Statistica Sinica **11**(2001), 553-604

FACTOR SCREENING AND RESPONSE SURFACE EXPLORATION

Shao-Wei Cheng and C. F. J. Wu

Academia Sinica and University of Michigan

Abstract: Standard practice in response surface methodology performs factor screening and response surface exploration sequentially, using different designs. A novel approach is proposed to achieve the two objectives on the same experiment, based on one design. Running a uni-stage experiment has the advantages of saving experimentation time and run size. The approach is based on a two-stage analysis that employs factor screening, projection and response surface exploration. Projection-efficiency criteria are defined to evaluate the performance of the projected designs. The projection-efficiency properties of the 3^{n-k} designs and three nonregular designs are studied, and comparisons with central composite designs are made. Nonregular designs appear to enjoy better projection properties. The strategy is illustrated with the analysis of a PVC insulation experiment.

Key words and phrases: Central composite design, D-optimality, fractional factorial design, orthogonal array, projection, response surface methodology.

1. Introduction and Review

Statistical design and analysis of experiments is an effective and commonly used tool in scientific and engineering investigation. Experimentation is used to understand and/or improve a system (i.e., a product or process). A system can have a large number of factors but usually only a handful of them are important or significant. A screening experiment employs a design of economic run size to identify the important factors. Once the important factors are identified, the next stage of the investigation focuses on the elucidation and optimization of the relationship between the response and the factors. For continuous factors, this relationship is referred to as a response surface. *Response surface methodology* (RSM), as pioneered by George Box, is often employed for factor screening and response surface exploration.

The standard RSM can typically be described as consisting of two parts. First, it conducts an experiment to screen out unimportant factors. Typically it is based on a first-order design such as the 2^{n-k} fractional factorial designs or Plackett-Burman designs. Second, it conducts a more intensive study of the response surface, typically with fewer factors and over a smaller region. It is based on a second-order design such as the central composite designs. If the two experiments are over very different regions, the move from one region to the other can be achieved by using steepest ascent search or other search methods. If the two experiments are over the same region (or one is expanded over the other), star points are added to the first-order design to form a second-order design. This is called *sequential assembly*. Details on RSM can be found in books like Box and Draper (1987), Myers and Montgomery (1995), and Khuri and Cornell (1996).

RSM has been a very effective tool and has seen many successful applications. Its sequential nature can, however, be a disadvantage, especially when the experimental preparation is time-consuming or its duration is long. To illustrate the former, running experiments on a production line may require change of the work schedule, training of operators, and trial runs. For the latter, consider an important step in the fermentation process for biological products called medium optimization. Several ingredients are fed to the bacteria to "grow" the right metabolites as new drugs, which can take 7 to 28 days. Instead of conducting experiments sequentially, it would be better to perform factor screening and response surface exploration on the *same* experiment. In this paper we propose a novel approach that can achieve the twin objectives of screening and surface exploration by using a *single* design. A key step is the projection of a larger factor space onto a smaller factor space, which serves as a link between screening and response surface exploration. As this differs from the standard RSM, new concepts, theory and analysis strategy are called for. Here we assume that screening and surface exploration are performed in the same region. This holds, for example, if the region chosen for factor screening is not far from the curved part of the response surface.

We now give a brief review of existing work. The 2^{n-k} designs are most commonly used for screening. When there are additional degrees of freedom for entertaining the estimation of interactions, the *minimum aberration criterion* is used for selecting an optimal 2^{n-k} fraction. Maximizing the number of *clear* main effects and clear two-factor interactions can be used as a supplementary criterion. The 2^{n-k} and 3^{n-k} series of designs are called regular. A design is called *regular* if it can be constructed through the defining contrast subgroup among its factors. Otherwise it is a *nonregular* design. Two-level nonregular designs like the Plackett-Burman designs can also be used for factor screening. Details on these concepts can be found in Wu and Hamada (2000). When it is adequate to use the quadratic approximation to the response surface, it is common to adopt the following *second-order model* to describe the relationship between the response y and n predictors $x_1, \ldots, x_n$:

