relationship between VDGs and the PVQ plots can also be seen in Figure 8. That is, within each plot the left and right endpoints correspond to  $VMIN_\rho$  and  $VMAX_\rho$  for that plot's value of  $\rho$ . There are advantages, however, of using PVQ plots instead of VDGs. For example, it is possible that two designs may have nearly identical VDGs but the distributions of the spherical prediction variances at varying radii may be very different. Khuri et al.<sup>29</sup> presented such an example and, based on the PVQ plots found that one design's spherical prediction variances were more stable than they were for a competing design. Thus, it could be determined from the PVQ plots (and not the VDGs) which design had superior prediction capability.

A second way to summarize the distribution of  $V(\mathbf{x})$  across various values of  $\rho$  is through the use of boxplots of values of  $V(\mathbf{x}|\rho)$  for specified  $\rho$ . Figure 9 summarizes the distributions of the 3-factor, 13-point SCD with  $\alpha = \sqrt{3}$  that was also plotted in Figure 8. The side-by-side boxplots provide the experimenter with information that may not be immediately evident using the PVQ plotting method of Figure 8. The minimum and maximum values associated with the whiskers of the boxplots correspond to  $VMIN_\rho$  and  $VMAX_\rho$  for that boxplot's  $\rho$  value. The medians also provide a measure of central tendency as  $\rho$  increases while the range and interquartile range provide information regarding the stability of  $V(\mathbf{x}|\rho)$ . Figure 9 highlights the increasing trend for the center of the distribution of  $V(\mathbf{x}|\rho)$  for this SCD as  $\rho \rightarrow \sqrt{3}$ . That is, there are positive location shifts as  $\rho$  increases. Also,  $VMAX_\rho$  and the third quartile Q3 increase much more rapidly with increasing  $\rho$  than either  $VMIN_\rho$ , the first quartile Q1, or the median. Hence, the boxplots also indicate that the distribution of  $V(\mathbf{x}|\rho)$  becomes increasingly skewed right with increasing variability as  $\rho$  increases. For design comparison purposes, similar boxplots would be produced for the CCD and BBD.

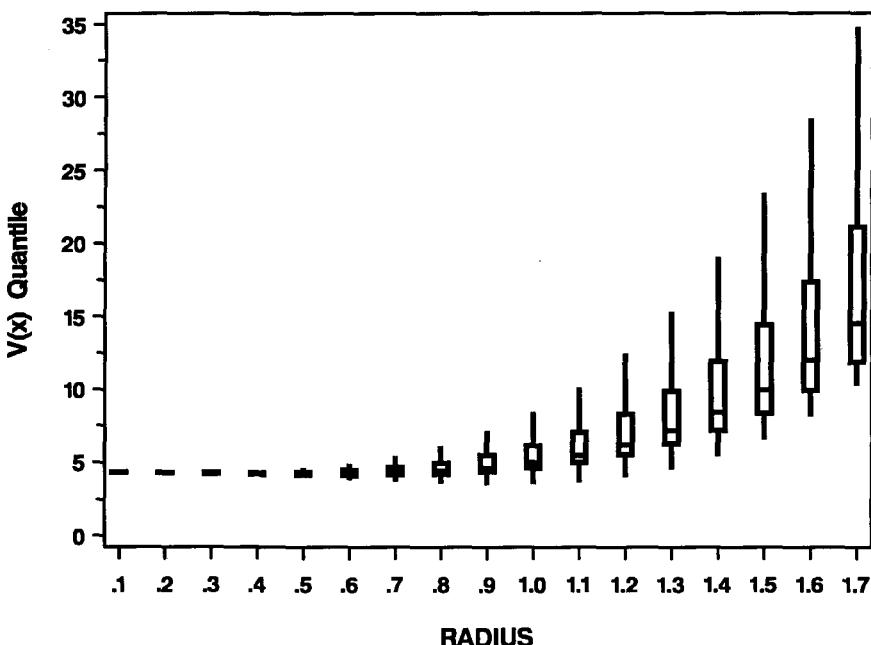


Fig. 9. Boxplots of values of  $V(x|\rho)$  for the 3-factor 13-point SCD at radii = 0.1, 0.2, ..., 1.7.

Khuri and his colleagues have adapted quantile plots to other statistical applications:

- (1) Kim et al.<sup>30</sup> developed quantile plots to study the distribution of the scaled average slope variance (SASV) on hyperspheres which are an extension of the SASV VDGs of Jang and Park.<sup>28</sup> These plots are graphical tools that can be used to compare designs based on their slope estimation capabilities.
- (2) Khuri<sup>31</sup> introduced quantile plots of the distribution of analysis of variance (ANOVA) estimator of a variance component. These plots, therefore, provide information concerning how often variance estimates would be negative. Although the methodology applies to any model provided normality assumptions are met, the focus is on the ANOVA estimator of a variance component in a balanced two-way model. These quantile plots depend on the true values of the model's unknown variance components. Khuri et al.<sup>31</sup> introduced the so-called *quantile dispersion graphs* (QDGs) to graphically describe the dependence of the

quantiles plots on the variance components. This was accomplished by computing the maximum and minimum of the  $p^{th}$  quantile value over a specified parameter space of the variance components for a given design and a given value of  $p$ . Plotting these extrema against  $p$  produces the QDGs for the given design. The QDGs enable one to compare several designs on the basis of their estimation capability while accounting for the dependence of the designs on the values of the variance components. In Lee and Khuri,<sup>32</sup> QDGs associated with two methods of estimation, ANOVA and maximum likelihood (ML), are studied. A further extension of QDGs to compare designs for an unbalanced random two-way model without interaction can be found in Lee and Khuri.<sup>33</sup> ANOVA and ML estimation were again studied.

- (3) For nonlinear models, it is common to study mean-squared error of prediction. Khuri and Lee<sup>34</sup> generated quantile plots of the *estimated scaled mean-squared error of prediction* (ESMSEP) on concentric surfaces inside the design region. For models with only one input variable, plots of the minimum and maximum *scaled mean-squared error of prediction* (SMSEP) over a subset of the parameter space were developed for use in comparison of nonlinear designs.
- (4) Applications of QDGs to logistic regression were developed by Robinson and Khuri.<sup>35</sup> Plots of the maxima and minima of the quantiles of the scaled mean-squared error of prediction over a subset of the parameter space were generated which can be used for evaluating or comparing designs for a logistic regression model.

## 6. The Fraction of Design Space (FDS) Plot

It is common for the experimenter not to have knowledge prior to running an experiment regarding where in the design region prediction is most important. It is also common for an experimental design to have regions associated with small  $V(\mathbf{x})$  values but with a tradeoff of other regions having large  $V(\mathbf{x})$  values. Thus, the experimenter would like to assess a design's prediction variance properties throughout the entire design region. Although both VDGs and PVQ plots provide some information for addressing this problem, the information is not complete. Although it can be determined from VDGs and PVQ plots the maximum and minimum values of  $V(\mathbf{x})$  and the distribution of  $V(\mathbf{x})$  values for a given  $\rho$ , they do not take into account the proportion of the volume of the design region for a sphere of radius  $\rho$ . This problem was addressed by Zahran et al.<sup>36</sup> when they

observed that the visual information provided by VDGs and PVQ plots are given the same weight at each  $\rho$ . They recommended that the information should be weighted by the proportion of the design region it represents.

The proportion of a spherical design region accounted for by all points within radius  $\rho$  of the design center is obviously an increasing function of  $\rho$ . As the dimension of the design increases in terms of the number of design factors ( $k$ ), the proportion of the design region becomes negligible for  $\rho$  close to zero but increases rapidly as  $\rho$  increases. Thus, the relatively small interval for  $\rho$  for those points near the design space boundary dominate the prediction capability of a design in the VDG. This is also true for the PVQ plots for the largest values of  $\rho$ .

Zahran et al.<sup>36</sup> developed a graphical method called the *fraction of design space* or FDS plot. An FDS plot involves plotting the quantiles  $Q$  of  $V(\mathbf{x})$  against  $P$ , the proportion of the volume of the design region  $R$  for which  $V(\mathbf{x}) \leq Q$  for any specified value  $Q$ .

For example, Figure 10 contains FDS plots for BBDs having 2, 3, 4, and 5 center points (which are the same designs in the VDGs in Figure 5). From these FDS plots we can compare designs. In Figure 10 we see that approximately 75% of the design region has  $V(\mathbf{x}) \leq 10$  (the optimal  $V(\mathbf{x})$ ) for the 5 center point BBD. The percentage increases to  $\approx 80\%$ ,  $85\%$  and  $90\%$  for the 4, 3, and 2 center point BBDs, respectively. Thus,  $V(\mathbf{x})$  is better than the ideal value (on a per point basis) for  $\approx 5\%$ ,  $10\%$  and  $15\%$  more of the spherical design space for the BBDs with 4, 3, and 2 center points than with 5 center points.

Figure 11 contains FDS plots for the 3-factor BBD, CCD, and SCD designs whose PVQ plots were presented in Figure 8. Based on the PVQ plots with  $\rho = .3$  and  $\rho = .5$  (Figures 8(a),(b)),  $V(\mathbf{x})$  is uniformly smaller for the SCD than it is for the BBD and CCD. It continues to be smaller than  $V(\mathbf{x})$  of the CCD for  $\approx 80\%$  of its distribution at  $\rho = .7$ . However, the proportion of the design region is small for these values of  $\rho$ . The FDS plots in Figure 11(a) indicate that in approximately 90% of the design region, values of  $V(\mathbf{x})$  for the BBD and CCD are less than the  $G$ -optimal value of 10. For the SCD, however, only 60% of the design region has  $V(\mathbf{x}) < 10$ . In Figure 11(b), the FDS plots of the unscaled prediction variance function  $V_1(\mathbf{x}) = V(\mathbf{x})/N$  are shown for the three designs. Because the design sizes are 17, 15, and 13, respectively, for the CCD, BBD, and SCD, the separation will be relatively larger than what was seen in Figure 11(a).

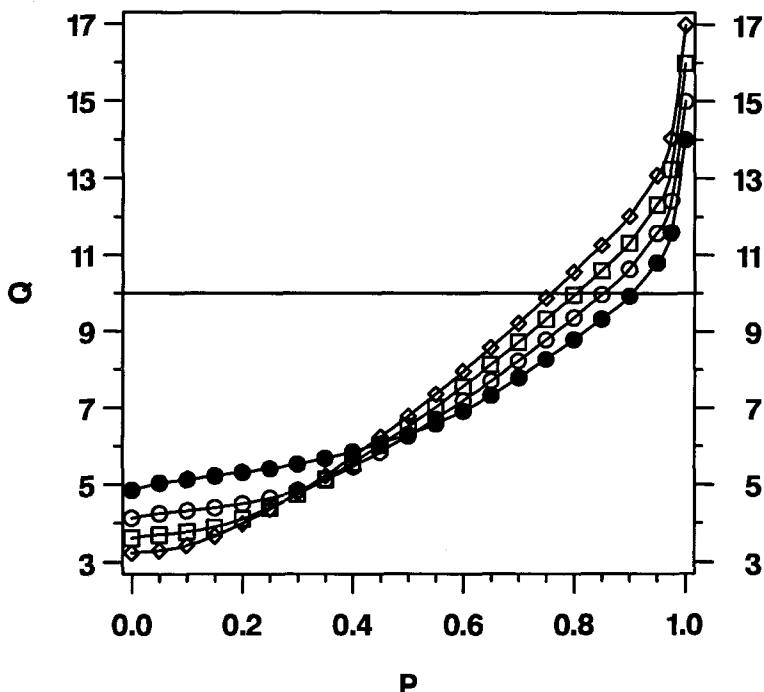


Fig. 10. FDS Plots for 3 factor BBDs with 2 center points (solid dots), 3 CPs (open circles), 4 CPs (open squares), and 5 CPs (open diamonds).

## 7. Applications to Experiments Involving Mixtures

So far in this chapter, and for all designs, the factors were process variables that could be set independently of each other. In experiments involving mixtures, however, this is not the case. In an experiment involving only mixture variables (often called *mixture components* or *mixture ingredients*), there is the constraint that the components sum to 1. That is, each experimental run in a  $q$ -component mixture design with component proportions

$x_1, \dots, x_q$  satisfies  $\sum_{i=1}^q x_i = 1$  and  $0 \leq L_i \leq x_i \leq U_i \leq 1$  for all  $i$  where

$L_i$  and  $U_i$  are lower and upper bounds for component proportion  $x_i$ . When all  $L_i = 0$  and all  $U_i = 1$ , the design region is the  $(q - 1)$ -dimensional simplex. If one or more of the  $L_i > 0$  or the  $U_i < 1$ , the design region a polyhedron (and often irregularly shaped). In some mixture experiments,

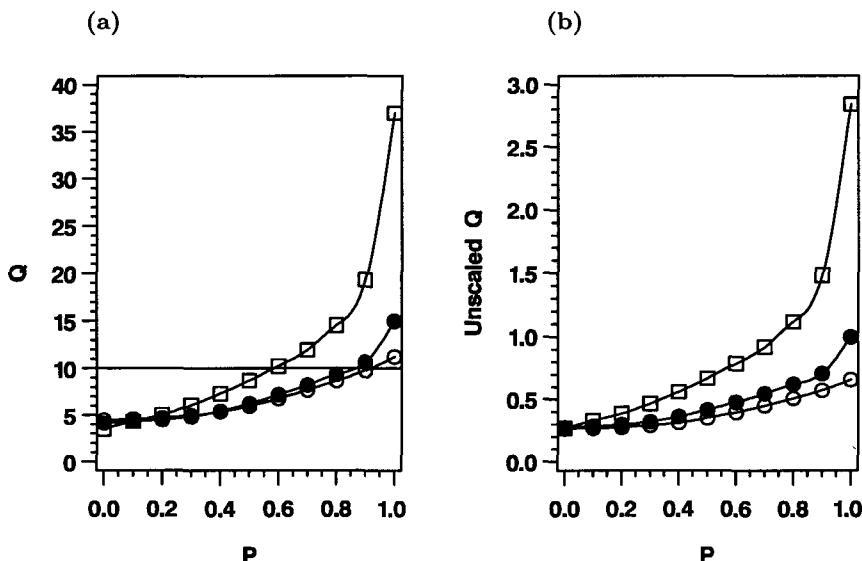


Fig. 11. FDS Plots for 3 factor Box-Behnken ( $N = 15$ ), central composite ( $N = 17$ ), and small composite ( $N = 13$ ) designs. Solid dots correspond to the BBD, open circles for the CCD, and open squares for the SCD.

additional constraints are placed on the  $x_i$  (e.g., a linear combination constraint such as  $a \leq \sum_{i=1}^q a_i x_i \leq A$  for some constants  $a$  and  $A$ ). For a review of experiments with mixtures, see Cornell.<sup>37</sup>

Whether the design regions is the  $(q - 1)$ -dimensional simplex or some irregularly-shaped subregion of the simplex, the problem arises as to how VDGs, PVQ plots, and FDS plots can be extended to experiments involving mixtures. Piepel and Anderson<sup>38</sup> developed VDGs for the average, minimum, and maximum for any design in a polyhedral design region  $R$ . Designs for mixture experiments naturally fall into this class of designs. They noted that although it is possible to generate VDGs based on spheres, the shape of the design region  $R$  is a problematic issue. As the radius of the sphere increases, an increasing portion of the surface will lie outside  $R$  causing unusual patterns in a VDG (like the cusps seen in Figure 6). Thus, their approach was to apply a *shrinkage factor*  $s$  to  $R$ . For each  $s$ , a "shrunken" version  $R_s$  of the polyhedral region  $R$  about a *shrinkage point*  $\mathbf{x}_0$  is created. Then, using subsets of points  $\mathbf{x} \in R_s$ , the average, minimum, and maximum of  $V(\mathbf{x})$  are estimated (which will be denoted as  $V_{R,s}$ ,  $VMIN_{R,s}$  and  $VMAX_{R,s}$ , respectively). The  $V_{R,s}$ ,  $VMIN_{R,s}$  and  $VMAX_{R,s}$  values

are then plotted against the shrinkage factor  $s$  yielding VDGs. In Piepel et al.<sup>39</sup>, the impact of the choice of shrinkage point  $\mathbf{x}_0$  on the VDGs was studied. See Ref. 40 for further examples of VDGs using shrinkage factors and shrunken polyhedral regions. Because the experimenter may be interested in comparing the performance of different designs near the center of  $R$ , Piepel et al.<sup>40</sup> introduced a VDG approach using shrunken spherical regions.

An alternative graphical method to VDGs for mixture designs, called *prediction variance trace* (PVT) plots, was proposed by Vining et al.<sup>41</sup> In a PVT plot, for each component  $x_i$ ,  $V(\mathbf{x})$  values are plotted along the *Cox directions*.<sup>42</sup> Cox directions are vectors passing through a reference mixture (often the centroid of  $R$ ) in the direction where the relative proportions of any two components remain the same as their ratio in the reference mixture. Because there is a Cox direction for each mixture component, there will be  $q$  VDGs. Because only  $q$  directions are considered, information about  $V(\mathbf{x})$  in all other directions is ignored.

A new VDG will now be proposed that summarizes  $V(\mathbf{x})$  throughout the entire polyhedral design region  $R$ . For the  $i^{th}$  mixture component, consider fixing its component proportion at  $x_i = a$ . Let  $V_{x_i=a}$ ,  $VMIN_{x_i=a}$ , and  $VMAX_{x_i=a}$  be the average, minimum, and maximum of  $V(\mathbf{x})$  over the space of mixtures with  $x_i = a$  that satisfy the set of imposed constraints on the component proportions. The plots of  $V_{x_i=a}$ ,  $VMIN_{x_i=a}$ , and  $VMAX_{x_i=a}$  across the set of feasible values of  $x_i$  will be called *componentwise variance dispersion graphs* (CVDGs) for component  $i$ .

Figure 12 contains the CVDGs for the extreme-vertices design for the 4-component 15-point flare experiment presented in McLean and Anderson.<sup>43</sup> The quadratic Scheffé mixture model

$$Y = \sum_{i=1}^q \beta_i x_i + \sum_{i=1}^{q-1} \sum_{j=i+1}^q \beta_{ij} x_i x_j + \epsilon \quad (12)$$

was assumed. In this experiment, the component constraints were given as

$$0.4 \leq x_1 \leq 0.6 \quad 0.1 \leq x_2, x_3 \leq 0.5 \quad 0.03 \leq x_4 \leq 0.08.$$

Thus, in Figure 12, the  $V_{x_1=a}$ ,  $VMAX_{x_1=a}$  and  $VMIN_{x_1=a}$  CVDGs begin at  $x_1 = 0.4$  and end at  $x_1 = 0.6$ . For  $x_4$ , the VDGs begin at  $x_4 = 0.03$  and end at  $x_4 = 0.08$ . Note, however, that  $0.10 \leq x_2, x_3 \leq 0.47$  form the actual bounds for  $x_2$  and  $x_3$  given the stated bounds for  $x_1$  and  $x_4$ . Thus, for  $x_2$  and  $x_3$ , the VDGs begin at 0.1 and end at 0.47. The CVDGs present a picture of the variability in the entire polyhedral design region  $R$ .

(unlike PVT plots). For multiple designs, CVDGs would be generated for each design and their prediction variance properties compared. Note that once data has been collected, analogous componentwise plots could also be generated for the predicted  $\hat{Y}(\mathbf{x})$  values. As the number of components  $q$  increases, however, comparison of designs using CVDGs will become more difficult.

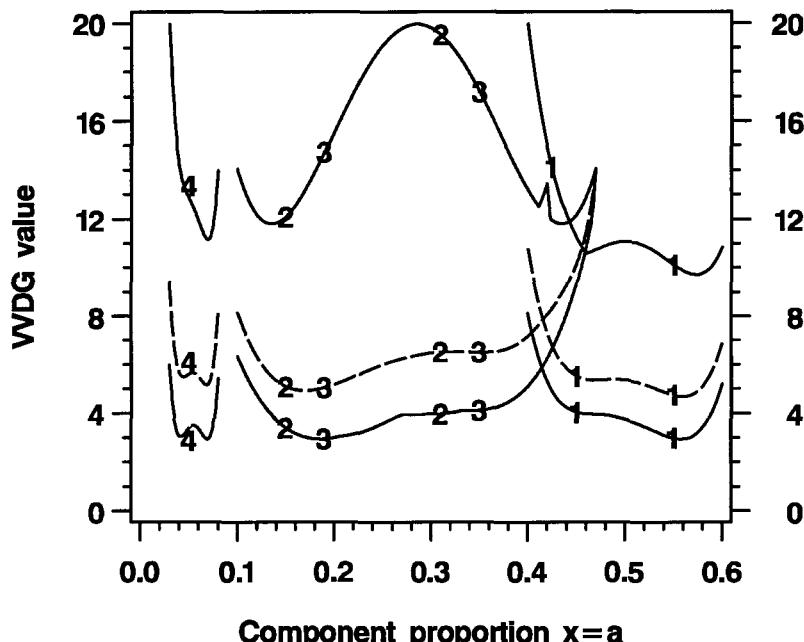


Fig. 12. CVDG for  $V_{x_i=a}$  (dashed line) and CVDGs for  $V_{MIN_{x_i=a}}$  and  $V_{MAX_{x_i=a}}$  (solid lines) for the 4-component 15-point mixture design.

Note that for the set of CVDGs, the maxima across the  $V_{MAX_{x_i=a}}$  will always be equal, and the minima across  $V_{MIN_{x_i=a}}$  will always be equal. In Figure 12, the maxima occur at  $(x_1, x_2, x_3, x_4) = (0.4, 0.285, 0.285, 0.03)$  which corresponds to the mixture in  $R$  having maximum  $V(\mathbf{x})$ , or, equivalently, the point  $\mathbf{x}$  yielding the design's  $G$ -value ( $\approx 20$ ). Analogously, from the CVDGs, the minimum of  $V(\mathbf{x}) \approx 2.94$  over  $R$  occurs at  $(x_1, x_2, x_3, x_4) \approx (0.55, 0.19, 0.19, 0.07)$ . The PVT plots (see Ref. 41), however, indicate the minimum  $V(\mathbf{x}) \approx 3.5$  and the maximum  $V(\mathbf{x}) \approx 10$ . From the CVDGs in

Figure 12, we can see that these PVT plot values do not reflect the minimum and maximum values over  $R$  (with the global maximum twice as large as the maximum in the PVT plots).

Another graphical technique has also been extended to mixture experiments. Khuri et al.<sup>44</sup> proposed *scaled prediction variance quantile* (SPVQ) plots for describing the distribution of  $V(\mathbf{x})$  within  $R$  using PVQ plots. Suppose the polyhedral design region  $R$  is shrunken with shrinkage factor  $s$ . In a SPVQ plot for a given  $s$ , the quantiles of  $V(\mathbf{x})$  on the surface of the shrunken region are plotted. To get an overall picture of  $V(\mathbf{x})$  in  $R$ , SPVQ plots are generated for multiple values of  $s$ . Like the CVDGs, the SPVQ plots provide information of the minimum and maximum of  $V(\mathbf{x})$  throughout  $R$ .

The fraction of design space (FDS) plots can also be extended to mixture experiments. For example, Figure 13 contains FDS plots for three 15-point mixture designs for the previously mentioned flare example of McLean and Anderson (M&A).<sup>43</sup> The three designs include the M&A extreme vertices design (see Figure 12), a *D*-optimal design, and a *IV*-optimal design. For

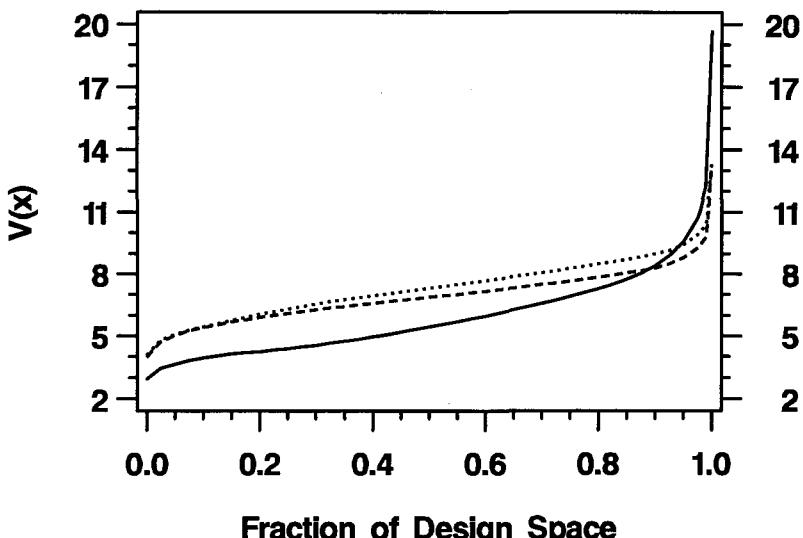


Fig. 13. FDS plots for three 4-component 15-point mixture designs for the McLean and Anderson flare example. The solid line is for the McLean and Anderson design, the dotted line for a *D*-optimal design, and the dashed line for the *IV*-optimal design.

the M&A design, in approximately 50% of the design space,  $V(\mathbf{x}) \leq 5$  while only 5% of  $V(\mathbf{x}) \leq 5$  for the  $D$  and  $IV$ -optimal designs. The distribution for the M&A design, however, has a longer right tail (although the proportion of large  $V(\mathbf{x})$  values (say,  $V(\mathbf{x}) > 13$ ) is very small).