$$y = \mu + \sum_{i=1}^{n} \beta_i x_i + \sum_{i=1}^{n} \beta_{ii} x_i^2 + \sum_{1=i<j}^{n} \beta_{ij} x_i x_j + \epsilon, \tag{1}$$

where ϵ is the error term, β_i are the linear effects, β_{ii} the quadratic effects, and β_{ij} the linear-by-linear interactions. A design is called a *second-order design* if all $(n+1)(n+2)/2$ parameters in (1) are estimable in the design. The most popular second-order designs are the *central composite designs* (CCDs) (Box and Wilson (1951); Box and Hunter (1957)). For consistency in design comparisons, we assume the design region to be cuboidal. For the CCDs, this amounts to assuming a *face center cube*, where the star points are placed in the center of the faces of the cube. We choose CCDs with *one* center point in the efficiency comparisons in later sections. The run sizes, cube points, and efficiency values of the chosen CCDs are listed in Table 1. Apart from the CCDs, various second-order designs are available in the literature. Details can be found in the aforementioned books on RSM.

Table 1. CCDs for efficiency comparison (n = # of factors, q = # of parameters, N = run size).

n	q	N	cube points*	D_{eff}	G_{eff}
2	6	9	2^2	.974	.828
3	10	15	2^3	.942	.836
4	15	25	2^4	.911	.780
5	21	27	2_V^{5-1}	.841	.749
6	28	45	2_{VI}^{6-1}	.852	.625
7	36	79	2_{VII}^{7-1}	.845	.442

* with resolution at least V

The paper is organized as follows. The proposed approach is described in Section 2. Its implementation is worked out for 3^{n-k} designs in Sections 3 and 4. With the new perspective provided by the proposed approach, a reassessment is made in Section 4.3 of a key rationale in Box and Wilson's (1951) work. The implementation continues in Sections 5 and 6 for two nonregular designs, the $OA(18, 3^7)$ and the $OA(36, 3^{12})$, and for a 27-run nonregular design. Nonregular designs appear to enjoy better projection properties than regular designs and an explanation is given. In Section 7 the proposed strategy is illustrated with an experiment (Taguchi (1987)) on PVC insulation. Conclusions and remarks are made in Section 8.

2. A Novel Approach to Factor Screening and Response Surface Exploration

An approach is proposed in this section that can achieve the twin objectives of factor screening and response surface exploration in *one* design. It is based on the following *two-stage analysis*:

Stage 1. Perform factor screening and identify important factors.

Stage 2. Fit a second-order model for the factors identified in Stage 1.

Although it is a two-stage analysis, it consists of three parts: *screening* analysis in Stage 1, *projection* that links Stages 1 and 2, and *response surface exploration* in Stage 2.

The first stage is a generic step for screening factors. Various screening analyses can be utilized. The conventional method is based on main effects estimation or ANOVA. A factor is identified as important if its main effects (which have $s-1$ degrees of freedom for s levels) are significant. A straight main-effect analysis may not be appropriate if the interactions are large enough to bias the main effect estimates. In this case, a more elaborate analysis like the Bayesian method due to Box and Meyer (1993) should be employed. In their approach, each subset of factors is associated with a linear model that contains main effects and two-factor interactions for these factors. By comparing the posterior probabilities of these models, the important factors can be identified.

A key step in the new approach is the projection of a larger factor space onto a smaller factor space, which serves as a link between screening a larger number of factors and the more intensive study of the response surface over a smaller number of factors. The concept of projection has been used in the design literature. It is well known (Box, Hunter and Hunter (1978)) that, for a regular fractional factorial design of resolution R, its projection onto any $R-1$ factors is a full factorial. This projection property was extended to nonregular designs by Lin and Draper (1992) and Wang and Wu (1995). Chen (1998) considered projection properties of 2^{n-k}_R designs when projected onto dimensions higher than R.