Other important developments of graphical techniques for displaying prediction variance properties are presented in Goldfarb et al.<sup>45,46</sup> who proposed VDGs and FDS plots for experiments that have both mixture and process variables. For experiments with both types of variables, Goldfarb et al.<sup>45</sup> consider two shrinkage factors:  $s_m$  for the mixture factors and  $s_p$  for the process factors. Then three-dimensional VDGs can be generated with  $(x, y) = (s_p, s_m)$  and  $z =$  the prediction variance function of interest. Both contour plots and 3-dimensional surface plots were presented. In Goldfarb et al.<sup>46</sup>, mixture-process designs where some of the process variables are noise variables were studied. They developed expressions for both the scaled and unscaled predictions variances for mean and slope models. VDG and FDS plots were then developed that accounted for the presence of noise factors.

## 8. Final Comments

Since the publication of the 1989 Giovannitti-Jensen and Myers paper,<sup>3</sup> there has been continued interest in graphical representations of the prediction variance properties associated with response surface designs. The primary techniques developed were VDGs, PVQ plots, and FDS plots.

Despite these advances in graphical methods, there is more work that can be done. An additional application of these types of plots can be to develop plots that incorporate potential model bias. For example, Vining and Myers<sup>47</sup> extended the VDG approach to the mean squared error of prediction and Piepel et al.<sup>40</sup> developed *bias dispersion graphs* (BDGs) for designs in irregularly-shaped design regions. Another potential application that has yet to be explored is in the area multiresponse experiments. Although Fogliatto and Algin<sup>48</sup> studied the prediction variance in multiresponse experiments, graphical techniques still need to be developed. Even though new methods will appear in the future, the graphical methods reviewed in this chapter will continue to assist the experimenter in the assessment of the properties of response surface designs.

## References

1. G.E.P. Box and J.S. Hunter, "Multi-Factor Experimental Designs for Exploring Response Surfaces", *The Annals of Mathematical Statistics* **22**, 195-241 (1957).
2. R.H. Myers and D.C. Montgomery, *Response Surface Methodology: Process and Product Optimization Using Designed Experiments* (John Wiley & Sons, New York, 2002).
3. A. Giovannitti-Jensen and R.H. Myers, "Graphical Assessment of the Prediction Capability of Response Surface Designs", *Technometrics* **31**, 159-171 (1989).
4. T.L. Hebble and T.J. Mitchell, "Repairing" Response Surface Designs", *Technometrics* **14**, 767-779 (1972).
5. A.I. Khuri, "A Measure of Rotatability for Response Surface Designs", *Technometrics* **30**, 95-104 (1988).
6. G.E.P. Box and N.R. Draper, "A Basis for the Selection of a Response Surface Design", *Journal of the American Statistical Association* **72**, 622-654 (1959).
7. G.E.P. Box and N.R. Draper, "The Choice of a Second Order Rotatable Design", *Biometrika* **50**, 335-352 (1963).
8. J.J. Borkowski, "A Comparison of Prediction Variance Criteria for Response Surface Designs", *Journal of Quality Technology*, **35**, 70-77 (2003a).
9. J.J. Borkowski, "Using a Genetic Algorithm to Generate Exact Small Response Surface Designs", *Journal of Probability and Statistical Science*, **1**, 65-88 (2003b).
10. J. Kiefer, "Optimal Design: Variation in Structure and Performance under Change of Criterion", *Biometrika* **62**, 277-288 (1975).
11. J.J. Borkowski and E.S. Valeroso, "Comparison of Design Optimality Criteria of Reduced Models for Response Surface Designs in the Hypercube", *Technometrics* **43**, 468-477 (2001).
12. L.M. Haines, "The Application of the Annealing Algorithm to the Construction of Exact Optimal Designs for Linear Regression Models", *Technometrics* **29**, 439-447 (1987).
13. L.M. Haines, "Optimal Design for Nonlinear Regression Models", *Communications in Statistics – Theory and Methods* **22**, 1613-1627 (1993).
14. A.C. Atkinson and A.N. Donev, *Optimum Experimental Designs* (Oxford University Press, New York, 1992).
15. G.E.P. Box and N.R. Draper, *Empirical Model-Building and Response Surfaces* (John Wiley & Sons, New York, 1987).
16. J.R. Hussey, R.H. Myers, and E.C. Houck "Correlated Simulation Experiments in First-Order Response Surface Design", *Operations Research* **35**, 744-758 (1987).
17. G.G. Vining, "A Computer Program for Generating Variance Dispersion Graphs", *Journal of Quality Technology* **25**, 45-47 (1993a).
18. G.G. Vining, Corrigenda for "A Computer Program for Generating Variance Dispersion Graphs", *Journal of Quality Technology* **25**, 333-335 (1993b).

19. J.J. Borkowski, "Spherical Prediction-Variance Properties of Central Composite and Box-Behnken Designs", *Technometrics* **37**, 399-410 (1995a).
20. J.J. Borkowski, "Minimum, Maximum, and Average Spherical Prediction Variances for Central Composite and Box-Behnken Designs", *Communications in Statistics – Theory and Methods* **24**, 2581-2600 (1995b).
21. A.E. Hoerl, "Optimum Solution of Many Variables Equations", *Chemical Engineering Progress*, **55**, 69-78 (1959).
22. A.E. Hoerl, "Ridge Analysis", *Chemical Engineering Progress*, **60**, 67-77 (1964).
23. G.E.P. Box and D.W. Behnken, "Some New Three Level Designs for the Study of Quantitative Variables", *Technometrics* **2**, 455-475 (1960).
24. R.H. Myers, G.G. Vining, A. Giovannitti-Jensen, and S.L. Myers, "Variance Dispersion Properties of Second-Order Response Designs", *Journal of Quality Technology* **24**, 1-11 (1992).
25. M.A. Rozum abd R.H. Myers, "Adaptation of Variance Dispersion Graphs to Cuboidal Regions of Interest", *Presented at Joint Statistical Meetings*, American Statistical Association, Atlanta, GA, 1991.
26. L.A. Trinca and S.G. Gilmour, "Variance Dispersion Graphs for Comparing Blocked Response Surface Designs", *Journal of Quality Technology* **30**, 314-327 (1998).
27. C.M. Borror, D.M. Montgomery, and R.H. Myers, "Evaluation of Statistical Designs for Experiments Involving Noise Variables", *Journal of Quality Technology* **34**, 54-70 (2002).
28. D.H. Jang and S.H. Park, "A Measure and a Graphical Method for Evaluating Slope Rotatability in Response Surface Designs", *Communications in Statistics: Theory and Methods* **22**, 1849-1863 (1993).
29. A.I. Khuri, H.J. Kim, and Y. Um, "Quantile Plots of the Prediction Variance for Response Surface Designs", *Computational Statistics & Data Analysis* **22**, 395-407 (1996).
30. H.J. Kim, Y. Um, and A.I. Khuri, "Quantile Plots of the Average Slope Variance for Response Surface Designs", *Communications in Statistics – Simulation and Computation*, **25**, 995-1014 (1996).
31. A.I. Khuri, "Quantile Dispersion Graphs for Analysis of Variance Estimates of Variance Components", *Journal of Applied Statistics* **24**, 711-722 (1997).
32. J. Lee and A.I. Khuri, "Graphical Technique for Comparing Designs for Random Models", *Journal of Applied Statistics* **26**, 933-947 (1999).
33. J. Lee and A.I. Khuri, "Quantile Dispersion Graphs for the Comparison of Designs for a Random Two-Way Model", *Journal of Statistical Planning and Inference* **91**, 123-137 (2000).
34. A.I. Khuri and J. Lee, "A Graphical Approach for Evaluating and Comparing Designs for Nonlinear Models", *Computational Statistics and Data Analysis* **27**, 433-443 (1998).
35. K.S. Robinson and A.I. Khuri, "Quantile Dispersion Graphs for Evaluating and Comparing Designs for Logistic Regression Models", *Computational Statistics and Data Analysis* **43**, 47-62 (2003).

36. A. Zahran, C.M. Anderson-Cook, and R.H. Myers, "Fraction of Design Space to Assess Prediction Capability of Response Surface Designs", *Journal of Quality Technology* **35**, 377-386 (2003).
37. J.A. Cornell, *Experiments with Mixtures* (John Wiley & Sons, New York, 2002).
38. G.F. Piepel and C.M. Anderson, "Variance Dispersion Graphs for Designs on Polyhedral Regions", *1992 Proceedings of the Section on Physical and Engineering Sciences* (American Statistical Association, Alexandria, VA, 1993) pp. 111-117.
39. G.F. Piepel, C.M. Anderson, and P.E. Redgate, "Variance Dispersion Graphs for Designs on Polyhedral Regions - Revisited", *1993 Proceedings of the Section on Physical and Engineering Sciences* (American Statistical Association, Alexandria, VA, 1994) pp. 102-107.
40. G.F. Piepel, C.M. Anderson, and P.E. Redgate, "Response Surface Designs for Irregularly-Shaped Regions - Parts I, II, III", *1993 Proceedings of the Section on Physical and Engineering Sciences* (American Statistical Association, Alexandria, VA, 1994) pp. 205-225.
41. G.G. Vining, J.A. Cornell, and R.H. Myers, "A Graphical Approach for Evaluating Mixture Designs", *Applied Statistics* **42**, 127-138 (1993).
42. D.R. Cox, "A Note on Polynomial Response Functions for Mixtures", *Biometrika* **58**, 155-159 (1971).
43. R.A. McLean and V.I. Anderson, "Extreme Vertices Design of Mixture Experiments", *Technometrics* **8**, 447-454. (1966).
44. A.I. Khuri, J.M. Harrison, and J.A. Cornell, "Using Quantile Plots of the Prediction Variance for Comparing Designs for a Constrained Mixture Region: An Application Involving a Fertilizer Experiment", *Applied Statistics* **48**, 521-532 (1999).
45. H.B. Goldfarb, C.M. Borror, D.C. Montgomery, and C.M. Anderson-Cook, "Three-Dimensional Variance Dispersion Graphs for Mixture-Process Experiments", *Journal of Quality Technology* **36**, 109-124 (2004).
46. H.B. Goldfarb, C.M. Borror, D.C. Montgomery, and C.M. Anderson-Cook, "Evaluating Mixture-Process Designs with Control and Noise Variables", *Journal of Quality Technology* **36**, 245-262 (2004).
47. G.G. Vining and R.H. Myers, "A Graphical Approach for Evaluating Response Surface Designs in Terms of the Mean Squared Error of Prediction", *Technometrics* **33**, 315-326 (1991).
48. F.S. Fogliatto and S.L. Albin, "Variance of Predicted Response as an Optimization Criterion in Multiresponse Experiments", *Quality Engineering* **12**, 523-533 (2000).

## **CHAPTER 15**

### **USING FRACTION OF DESIGN SPACE PLOTS FOR INFORMATIVE COMPARISONS BETWEEN DESIGNS**

Christine M. Anderson-Cook

*985 Los Pueblos Street, Los Alamos, NM 87544*

*E-mail: c-and-cook@lanl.gov*

Ayca Ozol-Godfrey

*59 New Holland Village, Building #6, Nanuet, NY 10954*

*E-mail: godfrea@wyeth.com*

Graphical methods are commonly sought to assess prediction performance of designs and making comparisons between competing designs. The Fraction of Design Space (FDS) plot allows for detailed information to be extracted from a single curve of the scaled prediction variance (SPV) for an assumed model and specified design region for a particular design. Measures of G- and V-efficiency are also easily available from the plot. Variations to the plot are presented which allow more specialized comparisons. For example, if additional experimental runs are not costly, studying the prediction variance directly, instead of the SPV, may be recommended. The FDS plot can also easily accommodate non-standard design regions. If the design variables can naturally be considered as two groups, such as control and noise variables, mixture and process variables, or whole-plot and sub-plot variables, then the design region can be partitioned into subsections which can be considered separately with several FDS curves to complement the global summary of the entire design space. If there is model uncertainty, adaptations to the FDS plot will allow several different models to be simultaneously compared. Examples illustrate a variety of different applications and compare competing designs in each case.

## 1. Introduction

Frequently in planned experiments, researchers have a good sense of which factors are worthy of further investigation, what ranges of values for each of those factors might be of interest, and have a tentatively proposed model which they wish to be able to fit. Several designs are typically possible given these conditions, and the researcher should choose between the options based on some strategy to get the best design given the particular priorities of the experiment. In Myers and Montgomery<sup>1</sup>, a list of some important characteristics of a good design is given. These include balancing cost, the ability to estimate all of the parameters in the model adequately, the ability to assess lack of fit and estimate pure error, and the ability to predict well in the design space. Typically not all of these good characteristics of a design are simultaneously possible, and so selecting a best design usually involves prioritizing what is most important for a given situation and making trade-offs between the various desirable design criteria.

Kiefer and Wolfowitz<sup>2</sup> were among the first authors who developed formal optimality criteria. These are single number criteria where each one intends to capture a different aspect of the 'goodness' of a design. Traditionally, many of the alphabetical criteria have focused on two major aspects of the design: the ability to estimate the model parameters, and the ability to predict new responses based on the model. D-optimality was the first alphabetical optimality criterion developed. It is also still among the most popular because of its simple computation, and the many available algorithms. The focus of D-optimality is on estimation of model parameters through good attributes of the moment matrix, which is defined as  $M = X^T X / N$  where  $X^T X$  is the information matrix for the model  $y = X\beta + \varepsilon$ , which assumes a completely randomized structure, and N, the total number of runs, is used as a penalty for larger designs.

Frequently one of the primary aims of the practitioner is to have good prediction at a particular location in the design space. Box and Hunter<sup>3</sup> defined the scaled prediction variance (SPV), which provides a measure of the precision of the estimated response at any point in the design space. The goal of a good design based on this criterion is to have the

distribution of the scaled prediction variance throughout the design space be reasonably stable. The SPV is defined as

$$\nu(\mathbf{x}_o) = \frac{N \operatorname{var}(\hat{y}(\mathbf{x}_o))}{\sigma^2} = N \mathbf{x}_o' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_o \quad (1)$$

where  $\mathbf{x}_o$  corresponds to the location in the design space written as a function of the chosen model,  $N$  is the total sample size,  $\operatorname{var}(\hat{y}(\mathbf{x}_o))$  is the variance of the estimated response at  $\mathbf{x}_o$ ,  $\mathbf{X}$  is the model matrix, and  $\sigma^2$  is the variance. For example, for a first order model with interaction,  $\mathbf{x}'_o = (1, x_{10}, x_{20}, x_{10}x_{20})$ . For a standard  $2^2$  factorial design with 1 center run ( $N=5$ ) using a first order model with interaction,

$$\begin{aligned} \nu(\mathbf{x}_o) &= N(1, x_{10}, x_{20}, x_{10}x_{20}) \begin{bmatrix} 1/5 & 0 & 0 & 0 \\ 0 & 1/4 & 0 & 0 \\ 0 & 0 & 1/4 & 0 \\ 0 & 0 & 0 & 1/4 \end{bmatrix} \begin{bmatrix} 1 \\ x_{10} \\ x_{20} \\ x_{10}x_{20} \end{bmatrix} \\ &= 1 + \frac{5}{4}(x_{10}^2 + x_{20}^2 + x_{10}^2 x_{20}^2). \end{aligned} \quad (2)$$

The goal of G-optimality is to minimize the maximum SPV throughout the region of the design. Intuitively, G-optimality tries to protect the experimenter against the worst-case scenario for prediction variance being too undesirable. An interesting and an important result is that the lower bound for the maximum SPV is equal to  $p$ , the number of parameters in the model<sup>4</sup>. Therefore, G-efficiency is defined as

$$\text{G-efficiency} = \frac{p}{\max_{x \in R} \nu(x)}. \quad (3)$$

V-optimality, also called I-, IV- and Q-optimality in the literature, is also based on properties of the scaled prediction variance, SPV. Instead of finding the maximum SPV in the design space, it makes use of the average of the SPV throughout the design space. Hence this gives a more commonly used statistical measure of the overall distribution of the SPV. Box and Draper<sup>5,6</sup> discuss the concept of average prediction variance. The aim of V-optimality is to minimize the average of the SPV throughout the whole region of interest. Even though, its computation is

the hardest, V-optimality may frequently be the most natural choice for optimality in terms of measuring an important characteristic of the design.

Even though the various optimality criteria above summarize important characteristics of a good design, they by necessity must reduce the information for each design to a single number. This may be too simplistic a summary for understanding the inherent properties of the design. Hence, graphical methods based around the scaled prediction variance can provide a more detailed way of comparing competing designs, for experiments where prediction of future observations is a priority.

Variance Dispersion Graphs (VDGs), developed by Giovannitti-Jensen and Myers<sup>7</sup>, are a graphical tool to study a design's prediction properties. It displays the SPV throughout a multidimensional region on a single two-dimensional graph. On the x-axis, the VDG uses the distance from the center of the design (scaled either by radius or by shrinkage factor<sup>8</sup>) and on the y-axis are the corresponding SPV values. The shrinkage concept allows for cuboidal or irregularly shaped regions to be summarized with the VDGs by shrinking the outer edge of the design space down towards the centroid, for shrinkages of 1 to 0. The VDG consists of three curves per design for standard linear models: the minimum, maximum and average SPV values for a given distance from the center. In addition to its three curves, two horizontal curves, the 100% and 50% G-efficiencies are frequently displayed which correspond to  $\nu(x) = p$  and  $\nu(x) = 2p$ . These graphs are effective gauging the prediction variance throughout the design space; however, because of the multiple curves per design they can sometimes be confusing for making comparisons between competing designs. In addition, because different proportions of the total design space are associated with different distances from the center depending on the number of factors considered, understanding the relative contributions from different sections of the graph requires some experience.

Quantile Dispersion Graphs<sup>9,10</sup> describe the distribution of the SPV using quantiles for a distance  $r$  from the center of the design space. These plots display the quantiles of the SPV including the minimum, and the maximum values.

Fraction of Design Space (FDS) plots were introduced by Zahran, Anderson-Cook and Myers<sup>11</sup> to complement the VDGs. They considered spherical and cuboidal design spaces, and show the fraction of the design space at or below a given SPV value. One key advantage of the FDS plot over the VDG is that the VDG hides the fact that different volumes are associated with each radii or shrinkage factor. FDS plots can be useful when comparing two or more designs, as they allow the researcher to compare the SPV distributions of designs with a single curve as well as their G-efficiency and V-average values. The FDS plot provides the experimenter information about the distribution of the SPV throughout the design space, including the minimum and the maximum SPVs. Similar to the VDG, the researcher can determine the approximate 50% and 100% G-efficiencies for a design by looking at an FDS plot. The idea is that the larger the fraction of the design space close to the minimum SPV value, the better the design will be. Moreover, the flatter the line, the more stable the SPV distribution for that design will be. The FDS plots should be considered a complement to the VDGs when it is possible to create both plots, as the two plots provide different views of the SPV information. The FDS plot helps summarize the range of SPV values, the relative proportions of values throughout the design space and make easy comparisons between designs with a single curve. On the other hand, the VDGs give more detailed information about where different SPV values occur in the design space.

Figure 1 shows an FDS plot for a Central Composite Design<sup>12</sup> with two center runs for an experiment involving three factors in a cuboidal region. It assumes that the model of interest is a second order model of the form

$$y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i=1}^k \beta_{ii} x_i^2 + \sum_{i < j} \sum \beta_{ij} x_i x_j + \varepsilon. \quad (4)$$

From the plot, we can see the minimum and maximum SPV values of 3.35 and 12.73. For any chosen SPV value, an approximate fraction of the design space at or below this value can be estimated. Hence, the point at approximately (0.48, 5) can be interpreted as 48% of the total design space has a SPV value at or below 5. Note that a flatter curve implies the maximum and minimum SPV are closer together giving a more stable

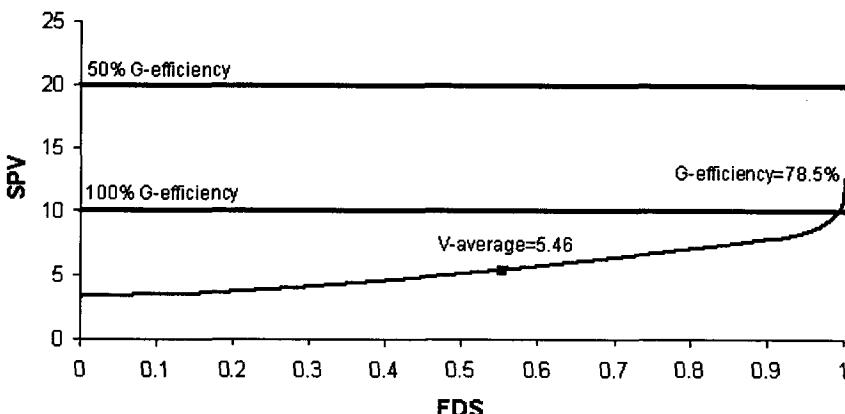


Fig. 1: Fraction of design space plot for a central composite design with 2 center runs on a cuboidal region.

distribution of the SPV. The number of parameters (G-optimal bound) of a second order model for  $k=3$ , is 10. Hence, the 100% G-efficiency = 10, and 50% G-efficiency = 20. From the above figure, the approximate G-efficiency is 78.5% ( $10/12.73$ ). The approximate V-average for this design is 5.46, which estimates the average of the SPV values throughout the entire cuboidal region. The corresponding FDS value for this point at approximately 0.56 indicates that the mean and median of the SPV distribution are not the same. This indicates that the distribution is positively skewed with a small number of large values influencing the average substantially.

The remainder of the chapter examines how to make informed comparisons between competing designs using the FDS plot for a variety of situations. Section 2 presents more details on the construction of the FDS plot and how this graphical method can be used for comparisons of precision and stability of prediction for different designs. Section 3 presents an example of selecting between several competitors to choose a best design when the design space is irregularly shaped. Section 4 gives an example of an experiment with two groups of factors, and how the FDS plot can be adapted to give more detailed information about the SPV distribution within sub-regions of the design. Section 5 considers model uncertainty and how the FDS plots can provide information about

assessing design performance across a variety of models. Section 6 discusses how the role of SPV changes and a second summary, the Penalized Prediction Variance (PPV), is required when a generalized linear model is considered, and how the FDS plots can be adapted to consider this situation as well.

## 2. Details about the Fraction of the Design Space Plot

The Fraction of Design Space plot can be constructed in a number of different ways. Originally, the SPV values for the plots were calculated analytically using the software package Mathematica. While feasible for up to second order models in lower dimensions of up to 5 factors, in higher dimensions this approach became computationally too cumbersome and slow. Hence, other possibilities were considered. Goldfarb et al.<sup>13</sup> looked at sampling on the perimeter of various shrinkage levels proportional to the relative volume associated with that portion of the total design space, as well as uniform sampling throughout the entire design space. Both of these approaches have been shown to give comparable results for a wide variety of applications at a fraction of the computational cost. For lower dimensional spaces, with 4 or fewer factors, an accurate FDS plot can be constructed with 2000-5000 points. For experiments involving 5 to 8 factors, simulations have shown that 10000 points is adequate. To get more precise estimates of the maximum value, and hence the G-efficiency, in the design space, the uniform sampling can be supplemented with some of the corners of the design space.

Once the sampled locations with their corresponding SPV values have been obtained, either from the various shrinkage levels or from the uniform sampling scheme, the FDS plot can be constructed by ordering the  $n$  SPV values from throughout the design space and plotting them against the quantiles ( $1/n, 2/n, \dots$ ). The process used here is similar to constructing an empirical cumulative distribution plot.

Using the FDS plots to make comparisons between designs can be quite straightforward. Figure 2 shows an FDS plot for three different designs all involving 3 factors on a cuboidal region with an assumed second order model. The three designs considered are the CCD presented

above, the Box and Draper<sup>14</sup> and Hoke D6<sup>15</sup> designs. The Box and Draper (B&D) design is a minimal point design for a cuboidal region, which consists of an augmentation to a subset of a  $2^k$  factorial. The Hoke design is one of a group of 7 designs available for 3 or more factors and is again intended for a second order model in a cuboidal region. The D6 design is one of the larger of the Hoke designs and these designs may be appropriate if there is restriction preventing one corner from being considered. The FDS plot in Figure 2 consists of 3 curves, one for each of the three designs under consideration, as well as the 100% G-efficiency line, which coincided with an SPV value of 10 for the three factor second order model with 10 terms (1 intercept, 3 linear terms, 3 two-way interactions, and 3 pure quadratic terms). Since all of the designs have maximum values less than the 50% G-efficiency line, this line was omitted. If we wish to decide on the best design based on G-efficiency, the CCD design is best with a maximum SPV value of 12.73 (78.5% G-efficient) compared to 15.36 for the Hoke D6 and 18.41 for the B&D designs.

When considering V-efficiency, again the CCD design is best with the smallest average SPV throughout the cuboidal region. In this

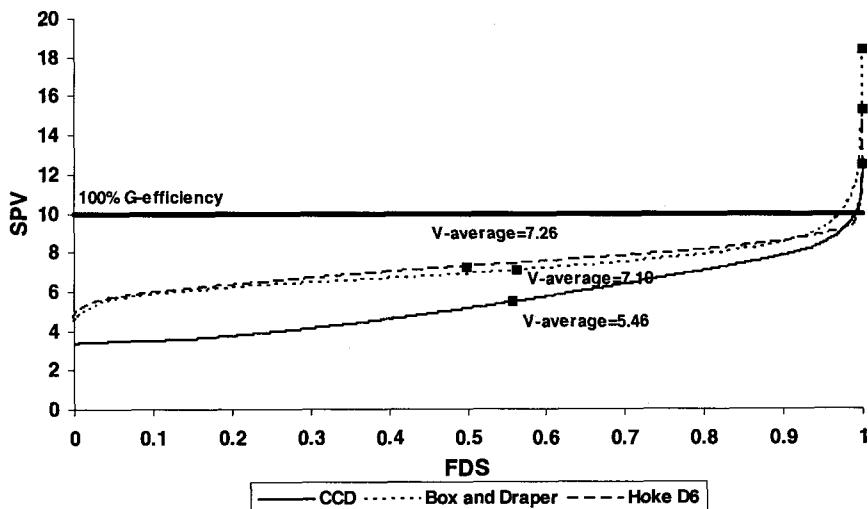


Fig. 2: Fraction of design space plot for 3 designs for second order models in a cuboidal designs space.

comparison the B&D design is slightly better than the Hoke D6 design, because a large portion of the design space has a slightly lower SPV, as can be seen by the B&D FDS curve being slightly below the Hoke D6 curve for the smallest 90% of the design space. If we wished to compare the median SPV values for the three designs, this is easily possible by looking at the 0.5 values on the x-axis, which corresponds to half of the total design space.