2.1. Projection-efficiency criteria

Each of the three steps has its own objective and should be evaluated by a different criterion. For each objective, the corresponding design property is given as follows:

Analysis Step		Design Property
factor screening	$\leftrightarrow$	orthogonality
		maximum number of factors
projection	$\leftrightarrow$	eligible projection
response surface exploration	$\leftrightarrow$	estimation efficiency

For *factor screening*, a good design should be able to accommodate a large number of factors with relatively few runs and still possess orthogonality. For *projection*, we propose a design property called *eligible projection*. A *projected design* obtained after Stage 1 is said to be *eligible* if it is a second-order design. Otherwise, it is said to be *ineligible*. A design can be projected onto many different combinations of factors. For example, for a design with 13 factors, there are 286 projected designs onto three factors and 715 projected designs onto four factors. Because it is not known which of the factors will turn out to be important from factor screening, a *larger* number of eligible projected designs should be preferred. Based on the factor sparsity principle (i.e., the number of relatively important factors is small), an eligible projection for lower dimension is more important than that for higher dimension. Furthermore, by Theorem 2 to be given in Section 4.1, an ineligible lower-dimensional projected design causes any higher-dimensional projected designs that contain it to be ineligible. If a design has fewer ineligible projected designs on lower dimensions, it should have a better chance of getting fewer ineligible projected designs on higher dimensions. Typically dimensions 3 to 6 should be considered. For *response surface exploration*, we adopt the D- and G-criteria to assess estimation efficiencies of projected designs and other second-order designs. The D- and G-efficiencies (see Myers and Montgomery (1995), Section 8.2.1 for definitions), denoted by D_{eff} and G_{eff} respectively, compare the performance of a design with the corresponding optimal designs. These efficiency values are adjusted for sample size and scale so that they can be used to compare designs with different run sizes and experimental regions. For the comparison of two designs in terms of the D- or G-criterion, the relative efficiency, which is the ratio of their D- or G-efficiencies, is adopted.

Because orthogonal designs (i.e., whose factor main effects can be estimated orthogonally) are ideal for factor screening, we consider *only orthogonal designs* for the proposed approach. Based on the previous discussion, orthogonal designs can be compared by the following criteria.

(i) The number of eligible projected designs should be large, and lower-dimension projection is more important than higher-dimension projection.

(ii) Among the eligible projected designs the estimation efficiency as measured by criteria like D and G should be high.

Collectively these criteria are referred to as the *projection-efficiency criteria*. In order to estimate second order effects, factors of the chosen designs must have at least three levels. In the next four sections, we study the projection-efficiency properties of some important regular and nonregular three-level designs. The study can be extended for factors with more than three levels and higher order polynomials (or even spline functions).

3. Classification of Projected Designs of 3^{n-k} Designs: Combinatorial Isomorphism and Model Isomorphism

Classification of projected designs is the first step in the study of projection. For a 3^{n-k} design, its projected designs are still three-level fractional factorial designs. The following theorem gives the relationship between the defining contrast subgroups of the original design and its projected designs.

Theorem 1. *Let $\mathfrak{D}$ be a 3^{n-k} design of factors $x_1, \ldots, x_n$ with the defining contrast subgroup $\mathcal{G}$. Let $\mathfrak{D}_\mathbf{p}$ be the projected design of $\mathfrak{D}$ onto p factors. Then, $\mathfrak{D}_\mathbf{p}$ is either a fractional factorial design or a 3^p design with $3^{n-k}/3^p$ replicates. If $\mathfrak{D}_\mathbf{p}$ is a fractional factorial design, the defining contrast subgroup $\mathcal{G}_\mathbf{p}$ of $\mathfrak{D}_\mathbf{p}$ is the set of words in $\mathcal{G}$ that are formed only by letters representing the projected factors. If $\mathcal{G}$ does not include any of the words mentioned above, then $\mathfrak{D}_\mathbf{p}$ is a ($3^{(n-k)-p}$)-replicate of a 3^p design.*

The proof is straightforward.

An example is given to illustrate the theorem.