From the location of the V-average values, it is clear that both the CCD and B&D designs have a small fraction of very large SPV values which skew the distribution of the SPV values and shift the mean away from the median. The Hoke D6 has a much more symmetric distribution. In making an overall assessment of the three designs, the FDS plot clarifies that not only is the CCD superior in both the G- and V-sense, but it also has the largest fraction of the design at or below a given SPV value for all values less than or equal to approximately 9, which corresponds to where the Hoke D6 line crosses the CCD line. Said another way, the CCD design has the smallest SPV value associated with any percentile less than 96% of the design space.

Occasionally, the cost of additional experimental runs is not a consideration. In these cases, another adaptation of the FDS plot allows for the prediction variance itself,  $x_o'(X'X)^{-1}x_o$ , to be plotted directly with no penalty incorporated for larger designs. This approach may also be helpful for studying how to augment designs, where viewing the actual improvement to the prediction variance with each added run can be seen directly, by looking at the prediction variance directly.

Hence, the FDS plot allows for richer comparisons based on prediction variance between designs than just the G- and V-based criteria, but also allows good graphical visualization of these characteristics.

### 3. The Use of Fraction of the Design Space Plots for Oddly-Shaped Regions

In this section we explore how to adapt the FDS plot for unusually shaped design spaces. Frequently some economic, practical, or physical constraints may exist on the factor settings resulting in an irregularly

shaped experimental region. One often encounters situations in which it is necessary to eliminate some portion of the design space where it is infeasible or impractical to collect experimental data. Hence, standard designs are not always feasible and the need arises for finding best designs under these restrictions. Variance Dispersion Graphs have been adapted to cope with this situation through the use of shrinkage factors, which maintain the shape of the outer region of the space, and shrink it down proportionately until it reduces just to the centroid of the design space. This approach can be effective, but is computationally challenging in some cases, and the interpretation of the results may be confusing if the centroid of the design space is not immediately clear.

Consider a simple example involving two factors, where we wish to fit a first order model with interactions. We wish to study the two factors in a square region, but in this case, we know that there are restrictions on collecting data for one of the corners of the design space. For example, in a drug interaction study, it may be not practical to simultaneously set the two factors at high (or low) levels, because it is known a priori that this combination has an undesirable effect. This example was first presented in Zahran, Anderson-Cook, and Myers and Smith<sup>16</sup>, where the authors found the G-optimal design. Figure 3 shows the design space, which can be adapted with various choices of  $r$ , to influence how much of the region is truncated.

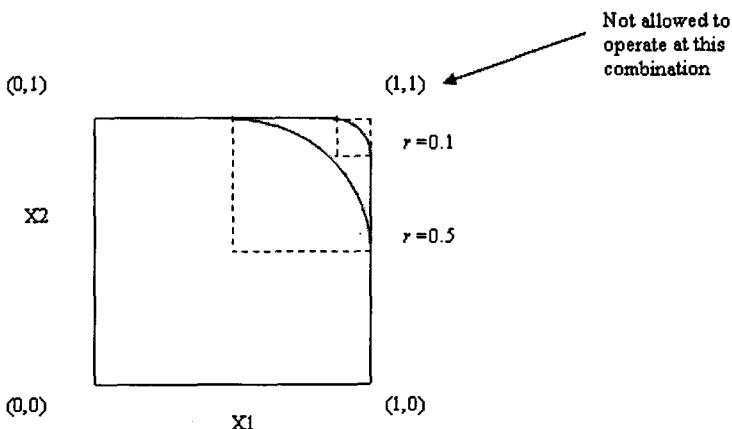


Fig. 3: Restricted design space for 2-factor experiment.

FDS plots can easily be adapted for unusually shaped design regions, by simply changing the sampling of the design space. In this case to construct the FDS plot, uniform sampling in the square region was performed for each factor between 0 and 1, and then the selected design locations sampled are compared to the condition to see whether they fall in the desired region. For the selected point, the SPV values are then calculated and the FDS plot can be constructed. This mechanism is easy to implement, does not increase the computational time for calculating the plots substantially, and provides great flexibility.

We consider the FDS plots for 4 possible designs for the case when  $r = 0.5$ , namely when the top right quarter square of the design space has been replaced by a quarter circle. The first design, labeled D1 in Figure 4, consists of the standard 4 observation factorial design with the one inadmissible value in the high-high corner shifted onto the outer boundary of the region of operability at (0.853, 0.853). It is the D- and G-optimal design for this region. Designs D2a and D2b are 5-point designs, with the points (0, 0), (0, 1) and (1, 0) supplemented by two points on the quarter circle. Design D2a has the two points at (1, 0.5) and (0.5, 1), while D2b has both points added in the same location as D1 at

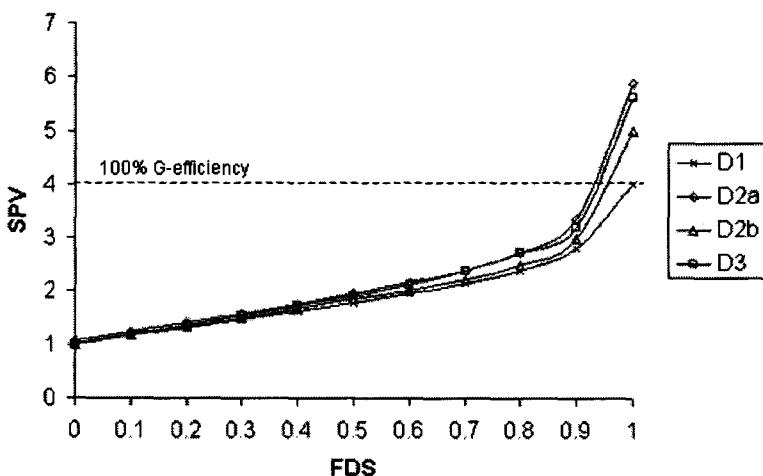


Fig. 4: Fraction of design space plots for a 2-factor experiment comparing 4 possible designs.

(0.853, 0.853). Finally D3 consists of the orthogonal factorial design, where the design points have been chosen to fit the design into the required design space. It has design points (0, 0), (0, 0.835), (0.835, 0) and (0.853, 0.853). For more details on the designs see Zahran and Anderson-Cook<sup>16</sup>.

In this case, the design space is no longer 1 square unit, but has been reduced to approximately 0.946 square units, since we replaced the square with a quarter circle. In the uniform sampling scheme used, 10000 points were selected at random, and 9540 locations satisfied the constraint and were used for the construction of the FDS plot for each of the designs considered.

From the plot, we see that D1 is the G-optimal design since the maximum SPV value is 4, which corresponds to the number of parameters in the first order model with interactions for the 2 factor case. Since the D1 design FDS curve is below the others throughout the entire range of the design space, it is the best design in terms of V-efficiency and best prediction by any criterion. Of the remaining designs, the next best design is D2b, which corresponds to the same design as D1, but with an additional observation at (0.853, 0.853). Designs D3 and D2a perform quite similarly with an intersection of the lines around (0.8, 2.7).

It would also be straightforward to use the FDS plot to determine the performance of various designs where the region of operability (where design points can be collected) and the region of interest (where the experimenter wishes to make prediction) are different. In this case, a design would be chosen within the region of operability, and then the region of interest would define where the sampling for the FDS plot was taken. This flexibility makes the FDS plots an adaptable tool for a variety of design comparisons.

#### **4. FDS Plots for Factors in Two-Groups**

In this section, we consider experiments with explanatory factors that can be grouped into two distinct categories. Examples of this situation include in robust parameter design studies where factors can be categorized as control and noise variables, in mixture-process experiments where some of the factors are components in a mixture and

others are process factors, and in split-plot designs where some variables are at the whole plot level and others are at the sub-plot level. In these cases, exploring the design space separately based on these categories can provide interesting insights into the relative performance of the designs.

For example, consider the mixture-process example presented in Goldfarb et al.<sup>13</sup> for the 3-mixture, 3-process factor experiment assuming a full second order model. The two designs considered, KCV31 with 31 runs, and KCV37 with 37 runs, were first presented by Kowalski, Cornell, and Vining<sup>17</sup>. From the FDS plot in Figure 5, it appears that the two designs are quite similar in performance with similar curves and G- and V-efficiencies (since the two designs have similar maxima and averages). Despite their similar minimum SPV values, there are some differences in the proportion of the design space at small SPV values.

Given the natural division of the factors into the two distinct groups, it may be natural to want to understand the relative performance of the SPV distribution conditional on where in the design space we are located. In the coded variables commonly used for modeling, the design space for this experiment consists of an equilateral triangle for the 3 mixture components crossed with a symmetric cube for the 3 process

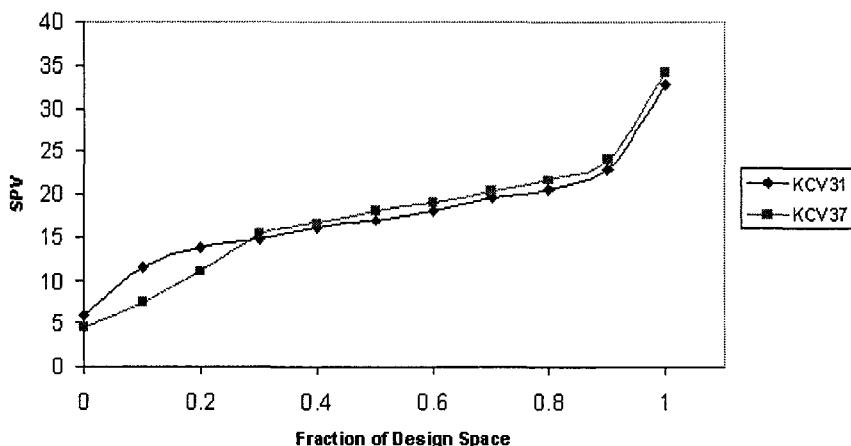


Fig. 5: Global fraction of design space plots for a 3-mixture, 3-process variable experiment.

variables. So an adaptation of the FDS plots allows separate curves to be drawn for different locations in the design space.

Figure 6 shows the modified FDS plot for the KCV31 design, which still has the FDS plot for the entire design space in bold and black, but now it has been supplemented with additional information about the SPV values. The three additional gray lines represent conditional FDS curves for sub-regions of the design space. The 0-line gives the SPV values for the 0-shrinkage value of the process variables, which corresponds to the centroid of the process variable cube crossed with the entire mixture triangle. The 1-line shows the SPV values for the outside edge of the process space crossed with the mixture triangle. Finally the 0.5-line gives the range of SPV values for a centered cube that has dimensions in each variable that are half as wide as the entire process range crossed with the mixture space.

Therefore from Figure 6, we can see that range of SPV values in the entire design space range from approximately 5 to 33. In addition, the worst prediction of the mixture space occurs on the outside edge of the process space and ranges from values of 11 to 33. At the center of the process space, the range of SPV values throughout the mixture pace range from approximately 8 to 28, while the best prediction of the

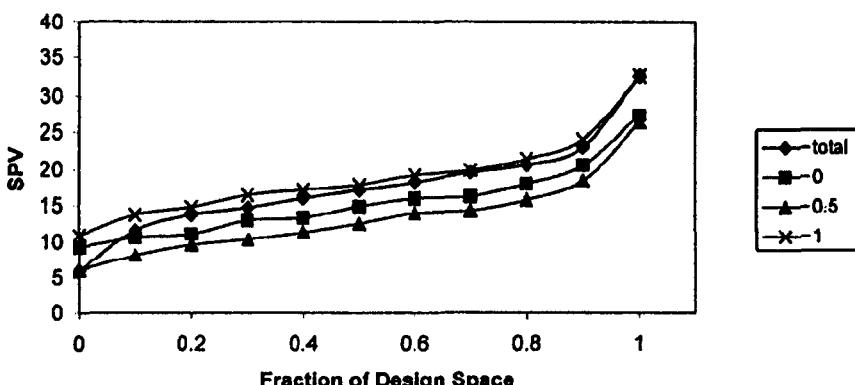


Fig. 6: Detailed fraction of design space plots for the KCV31 design with slices of design space for process shrinkages of 0, 0.5 and 1. FDS curves for slices represent sampling over the entire mixture space at the specified process shrinkage value.

mixture space occurs on a cube half the size of the total process space, with SPV values ranging from 5 to 26.

Figure 7 shows the KCV37 design with the supplemental FDS lines. In this plot we see that the best prediction occurs at the center of the process variable space, with SPV values ranging from 5 to 10. Similar values are observed for half way out in the process space with the 0.5-line giving SPV values in the 5 to 12 range. Finally on the outside edges of the process space, the prediction variance has considerably higher values ranging from approximately 12 to 35. On closer examination of the global FDS curve, we see that the small SPV values observed on the inner portion of the process space, with shrinkages of 0.5 and less, represent a relatively small portion of the total design space. As with the global FDS curve, less than 25% of the total design space has SPV values less than or equal to 12.

When comparing the two designs, we can now see important differences in where they predict well. The KCV31 design in Figure 6 has supplemental FDS plot lines close to each other, indicating that which location in the process space we are considering makes relatively little difference to how well we will predict. However, the range of SPV values covered by each individual conditional line is quite large, indicating that where we are in the mixture space is important for influencing prediction variance. In contrast to that, the KCV37 design in Figure 7 has relatively horizontal lines, indicating that where in the mixture space we wish to predict does not have a large influence on SPV values. However, the large spaces between the conditional curves indicate that where in the process space we wish to predict is an important factor affecting prediction variance. Hence, although the global FDS curves of the designs were quite similar, the two designs target good prediction in quite different ways and an experimenter may find this additional information helpful in making an appropriate choice between the designs.

In constructing the supplemental FDS lines for the two categories of the experimental factors, there is flexibility on how to allocate the roles of the categories to the additional curves. For the example above, it would also have been possible to condition on the shrinkage value in the mixture space, and then sample SPV values throughout the entire process

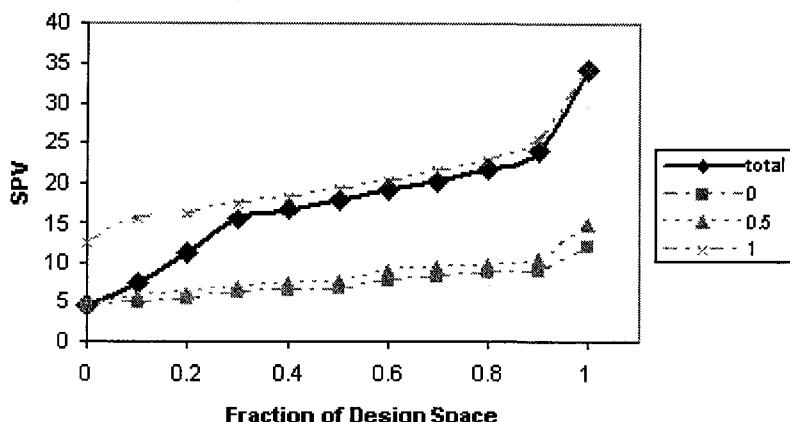


Fig. 7: Detailed fraction of design space plots for the KCV37 design with slices of design space for process shrinkages of 0, 0.5 and 1. FDS curves for slices represent sampling over the entire mixture space at the specified process shrinkage value.

design space cube for each value of the shrinkage for the mixture space. While the choice is subjective, it does seem that some allocations may be more natural than others, based on the interpretation of the different components or the shape of the corresponding portion of the design space.

For example, in the robust parameter design case with control and noise variables as the two categories of factors, it would likely be more desirable to condition on the control variable location, and examine the range of SPV values across the noise space. This may be more sensible, since in production situations the control variables are typically set at a particular value, while the noise variables are allowed to vary across their entire range. For the split plot case, with whole-plot and sub-plot factors, conditioning on the whole-plot location for each FDS line may be more intuitive for interpretation of results.

## 5. Examining Model Robustness with FDS Plots

So far in this chapter, we have assumed that the experimenter knows both the design space of interest as well as the model that should be fit to the data. While the boundaries of the design space may initially take

some expertise to ascertain, once determined they are reasonably assumed to stay fixed throughout the analysis of the experiment. However, this may not be the case for the model selected. A common scenario for design of experiments and response surface methodology is to assume the most complicated possible model for the underlying surface to be modeled, collect data from a design suitable for this maximal model, and then during the analysis phase consider whether the assumed complexity is actually required.

For example, it is quite common in response surface applications to assume a second order model will be adequate to describe the underlying relationship between the design factors and the response. Once, the data has been collected, testing of various terms in the model can be performed to determine if some of the terms are not statistically significant at the desired level, and then superfluous terms can be removed. The reduced model may only have a fraction of the quadratic terms, it may be a first order model with some of the interaction terms included, or it may be that one of the factors does not influence the response at all. It could also be such that the principle of hierarchy may not be appropriate. With this new chosen model, the design may no longer have its desired properties of the prediction variance. The distribution of the SPV may change drastically depending on which terms are excluded from the model.

The VDGs could be used to make comparisons between different assumed models, but since for each assumed model and design combination, three curves are produced, it would be difficult to determine which design performs best. By making adaptations to the FDS plots, it is possible for the experimenter to take into account design performance for a variety of nested models within the maximal model for which the design was selected. The single curve of the FDS for each design and model combination facilitates easier comparisons. These new FDS plots allow the experimenter to study a design's prediction characteristics for any reduced model. The robustness of the design to model changes will be determined by examining the behavior of the various curves in the FDS plot.

Consider an experiment to investigate the relationship between 5 factors on a response in a coded hypercube. The experimenter thinks that

a second order model will be adequate to model the underlying relationship. Possible designs to consider include the CCD<sup>12</sup> with 4 center runs ( $N=30$ ) and two of the Hoke<sup>15</sup> designs, D6 and D7 ( $N=26$  for both). In Figure 8, which shows the FDS plots for these three designs, we can see that the designs are quite similar in performance throughout the design space with the CCD having the smallest SPV values for about half of the total design space, and the Hoke D6 design being the most G-efficient (with maximum SPV value of 29.30 compared to 29.32 for the Hoke D7 and 31.13 for the CCD). For the five factor second order model, there are a total of 21 terms (1 intercept, 5 linear, 10 2-way interaction and 5 pure quadratic terms). Hence the G-efficiencies of the three designs are approximately 72% for the Hoke D6 and D7 designs, and 67% for the CCD. Since the Hoke D6 and D7 designs are quite similar in performance, in subsequent plots we will only present the curves for the Hoke D6 and CCD designs.

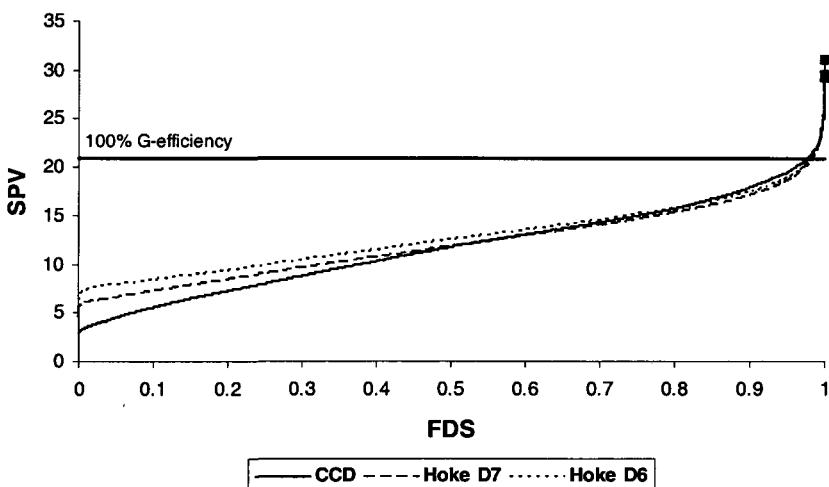


Fig. 8: FDS plots for 5-factor experiment for second order model on a hypercube.

We now consider adapting the FDS plot to examine model robustness, since the three designs above appear to be relatively similar in performance based on the assumed second order model. Hence, we consider how the SPV values for each design would change if some of

the terms of the model were not significant, and hence the experimenter wanted to eliminate some of them from the final model. With 21 possible terms in the model, there are clearly a very large number of possible nested models to consider (if we only assume that the intercept should always be included and do not require hierarchical assumptions, there are  $2^{20}=1048576$  possible models). Each of which could potentially have its own FDS line on the plot.

To help understand how we should best choose from among the vast models, some characteristics of the SPV will be helpful. Recall that adding additional terms to the model can only increase the SPV values of equation (1). This is one of the central aspects to the bias-variance trade-off that is commonly presented in model selection. Smaller models have less prediction variance, but if they do not adequately describe the underlying relationship, they may suffer from bias problems. Hence, for any of the smaller nested models, they will have SPV values less than or equal to those of the second order model. In addition, the G-optimal values for each of the reduced models will also change as the number of parameters in the model is altered.

For some applications there may be a natural subset of models nested within the original model to be considered for model reduction. If this were the case then it would be sensible to create an FDS curve for each of the models of potential interest on a single curve. This will allow for easy comparison between possible models for a given design, and also for comparisons between competing designs for a given model.

Frequently, the experimenter does not know *a priori* which terms in the model may turn out to be insignificant. In these cases, since all possible models cannot be considered for practical reasons, some systematic strategy for examining important subsets of models may be needed. One way of obtaining some understanding of the behavior of some of the designs for some of the reduced models is to consider only strategic sub-models.

One option would be to consider reducing the number of factors that are important. This would correspond to one or more of the factors in the experiment having no influence on the response either through linear, quadratic or interaction terms. Figure 9 shows how both the CCD and the Hoke D6 designs perform when we consider the designs with full second

order models with 5, 4 and 3 factors respectively. From this plot, we can see that the pattern observed for the full model in 5 factors remains unchanged for smaller numbers of factors. In each case the CCD has better SPV properties for much of the design space, but the Hoke D6 design has slightly better G-efficiency.

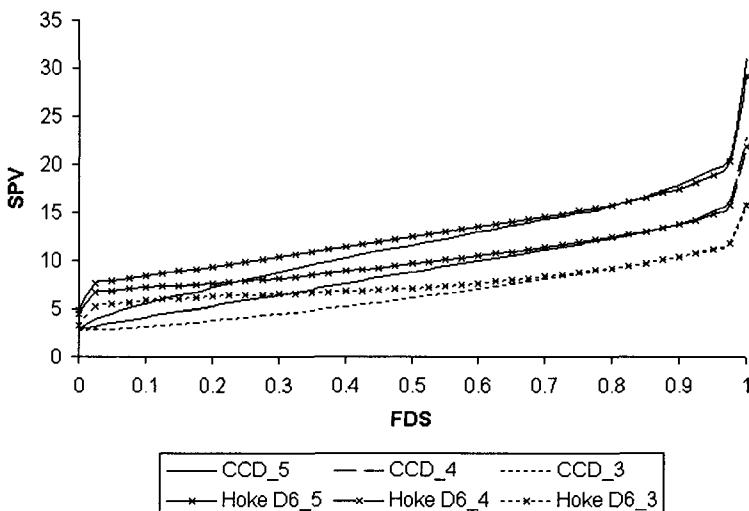


Fig. 9: Model robustness FDS plot for CCD and Hoke D6 designs considering second order models with 5, 4 and 3 factors.

Next we examine reducing the order of the model. If we look only at models that are hierarchical and symmetric in all of the factors, this leads to 3 curves: one for the second order model with 21 terms, one for the first order model with all two-way interactions with 16 terms, and one for the first order model with 6 terms. These are frequently considered classes of models in Response Surface Methodology, and are intuitively pleasing summaries of the level of complexity needed to summarize a particular relationship between factors and response. Figure 10 shows the three FDS curves for each of the CCD and the Hoke D6 designs. Since many of the other non-symmetric models involving the 5 factors are likely to contain a number of parameters that somewhere between the simplest first order model, and the most complicated second order model,

we can treat the three curves as forming an “envelope” around the range of FDS curves likely for a broad range of other models. Ozol-Godfrey, Anderson-Cook and Montgomery<sup>18</sup> give some examples where a variety of reduced models are considered as well as their relationship to the symmetric hierarchical models.

Figure 10 shows that although the CCD design has some advantage in lower SPV values for the second order model, the Hoke D6 design performs better with lower FDS curves for both the first order with interaction model and the first order model. Hence, with the information in Figures 9 and 10 we are able to make a much more detailed comparison between the designs based on how likely the experimenter feels that model reduction might be, and what type of reduction might be expected.

Since the G-optimal values for each of the models is different, it is now difficult to estimate the G-efficiency of each design for the various models. In addition as the number of factors in the experiment increases,

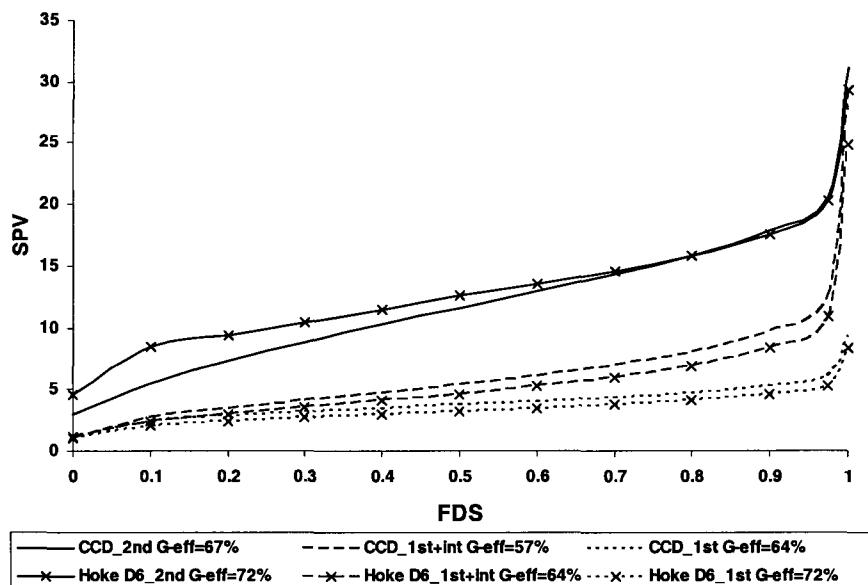


Fig. 10: Model robustness FDS plot for CCD and Hoke D6 designs considering second order, first order with interaction and first order models.

the FDS plots for various nested designs become increasingly separated as the G-optimal values for the different design differ more and more. Hence, one final modification to the FDS plots is considered to allow clearer presentation of various nested models. Figure 11 shows a G-scaled FDS plot, where each of the SPV values from the curves in Figure 10 has been divided by the appropriate G-optimal value. Hence in this case, the second order model SPV values would be divided by 21, the first order with interaction terms divided by 16 and the first order model terms by 6. This reduces the problem of the wide separation between curves as the number of factors increases and allows for G-efficiency comparisons to be made more easily. From this plot, we can make better comparisons between the relatively efficiencies of the designs for the various nested models. If we are interested in model robustness, the G-scaled FDS plot allows us to look directly at how closely various FDS curves for a given design remain across various possible models, giving us a direct assessment of robustness.

From the plot, it now becomes clear that the first order models with interactions have low SPV values relative to the G-optimum for small

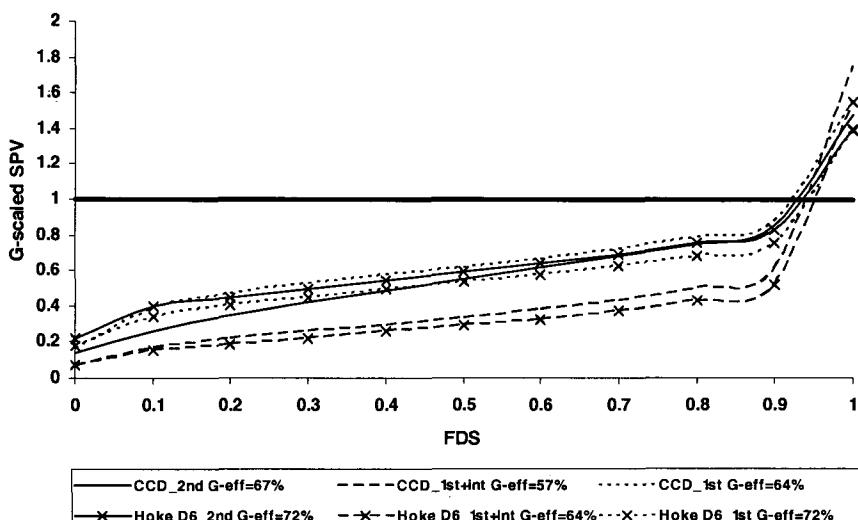


Fig. 11: G-scaled model robustness FDS plot for CCD and Hoke D6 designs considering second order, first order with interaction and first order models.

SPV values, but their maximum SPV values are relatively larger lowering their overall G-efficiencies to 57% and 64% for the CCD and Hoke D6, respectively. Not surprisingly, the designs both perform best for the second order model, for which they were specifically designed.

Therefore, the adaptation of the FDS plots allows the practitioner to study how various designs will perform under a number of different types of model reduction. It can be reassuring that the largest model, for which the designs are typically selected, represent the worse-case scenario for SPV values. While it may be unknown ahead of time what type of reduction may be needed, by selecting an appropriate set of nested models within the largest model considered, the plots provide useful information to help with the selection of designs that are relatively robust to changes in the form of the model.

## 6. Using FDS Plots with Generalized Linear Models

Up to this point in the chapter, we have considered evaluating designs for standard linear models of the form,  $y = X\beta + \varepsilon$ . In this section, we explore what adaptations are needed to study prediction variance properties in the generalized linear model (GLM) case, which are a general case of the linear models where the error term belongs to an exponential family type distribution, and is typically not constant throughout the design space. These models were introduced by Nelder and Wedderburn<sup>19</sup>. The observations,  $y_1, \dots, y_n$ , are independently distributed each with mean  $E(y_i) = \mu_i$  respectively. Each  $y_i$  has a distribution that is a member of the exponential family, and  $\eta_i = \mathbf{x}_i'\boldsymbol{\beta}$  is the linear predictor of some function of  $E(y_i) = \mu_i$ . It involves the regressors,  $x_1, \dots, x_k$ . The link function,  $g$ , is used to define the model and is given by  $\eta_i = g(\mu_i)$  for  $i = 1, \dots, n$ . For more details and examples of GLMs, see Myers, Montgomery and Vining<sup>20</sup> or Dobson<sup>21</sup>. Khuri<sup>22</sup> presents some of the special issues of using GLMs for response surface applications.

Two common cases of GLMs are logistic regression, and Poisson regression. For both of these cases, the variances of the model parameter estimates are  $\text{var}(\boldsymbol{b}) = (X'VX)^{-1}$ , where  $V = \text{diag}\{\sigma_i^2\}$ , and  $\sigma_i^2$  is a function of  $\mu_i$ . For the logistic case,  $\sigma_i^2 = p_i(1 - p_i)/n_i$  where  $n_i$  is the

number of observations taken at  $\mathbf{x}_i$ , and  $p_i$  is the probability of success/failure. For the Poisson case,  $\sigma_i^2 = e^{\mathbf{x}_i \beta} = \mu_i$  where  $\mu_i$  is the expected number of counts at  $\mathbf{x}_i$ .

Design optimality for GLMs is different than for the standard linear models because in the GLM case, the information and hence the variance matrices are functions of the unknown parameters. The  $V$  matrix above is a function of the unknown parameters that are typically estimated only after data has been collected. Hence design optimality and criteria for good designs become more complex.

When we consider how to study prediction variance for the GLM case, two measures of the quality of prediction emerge. First, we consider the SPV and how to adapt it for generalized linear models. The SPV for the linear case is given by equation (1). Here the variance of the estimated response at  $\mathbf{x}_o$ ,  $\text{var}(\hat{y}(\mathbf{x}_o))$ , is defined as  $\text{var}(\hat{y}(\mathbf{x}_o)) = \sigma^2 \mathbf{x}'_o (\mathbf{X}' \mathbf{X})^{-1} \mathbf{x}_o$  where  $\sigma^2$  is the variance of the observation at  $\mathbf{x}_o$ . Since the variance,  $\sigma^2$ , is constant for any observation, it can be divided out of the equation leaving only the  $\mathbf{x}'_o (\mathbf{X}' \mathbf{X})^{-1} \mathbf{x}_o$  term. In this case, the shape of the prediction is not influenced by removing the  $\sigma^2$  term, and means that the quantity can be evaluated before any data is collected. The SPV is also multiplied by  $N$ , the number of runs, to penalize larger designs, and give information on a per observation basis.

When the variance for different observations throughout the design space is not a constant, the form of the SPV becomes different from the linear case. For the GLM case, the prediction variance of the estimated response at location  $\mathbf{x}_o$  is defined as

$$\text{var}(\hat{y}(\mathbf{x}_o)) = \text{var}(y(\mathbf{x}_o))^2 \mathbf{x}'_o (\mathbf{X}' \mathbf{V} \mathbf{X})^{-1} \mathbf{x}_o, \quad (4)$$

where  $\text{var}(y(\mathbf{x}_o))$  corresponds to the variance of the observation at  $\mathbf{x}_o$ . When the variance of the estimated mean response is scaled by the specific variance of any observation and multiplied by  $N$ , the SPV for a GLM becomes

$$\nu(\mathbf{x}_o) = N \text{var}(y(\mathbf{x}_o)) \mathbf{x}'_o (\mathbf{X}' \mathbf{V} \mathbf{X})^{-1} \mathbf{x}_o \quad (5)$$

where  $\nu(\mathbf{x}_o)$  is a function of the number of runs, the location, the design and also now the parameters in the model. Note that this scaling now involves dividing by different values for each location in the design

space. For the GLM case, the G- and the V-optimal values are still in the correct scale as the SPV with the G-optimal value being  $p$ , the number of parameters in the model. This can be justified by using a transformation of the response variable to return to the standard linear models case.

The penalized prediction variance (PPV) is considered next for the case of a GLM. The PPV for the standard linear model situation is defined as

$$\rho(\mathbf{x}_o) = N \operatorname{var}(\hat{y}(\mathbf{x}_o)) = N\sigma^2 \mathbf{x}'_o (\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}_o \quad (6)$$

which is nothing more than the prediction variance of an estimated mean response,  $\operatorname{var}(\hat{y}(\mathbf{x}_o))$ , multiplied by  $N$ , the number of runs in the design. The distinction between a PPV and a SPV for a linear model is that the PPV is *not* scaled by dividing by  $\sigma^2$ . For the linear model case, the PPV is just a fixed multiple of the SPV, so the shape and stability of both of these for a given design will be similarly conveyed by either quantity.

The PPV for the GLM case is very important because the researcher is able to study the quality of the actual prediction throughout the design space with initial parameter estimates before the data is collected. The PPV for GLM is defined as

$$\rho(\mathbf{x}_o) = N \operatorname{var}(\hat{y}(\mathbf{x}_o)) = N \operatorname{var}(y(\mathbf{x}_o))^2 \mathbf{x}'_o (\mathbf{X}'\mathbf{V}\mathbf{X})^{-1} \mathbf{x}_o, \quad (7)$$

where both  $\mathbf{V}$  and  $\operatorname{var}(y(\mathbf{x}_o))$  are dependent on the initial parameter estimates. Hence for the GLM case, there are now two quantities of interest for summarizing a design's prediction capability: the SPV, which has a known theoretical bound for G-optimality, and the PPV, which is of more direct interest to the practitioner, since it models the actual prediction behavior in the design space.

To obtain the FDS plot for a design with a GLM, the researcher first needs to specify a model of interest, the design space, one or more designs, and the initial parameter estimates. Using this information the SPV and PPV values throughout the design space can be calculated to produce the FDS plots. For example, consider a manufacturing experiment to investigate the proportion of defects being produced for a variety of factor combinations. A logistic model will be appropriate for the binomial response that is obtained. The experimenters decide to focus

on two important factors  $x_1$  (temperature), and  $x_2$  (time), and their interaction. They consider some variation of a  $2^2$  factorial design in a square region coded so that each variable is in the range  $[-1, 1]$ . A first order model with interaction in the linear predictor with the logistic link is used to give the form of the model as

$$\ln\left(\frac{P_i}{1-P_i}\right) = \beta_o + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{12} x_{1i} x_{2i}, \quad (8)$$

which gives the form  $P_i = \frac{1}{1 + e^{-(\beta_o + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{12} x_{1i} x_{2i})}}$  for the probability of success at a given design location. Initial estimates of the probability of non-defects at each of the corners are obtained as (0.93, 0.60, 0.95, 0.88) for  $\{(-1, -1), (-1, 1), (1, -1), (1, 1)\}$  respectively. This corresponds to  $\hat{\beta}' = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_{12}) = (2, 0.5, -0.8, 0.3)$  as parameter estimates from equation (8).

With the available resources, two possible designs each with 400 observations are considered. The first design, D1, is an equal allocation design that assigns 100 observations to each of the four corners. The second design, D2, strives to improve estimation at the locations that are thought to have larger variances based on the estimated proportion of defects at each location. Since the variance at each design point is  $\sigma_i^2 = p_i(1-p_i)/n_i$  where  $n_i$  is the number of observations at that point, D2 selects sample sizes at each location to make  $\sigma_i^2$  as close to constant as possible. This yields an allocation of (57, 210, 41, 92) for the design locations  $\{(-1, -1), (-1, 1), (1, -1), (1, 1)\}$  respectively.

Figures 12 and 13 show the FDS plots for SPV and PPV, respectively, for the two designs. From the Figures we can see that the two designs compare quite differently depending on which of the criteria we wish to focus on. For the practitioner interested in optimizing prediction of the rate of defects throughout the region, D2 is the better choice with a more stable range of PPV values throughout the design space, with a lower maximum value. The flatness of the line reflects the stability of the design across the entire space. The equal allocation design, D1, has better prediction in the majority of the design space, but

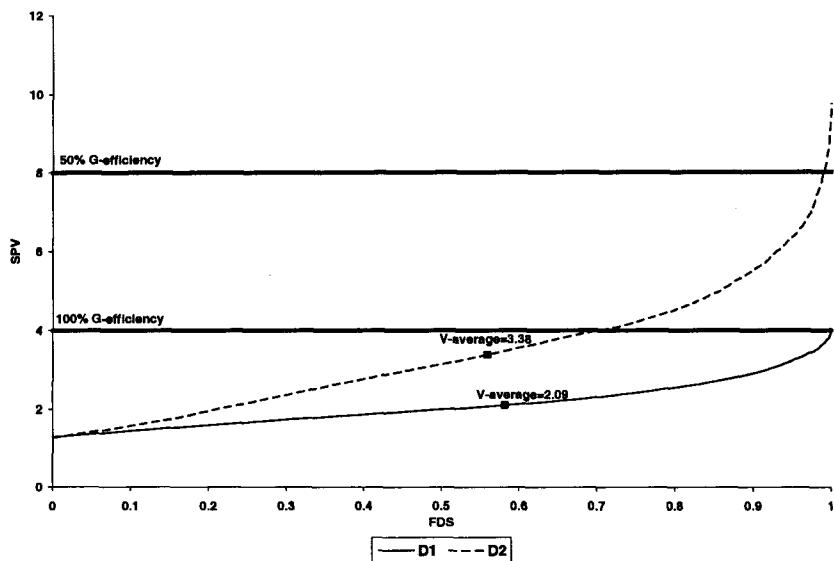


Fig. 12: FDS plot for SPV for comparing two designs for a logistic model involving two-factors on a square design space.

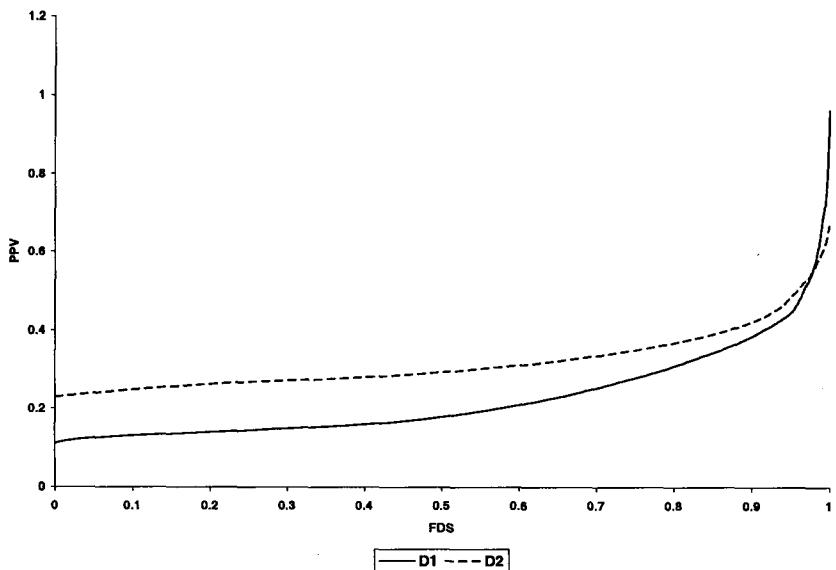


Fig. 13: FDS plot for PPV for comparing two designs for a logistic model involving two-factors on a square design space.

then has a small region of quite poor prediction at the corner of the design space with probability of non-defects close to 0.60. However, from a theoretical standpoint, D1 is a G-optimal design with minimum possible maximum SPV value. Clearly in the GLM case, the distinction between theoretical and practical assessment of the design is more complicated than in the standard linear model case.

An additional consideration in assessing the differences between the two designs in this case, is to take into account the precision of the initial parameter estimates. Typically, those parameter estimates are merely educated guesses, which could vary widely from the actual parameter values. After all, it is one of the purposes of the experiment to estimate these parameters more precisely. Ozol-Godfrey, Anderson-Cook and Myers<sup>23</sup> consider a number of approaches for using FDS plots to assess design robustness to a number of types of parameter misspecification. The various approaches involve having several FDS curves for a variety of different parameter estimate choices, which may help quantify the uncertainty inherent in the single best guess of the parameters that is usually provided.

Design assessment for GLMs is still developing as various distinct aspects of this problem are being studied and better understood. The FDS plots can potentially be a helpful tool for aiding in design assessment and comparisons between competing designs.

## 7. Conclusions

In this chapter, a number of different FDS plots have been presented for a diverse collection of design situations. The goal of incorporating graphical summaries of prediction characteristics of designs should be to help make a more informed decision about the relative advantages and disadvantages of each design. These graphical methods help enrich the information available and allow exploration of a greater breadth of optimality criteria.

The FDS plots are relatively easy to create, and a wide variety of Visual Basic programs in Microsoft Excel are available by request from the authors. The output from the program can be exported into any software package that is compatible with Excel spreadsheets.

## Acknowledgments

The authors would like to express their thanks to Raymond H. Myers and Douglas C. Montgomery for their insights and contributions in helping with various aspects of the development of the FDS plots.

## References

1. R.H. Myers and D.C. Montgomery, *Response Surface Methodology: Process and Product Optimization using Designed Experiments* (Wiley, New York, 2002), p. 304.
2. J. Kiefer and J. Wolfowitz, *Canadian Journal of Mathematics* 363 (1959).
3. G.E.P. Box and J.S. Hunter, *Annals of Mathematical Statistics* 195-241 (1957).
4. R.H. Myers and D.C. Montgomery, in Ref. 15, p. 396.
5. G.E. P. Box and N.R. Draper, *Journal of the American Statistical Association* 622 (1959).
6. G.E. P. Box and N.R. Draper, *Biometrika* 335 (1963).
7. Giovannitti-Jensen and R.H. Myers, *Technometrics* 159 (1989).
8. G. Piepel, C.M. Anderson and P.E. Redgate, *Proceedings of the Section on Physical and Engineering Sciences* 205 (American Statistical Association, 1993).
9. A.I. Khuri, H.J. Kim and Y. Um Y, *Computational Statistics & Data Analysis* 395 (1996).
10. A.I. Khuri, J.M. Harrison and J.A. Cornell, *Applied Statistics* 521 (1999).
11. A. Zahran, C.M. Anderson-Cook and R.H. Myers, *Journal of Quality Technology* 377 (2003).
12. G.E.P. Box and K.B. Wilson, *Journal of the Royal Statistical Society B* 1 (1951).
13. H.B. Goldfarb, C.M. Anderson-Cook, C.M. Borror and D.C. Montgomery, *Journal of Quality Technology* 169 (2004).
14. G.E.P. Box and N.R. Draper, *Technometrics* 613 (1974).
15. A.T. Hoke, *Technometrics* 375 (1974).
16. A. Zahran, C.M. Anderson-Cook, R.H. Myers and E.P. Smith, *Journal of Quality Technology* 387 (2003).
17. S.M. Kowalski, J.A. Cornell and G.G. Vining, *Technometrics* 72 (2000).
18. A. Ozol-Godfrey, C.M. Anderson-Cook and D.C. Montgomery, *Journal of Quality Technology* 223 (2005).
19. J.A. Nelder and R.W.M. Wedderburn, *Journal of the Royal Statistical Society, Series A* 370 (1972).
20. R.H. Myers, D.C. Montgomery and G.G. Vining, *Generalized Linear Models with Applications in Engineering and the Sciences* (Wiley, New York, 2002).
21. A.J. Dobson, *An Introduction to Generalized Linear Models, 2nd Edition* (Chapman & Hall, Boco Raton, 2002).
22. A.I. Khuri, *Nonlinear Analysis* 2023 (2001).
23. A. Ozol-Godfrey, C.M. Anderson-Cook and R.H. Myers, Virginia Tech Department of Statistics Technical Report 04-3.

This page is intentionally left blank

## **CHAPTER 16**

### **CONCEPTS OF SLOPE-ROTatability FOR SECOND ORDER RESPONSE SURFACE DESIGNS**

Sung H. Park

*Department of Statistics, College of Natural Sciences  
Seoul National University, San 56-1, Shinrim-dong, Kwanak-ku  
Seoul, 151-747, Korea  
E-mail: parksh@plaza.snu.ac.kr*

This paper reviews the design concepts of slope-rotatability for second order response surfaces. Two basic concepts and two modified concepts of slope-rotatability are reviewed. The basic concepts are slope-rotatable designs over axial directions (Hader and Park<sup>1</sup>) and slope-rotatable designs over all directions (Park<sup>2</sup>). The modified concepts are slope-rotatable designs with equal maximum directional variance ((Park and Kwon<sup>3</sup>) and slope-rotatable minimax designs (Park and Kwon<sup>3</sup>). Also two measures to assess the slope-rotatability are reviewed. One measure is to assess the slope-rotatability over axial directions (Park and Kim<sup>4</sup>) and the other is to assess the slope-rotatability over all directions (Jang and Park<sup>5</sup>).

#### **1. Introduction**

The usual response surface model assumes that the dependent variable  $\eta$  is adequately approximated by a low order polynomial in  $k$  independent variables  $x_1, x_2, \dots, x_k$  which have been coded so that their origin is the center of some region of interest. We will concentrate on the second order polynomial model

$$\eta(\mathbf{x}) = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i=1}^k \beta_{ii} x^2 + \sum_{i < j} \beta_{ij} x_i x_j$$

which may be written in matrix notation as

$$\eta(\mathbf{x}) = \mathbf{x}_s' \boldsymbol{\beta}$$

in which  $\mathbf{x} = (x_1, x_2, \dots, x_k)'$  is the  $k \times 1$  vector of a design point,

$$\mathbf{x}_s = (\mathbf{1}, x_1, x_2, \dots, x_k, x_1^2, x_2^2, \dots, x_k^2, \dots, x_1 x_2, x_1 x_3, \dots, x_{k-1} x_k)'$$

is an  $m \times 1$  vector and  $\boldsymbol{\beta}$  is the  $m \times 1$  column vector of the corresponding regression coefficients, where  $m = (k+1)(k+2)/2$ . The coefficients in the polynomial are to be estimated, by the method of least squares, from observations on the response variable,

$$y_u = \eta(\mathbf{x}_u) + \varepsilon_u, \quad u = 1, 2, \dots, N$$

where the observations are taken at  $N$  selected combinations of the  $x$  variables. The  $\varepsilon_u$ 's are assumed to be uncorrelated random errors with zero means and constant variance,  $\sigma^2$ . The  $\beta$ 's are then estimated by the method of least squares as

$$\mathbf{b} = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y}, \quad (1)$$

in which  $\mathbf{X}$  is the  $N \times m$  matrix of values of the  $m$  elements of  $\mathbf{x}_s$  taken at the design points and  $\mathbf{y}$  is the  $N \times 1$  vector of  $y$  observations.

When the fitted equation

$$\hat{y}(\mathbf{x}) = \mathbf{x}_s' \mathbf{b}$$

is to be used to estimate  $\eta(\mathbf{x})$ , it is well known that the variance of  $\hat{y}(\mathbf{x})$  is

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

$Var[\hat{y}(\mathbf{x})]$  thus depends on the particular values of the independent variables through the vector  $\mathbf{x}'_s$ . It also depends on the design through the matrix  $(\mathbf{X}'\mathbf{X})^{-1}$ .

There are a number of desirable properties for response surface experimental designs to have. Among others, an interesting and important property is that of rotatability. A design is said to be rotatable if the variance of  $\hat{y}(\mathbf{x})$  is a function only of the distance  $\rho = (x_1^2 + x_2^2 + \dots + x_k^2)^{1/2}$  from the center of the design to the point  $\mathbf{x}$ . Thus, with a rotatable design, the prediction variance,  $Var[\hat{y}(\mathbf{x})]$ , is the same at all points  $\mathbf{x}$  that are equidistant from the design center. The concept of rotatability was first introduced by Box and Hunter<sup>6</sup>, and explained in detail in Myers<sup>7</sup>.

The necessary and sufficient conditions for a design to be rotatable are as follows.

[C1] All odd-order moments of order  $\leq 4$  must be zero.

[C2] [ii] are equal for all  $i$ , and [iiii] are equal for all  $i$ .

[C3] [iiii] = 3 [iiji] for all  $i \neq j$ . (2)

Here [ii], [iiii] and [iiji] denote the pure second order moments, pure fourth order moments and mixed fourth order moments, respectively. That is,

$$[ii] = \frac{1}{N} \sum_{u=1}^N x_{iu}^2, [iiii] = \frac{1}{N} \sum_{u=1}^N x_{iu}^4, [iiji] = \frac{1}{N} \sum_{u=1}^N x_{iu}^2 x_{ju}^2.$$

## 2. Slope-rotatability over axial directions

In the design of experiments for response surface analysis, it is also meaningful to focus on the estimation of differences in response rather than absolute value of the response variable  $\eta$ . Estimation of differences in response at different points in the factor space will often be of great importance. Herzberg<sup>8</sup>, Box and Draper<sup>9</sup>, and Huda and Mukerjee<sup>10</sup> deal with differences in response at different points. If differences at points

close together in the factor space are involved, estimation of the local slopes (the rates of change) of the response surface is of interest.

This problem, estimation of slopes, occurs frequently in practical situations. For instance, there are the cases in which one wants to estimate rates of reaction in chemical experiments, rates of change in the yield of a crop to various fertilizers, rates of disintegration of radioactive material in an animal, and so forth. Since the first work of Atkinson<sup>11</sup> in this slope area, many research papers have been subsequently published such as Ott and Mendenhall<sup>12</sup>, Murty and Studden<sup>13</sup>, Myers and Lahoda<sup>14</sup>, Hader and Park<sup>1</sup>, Mukerjee and Huda<sup>15</sup>, Park<sup>2</sup>, Park and Kim<sup>4</sup>, Huda and Al-Shiha<sup>16</sup>, Huda and Shafiq<sup>17</sup>, Draper and Ying<sup>18</sup>, Ying, Pukelsheim and Draper<sup>19,20</sup>, Kim, Um and Khuri<sup>21</sup>, Park and Kwon<sup>3</sup> and so on. Especially, as an analogue to variance dispersion graph in Giovannitti-Jesen and Myers<sup>22</sup>, Jang and Park<sup>5</sup> proposed the slope variance dispersion graph for estimating the slope of a response surface.

Suppose that the estimation of the first derivatives of  $\eta(\mathbf{x})$  is of interest. For the second order model the first derivative of  $\hat{y}(\mathbf{x})$  with respect to  $x_i$  is

$$s_i(\mathbf{x}) = \frac{\partial \hat{y}(\mathbf{x})}{\partial x_i} = b_i + 2b_{ii}x_i + \sum_{i < j}^k b_{ij}x_j. \quad (3)$$

And the variance of this derivative is a function of the point  $\mathbf{x}$  at which the derivative is estimated and also a function of the design through the relationship

$$Var(b) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}.$$

Hader and Park<sup>1</sup> first proposed an analog of the Box-Hunter rotatability criterion. This requires the following:

[C1] For each  $i = 1, 2, \dots, k$ , the variance of  $\hat{y}(\mathbf{x})/\partial x_i$  are equal for all  $\mathbf{x}$  that are equidistant from the design origin, that is,  $\hat{Var}[\partial \hat{y}(\mathbf{x})/\partial x_i]$  is a function only of

$$\rho = (x_1^2 + x_2^2 + \dots + x_k^2)^{1/2}$$

[C2] The variance of  $\partial\hat{y}(\mathbf{x})/\partial x_i$ , are equal, that is,

$$\text{Var}[\partial\hat{y}(\mathbf{x})/\partial x_1] = \text{Var}[\partial\hat{y}(\mathbf{x})/\partial x_2] = \dots = \text{Var}[\partial\hat{y}(\mathbf{x})/\partial x_k]$$

for any point  $\mathbf{x}$ .

If a design satisfies [C1] and [C2], estimation of the slopes over axial directions will be equally reliable for all points  $\mathbf{x}$  equidistant from the design origin. Hader and Park<sup>1</sup> referred to this property as slope-rotatability. They also presented slope-rotatable central composite designs. Park<sup>2</sup> extended the concept of slope-rotatability to ‘slope-rotatability over all directions’, and this slope-rotatability is referred as ‘slope-rotatability over axial directions’.

### 3. Slope-rotatability over all directions

Note that the equation (3) can be written in matrix notation as

$$\frac{\partial\hat{y}(\mathbf{x})}{\partial x_i} = d_i(\mathbf{x})\mathbf{b} \quad (4)$$

Here,  $\mathbf{b}$  is given in (1) and  $d_i(\mathbf{x})$  is  $1 \times p$  row vector which equals to  $\partial\mathbf{x}_s'/\partial x_i$ . For example, when  $k = 2$ ,  $d_1(\mathbf{x}) = (0, 1, 0, 2x_1, 0, x_2)$  and  $d_2(\mathbf{x}) = (0, 0, 1, 0, 2x_2, x_1)$ . Then the variance of the estimated derivative over the  $i^{th}$  axial direction is given by

$$\text{Var}\{s_i(\mathbf{x})\} = \sigma^2 d_i(\mathbf{x})(\mathbf{X}'\mathbf{X})^{-1} d_i'(\mathbf{x})$$

The analogue of the Box-Hunter rotatability criterion is a requirement that  $\text{Var}\{s_i(\mathbf{x})\}$  be constant on circles ( $k = 2$ ), spheres ( $k = 3$ ) or hyperspheres ( $k > 3$ ) centered at the design origin.

It is often of interest to estimate the slope of a response surface at a point  $\mathbf{x}$ , not only over axial directions, but also over any specified direction. Let the estimated slope vector be

$$\mathbf{s}(\mathbf{x}) = \begin{pmatrix} \partial \hat{y}(\mathbf{x}) / \partial x_1 \\ \partial \hat{y}(\mathbf{x}) / \partial x_2 \\ \vdots \\ \partial \hat{y}(\mathbf{x}) / \partial x_k \end{pmatrix} = \begin{pmatrix} d_1(\mathbf{x})\mathbf{b} \\ d_2(\mathbf{x})\mathbf{b} \\ \vdots \\ d_k(\mathbf{x})\mathbf{b} \end{pmatrix} = D(\mathbf{x})\mathbf{b} \quad (5)$$

where  $D(\mathbf{x})$  is the matrix arising from the differentiation of  $\hat{y}(\mathbf{x})$  with respect to each of  $k$  independent variables, that is,  $D'(\mathbf{x}) = (d'_1(\mathbf{x}), d'_2(\mathbf{x}), \dots, d'_k(\mathbf{x}))'$

The estimated derivative at any point  $\mathbf{x}$  in the direction specified  $k \times 1$  vector of a direction cosine  $\mathbf{c} = (c_1, c_2, \dots, c_k)'$  is given by  $\mathbf{c}'\mathbf{s}(\mathbf{x})$  where  $\sum_{i=1}^k c_i^2 = 1$ . Then the variance of this specified slope can be written as

$$\begin{aligned} V_c(\mathbf{x}) &= \text{Var}\{\mathbf{c}'\mathbf{s}(\mathbf{x})\} \\ &= \mathbf{c}'\text{Var}\{\mathbf{s}(\mathbf{x})\}\mathbf{c} \\ &= \sigma^2 \mathbf{c}' D(\mathbf{x})(\mathbf{X}'\mathbf{X})^{-1} D'(\mathbf{x})\mathbf{c} \\ &= \sigma^2 \mathbf{c}' M(\mathbf{x})\mathbf{c} \end{aligned} \quad (6)$$

where  $M(\mathbf{x}) = D(\mathbf{x})(\mathbf{X}'\mathbf{X})^{-1} D'(\mathbf{x})$ . For example, when  $\mathbf{c}$  is  $(1, 0, \dots, 0)'$ ,  $V_c(\mathbf{x})$  is the variance of the estimated slope over the first axial direction.

Atkinson<sup>11</sup> and Park<sup>2</sup> show that the averaged value of  $V_c(\mathbf{x})$  over all possible directions can be written as

$$\bar{v}(\mathbf{x}) = \frac{\sigma^2}{k} \text{tr}[M(\mathbf{x})], \quad (7)$$

where  $\text{tr}$  denotes the trace of a matrix. For some designs, it is possible to make this averaged variance be constant for all points  $\mathbf{x}$  that are equidistant from the design center. In that case,  $\bar{v}(\mathbf{x})$ , is a function of

only  $\rho^2 = \sum_{i=1}^k x_i^2$ . Park<sup>2</sup> called this property slope rotatability over all directions and gave the necessary and sufficient conditions for a design to be slope rotatable over all directions.

If a design satisfies the following conditions, it is slope rotatable over all directions.

[C1] All odd-order moments up to order 4 are zero.

[C2] [ii] are equal for all  $i$ , and [iiii] are equal for all  $i$ .

[C3] [ijjj] are equal for all  $i \neq j$ . (8)

In general, the designs which satisfy the three conditions in (8) are called symmetric permutation invariant designs. Actually, most commonly used response surface designs such as central composite designs with any positive  $\alpha$  value,  $3^n$  factorial designs and  $3^{n-k}$  fractional factorial designs belong to the class of slope-rotatable design over all directions. Comparing the conditions in (2) and (8), it is clear that the class of rotatable designs is a subset of the class of slope rotatable designs over all directions.

#### 4. Measures of slope-rotatability

There are several measures proposed to assess the degree of rotatability for a given response surface design. They are, for instance, Khuri<sup>23</sup>, Draper and Guttman<sup>24</sup>, Draper and Pukelsheim<sup>25</sup> and Park, Lim and Baba<sup>26</sup>. For assessing the degree of slope-rotatability, Park and Kim<sup>4</sup> proposed a measure for a design to be slope-rotatable over axial directions, and Jang and Park<sup>5</sup> proposed a measure for a design to be slope-rotatable over all directions.

#### 4.1 Measure of slope-rotatability over axial directions

Returning to equation (3), the variance of this derivative is written as

$$\begin{aligned}
 Var\left(\frac{\partial \hat{y}(\mathbf{x})}{\partial x_i}\right) &= Var(b_i) + 4x_i^2 Var(b_{ii}) + \sum_{\substack{j=1 \\ j \neq i}}^k x_j^2 Var(b_{ij}) \\
 &\quad + 4x_i Cov(b_i, b_{ii}) + 2 \sum_{\substack{j=1 \\ j \neq i}}^k x_j Cov(b_i, b_{ij}) \\
 &\quad + 4x_i \sum_{\substack{j=1 \\ j \neq i}}^k x_j Cov(b_{ii}, b_{ij}) + 2 \sum_{\substack{j < l \\ j, l \neq i}}^k x_j x_l Cov(b_{ij}, b_{il}) \quad (9)
 \end{aligned}$$

From the above equation, it can be seen that the necessary and sufficient conditions for [C1] and [C2] in equation (4) are:

$$\begin{aligned}
 v_1 &= v_2 = \dots = v_k \\
 4v_{11} &= 4v_{22} = \dots = 4v_{kk} = v_{12} = v_{13} = \dots v_{k-1,k} , \\
 c_{i,ii} &= c_{i,ij} = c_{ii,ij} = c_{ij,il} = 0 \quad (i \neq j \neq l \neq i) . \quad (10)
 \end{aligned}$$

Here we have used the following notation for the sake of simplicity. This notation will be used throughout the paper:

$$\begin{aligned}
 v_i &= Var(b_i), v_{ii} = Var(b_{ii}), v_{ij} = Var(b_{ij}) \\
 c_{i,ii} &= Cov(b_i, b_{ii}), c_{i,ij} = Cov(b_i, b_{ij}), c_{ii,ij} = Cov(b_{ii}, b_{ij}), c_{ij,il} = Cov(b_{ij}, b_{il}) .
 \end{aligned}$$

Hence, the equations in (10) are the necessary and sufficient conditions for a design to be slope-rotatable over axial directions.

Park and Kim<sup>4</sup> proposed that, if the conditions in (8) are met, then the following measure assesses the degree of slope-rotatability over axial directions for a design  $D$  with  $k$  independent variables.

$$Q_k(D) = \frac{1}{\sigma^4} (4v_{11} - v_{12})^2 \quad (11)$$

If this measure is zero, it is slope-rotatable over axial directions. If it becomes large, it becomes different from being slope-rotatable over axial directions.

For the class of central composite designs with  $n_0$  center points, the total number of  $N$  of experimental points can be written as

$$N = F + T,$$

Where  $F$  is the number of factorial points and  $T = 2k + n_0$  and  $F = 2^k$  if a full factorial is used and  $F = 2^{k-p}$  if a fractional factorial is used. For this design, the measure is found to be

$$Q_k(D) = \left( \frac{F + 2\alpha^2}{N} \right)^4 \left( 4e - \frac{1}{F} \right)^2,$$

where  $e = \frac{(k-1)FT - 4(k-1)F\alpha^2 + 2[N-2(k-1)]\alpha^4}{2\alpha^4[kFT - 4kF\alpha^2 + 2(N-2k)\alpha^4]}$

Table 1 gives the value of  $Q_k(D)$  for the central composite designs for various  $k$ ,  $n_0$ , and  $\alpha$ , and the values of  $Q_k(D)$  which make central composite designs slope-rotatable over axial directions. In this table, it is assumed that  $p = 0$ .

#### 4.2 Measure of slope-rotatability over all directions

Jang and Park<sup>5</sup> proposed a measure for evaluating slope rotatability over all directions in second order response surface designs. This measure can be used to form a slope variance dispersion graph evaluating the overall slope rotatability and the slope estimation capability of an experimental design throughout the region of interest.

Table 1. Values of  $Q_k(D)$  for central composite designs.(a)  $k = 2$ 

$\alpha$	$n_0$				
	1	2	3	4	5
1.0	0.6049	0.2779	0.1563	0.0977	0.0652
1.3	0.8156	0.1944	0.0792	0.0404	0.0234
1.6	0.4660	0.0846	0.0265	0.0106	0.0050
1.9	0.0582	0.0043	0.00003	0.0003	0.0008
2.2	0.0171	0.0300	0.0296	0.0255	0.0209
2.5	0.2420	0.1927	0.1460	0.1099	0.0833
2.8	0.8180	0.5782	0.4126	0.3000	0.2222
3.1	1.9863	1.3513	0.9440	0.9768	0.4969
*	2.090	1.984	1.911	1.859	1.820

(b)  $k = 3$ 

$\alpha$	$n_0$				
	1	2	3	4	5
1.0	0.4042	0.2958	0.2240	0.1739	0.1376
1.3	0.1585	0.0883	0.0577	0.0409	0.0304
1.6	0.1941	0.0568	0.0271	0.0158	0.0103
1.9	0.1185	0.0261	0.0096	0.0045	0.0024
2.2	0.0163	0.0025	0.0003	0.00008	0.00003
2.5	0.0010	0.0031	0.0039	0.0041	0.0039
2.8	0.0262	0.0245	0.0215	0.0183	0.0016
3.1	0.0850	0.0701	0.0574	0.0471	0.0388
*	2.432	2.339	2.268	2.213	2.172

\*: value of  $\alpha$  for slope-rotatability

Let us define that the spherical average slope variance,  $\bar{V}(r)$ , the average of the estimated average slope variances over the surface of a sphere with the radius  $r$ , is a quantity given by

$$\bar{V}(r) = \frac{\varphi}{\sigma^2} \int_{U_r} \bar{V}(\mathbf{x}) d\mathbf{x} \quad (12)$$

Where  $\bar{V}(\mathbf{x})$  is in (7), and  $U(r) = \{\mathbf{x} : \sum_{i=1}^k x_i^2 = r^2\}$  and  $\varphi^{-1} = \int_{U_r} d\mathbf{x}$  is the surface area of  $U(r)$ . The spherical average slope variance,  $\bar{V}(r)$  can be obtained as

$$\begin{aligned}\bar{V}(r) &= \frac{\psi}{\sigma^2} \int_{U_r} \frac{\sigma^2}{k} \text{tr}[\mathbf{D}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{D}'] d\mathbf{x} \\ &= \frac{\psi}{k} \text{tr} \left[ \int_{U_r} \mathbf{D}'\mathbf{D}(\mathbf{X}'\mathbf{X})^{-1} d\mathbf{x} \right] \\ &= \frac{1}{k} \text{tr}[S(\mathbf{X}'\mathbf{X})^{-1}] \quad (13)\end{aligned}$$

where  $S = \psi \int_{U_r} \mathbf{D}'\mathbf{D} d\mathbf{x}$  is the matrix of spherical region moments.

Then the slope variance dispersion measure, the range of  $\bar{V}(\mathbf{x})/\sigma^2$  on the sphere of radius  $r$ , as

$$RV(r) = V_{\max}(r) - V_{\min}(r) \quad (14)$$

$$\text{where } V_{\max}(r) = \max_{\mathbf{x} \in U_r} \frac{\bar{V}(\mathbf{x})}{\sigma^2} \quad \text{and} \quad V_{\min}(r) = \min_{\mathbf{x} \in U_r} \frac{\bar{V}(\mathbf{x})}{\sigma^2}$$

Since the form of  $\bar{V}(\mathbf{x})/\sigma^2$  depends on the chosen model, the form of the range will also depend on the model. Note that  $RV(r)$  is zero if and only if a design is slope-rotatable over all directions.  $RV(r)$  become larger as a design deviates from a slope-rotatable design.

A plot of the spherical average slope variance in (12), and the maximum and minimum average slope variances for locations on a sphere against the radius of the sphere, a slope variance dispersion graph (*SVDG*), can be used to give a comprehensive picture of the behavior of the average slope variances throughout a region and hence of the quality of the estimated slopes obtained with a particular design. Such plots can be used to investigate and compare the slope rotatability properties of

certain response surface designs currently available to the researcher. For more details, see Jang and Park<sup>5</sup>.

## 5. Slope-rotatability with equal maximum directional variance

Park and Kwon<sup>3</sup> proposed another concept of slope rotatability, which is called ‘slope rotatability with equal maximum directional variance’. Let  $V_{\max}(\mathbf{x})$  be the maximum directional variance at a point  $\mathbf{x}$  in the region of interest, that is,

$$V_{\max}(\mathbf{x}) = \max_{\mathbf{c}: \mathbf{c}'\mathbf{c}=1} V_{\mathbf{c}}(\mathbf{x}) = \max_{\mathbf{c}: \mathbf{c}'\mathbf{c}=1} \mathbf{c}'\mathbf{D}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{D}'\mathbf{c} \quad (15)$$

A design is said to be slope rotatable with equal maximum directional variance, if  $V_{\max}(\mathbf{x})$  would be equal for all points  $\mathbf{x}$  that are equidistant from the design center, that is,  $V_{\max}(\mathbf{x})$  is a function of only  $\rho^2 = \sum_{i=1}^k x_i^2$ .

The necessary and sufficient conditions for a design to be slope rotatable with equal maximum directional variance in the case of  $k=2$  are as follows.

$$[C1] \quad c_{i,j} = c_{i,ii} = c_{i,jy} = c_{j,ji} = c_{ii,jy} = 0, \quad (i \neq j, i, j = 1, 2)$$

$$[C2] \quad v_1 = v_2 .$$

$$[C3] \quad v_{11} = v_{22} .$$

$$[C4] \quad 2v_{11}^2 - v_{11}v_{12} - 2c_{11,22}^2 - c_{11,22}v_{12} = 0 . \quad (16)$$

For the cases of  $k > 2$  the conditions are not given here since they are lengthy. It was shown in Park and Kwon<sup>3</sup> that any rotatable permutation invariant symmetric second order design which satisfies the conditions (8) is also slope-rotatable with equal maximum directional variance. It was also shown that the conditions for a design to be rotatable are

sufficient for a design to be slope rotatable with equal maximum directional variance.

## 6. Slope-rotatable minimax design

An A-optimal design for  $\beta$  minimizes  $tr[(\mathbf{X}'\mathbf{X})^{-1}]$ , a D-optimal design minimizes  $|(\mathbf{X}'\mathbf{X})^{-1}|$ , and E-optimal design minimizes the maximum eigenvalue of  $(\mathbf{X}'\mathbf{X})^{-1}$ . If we are interested in the slope estimation rather than the estimation of  $\beta$ , it is natural to consider criteria based on the matrix  $\mathbf{M}=\mathbf{D}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{D}'$  in (6) instead of the matrix  $(\mathbf{X}'\mathbf{X})^{-1}$ . Consider the following quantities:

$$\begin{aligned} tr[\mathbf{D}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{D}'] &= \sum_{i=1}^k \lambda_i, \\ |\mathbf{D}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{D}'| &= \prod_{i=1}^k \lambda_i, \\ V_{\max} &= \max(\lambda_1, \lambda_2, \dots, \lambda_k). \end{aligned} \quad (17)$$

where  $\lambda_1, \lambda_2, \dots, \lambda_k$  are eigenvalues of  $\mathbf{D}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{D}'$ .

A design that minimizes one of the above quantities may be called an A-optimal, D-optimal or E-optimal design for slope estimation, respectively. The works of Huda and Al-Shiha<sup>16,27</sup> and Al-Shiha and Huda<sup>28</sup> deal with designs for estimating slopes of second order response surfaces.

It should be noted that the eigenvalues of the matrix  $\mathbf{D}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{D}'$  depend on the point  $\mathbf{x}$  through  $\mathbf{D}$ , and the design through  $(\mathbf{X}'\mathbf{X})^{-1}$ . In the region of interest  $R$ , if a design minimizes the maximum directional variance in the whole region  $R$ , it is called the “slope-directional minimax design (SDMD)”. Observe that the SDMD minimizes

$\max_{\mathbf{x} \in R} \max_{\mathbf{c}: \mathbf{c} \cdot \mathbf{c} = 1} \mathbf{c}' \mathbf{D}(\mathbf{X}' \mathbf{X})^{-1} \mathbf{D}' \mathbf{c}$ , which is from the equation (15), equal to

$$\max_{\mathbf{x} \in R} V_{\max}(\mathbf{x}).$$

We will illustrate how to find the SDMDs when the second order polynomial model is used for two independent variables. Suppose the region of interest is the unit circle

$$R = \{(x_1, x_2) : x_1^2 + x_2^2 \leq 1\}.$$

We want to find the SDMDs for the class of equiradial designs. Suppose the equiradial designs are arranged with  $n_0$  points at the design center and  $n_1$  points equally spaced on a circle of radius  $\rho_1$ , where  $0 < \rho_1 \leq 1$  and  $n_1 \geq 5$ . The total number of experimental points is  $N = n_0 + n_1$ . For these designs the moment matrix has the form

$$\mathbf{X}' \mathbf{X} = \begin{pmatrix} N & 0 & 0 & 4a\rho_1^2 & 4a\rho_1^2 & 0 \\ 0 & 4a\rho_1^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4a\rho_1^2 & 0 & 0 & 0 \\ 4a\rho_1^2 & 0 & 0 & 3a\rho_1^4 & a\rho_1^4 & 0 \\ 4a\rho_1^2 & 0 & 0 & a\rho_1^4 & 3a\rho_1^4 & 0 \\ 0 & 0 & 0 & 0 & 0 & a\rho_1^4 \end{pmatrix} \quad (18)$$

where  $a = n_1/8$ . Note that these equiradial designs are rotatable as well as slope-rotatable with equal maximum directional variance. From the inverse matrix of  $(\mathbf{X}' \mathbf{X})^{-1}$ , we can obtain that

$$v_1 = \frac{1}{4a\rho_1^2},$$

$$v_{11} = \frac{3N - 16a}{8a\rho_1^2(N - 8a)}.$$

Thus,  $V_{\max}(\mathbf{x})$  for these designs is given by

$$V_{\max}(\mathbf{x}) = v_1 + 4v_{11}\rho^2 \\ = \frac{1}{4a\rho_1^2} + \frac{4(3N-16a)}{8a\rho_1^4(N-8a)} \cdot \rho^2 , \quad (19)$$

where  $\rho^2 = x_1^2 + x_2^2$ . Since  $V_{\max}(\mathbf{x})$  is an increasing function of  $\rho^2$ , the value of  $V_{\max}(\mathbf{x})$  is maximized at  $\rho=1$ . On the other hand,  $V_{\max}(\mathbf{x})$  is a decreasing function of  $\rho_1^2$ , which depends on the selected design. Therefore, the minimum value of the equation (19) can be obtained by taking  $\rho_1=1$  and  $\rho=1$ , simultaneously.

That is,

$$\min_{0 < \rho_1 \leq 1} \max_{\mathbf{x} \in R} V_{\max}(\mathbf{x}) = \frac{14N-10n_1}{n_1(N-n_1)}. \quad (20)$$

Simple differentiation shows that for fixed  $N$ , the right hand side of equation (20) is minimized at

$$n_1 = (7 - \sqrt{14})N/5.$$

If  $N$  is sufficient large, the minimum value of  $\max_{\mathbf{x} \in R} V_{\max}(\mathbf{x})$  is obtained by locating about 34.8% of the design points at the center and the remainder equally spaced on the circle of radius 1. If  $N$  is small, we can find the optimal assignment for  $n_0$  and  $n_1$  as shown in Table 2.

Table 2. SDMD-optimal assignment of equiradial designs for the unit circle.

$N$	6	7	8	9	10	11	12	13	14	15
$n_0$	1	2	3	3	3	4	4	5	5	5
$n_1$	5	5	5	6	7	7	8	8	9	10

It is of interest to compare the equiradial SDMDs with D-optimal designs. It is well known that, when the unit  $k$ -ball is used as the region of interest, the D-optimal designs for the second order model put the entire mass at the center and the surface of the  $k$ -ball. Thus, we can

obtain the D-optimal designs for the unit circle region of interest in case of  $k = 2$  by locating all experimental points at the design center and at the circle of radius  $\rho_1 = 1$ . The equiradial designs whose design points are equally spaced on the unit circle are such designs, and their  $\mathbf{X}'\mathbf{X}$  matrix is given in the equation (18). The determinant of this  $\mathbf{X}'\mathbf{X}$  matrix is given by

$$|\mathbf{X}'\mathbf{X}| = \frac{1}{2^8} n_1^5 (N - n_1).$$

When  $N$  is fixed, this determinant is maximized at  $n_1 = 5N/6$  and  $n_0 = N/6$ , which means that the D-optimal design is obtained by locating about 16.7% of design points at the design center and the remainder on the circle of radius 1.

Table 3. D-optimal assignment of equiradial designs for the unit circle.

$N$	6	7	8	9	10	11	12	13	14	15
$n_0$	1	1	1	2	2	2	2	2	2	3
$n_1$	5	6	7	7	8	9	10	11	12	12

Table 3 shows that the D-optimal equiradial design for the unit circle. Note that the SDMDs suggest more observations at the center point than the D-optimal designs do.

## 7. Concluding remarks

Several concepts of slope-rotatability were reviewed. These design concepts seem to be useful in deciding a proper design for second order response surface polynomial models. It is often necessary that the experimental points of a response surface design are arranged in blocks. Khuri<sup>29</sup> shows that there are some designs which are arranged in blocks, but still maintain the rotatable property. It is currently under study to find some useful response surface designs which are arranged in blocks, but

have the slope-rotatable property. We hope to report good results on this soon.

## Acknowledgement

The author is very grateful to the referee and the editor for helpful comments. This work was partially supported by Korean Research Foundation Grant (KRF -2003-041-C00056).

## References

1. R. J. Hader and S. H. Park, *Technometrics*, **20**, 413-417 (1978).
2. S. H. Park, *Technometrics*, **29**, 449-453 (1987).
3. S. H. Park and H. T. Kwon, *Communications in Statistics-Theory and Method*, **27**, 2837-2851 (1998).
4. S. H. Park and H. J. Kim, *Journal of Applied Statistics*, **19**, 391-404 (1992).
5. D. H. Jang and S. H. Park, *Communications in Statistics-Theory and methods*, **22**, 1849-1863 (1993).
6. G. E. P. Box and J. S. Hunter, *Annals of mathematical Statistics*, **28**, 195-241 (1957).
7. R. H. Myers, "Response surface methodology", Blacksburg, VA, (Distributed by Edwards Brothers, Ann Arbor, MI), (1976).
8. A. M. Herzberg, *Journal of the Royal Statistical Society*, **B29**, 174-179 (1967).
9. G. E. P. Box and N. R. Draper, *Journal of the Royal Statistical Society*, **B42**, 79-82 (1980).
10. S. Huda and R. Mukerjee, , *Biometrika*, **71**, 381-385 (1984).
11. A. C. Atkinson, *Biometrika*, **57**, 319-328 (1970).
12. L. Ott and W. Mendenhall, *Technometrics*, **14**, 341-353 (1972).
13. V. N. Murty and W. J. Studden, *Journal of the American Statistical Association*, **67**, 869-873 (1972).
14. R. H. Myers and S. J. Lahoda, *Technometrics*, **17**, 481-486 (1975).
15. R. Mukerjee and S. Huda, *Biometrika*, **72**, 173-178 (1985).
16. S. Huda and A. A. Al-Shiha, *Sankhya, Ser. B*, **61**, 488-495 (1999).
17. S. Huda and M. Shafiq, *Journal of Applied Statistics*, **19**, 501-507 (1992).
18. N. R. Draper and L. H. Ying, *Journal of Statistical Planning and Inference*, **41**, 113-119 (1994).
19. L. H. Ying, F. Pukelsheim, and N. R. Draper, *Journal of Applied Statistics*, **22**, 331-341 (1995).
20. L. H. Ying, F. Pukelsheim, and N. R. Draper, *Journal of Applied Statistics*, **22**, 343-354 (1995).
21. H. J. Kim, Y. H. Um, and A. I. Khuri, *Communications in Statistics-Simulation and Computation*, **25**, 995-1014 (1996).

22. A. Giovannitti-Jensen and R. H. Myers, *Technometrics*, **31**, 159-171 (1989).
23. A. I. Khuri, *Technometrics*, **30**, 95-104 (1988).
24. N. R. Draper and I. Guttman, *I. Technometrics*, **30**, 105-111 (1988).
25. N. R. Draper and F. Pukelsheim, *Technometrics*, **32**, 195-202 (1990).
26. S. H. Park, J. H. Lim, and Y. Baba, *Ann. Inst. Statist. Math.*, **45**, 655-664 (1993).
27. S. Huda and A. A. Al-Shiha, *Communications in Statistics – Theory and Methods*, **92**, 1827-1849 (2000).
28. A. A. Al-Shiha and S. Huda, *Applied Statistical Sciences*, **10**, 357-364 (2001).
29. A. I. Khuri, *Technometrics*, **34**, 26-37 (1992).

## **CHAPTER 17**

### **DESIGN OF EXPERIMENTS FOR ESTIMATING DIFFERENCES BETWEEN RESPONSES AND SLOPES OF THE RESPONSE**

Shahariar Huda

*Department of Statistics & O.R., Kuwait University  
P.O. Box 5969, Safat-13060, Kuwait  
E-mail: shuda@kuc01.kuniv.edu.kw*

Design of experiments is considered for situations where estimating the difference between responses or slopes of the response rather than estimating the absolute response is the primary objective of the experimenter. Literature on experimental designs for estimating the differences between responses is reviewed. Available results are summarized and some open problems are discussed. Experimental designs for estimating the slopes of a response surface are also investigated. An unified treatment of the various optimality criteria in this context is provided. The concepts of slope-rotatability are discussed and some new results are presented.

#### **1. Introduction**

In statistical work on design of experiments usually the primary objective is to compare the effects of different treatments and, especially, estimate treatment contrasts. An exception is the field of response surface designs in which the treatments are various combinations of different levels of factors that are quantitative and the experimenter is mainly concerned with estimating the absolute response or the parameters of a model describing the smooth functional relationship between the response and the factors. However, in recent years it has been recognized that even in response surface designs the main interest of the experimenter may not

always be in the response at individual locations. Often the difference between responses at two locations may be of greater interest. If that be the case it is only natural that any design criterion used to select a design from the many competing ones should reflect this change of emphasis.

If differences in response at points close together in the factor space are of interest, the estimation of the local slopes of the response surface becomes important. Estimation of the slopes is particularly important in situations where the experimenter wishes to determine the optimal settings of the factors that produce the maximum (or minimum) value of the response. When estimation of slopes is the main objective of an experiment the design criterion used should reflect this objective.

In what follows a review of past work on designs for estimating differences between responses and designs for estimating slopes is provided along with some new results.

## 2. Preliminaries

Consider the typical response surface design set-up in which the univariate response  $y$  depends upon  $k$  quantitative factors  $x_1, \dots, x_k$  through a smooth functional relationship  $y = \varnothing(x, \theta)$  where  $x = (x_1, \dots, x_k)'$  and  $\theta = (\theta_1, \dots, \theta_p)'$  is a  $p$ -component column vector of unknown parameters. A design  $\xi$  is a probability measure on the experimental region  $\chi$  which is that part of the factor space in which experimentation is permissible. Let  $y_i$  be the observation at the point  $x_i = (x_{i1}, \dots, x_{ik})'$  ( $i = 1, \dots, N$ ) selected according to the design. It is assumed that  $y_i = \varnothing(x_i, \theta) + e_i$  where the  $e_i$ 's are uncorrelated zero-mean random errors with a constant variance  $\sigma^2$ .

The estimation of  $\theta$  is usually done by the method of least squares. If  $\hat{\theta}$  is the estimate of  $\theta$  then  $\hat{y}(x) = \varnothing(x, \hat{\theta})$  is the corresponding estimate of the response at the point  $x$  and  $\hat{y}(z) - \hat{y}(x)$  is the estimated difference between responses at points  $z$  and  $x$ . The column vector of estimated slopes along the factor axes at a point  $x$  is given by  $d\hat{y}/dx = (\partial \hat{y}(x)/\partial x_1, \dots, \partial \hat{y}(x)/\partial x_k)'$ . Let  $V(\xi, x)$  denote  $(N/\sigma^2) \text{ cov}(d\hat{y}/dx)$ ,

the normalized covariance matrix of the slopes, indicating its dependence on the point at which slopes are estimated as well as the design used. It is worth noting that the vector  $dy/dx$  not only displays the rates of change along the axial directions but also provides information about the rates of change in other directions. For example, the estimated directional derivative at point  $x$  in the direction specified by the vector of direction cosines  $c = (c_1, \dots, c_k)'$  is  $c' d\hat{y}/dx$ . Further,  $\{(d\hat{y}/dx)' (d\hat{y}/dx)\}^{-1/2}$   $d\hat{y}/dx$  is the estimate of the direction in which the derivative is largest.

Most of the available work on response surface designs is concerned with situations where the model is linear, that is,  $\mathcal{O}(x, \theta) = f'(x) \theta$  with  $f'(x) = (f_1(x), \dots, f_p(x))$  containing  $p$  linearly independent functions of  $x$ . In this case the least squares estimate  $\hat{\theta}$  has the normalized covariance matrix given by  $(N/\sigma^2) \text{ cov } (\hat{\theta}) = M^{-1}(\xi)$  where  $M(\xi) = \int_x f(x) \xi (dx)$  is the information matrix of  $\xi$ . Then  $(N/\sigma^2) \text{ var } \{\hat{y}(x)\} = f'(x) M^{-1}(\xi) f(x)$ ,  $(N/\sigma^2) \text{ var } \{\hat{y}(z) - \hat{y}(x)\} = \{f(z) - f(x)\}' M^{-1}(\xi) \{f(z) - f(x)\}$  and  $V(\xi, x) = H(x) M^{-1}(\xi) H'(x)$  where  $H(x)$  is a  $k \times p$  matrix whose  $i$ -th row is  $\partial f'(x)/\partial x_i = (\partial f_1(x)/\partial x_i, \dots, \partial f_p(x)/\partial x_i)$ .

The linear models that seem to have the widest application in the context of response surface design set up are the polynomial models for which  $f'(x)$  contains the terms of a polynomial of degree (order)  $d$  in  $x$ . When  $d=1$  the model is called a first-order model, when  $d=2$  the model is a second-order model and so on. If all the terms of a polynomial of degree  $d$  are included in the model then  $f(x)$  (and  $\theta$ ) contains  $\binom{k+d}{d}$  components. For example, in the first-order model one may write  $f'(x) = (1, x_1, \dots, x_k)$ . For the second-order model it is often convenient to write  $f'(x) = (1, x_1^2, \dots, x_k^2, x_1, \dots, x_k, x_1 x_2, \dots, x_{k-1} x_k)$ . For the third-order model the most convenient expression for  $f'(x)$  seems to be  $f'(x) = (1, x_1^2, \dots, x_k^2, x_1 x_2, \dots, x_{k-1} x_k, x_1 x_2 x_3, \dots, x_{k-2} x_{k-1} x_k, g'_1(x), \dots, g'_k(x))$  where  $g'_i(x) = (x_i, x_i^3, x_i x_1^2, \dots, x_i x_{i-1}^2, x_i x_{i+1}^2, \dots, x_i x_k^2)$  ( $i = 1, \dots, k$ ). In what follows only the polynomial models are to be

considered. A design  $\xi$  is called a d-th order design if it permits estimation of all the parameters of a d-th order model. A design  $\xi$  of order d is called symmetric if all the "odd moments" up to order 2d are zero, that is, if  $\int_{\chi} x_1^{d_1} \dots x_k^{d_k} \xi(dx) = 0$  whenever one or more of the  $d_i$ 's are odd integers and  $\sum_{i=1}^k d_i \leq 2d$ . A design is balanced (permutation invariant) if the moments are invariant with respect to permutations of the factors  $x_1, \dots, x_k$ . The class of "symmetric balanced" designs is easier for analysis and also very rich in the sense that it contains optimal designs under many commonly used design criteria. For example, in case of a first-order symmetric balanced design the information matrix is a simple diagonal matrix that may be written as  $M(\xi) = \text{Diag}\{1, \alpha_2 I_k\}$  where  $I_k$  is the identity matrix of order k and  $\alpha_2 = \int_{\chi} x_i^2 \xi(dx)$  ( $i = 1, \dots, k$ ). For a second-order symmetric balanced design one may write  $M(\xi) = \text{Diag}\{M, \alpha_2 I_k, \alpha_{22} I_{k'}\}$  with  $k' = k(k-1)/2$ ,  $\alpha_{22} = \int_{\chi} x_i^2 x_j^2 \xi(dx)$  ( $i \neq j = 1, \dots, k$ ),

$$M = \begin{bmatrix} 1 & \alpha_2 l_k' \\ \alpha_2 l_k & (\alpha_4 - \alpha_{22}) I_k + \alpha_{22} E_k \end{bmatrix},$$

$\alpha_4 = \int_{\chi} x_i^4 \xi(dx)$  ( $i = 1, \dots, k$ ),  $l_k$  being the k-component column vector of 1's and  $E_k = l_k l_k'$ .

For a third-order symmetric balanced design  $\xi$ ,  $M(\xi)$  is a block diagonal matrix given by  $M(\xi) = \text{Diag}\{M, \alpha_{22} I_{k'}, \alpha_{222} I_{k''}, M_1, \dots, M_k\}$  with  $k'' = k(k-1)(k-2)/6$  and

$$M_i = M^* = \begin{bmatrix} \alpha_2 & \alpha_4 & \alpha_{22} l_{k-1}' \\ \alpha_4 & \alpha_6 & \alpha_{24} l_{k-1}' \\ \alpha_{22} l_{k-1} & \alpha_{24} l_{k-1} & (\alpha_{24} - \alpha_{222}) I_{k-1} + \alpha_{222} E_{k-1} \end{bmatrix}$$

$$\text{where } \alpha_6 = \int_x x_i^6 \xi(dx), \alpha_{24} = \int_x x_i^2 x_j^4 \xi(dx) \text{ and } \alpha_{222} = \int_x x_i^2 x_j^2 x_l^2$$

$\xi(dx)$  ( $i \neq j \neq l = 1, \dots, k$ ), the other quantities being as defined earlier. Note that there is very little literature on models or designs of order greater than four except when  $k=1$ .

### 3. Designs for the differences

The importance of differences between estimated responses in the context of response surface designs was first realized by Professor Sir David Cox whose encouragement led Herzberg<sup>1</sup> to investigate the behaviour of the variance function of the difference between two estimated responses for first- and second-order rotatable designs. Rotatable designs were introduced by Box and Hunter<sup>2</sup> and are such that the variance of the estimated response at a point is a function of the distance of that point from the origin (centre of the design) and hence invariant under orthogonal rotations of the design. Rotatability is clearly a highly desirable property. In their seminal paper Box and Hunter<sup>2</sup> also derived the moment requirements of a  $d$ -th order rotatable design. From these requirements it was seen that the rotatable designs form a sub-class of the symmetric balanced designs, with special relationships between the non-zero moments of the same order to ensure rotatability. For example the first-order symmetric balanced designs are automatically rotatable while, with second-order designs  $\alpha_4 = 3$   $\alpha_{22}$  is an additional requirement and for third-order designs  $\alpha_6 = 3 \alpha_{24} = 15 \alpha_{222}$  is a further requirement for rotatability. Herzberg<sup>1</sup> showed that for a rotatable design the variance function of the difference between two estimated responses is a function of the distances of the points from the origin of the design and the angle subtending the points at the origin. Box and Draper<sup>3</sup> reviewed and simplified the findings of Herzberg<sup>1</sup> and also extended the results by investigating the behaviour of the variance function for a third-order rotatable design. Koske<sup>4</sup> further extended these results by obtaining the variance function of the difference between two estimated responses for a fourth-order rotatable design.

The moment requirements of a rotatable design are rather restrictive for implementation in all situations. Herzberg<sup>5,6</sup> introduced various types of cylindrically rotatable designs which are less restrictive but have somewhat similar properties. Huda<sup>7</sup> introduced m-grouped cylindrically rotatable designs of type 3 which allow grouping the factors into m ( $\geq 2$ ) groups retaining the property of cylindrical rotatability of type 3 considered in Herzberg<sup>6</sup> and Das and Dey<sup>8</sup>. Huda<sup>9</sup> provided formula for the variance function of the difference between two estimated responses for m-grouped first-order and 2-grouped second- and third-order cylindrically rotatable designs of type 3. The expression for variance function for second-order 3-grouped cylindrically rotatable designs of type 3 was presented in Huda<sup>10</sup>

When differences between estimated responses is more important than the estimated response at individual locations it is desirable to use a design chosen according to a criterion based on the differences. Encouraged by Professor Sir David Cox, Huda<sup>11</sup> introduced minimization of the maximum and average variance of the difference between estimated responses at two points in the region of interest R as design criteria to take account of such situations. Thus when  $R = \chi$  and a linear model is assumed, under the minimax criterion the objective is to minimize with respect to the design  $\xi$  the value of  $\{f(x) - f(z)\}' M^{-1}(\xi) \{f(x) - f(z)\}$  maximized with respect to the points x and z in  $\chi$  while under the average variance criterion the objective is to minimize  $\int_{\chi} \int_{\chi} \{f(x) - f(z)\}' M^{-1}(\xi) \{f(x) - f(z)\} d\eta(x) d\eta(z)$  where  $\eta$  is some averaging measure on  $\chi$ .

Huda<sup>12</sup> derived the optimal second-order design under the minimax criterion and a class of nearly optimal designs of order d under the average variance criterion for univariate polynomial regression over an interval. Huda and Mukerjee<sup>13</sup> derived the optimal second-order designs under the minimax criterion for regression over hyperspheres, restricting to rotatable designs since for polynomial regression over hyperspheres the optimal designs under this type of criterion are known to be rotatable (Kiefer<sup>14</sup>). The corresponding minimax problem for extrapolation

outside a hypersphere was tackled in Huda and Mukerjee<sup>15</sup>. Mukerjee and Huda<sup>16</sup> obtained the optimal second-order design under average variance criterion for regression over hyperspheres. The centre of the design space often plays the role of a control treatment. In such situations interest may lie in estimating difference between the responses at two points with one point being the origin. The minimax second-order design when one point is the origin and the second anywhere in a cocentric hyperspherical region of interest is provided in Huda<sup>17</sup>.

The derivation of optimal designs for the difference between estimated responses for regression over hypercubes is more difficult than the corresponding problem over hyperspheres. Huda<sup>18</sup> obtained the second-order designs over hypercubic regions that minimize average variance of the difference subject to one point being the origin. The minimax second-order designs over hypercubes for the difference between estimated responses at a point and at the centre was obtained in Huda<sup>19</sup> while Huda<sup>20</sup> provided the minimax second-order designs over hypercubes for the difference between estimated response at two points on a straight line through the centre. The unrestricted minimax problem for second-order designs over hypercubes is yet to be solved.

The minimax problem for third-order designs is extremely difficult to tackle and so far solutions have only been obtained in some simplified special cases. The minimax third-order designs for univariate regression over an interval was derived in Huda<sup>21</sup>. For regression over hyperspheres, assuming the region of interest to be a cocentric hypersphere, Huda<sup>22</sup> obtained the minimax third-order design when one of the points is taken to be the origin. The unrestricted minimax problem for hyperspherical regions is yet to be solved. The minimax problem for third-order designs over Hypercubic regions is even more formidable and remains to be investigated.

In none of the papers referred to above the possibility of inadequacy in the assumed model had been taken into consideration. Park<sup>23</sup> considered designs for estimating the differences between two responses and obtained optimal first and second-order designs that minimize the average (integrated) mean squared error of which the bias component is taken to arise from a polynomial of degree two and three, respectively.

The conclusions arrived at were analogous to those in Box and Draper<sup>24,25</sup> who considered minimization of integrated mean squared error of estimated response (at a point) as a design criterion.

#### 4. Designs for estimating slopes

The importance of designing experiments for estimating the slopes of a response surface was also first recognized by Professor Sir David Cox whose suggestion led Atkinson<sup>26</sup> to initiate research in the field. Atkinson<sup>26</sup> proposed minimization of the trace of the "integrated mean squared error matrix" of the estimated slopes as a design criterion and investigated first-order designs when the true model may be a second-order model. Since the pioneering work of Atkinson<sup>26</sup> the problem of optimal experimental design for estimating slopes has been investigated by many other researchers. Ott and Mendenhall<sup>27</sup> considered univariate second-order model over an interval and obtained the optimal design under the criterion of minimizing the variance of the estimated slope maximized over the design region. Murty and Studen<sup>28</sup> considered univariate polynomial model of order d over an interval and obtained optimal designs to minimize the variance of the estimated slope at a fixed point in the interval as well as averaged over the interval. Myers and Lahoda<sup>29</sup> extended the work of Atkinson<sup>26</sup> by considering the integrated mean squared error criterion for first- and second-order designs in the presence of second- and third-order terms in the true model, respectively when the integration is done with respect to an uniform measure.

Under the traditional A-, D- and E-optimality criteria the objective, assuming the specified model to be correct, is to minimize  $(1/p) \operatorname{tr} \operatorname{cov}(\hat{\theta}) (= \sum_{i=1}^p \mu_i/p)$ ,  $|\operatorname{cov}(\hat{\theta})|^{1/p}$  ( $= (\prod_{i=1}^p \mu_i)^{1/p}$ ) and  $\max(\mu_1, \dots, \mu_p)$ , respectively where  $\mu_i$  ( $i = 1, \dots, p$ ) are the e-values of  $\operatorname{cov}(\hat{\theta})$ . The parameter based D-criterion is equivalent to the response based G-criterion under which the objective is minimization of

$\text{var}(\hat{y}(x))$  maximized with respect to  $x \in \chi$  (Kiefer and Wolfowitz<sup>30</sup>). Similar equivalence results exist connecting other parameter based criteria with response based criteria (Silvey<sup>31</sup>). If the primary goal of the experimenter is estimation of slopes, it is natural to consider design criteria based on  $V(\xi, \eta)$  rather than  $\text{cov}(\hat{\theta})$  or  $\text{var}(\hat{y}(x))$  where  $V(\xi, \eta) = \int_R V(\xi, x) \eta(dx)$  with  $R$  being the region of interest and  $\eta$  a measure

reflecting the pattern of the experimenters interest. In analogy with the traditional set-up we may then define A-, D- and E-average optimal designs for estimating slopes as those minimizing

$$\beta = \begin{cases} \text{tr } V(\xi, \eta)/k = \sum_{i=1}^k \beta_i/k, \\ |V(\xi, \eta)|^{1/k} = (\prod_{i=1}^k \beta_i)^{1/k}, \\ \text{Max } (\beta_1, \dots, \beta_k), \end{cases}$$

respectively, where  $\beta_i$  are e-values of  $V(\xi, \eta)$ .

A particular choice is to take  $\eta$  as the measure putting all its mass at a single point  $x^*$  where  $\text{tr } V(\xi, x)$ ,  $|V(\xi, x)|$  and  $\max(\beta_1(\xi, x), \dots, \beta_k(\xi, x))$  is maximized, respectively, the  $\beta_i(\xi, x)$  being e-values of  $V(\xi, x)$ . It is to be noted that these objective functions are not necessarily maximized at the same point and hence the measure  $\eta$  may be different in each case. This is the minimax approach and we may define A-, D- and E-minimax optimality criteria of designs for estimating slopes as minimization of

$$\beta = \begin{cases} \text{tr } V(\xi, x)/k, \\ |V(\xi, x)|^{1/k}, \\ \text{Max } (\beta_1(\xi, x), \dots, \beta_k(\xi, x)), \end{cases}$$

maximized with respect to  $x \in R$ , respectively. It is possible to think of other options with the minimax approach. For example, in analogy with an optimality criterion due to Takeuchi<sup>32</sup> and called MV-optimality by Jacroux<sup>33</sup>, we could have the “MV-minimax optimality” criterion under which the objective would be to minimize the largest diagonal element of  $V(\xi, x)$  maximized with respect to  $x \in R$ . Note that the  $i$ -th diagonal element of  $V(\xi, x)$  is the variance of the estimated slope along the  $i$ -th axis at the point  $x$ .

The terms A-, D- and E-minimax optimality criteria were coined by Huda and Al-Shiha<sup>34</sup> who considered the linear model set-up with  $R = \chi$ . The A-criterion is the easiest to handle and has been extensively used in the past. The A-minimax second- and third-order designs for regression over hyperspheres were derived in Mukerjee and Huda<sup>35</sup>, some results for the set-up with  $R \neq \chi$  being given in Huda<sup>36</sup>. For the hypercubic regions the A-minimax second-order designs were presented in Huda and Shafiq<sup>37</sup> while the third-order designs were derived in Huda and Al-Shiha<sup>38</sup>. The D-criterion is much more complicated to tackle. For second-order models over spherical regions the D-minimax designs were derived in Huda and Al-Shiha<sup>39</sup> while the E-minimax designs were presented in Al-Shiha and Huda<sup>40</sup>. More recently, Huda and Al-Shingiti<sup>41</sup> obtained A-, D- and E-minimax designs over spherical regions when  $R \neq \chi$ . Huda and Al-Shiha<sup>34</sup> provided the D- and E-minimax second-order designs over hypercubic regions.

A-average optimal second- and third-order designs for spherical regions were studied in Huda<sup>42,43</sup>. The A-average second-order designs for cubic regions were studied in Huda<sup>44</sup>. Some investigations into the efficiencies of designs under various criteria have been carried out in Huda<sup>45,46</sup>, Huda and Ali<sup>47,48</sup>, Huda and Al-Shiha<sup>49</sup>, Huda and Al-Shingiti<sup>50,51,52</sup>.

The D- and E-minimax criteria have not yet been applied for third- and higher-order designs. The derivation of optimal designs for models of order three and higher is likely to be very difficult but certainly deserves the attention of researchers in the field.

Vaughan<sup>53</sup> obtained a new optimal second-order design for estimating slopes near a stationery region, taking account of both variance and bias in the estimation.

## 5. Slope-rotatable designs

Ever since its introduction by Box and Hunter<sup>2</sup>, rotatability has been recognized as a highly desirable property for response surface designs to possess. When considering designs for estimating slopes of a response surface it is therefore pertinent to look for designs having the property of rotatability in terms of estimated slopes rather than the estimated response. Since in this situation the experimenter is concerned with the  $k \times k$  matrix  $V(\xi, x)$ , many different concepts of slope-rotatability (rotatability in terms of slopes) are possible, each corresponding to a different scalar function of the matrix being considered for constancy at points equidistant from the centre or origin. In analogy with developments in other areas of experimental design we may define the following concepts of slope-rotatability. A design  $\xi$  is to be called

- (a) A-rotatable if and only if  $\text{tr } V(\xi, x)$ ,
- (b) D-rotatable if and only if  $|V(\xi, x)|$ ,
- (c) E-rotatable if and only if  $\max (\beta_1(\xi, x), \dots, \beta_k(\xi, x))$ ,
- (d) E<sub>all</sub>-rotatable if and only if  $\beta_i(\xi, x)$  ( $i = 1, \dots, k$ ),
- (e) MV-rotatable if and only if  $\max ((V(\xi, x))_{ii})$  ( $i = 1, \dots, k$ )),
- (f) Axial<sub>all</sub>-rotatable if and only if  $(V(\xi, x))_{ii}$  ( $i = 1, \dots, k$ ),

depends on  $x$  only through  $\rho = (x' x)^{1/2}$ , respectively where the quantities above are as defined earlier in Section 3 and Section 4.

Clearly (d) implies (a), (b) and (c). Also (f) implies (a) and (e). The earliest work on slope rotatability is by Hader and Park<sup>54</sup> who introduced the concept (f) for dealing with central composite designs in the context of second-order models only. In what follows we consider the concepts of A-, D- and E-rotatability in some details.

### 5.1. A-rotatability

Among the various concepts of slope-rotatability mentioned above, the A-rotatability appears to be the easiest to handle. It has also received the maximum attention since its introduction, in the context of second-order designs, by Park<sup>55</sup> who termed it “slope-rotatability over all directions” (SROAD). Earlier Mukerjee and Huda<sup>35</sup> derived that for a second-order symmetric permutation invariant (balanced) design  $\xi$ ,

$$V(\xi, x) = (1/\alpha_2 + \rho^2/\alpha_{22}) + \{4/(\alpha_4 - \alpha_{22}) - 2/\alpha_{22}\} \text{ diag } \{x_1^2, \dots, x_k^2\} \\ + (1/\alpha_{22} + 4 [1/\{\alpha_4 + (k-1)\alpha_{22} - k\alpha_2^2\} - 1/(\alpha_4 - \alpha_{22})]/k) x x' \quad (1)$$

and from (1) they observed that such a design is automatically A-rotatable. This observation is restated as Corollary 1 of Park<sup>55</sup> whose Theorem 1 provided the necessary and sufficient conditions for an arbitrary second-order design to be A-rotatable. These conditions were stated in terms of the elements of  $M^{-1}(\xi)$ . Equivalent conditions in terms of the elements of  $M(\xi)$  would provide greater insight into the structure of the designs but may be impossible to state for large values of  $k$  without imposing some restrictions on the extent of asymmetry or imbalance in the design. Since the rotatable designs are a subclass of the symmetric balanced designs, it follows from (1) that a second-order rotatable design is also A-rotatable, as stated in Corollary 2 of Park<sup>55</sup>. In fact, Mukerjee and Huda<sup>35</sup> found that for regression over spherical regions the A-minimax (optimal) second-order designs belong to the subclass of rotatable designs which suggests the same to be true for higher order designs.

It is possible to construct A-rotatable designs that are not symmetric and balanced. When dealing with a general design that might be asymmetric and unbalanced, it is convenient to denote the design moments by  $[i] = \int_x x_i \xi(dx)$ ,  $[ij] = \int_x x_i x_j \xi(dx)$ ,  $[iij] = \int_x x_i^2 x_j \xi(dx)$ ,

and so on. It can be shown that when  $d = 2$  and  $k = 2$ , the necessary and sufficient condition for a symmetric design to be A-rotatable is that  $[2222] + [11]^2 = [1111] + [22]^2$ . Thus for example a design with one

trial at each of  $(\pm a, \pm b)$ ,  $(\pm c, 0)$ ,  $(0, \pm c)$  and  $n_0$  trials at the origin  $(0, 0)$  is A-rotatable if  $c^2 = (a^2 + b^2)(n_0 + 4)/4$ . This two-dimensional design is clearly not balanced (permutation invariant) for arbitrary values of  $a$  and  $b$ . Another example of symmetric unbalanced two dimensional A-rotatable design is given by one trial at each of  $(\pm a, \pm a)$ ,  $(\pm b, 0)$ ,  $(0, \pm c)$  and  $n_0$  trials at the origin  $(0, 0)$  with  $a^2 = (b^2 + c^2)(n_0 + 6)/8$ , which is not balanced for arbitrary values of  $b$  and  $c$ . Ying, Pukelsheim and Draper<sup>56</sup> studied the problem in depth and derived several asymmetric balanced and asymmetric unbalanced second-order A-rotatable designs in two and three dimensions. The corresponding problem in higher dimensions ( $k \geq 4$ ) was investigated in Ying, Pukelsheim and Draper<sup>57</sup>.

Very little is known about A-rotatability of third-order designs. Mukerjee and Huda<sup>35</sup> observed that for third-order designs symmetry and balance are not sufficient to ensure A-rotatability but a rotatable design is automatically A-rotatable. Huda and Ali<sup>58</sup> derived some sufficient conditions for a symmetric balanced third-order design to be A-rotatable but without presentation of any example from outside the class of rotatable designs. No investigation seems to have been done into A-rotatability of fourth-order designs. More efforts are needed for investigation into A-rotatability of third- and higher-order designs.

## 5.2. D-rotatability

As stated earlier a design  $\xi$  is to be called D-rotatable if and only if the determinant of  $V(\xi, x)$  depends on the point  $x$  only through  $\rho^2 = x'x$ . The moment requirements of a D-rotatable design are quite difficult to derive even for low-order models. However, the following results are easy to obtain.

**Lemma 1:** Any second-order rotatable design is also D-rotatable.

**Proof:** If  $\xi$  is a symmetric balanced second-order design,  $V(\xi, x)$  is given by the expression (1) in Section 5.1. If the design is also rotatable

(i.e.  $\alpha_4 = 3\alpha_{22}$ ) then the second term on right hand side of (1) vanishes and the e-values of  $V(\xi, x)$  become  $(1/\alpha_2 + \rho^2/\alpha_{22})$  with multiplicity  $k-1$  and  $1/\alpha_2 + [2(k-1)/\alpha_{22} + 4/\{(k+2)\alpha_{22} - k\alpha_2^2\}] \rho^2/k$ . Hence the result.

**Lemma 2:** A second-order symmetric design in two dimensions is D-rotatable if and only if the following hold.

$$[1111] + [22]^2 = [2222] + [11]^2, \quad (2)$$

$$\{([1111] - [11])^2 ([2222] - [22])^2\}^{1/2} = 3 [1122] - [11][22], \quad (3)$$

$$\begin{aligned} & [11][1122]^2 + 2[11]^2[22][1122] + [11]^3[2222] + [11][22]^2[1111] \\ & + 4[11][2222][1122] + [22][1111][2222] \\ & = [22][1122]^2 + 2[11][22]^2[1122] + [22]^3[1111] + [11]^2[22][2222] \\ & + 4[22][1111][1122] + [11][1111][2222]. \end{aligned} \quad (4)$$

**Proof:** For a two-dimensional second-order symmetric design the expression for determinant of  $V(\xi, x)$  is a linear combination of the terms  $x_1^4, x_2^4, x_1^2x_2^2, x_1^2, x_2^2$  and a constant term. The constraints on the coefficients of these terms needed to make the expression a function of  $\rho^2 (= x_1^2 + x_2^2)$  only are equivalent to the three conditions (2), (3) and (4) stated in the lemma.

**Theorem 1:** A two-dimensional second-order symmetric balanced design is D-rotatable if and only if the design is rotatable.

**Proof:** The conditions (2) and (4) of Lemma 2 are satisfied by any symmetric balanced design (i.e. design with  $[11] = [22]$ ,  $[1111] = [2222]$ ) and then (3) can be satisfied if and only if  $[1111] = 3[1122]$ , i.e. if and only if the design is rotatable.

In view of Theorem 1, we are led to make the following conjecture.

**Conjecture 1:** A symmetric balanced second-order design is D-rotatable if and only if it is rotatable.

For  $k \geq 3$ , we are yet to actually investigate the necessary conditions for a second-order symmetric balanced design to be D-rotatable. However, in the light of results for  $k = 2$  given in the above theorem, the conjecture is highly likely to be true.

It may be possible to obtain D-rotatable designs from outside the class of symmetric balanced designs. The following lemma considers a class of asymmetric designs in two dimensions.

**Lemma 3:** For the second-order two-dimensional case, suppose all odd moments of order 4 or less are 0, except  $[12]$ ,  $[1112]$ , and  $[1222]$  (i.e.,  $[i] = [ij] = [iii] = [iji] = 0$  ( $i \neq j = 1, 2$ )). Then the design is D-rotatable if and only if condition (2) and the following hold.

$$\{([1111] - [11]^2)([2222] - [22]^2)\}^{1/2} = 3[1122] - 2[12]^2 - [11][22], \quad (5)$$

$$[1112] = [11][12], \quad (6)$$

$$[1222] = [22][12], \quad (7)$$

$$\begin{aligned} & 8[22][1112][12]^2 + 4[22][1222][1111] + 10[12][1122]^2 + 4[11] \\ & [1112][2222] + 8[11][1222][12]^2 \\ & = 4[22][1122][1112] + 8[1112][12][1222] + 2[12][22]^2[1111] \\ & + 8[1122][12]^3 + 2[12][1111][2222] + 4[11][1222][1122] \\ & + 4[11][1122][12][22] + 2[11]^2[2222][12], \end{aligned} \quad (8)$$

$$\begin{aligned} & 4[22]^2[1112][12] + [1111][22][2222] + 4[1222][1122][12] + 4[11] \\ & [1122][2222] + 4[11][22][1222][12] + [11]^3[2222] + [11][1122]^2 \\ & + 4[1222][12][1111] + 4[11][1122][12]^2 \\ & + [11][22]^2[1111] + 2[11]^2[1122][22] + 4[22][1112]^2 \\ & = 4[1112][12][2222] + 4[1122][22][12]^2 + 4[11][1222]^2 \\ & + 2[11][1122][22]^2 + [11]^2[2222][22] + [22]^3[1111] \\ & + [22][1122]^2 + 4[22][1111][1122] + 4[1112][12][1122] \\ & + [11][2222][1111] + 4[11]^2[1222][12] \\ & + 4[11][22][1112][12]. \end{aligned} \quad (9)$$

**Proof:** For a second-order design of the type considered the determinant of  $V(\xi, x)$  is a polynomial of degree 4 with non-zero coefficients for  $x_1^4$ ,  $x_2^4$ ,  $x_1^3 x_2$ ,  $x_1 x_2^3$ ,  $x_1^2 x_2^2$ ,  $x_1^2$ ,  $x_2^2$ ,  $x_1 x_2$  and a constant term. Conditions (2) and (5) are obtained by making the terms involving  $x_1^4$ ,  $x_2^4$ , and  $x_1^2 x_2^2$  a function of  $\rho^2 = x_1^2 + x_2^2$  while (6) and (7) are obtained by equating to zero the coefficients of  $x_1^3 x_2$  and  $x_1 x_2^3$ . Condition (8) arises from the need for vanishing of the coefficient of  $x_1 x_2$  while (9) is obtained by equating the coefficients of  $x_1^2$  and  $x_2^2$ .

**Lemma 4:** A two-dimensional non-singular balanced second-order design within the class considered in Lemma 3 cannot be D-rotatable.

**Proof:** For a balanced design in the class considered in Lemma 3, conditions (2) and (9) are automatically satisfied. Condition (5) reduces to  $[1111] (= [2222]) = 3 [1122] - 2 [12]^2$  while (6) and (7) are equivalent to  $[1112] (= [1222]) = [11] [12] (= [22] [12])$ . Further, (8) reduces to  $8 [11] [1112] [12]^2 + 4 [1112] [11] [1111] + 5 [12] [1122]^2 = 4 [1112] [1122] [11] + 2 [11]^2 [12] [1111] + 4 [1112]^2 [12] + 4 [1122] [12]^3 + [12] [1111]^2 + 2 [11]^2 [1122] [12]$ . Substituting  $[1112]$  and  $[1111]$  in terms of  $[1122]$  and  $[12]$  in the reduced version of (8) leads to  $[1122] = [12]^2$  and thus the solution to equations (2) and (5)-(9) is given by  $[1111] = [1122] = [12]^2$ ,  $[1112] = [11] [12]$  which implies that the design must be singular.

Nothing is yet known about D-rotatability of designs of order three or higher. Further investigation is also needed to obtain the necessary and sufficient conditions for D-rotatability of second-order designs. It would be interesting to find asymmetric and unbalanced designs that are D-rotatable. In view of the results obtained above for the two-dimensional case, it is highly unlikely that there exist any asymmetric balanced design in higher dimensions that is D-rotatable.

### 5.3. E-rotatability

A design is to be called E-rotatable if and only if the largest e-value of  $V(\xi, x)$  depends on the point  $x$  only through  $\rho^2 = x' x$ . In general, like

D-rotatability, the E-rotatability also may not be very easy to deal with, particularly for higher order models and for  $k \geq 3$ . For the second-order models, some results are obvious. For example, if a second-order design is symmetric and balanced then from Lemma 1 it immediately follows that a rotatable design is also E-rotatable. This is also the result of Theorem 2 of Park and Won<sup>59</sup> who used the term "slope-rotatability with equal maximum directional variance" to describe E-rotatability in the context of second-order designs. In Theorem 1 of the paper, Park and Kwon<sup>59</sup> provided the necessary and sufficient conditions for a two-dimensional second-order design to be E-rotatable but the conditions were stated in terms of the elements of  $M^{-1}(\xi)$  and do not provide much insight into the structure of designs that may possess E-rotatability.

It would be interesting to find out if second-order E-rotatable designs exist outside the class of rotatable designs. In particular, the existence of any symmetric balanced design that is E-rotatable but not rotatable may have useful applications. The matter is currently under investigation.

## 6. Concluding remarks

As mentioned in Section 3 and Section 4 there are many interesting open problems, yet to be solved, in optimal design of experiments for estimating the differences between responses as well as estimating the slopes. Very little work has been done for third-order designs and practically nothing done for higher order designs in more than one variable. More efforts are also needed to consider situations that take account of the possibility of bias in the assumed models.

The various concepts of slope-rotatability also require further investigation. Necessary and sufficient conditions of slope-rotatability in terms of the moments of the design need to be derived. Existence or non-existence of slope-rotatable designs within classes of commonly used designs need to be proved and actual examples constructed when existence is assured. The recent work by Ying, Pukelsheim and Draper<sup>56,57</sup> should be inspiration for those keen to research in this area.

Rotatability is a highly desirable property for response surface designs. However, physical or other constraints may sometimes prevent

an experimenter from employing a rotatable design. In such situations it is of interest to know how close to (or how far from) being rotatable the design used is. The pioneering work on measures of rotatability was by Khuri<sup>60</sup>, followed by Draper and Guttman<sup>61</sup> and Draper and Pukelsheim<sup>62</sup>. More recently, researchers have been developing analogous measures of slope-rotatability. Park and Kim<sup>63</sup> and Jang<sup>64</sup> dealt with measures of Axial *all*-rotatability while Jang and Park<sup>65</sup>, Draper and Ying<sup>66</sup>, Kim, Um and Khuri<sup>67</sup> developed measures of A-rotatability. These papers deal with second-order designs only. Measures of rotatability or slope-rotatability for higher order models are yet to be investigated. It would also be interesting to develop numerical or graphical measures of D- and E-rotatability.

## Acknowledgments

This research was supported by Grant SS01/04 from Kuwait University.

## References

1. A. M. Herzberg, *Jour. Roy. Statist. Soc., Ser. B*, 174 (1967).
2. G. E. P. Box and J. S. Hunter, *Ann. Math. Statist.*, 195 (1957).
3. G. E. P. Box and N. R. Draper, *Jour. Roy. Statist. Soc., Ser. B*, 79 (1980).
4. J. K. A. Koske, *J. Statist. Plann. Inf.*, 263 (1989).
5. A. M. Herzberg, *Ann. Math. Statist.*, 242 (1966).
6. A. M. Herzberg, *Ann. Math. Statist.*, 167 (1967)
7. S. Huda, *Biometrical J.*, 469 (1982).
8. M. N. Das and A. Dey, *Ann. Inst. Statist. Math.*, 331 (1967) (Corr., 337 (1968)).
9. S. Huda, *Ann. Inst. Statist. Math.*, 481 (1984).
10. S. Huda, *Pakistan J. Statist.*, 245 (1989).
11. S. Huda, Ph. D. Thesis, Imperial College, University of London (1981).
12. S. Huda, *J. Statist. Plann. Inf.*, 89 (1985).
13. S. Huda and R. Mukerjee, *Biometrika*, 381 (1984).
14. J. Kiefer, *Ann. Math. Statist.*, 298 (1961).
15. S. Huda and R. Mukerjee, *Statist. Prob. Letters*, 43 (1987).
16. R. Mukerjee and S. Huda, *J. Statist. Research*, 41 (1985).
17. S. Huda, *Pakistan J. Statist.*, Ser. B, 27 (1987).
18. S. Huda, *Parisankhyayan Samikkha*, 53 (1995).
19. S. Huda, *Statist. Prob. Letters*, 193 (1997).

20. S. Huda, *Calcutta Statist. Assoc. Bull.*, 101 (1998).
21. S. Huda, *J. Statist. Research*, 65 (1986).
22. S. Huda, *Biometrical J.* 827 (1989).
23. J. Y. Park., *Commun. Statist.-Theory Meth.*, 4773 (1990).
24. G. E. P. Box and N. R. Draper, *J. Amer. Statist. Assoc.*, 622 (1959).
25. G. E. P. Box and N. R. Draper, *Biometrika*, 335 (1963) (Corr., 305 (1965)).
26. A. C. Atkinson, *Biometrika*, 319 (1970).
27. L. Ott and W. Mendenhall, *Technometrics*, 341 (1972).
28. V. N. Murty and W. J. Studden, *J. Amer. Statist. Assoc.*, 869 (1972).
29. R. H. Myers and S. J. Lahoda, *Technometrics*, 481 (1975).
30. J. Kiefer and J. Wolfowitz, *Canad. J. Math.*, 363 (1960).
31. S. D. Silvey, Optimal Design: An Introduction to the Theory for Parameter Estimation, Chapman and Hall (1980).
32. K. Takeuchi, *Appl. Res. Un. Japan Sci. Engrs.*, 140 (1961).
33. M. Jacroux, *Jour. Roy. Statist. Soc.*, Ser. B, 70 (1983).
34. S. Huda and A. A. Al-Shiha, *Commun. Statist.-Theory Meth.*, 1827 (2000).
35. R. Mukerjee and S. Huda, *Biometrika*, 173 (1985).
36. S. Huda, *Biometrical J.*, 189 (1990).
37. S. Huda and M. Shafiq, *J. Appl. Statist.*, 501 (1992).
38. S. Huda and A. A. Al-Shiha, *Commun. Statist.-Theory Meth.*, 345 (1998).
39. S. Huda and A. A. Al-Shiha, *Sankhya, Ser. B*, 488 (1999).
40. A. A. Al-Shiha and S. Huda, *J. Appl. Statist. Sciences*, 357 (2001).
41. S. Huda and A. M. Al-Shingiti, *Commun. Statist.- Simul. Comp.*, 773 (2004).
42. S. Huda, *Pakistan J. Statist.*, 31 (1986).
43. S. Huda, *J. Appl. Statist. Sciences*, 217 (1999).
44. S. Huda, *Parisankhyayan Samikkha*, 11 (1998).
45. S. Huda, *J. Ind. Soc. Agr. Statist.*, 39 (1987).
46. S. Huda, *J. Ind. Soc. Agr. Statist.*, 163 (1991).
47. S. Huda and H. Ali, *Pakistan J. Statist.*, 125 (1990).
48. S. Huda and H. Ali, *J. Statist. Studies*, 6 (1993).
49. S. Huda and A. A. Al-Shiha, *J. Statist. Studies*, 1 (2001).
50. S. Huda and A. M. Al-Shingiti, *Pakistan J. Statist.*, 399 (2003).
51. S. Huda and A. M. Al-Shingiti, *J. Statist. Theory and Appl.*, 223 (2003).
52. S. Huda and A. M. Al-Shingiti, *Aligarh J. Statist.*, 143 (2004).
53. T. S. Vaughan, *Commun. Statist.-Theory Meth.*, 1335 (1993).
54. R. J. Hader and Park, S. H., *Technometrics*, 413 (1978).
55. S. H. Park, *Technometrics*, 449 (1987).
56. L. H. Ying, F. Pukelsheim and N. R. Draper, *J. Appl. Statist.*, 331 (1995).
57. L. H. Ying, F. Pukelsheim and N. R. Draper, *J. Appl. Statist.*, 343 (1995).
58. S. Huda and H. Ali, *J. Statist. Research*, 27 (1991).
59. S. H. Park and H. T. Kwon, *Commun. Statist.-Theory Meth.*, 2837 (1998).
60. A. I. Khuri, *Technometrics*, 95 (1988).

61. N. R. Draper and I. Guttman, *Technometrics*, 105 (1988).
62. N. R. Draper and F. Pukelsheim, *Technometrics*, 195 (1990).
63. S. H. Park and Kim, H. J., *J. Appl. Statist.*, 391 (1992).
64. D. H. Jang, *Comp. Statist. Data Anal.*, 343 (2002).
65. D. H. Jang and S. H. Park, *Commun. Statist.-Theory Meth.*, 1849 (1993).
66. N. R. Draper and L. H. Ying, *J. Statist. Plann. Infer.*, 113 (1994).
67. H. J. Kim, Y. H. Um and A. Khuri, *Commun. Statist.-Simul. Comp.*, 995 (1996).

## Subject Index

- $2^6$  design, 3  
 $2^k$  factorial, 252, 254, 255  
 $3^2$  factorial, 261, 262, 269  
 $3^2$  factorial design, 256, 276  
 $3^k$  design, 254  
 $3^k$  factorial, 252  
3D surface plots, 300, 302  
50th anniversary of published mixture experiment research, 283
- A*-optimality, 258, 260  
*A*-, *D*- and *E*-average optimal design, 435  
*A*-, *D*- and *E*-minimax design, 436  
*A*-, *D*- and *E*-minimax optimality, 435, 436  
*A*-, *D*- and *E*-optimality, 434  
*A*-, *D*- and *E*-rotatability, 437  
*A*-criterion, 436  
*A*-rotatability, 438, 439  
*A*-rotatable, 437–439  
*A*-rotatable design, 438, 439  
additive cubic model, 291  
additive heteroscedastic error, 189  
additive models, 296, 298  
additive quadratic model, 291, 299  
additive spline partial least squares, 297  
agrochemicals, 19  
algorithm, 176  
alias, 146  
alphabetic optimality criteria, 330  
alphabetic-optimal design, 289, 290  
alternative generators, 4
- analysis of symmetric simplex designs, 301  
analyzed component proportions, 301, 302  
assessing and comparing designs, 283  
assessing component effects, additivity, and interactions, 300  
augmentation design, 291  
augmented simplex-centroid design, 288, 289  
authors of mixture experiment publications, 286  
average prediction variance, 337, 354, 355, 362  
average spherical prediction variance, 337, 356  
average variance, 54  
average variance criterion, 432  
axial design, 288, 290  
axial<sub>all</sub>-rotatable, 437
- balanced incomplete block design, 30  
Bayesian, 274  
Bayesian analysis, 44  
Bayesian credible region, 81  
Bayesian *D*-optimal design, 291  
Bayesian design, 44, 200  
Bayesian model averaging, 81  
Bayesian posterior predictive approach, 302  
Bayesian reliability, 81  
Bayesian two-stage optimal design, 291  
BBD, 359–361, 365, 366, 371

- BBDs, 362, 364, 369, 370  
 Becker models, 304  
 bias, 49  
 bias component, 433  
 bias design criteria, 304  
 bias dispersion graphs (BDGs), 294, 375  
 bias properties of a design, 292  
 bias-variance trade-off, 397  
 binary data, 181  
 binary response, 174  
 binary response model, 297, 299, 301  
 binomial distribution, 175  
 bioprocessing, 19  
 biplot display, 301, 302  
 blending properties, 299  
 block, 2, 153  
 block design, 40, 292  
 block effects, 300  
 block generators, 4, 5, 9, 11, 14, 15  
 block mixture design, 289  
 block mixture-process variable design, 289  
 block structure, 19, 21, 31  
 blocked experiment, 49  
 blocking, 19, 22, 44, 300, 305, 363  
 blocking arrangements, 4  
 blocks, 363, 364  
 blocks of size two, 1, 3  
 bootstrap approach, 75  
 boundary of the experimental region, 292, 293, 304  
 Box and Cox, 174  
 Box-Behnken design (BBD), 27, 29, 252, 254, 268, 357, 358  
 Box-Draper design, 360  
 boxplots, 366  
 branch-and-bound, 278  
 c-optimal design, 228, 233  
 candidate design points, 292, 304  
 candidate-point-free optimal design, 291  
 canonical analysis, 44  
 canonical form, 185  
 canonical parameter, 206  
 canonical polynomial model, 283, 295  
 Carathéodory's Theorem, 183  
 categorical variables design, 289  
 categorized components & process variables, 288, 290  
 categorized components design, 290, 293, 304  
 categorized components mixture model, 296  
 centered-and-scaled intercept model, 297, 299  
 central composite analogue design, 291, 293  
 central composite design (CCD), 22, 27, 42, 148–151, 252, 254, 255, 265, 268, 355, 357–361, 364–366, 369, 371  
 centroids, 288, 290  
 check points for model lack-of-fit, 290  
 chemical kinetics, 193  
 chronological listing, 287, 289, 294, 296, 300, 303  
 chronological order, 288  
 code, 67  
 coded form, 68  
 collinearity, 300, 302  
 collinearity diagnostics, 293, 294  
 combination design, 6–8, 11, 13  
 combined array, 129, 135, 137, 147  
 comparing design efficiencies, 294  
 comparing mixture experiment designs, 293  
 completely randomized, 42  
 component additivity, 302  
 component categories, 293  
 component correlations, 293  
 component effects, 302  
 component interaction plots, 302  
 component interactions, 302  
 component proportions, 284, 285, 288  
 component slope linear mixture model, 297, 298

- component slope variance trace plots, 294  
component transformation, 300  
component-amount design, 291, 292  
component-amount model, 296  
component-wise mixture model, 296  
components, 284  
componentwise variance dispersion graphs (CVDGs), 372  
composite mixed resolution design (CMRD), 150, 152  
composition space, 285  
computer generated design (CGD), 152, 153  
concave, 76  
confidence band, 75, 80  
confidence cone, 69  
confidence region, 69, 76, 80  
confidence region for optimum point, 301  
confounding patterns, 1, 11  
consistent constraints, 294  
constrained canonical mixture polynomial model, 297, 299  
constrained nonlinear optimization, 302  
constrained optimal factor level, 76  
constrained region design, 288, 304  
constraints, 302  
constructed variable, 192  
continuation-ratio model, 225, 230  
continuous response, 189  
contour plots, 44, 300, 302, 352, 375  
control factor (variable), 124–126, 128–133, 135–137, 140–142, 144–148, 150–154  
controllable factors, 99  
convergence of MLE's, 238  
convex, 76  
convex polyhedron, 285  
convexity, 93  
cost, 49  
count response model, 299, 305  
covariance structure, 31  
Cox directions, 372  
Cox polynomial model, 296  
crossed array, 124, 128, 129, 132, 135, 147, 148, 152  
crossed blocking, 37  
cuboidal, 265, 266, 268, 277  
cuboidal region, 264  
CVDGs, 373, 374  
cylindrical rotatability, 432  
cylindrically rotatable design, 432  
  
 $D$ -criterion, 434, 436  
 $D$ -criterion (efficiency), 151–154  
 $D$ -efficiency, 28, 259, 261, 277  
 $D$ -minimax design, 436  
 $D$ -optimal, 22, 259, 261, 262  
 $D$ -optimality, 176, 258, 334  
 $D$ -optimal design, 228, 291  
 $D$ -optimal split-plot mixture-process variable design, 291  
 $D$ -rotatability, 442  
 $D$ -rotatable, 437, 439–442  
 $D$ -rotatable design, 439, 441  
 $D$ - and  $E$ -minimax criteria, 436  
 $D_s$ -efficiency, 34, 42  
 $D_s$ -optimal, 35, 260, 262, 273  
data analysis topics, 283, 299, 300  
delta method, 72, 75  
design  
choice of, 209  
dependence problem, 204, 205, 209  
GLM, 203  
locally-optimal, 205  
response surface, 203  
design criterion, 428, 430, 432, 434, 435  
design matrix, 350  
design of experiment, 427  
design optimality, 353  
design optimality criteria, 288, 292, 293  
design region, 148, 151, 152, 178, 434

- designs for additive quadratic or cubic models, 291  
 designs for Cox models, 290  
 designs for log-contrast models, 291  
 designs for Scheffé models with inverse terms, 290  
 designs for Scheffé models with logarithmic terms, 290  
 desirability function, 79, 80, 144, 301, 302  
 difference between estimated responses, 432, 433  
 difference between two estimated responses, 431, 432  
 differences between responses, 428, 443  
 direction  
     Joshi et al., 93, 94  
     Newton, 94, 95  
     orthogonal, 94  
     steepest ascent, 89–91  
     steepest descent, 90  
 direction cosines, 69  
 directional derivative, 176, 429  
 dispersion effect, 130, 132  
 dispersion parameter, 175, 206, 211  
 distance-based criteria, 292  
 distance-based design, 290  
 dose-response surface, 159  
 double linear regression, 72  
 drug synergy, 227  
 dual response approach, 90, 93  
 dual response problem, 106  
*E*-rotatable, 437, 442, 443  
 e-value, 434, 435, 442  
 $E_{all}$ -rotatable, 437  
*E*-minimax design, 436  
*E*-rotatability, 443  
 easy to change (ETC), 49  
 effects plot, 300, 302  
 efficiencies, 49  
 efficiency, 187  
 efficiency factor, 34  
 efficiency measures, 293, 294  
 eigenanalysis, 293, 294  
 eigenvalues, 71  
 empirical method, 66  
 equivalence result, 435  
 estimated scaled mean-squared error of prediction (ESMSEP), 368  
 evolutionary operations procedure, 159  
 expected mean squares, 51  
 experimental design, 146, 154  
 experimental region, 285, 305  
 experimental units, 21  
 experiments with mixtures, 283, 371  
 exponential decay, 195  
 exponential family, 175, 205, 206, 211  
 extended hexagon lattice design, 291  
 extrapolation outside a hypersphere, 433  
 extreme vertices design, 288, 289, 372, 374  
 face-centered cube, 214  
 face-centered cube (FCC) design, 148, 151, 360–363  
 factorial design, 125, 147, 189  
 factorwise balance, 24, 28  
 FDS plot, 338, 369–371, 375  
 first-order decay, 193  
 first-order design, 434  
 first-order model, 20, 67, 146, 150, 429  
 first-order polynomial, 251, 253, 254, 257, 258, 262, 263, 265, 277  
 Fisher information matrix, 227  
 fixed effects, 53  
 foldover, 4, 14  
 foldover pairs, 1, 2, 14, 16  
 food engineering, 19  
 fourth-order rotatable design, 431  
 fraction of design space (FDS), 277, 349, 383  
 fraction of design space (FDS) plot, 269, 294, 332, 356, 368, 369, 374

- fractional factorial design, 147, 148  
fractional factorial split-plot design, 154  
fractional replicates, 29  
fractionating mixture-process variable designs, 292  
full-cubic model, 295  
full-quartic model, 295  
future mixture experiment research, 283  
fuzzy set approach, 301  
*G*, 353, 354, 358, 369, 373  
*G*-criterion, 434  
*G*-criterion (efficiency), 151  
*G*-efficiency, 266  
*G*-optimal, 265, 266  
*G*-optimality, 334, 381  
gamma model, 174  
gasoline blending model, 296  
generalized linear mixed model (GLMM), 145, 297  
generalized linear model (GLM), 74, 129, 132, 139–142, 144, 145, 173, 203, 205, 297, 401  
generalized standard deviation, 259  
generalized standardized deviation, 277  
genetic algorithm, 36, 84, 266, 278, 291, 354  
Gibbs sampling, 81  
global optimal point, 76  
global optimization, 84  
gradient plot, 300, 302  
graphical methods, 302, 379  
Großner basis methods, 301  
group confounding pattern, 6  
group patterns, 8, 11, 13  
  
hard to change (HTC), 49  
hard-to-change factor, 21, 304  
Hessian, 92  
Hessian matrix, 93  
heteroscedastic linear model, 192  
hierarchical, 398  
Hoke design, 360  
homogeneous-of-degree-one model, 296, 297  
hybrid designs, 360  
hyperarcs, 361, 362  
hypercube, 355  
hypercubic region, 433, 436  
hypersphere, 265, 267, 268  
hyperspherical, 266, 277  
hyperspherical coordinates, 364  
hyperspherical design spaces, 264  
hyperspherical region of interest, 433  
  
ill-conditioning, 300  
important mixture components, 292  
improving direction, 92  
inactive components, 296  
independently resetting factor level, 49  
indices of rotatability, 263  
induced design region, 185  
information matrix, 176, 429, 430  
ingredients, 284  
initial arrangement, 4  
initial parameter estimates, 403  
integrated mean squared error, 271, 272  
integrated mean squared error matrix, 434  
integrated squared bias, 272  
interaction term, 186  
intercept mixture model, 296, 299  
interchange algorithm, 36, 39, 40, 42  
interior of the experimental region, 293, 304  
inverse polynomial model, 296, 305  
inverse-term model, 296, 298, 304  
irregularly shaped experimental region, 388  
iterative weight, 181  
iterative weighted least squares, 175  
*IV*, 353, 354, 358  
*IV*-efficiency, 264, 265, 269

- IV*-optimal, 264, 265
- Kiefer ordering, 294
- Kronecker homogeneous polynomial model, 297, 299
- lack of fit, 21, 23
- Lagrange multiplier, 74
- larger the better, 83
- layered design, 291, 293
- least squares estimate, 429
- least squares estimation, 3
- leverage, 293, 294
- limiting optimal design, 236
- linear blending model, 295
- linear equality constraints, 305
- linear inequality constraints, 285, 305
- linear logistic model, 181
- linear model, 204, 429, 436
- linear predictor, 175, 206, 212, 213
- link
- canonical, 207, 217
  - function, 140, 142, 175, 206
  - logarithmic, 211
  - logit, 206
  - square root, 212, 217
- locally optimum design, 177
- locally *D*-optimum design, 193
- log contrast model, 296, 299, 304
- log link, 177, 178
- log-linear model, 213
- logistic model, 181, 297
- logistic regression, 368, 401
- logistic regression model, 207
- loglinear canonical model, 296
- loss function, 83
- lower and upper bounds, 285, 288
- main effects, 1
- major components, 293
- marginal, 23
- mathematically independent variable designs, 288
- mathematically independent variables, 295
- matrix transformation, 288
- maximum likelihood (ML), 130, 139
- maximum likelihood estimate, 207
- maximum prediction variance, 357, 358, 362
- maximum tolerated dose, 159
- mean response, 206, 207, 212
- mean squared error, 83, 433, 434
- mean-squared error of prediction, 209, 210, 212
- mean-squared-error properties of a design, 292
- mean-variance relationship, 173, 189
- measure of slope-rotatability over all directions, 417
- measure of slope-rotatability over axial directions, 416
- measures of *A*-rotatability, 444
- measures of axial<sub>all</sub>-rotatability, 444
- measures of *D*- and *E*-rotatability, 444
- measures of rotatability, 444
- measures of slope-rotatability, 444
- measures to assess the slope-rotatability, 409
- mechanistic model, 66, 173, 193, 297
- method of least squares, 428
- method of Newton, 92
- minimax approach, 435, 436
- minimax criterion, 432
- minimax problem, 432, 433
- minimax second-order design, 433
- minimax third-order design, 433
- minimum aberration, 304
- minimum bias & MSE designs, 289, 291
- minimum prediction variance, 357, 358, 362
- minor components, 293
- mirror image, *see* foldover
- mixed effects, 53
- mixed model, 22, 32, 34, 40, 43

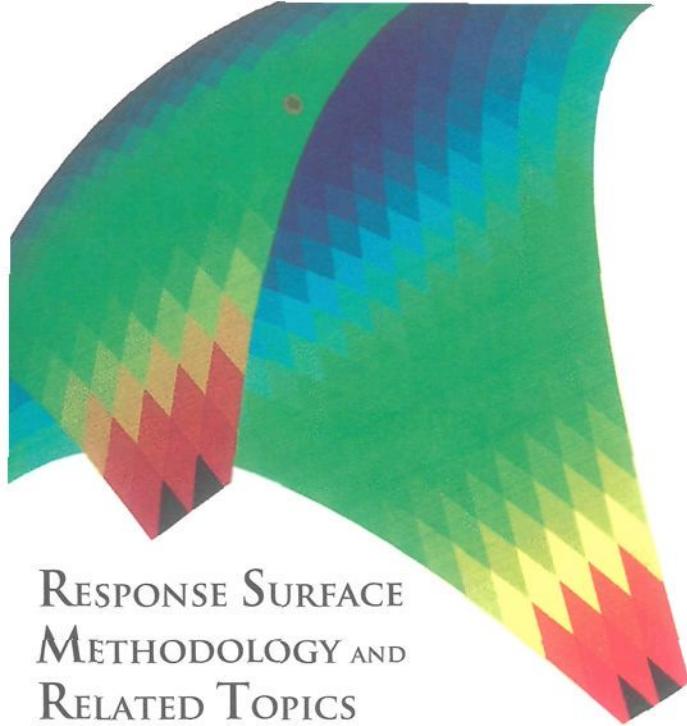
- mixed resolution, 147, 150  
mixture components, 299, 331  
mixture constraints, 285  
mixture experiment, 67, 74, 144, 329,  
    370, 371, 374  
mixture experiment bibliography, 284  
mixture experiment design, 283, 288  
mixture experiment models, 283, 286,  
    295  
mixture experiment problem, 284  
mixture experiment research, 283, 284  
mixture experiment research  
    chronology, 286  
mixture-amount design, 291, 292  
mixture-amount experiment, 285  
mixture-amount model, 296, 299  
mixture-of-mixtures design, 290, 293,  
    304  
mixture-of-mixtures model, 296  
mixture-process, 390  
mixture-process design, 375  
mixture-process experiment, 329  
mixture-process variable design, 289,  
    292, 304  
mixture-process variable experiment,  
    285  
mixture-process variable model, 296,  
    299  
ML estimators, 96  
model lack of fit, 300  
model matrix, 252, 254, 259, 262,  
    265, 271, 274, 350  
model reduction, 300, 301  
model robustness, 394  
model selection, 301  
model-free design, 304  
model-robust design, 291, 292, 304  
models homogeneous-of-degree-one,  
    296, 297  
models with inverse terms, 296, 298,  
    304  
models with logarithmic terms, 296,  
    298  
modified central composite design  
    (MCCD), 148, 150, 151  
modified L-pseudocomponent model,  
    297, 299  
modified small composite design  
    (MSCD), 150, 152  
moment matrix, 257, 259–261, 263,  
    350  
MPV experiments with noise factors,  
    302  
MPV model reduction, 302  
 $(M, S)$  criterion, 35  
MSE design criteria, 304  
multi-stratum designs, 19, 21, 39, 41,  
    44  
multicomponent constraints, 288  
multicriteria steepest ascent, 301  
multifactor mixture design, 290, 293,  
    304  
multifactor mixture model, 296  
multinomial response model, 297,  
    299, 305  
multiple comparison, 68  
multiple comparisons with a control,  
    77  
multiple comparisons with the best,  
    78  
multiple trace plots, 300  
multiple-response surface, 68, 79  
multiresponse experiments, 375  
multiresponse optimization, 301, 302  
 $MV$ -minimax optimality, 436  
 $MV$ -optimality, 436  
 $MV$ -rotatable, 437  
 $M_w$  criterion, 34  
near-rotatability, 268  
nested blocking, 39, 41  
noise factor (variable), 123–125, 129,  
    130, 135–138, 140, 142–148,  
    150–154  
noise factors, 90, 91, 99, 291, 301  
noise in mixture compositions, 302

- noise variable, 81, 82  
 non-normal response, 129, 140, 145  
 nonlinear models, 304  
 nonlinear programming, 133, 134, 144  
 nonlinear regression, 176  
 nonparametric regression, 84  
 nonrotatable design, 71  
 normalized covariance matrix, 429  
 Notz design, 360  
 number of support points, 183  
 oddly-shaped regions, 387  
 OLS estimator, 96, 98  
 optimal block design, 40  
 optimal coordinate plot, 73, 74  
 optimal design, 9, 288, 292, 304, 430,  
   432–434, 436, 443  
 optimal design criteria, 288, 292, 293  
 optimal designs of order  $d$ , 432  
 optimal operating condition, 133, 144  
 optimal response plot, 73, 74  
 optimal response trace, 75  
 optimal ridge trace, 80  
 optimal second-order design, 432,  
   433, 437  
 optimal simplex designs, 288  
 optimality, 354  
 optimality criteria, 355, 356, 358, 380  
 optimum experimental design, 174  
 optimum mixture, 285  
 orders of magnitude, 193  
 ordinary least squares, 51  
 orthogonal design, 124, 125  
 orthogonality, 24, 28, 256–258  
 orthogonally blocked, 33  
 overall confounding pattern, 6  
 overdispersion, 132, 142  
 overdispersion parameter, 211  
 overlaid contour plots, 301, 302  
 overlapping contour plot, 79  
 overparameterization, 298  
 paired-comparison experiments, 291  
 parameter sensitivity, 176, 194  
 parameter space, 210, 215  
 parameterized variance function, 174,  
   189  
 partial least squares, 297, 301, 302  
 partial quadratic mixture models,  
   297, 298  
 path of steepest ascent, 20  
 penalized prediction variance, 403  
 pharmaceuticals, 19  
 pharmacokinetic reaction, 195  
 physical relationship, 174  
 piecewise linear model, 296, 299  
 Poisson distribution, 175  
 Poisson regression, 401  
 Poisson response, 211, 212  
 polar coordinate transformation, 84  
 polyhedral region, 288  
 polynomial model, 295, 429  
 polynomial regression, 432  
 posterior predictive distribution, 81  
 power link, 178  
 power model, 296, 299  
 power studies, 24  
 power transformation, 174, 179  
 precision matrix, 350  
 precision of parameter estimates, 252,  
   253, 258, 262, 277  
 prediction bias, 208, 209  
 prediction bias trace plots, 294  
 prediction variance, 208, 252, 253,  
   256, 262, 263, 265–269, 272, 277,  
   351–356, 358, 363, 364, 368, 373,  
   375, 387  
 prediction variance quantile (PVQ),  
   349  
 prediction variance quantile (PVQ)  
   plots, 294, 356, 364–366, 368, 369,  
   375  
 prediction variance trace (PVT)  
   plots, 294, 372  
 principal components analysis, 301,  
   302  
 probit, 181  
 PROC GENMOD, 215

- process robustness, 364  
process variables, 285, 288, 299  
product of the design criteria, 190  
projection, 95, 96  
projection design, 288, 289, 300  
projective properties, 275  
proportion response model, 299  
proportional odds model, 225, 229  
pseudocomponent transformation,  
    300, 302  
pure error, 21, 24, 28
- quadratic loss function, 81  
quadratic models, 68  
quadratic nonlinear blending model,  
    295  
qualitative factors, 19  
quantal response data, 297  
quantile, 364, 368, 369  
quantile dispersion graphs (QDGs),  
    203, 205, 209, 210, 367, 368, 382  
quantile plots, 367  
quantiles, 210, 215  
    maximum, 215  
    minimum, 217
- radial design, 288, 290  
random effects, 22  
random factor (effect), 145, 146, 153,  
    154  
random run order, 49  
randomization, 19, 22, 41, 44, 48  
randomization restriction, 53, 304  
randomized complete block design  
    (RCBD), 153  
randomized incomplete block design  
    (RICBD), 153  
rank-based regression, 84  
ratio variable design, 288  
ratios of component proportions, 295,  
    296  
reference mixture, 298  
region moment matrix, 264, 268, 271  
region of interest, 390  
region of operability, 390  
regression over hypercubes, 433  
regression over hyperspheres, 432,  
    433, 436  
regression over spherical region, 438  
regular fractional replicates, 22  
Reinsel, Gregory, 18  
replication, 19, 44  
residual maximum likelihood  
    (REML), 43  
resolution, 29, 146, 147, 149, 150, 152  
resolution of design, 304  
resolved block designs, 31  
resource equation, 24  
response surface, 173  
response surface design, 203, 427–429,  
    431, 437, 443  
response surface design in  $q - 1$   
    independent variables, 289  
response trace plots, 300, 302  
response transformation, 191  
restricted maximum likelihood  
    (REML), 139, 142  
restricted region simplex design, 288,  
    290  
restricted simplex design, 290  
review publications, 283, 303  
ridge analysis, 68, 73, 74, 83, 84, 301  
ridge regression, 301  
ridge trace path, 69  
rising ridge, 71  
robust design, 291, 293  
robust parameter design (RPD), 89,  
    123, 124, 127–129, 137, 138, 140,  
    142, 143, 145–148, 150, 152–154,  
    390  
robustness, 106  
robustness to noise factors, 302  
rotatability, 263, 267, 351, 353, 431,  
    437, 443  
rotatable CCD, 359, 360, 364  
rotatable design, 71, 351, 411, 431,  
    432, 438, 439, 443, 444  
row-column design, 31, 37, 40

- RPD, 90, 91  
 RRO, 49  
 run size, 19
- saddle surface, 71, 76  
 scaled average slope variance (SASV), 364, 367  
 scaled mean-squared error of prediction (SMSEP), 368  
 scaled prediction variance, 262, 264, 330, 351, 380  
 scaled prediction variance quantile (SPVQ) plots, 374  
 scatterplot matrix, 293, 294  
 SCD, 365, 366, 369, 371  
 Scheffé canonical polynomial model, 296  
 Scheffé models with inverse terms, 296, 298, 304  
 Scheffé models with logarithmic terms, 296, 298  
 Scheffé quadratic model, 299  
 screening components, 300  
 screening design, 290, 292  
 search algorithm, 33  
 search direction, 91, 93, 96, 97  
 second-order design, 431, 433, 434, 436, 439, 442, 443  
 second-order model, 20, 23, 28, 42, 67, 89, 146–148, 150, 151, 429, 434, 443  
 second-order polynomial, 251–255, 257–263, 265, 269, 276, 277  
 seemingly unrelated regression model, 81  
 selecting optimal formulations, 301  
 sequential combination, 4, 11  
 sequential design, 16, 17, 226, 290  
 shape of the quadratic model, 71  
 shrinkage factor, 371, 372, 374, 375  
 shrinkage point, 371, 372  
 signal-to-noise ratio (SNR), 125–128, 133  
 simple orthogonal block structure, 31  
 simplex, 285, 370, 371  
 simplex coordinates, 294  
 simplex design, 288, 289  
 simplex-centroid design, 288, 289  
 simplex-lattice design, 288, 289, 300  
 simulated annealing, 36, 266, 278  
 simultaneous estimation, 192  
 single-component constraints, 288  
 slack variable model, 296  
 slope, 428, 429, 434, 435, 437, 443  
 slope-rotatability, 409, 413, 437, 438, 443, 444  
 slope rotatability over all directions, 415, 438  
 slope-rotatability over axial directions, 413  
 slope-rotatability with equal maximum directional variance, 420, 443  
 slope-rotatable design, 290, 443  
 slope-rotatable designs over all directions, 409  
 slope-rotatable designs over axial directions, 409  
 slope-rotatable designs with equal maximum directional variance, 409  
 slope-rotatable minimax design, 409, 421  
 slope variance dispersion graphs (SVDGs), 364  
 small composite design (SCD), 149, 151–153, 360, 364  
 small samples, 243  
 smaller the better, 83  
 space-filling, 278  
 space-filling design, 290, 292, 304  
 special-cubic model, 295  
 special-quartic model, 295  
 spherical prediction variance, 356, 357, 360, 366  
 spherical region, 436  
 split plot design (SPD), 31, 41, 128, 153, 154, 391  
 split-plot mixture-process variable design, 291, 292

- split-plot mixture-process variable experiment, 301, 302  
standard polynomials model, 296  
stationary point, 71, 76  
stationary ridge, 71  
steepest ascent (descent), 68, 69  
stepwise regression, 305  
stepwise variable selection, 23  
stopping rule, 98  
strata, 41  
structured parameter, 173, 194  
subset designs, 27  
superblocks, 40  
symmetric balanced design, 430, 431, 438, 441  
symmetric-simplex design, 290
- t*-distributed residual error, 81  
*T*-optimal, 273  
Taguchi, 124–128, 133  
target value, 83  
target-is-best, 90, 91  
test statistics for mixture models, 300, 302  
third-order design, 431, 433, 443  
third-order model, 429  
third-order rotatable design, 431  
three-dimensional VDGs, 331, 340  
three-level designs, 19  
three-level factors, 30  
tolerance range, 98, 102  
topics for future research, 304  
total amount of the mixture, 285, 288  
transforming both sides, 193  
treatment contrast, 427  
treatments, 21  
trinomial response, 225  
trivial block structure, 31
- tuning factor, 126, 128, 136  
tutorial publications, 283, 303  
two-factor interactions, 1  
two-level factor, 26, 30, 33  
two-level factorial designs, 1, 29  
two-level fractional factorial, 1
- uniform design, 291, 292  
unrestricted minimax problem, 433
- $V_\rho$ , 356–360  
 $V$ -efficiency, 264  
 $V$ -optimal, 264, 265  
 $V$ -optimality, 381  
variable selection, 28  
variable selection methods, 305  
variance component, 367, 368  
variance dispersion graphs, 253, 263, 268, 269, 272, 277, 294, 330, 349, 356–358, 361, 363, 364, 366, 368, 369, 371, 372, 375, 382  
variance function, 431, 432  
variance inflation factor, 293, 294  
variance modeling, 130, 131  
variance of prediction, 179  
variance quantile plot, 294  
variance trace plot, 293  
variance-optimal design, 292  
vertices, 288  
 $VMAX_\rho$ , 357–360, 366  
 $VMIN_\rho$ , 357–359, 366
- weighted  $A$ -efficiency, 43  
weighted least squares, 130–133, 139, 142, 190, 200  
weighted least squares regression, 74  
whole plot, 49  
Wilkinson-Rogers notation, 31



## RESPONSE SURFACE METHODOLOGY AND RELATED TOPICS

This is the first edited volume on response surface methodology (RSM). It contains 17 chapters written by leading experts in the field and covers a wide variety of topics ranging from areas in classical RSM to more recent modeling approaches within the framework of RSM, including the use of generalized linear models. Topics covering particular aspects of robust parameter design, response surface optimization, mixture experiments, and a variety of new graphical approaches in RSM are also included. The main purpose of this volume is to provide an overview of the key ideas that have shaped RSM, and to bring attention to recent research directions and developments in RSM, which can have many useful applications in a variety of fields. The volume will be very helpful to researchers as well as practitioners interested in RSM's theory and potential applications. It will be particularly useful to individuals who have used RSM methods in the past, but have not kept up with its recent developments, both in theory and applications.