Example 1. Let $\mathfrak{D}$ be the 3^{6-3} design with the defining generators D = AB, E = AB^2C, F = AB^2C^2. Its defining contrast subgroup $\mathcal{G}$ is {**I**, ABD^2, CEF^2, AB^2CE^2, $AB^2C^2F^2$, AB^2EF, AC^2DE, ACDF, ADE^2F^2, $BCDE^2$, BC^2DF^2, BDEF, $ABCD^2EF^2$, $ABC^2D^2E^2F$}. Let $\mathfrak{D}_{\mathbf{p_1}}$ be the projected design of $\mathfrak{D}$ onto the factors B, C, D, E, and F. The words in $\mathcal{G}$ that contain only letters from {B, C, D, E, F} are {CEF^2, $BCDE^2$, BC^2DF^2, BDEF}. Therefore, $\mathfrak{D}_{\mathbf{p_1}}$ is a 3^{5-2} design with the defining relation $\mathbf{I} = CEF^2 = BCDE^2 = BC^2DF^2 = BDEF$. For another instance, let $\mathfrak{D}_{\mathbf{p_2}}$ be the projected design of $\mathfrak{D}$ onto any three factors except for the combinations {A, B, D} and {C, E, F}. Because there are no words in $\mathcal{G}$ that are formed by three letters except for ABD^2 and CEF^2, $\mathfrak{D}_{\mathbf{p_2}}$ is a 3^3 design.

A distinction is made between the *design matrix* and the *model matrix*. For a design with n factors and r runs, its design matrix is an $r \times n$ matrix with r rows for the experimental runs and n columns for the factors. The model matrix is a coded matrix for the design and has columns for each of the factorial effects. Suppose an $r \times n$ design matrix is used to fit a model with q parameters. Its model matrix is an $r \times q$ matrix whose columns represent the effects in the model.

Two designs are said to be *combinatorially isomorphic* if the design matrix of one design can be obtained from that of the other by permutations of rows, columns, and levels in the columns. For example, the two 3^{4-1} designs with the defining relation $\mathbf{I} = AB^2CD$ and $\mathbf{I} = AB^2CD^2$ are combinatorially isomorphic because the latter can be obtained from the former by interchanging the levels "0", "1" and "2" in column D of the design matrix with "0", "2" and "1" (i.e., changing D to D^2). Another criterion is *model isomorphism*. Two designs are said to be equivalent in terms of model isomorphism if the model matrix of one design can be obtained from that of the other by permutations of rows, columns and changes of signs in the columns. (The distinction between combinatorial isomorphism and model isomorphism is fundamental to the proposed approach.) For 2^{n-k} designs, combinatorial isomorphism is equivalent to model isomorphism.

For designs with more than two levels this equivalence does not hold, as can be seen in the following example.

Example 2. All 3_{IV}^{4-1} designs are combinatorially isomorphic, i.e., there is only one type of design with the defining relation $\mathbf{I}$ = ABCD. However, in terms of model isomorphism, there are two types of designs, labeled as 27-4.1 and 27-4.2. Design 27-4.1 is any 3^{4-1} design with its defining word taken from any word in the set

$$\mathcal{K}_1 = \{\text{ABCD, AB}^2\text{CD, ABC}^2\text{D, ABCD}^2\text{, AB}^2\text{C}^2\text{D}^2\}.$$

Design 27-4.2 is any 3^{4-1} design with its defining word taken from any word in the set

$$\mathcal{K}_2 = \{\text{AB}^2\text{C}^2\text{D, AB}^2\text{CD}^2\text{, ABC}^2\text{D}^2\}.$$

This is an interesting finding as it is a marked departure from traditional thinking which, by taking a combinatorial perspective, treats the designs in $\mathcal{K}_1$ and $\mathcal{K}_2$ as the same. From a geometric viewpoint, the two types of designs are different. For example, the collection of design points in $\mathcal{K}_1$ cannot be mapped to that in $\mathcal{K}_2$ by rotation and reflection through the origin of the design region. The model non-isomorphism between $\mathcal{K}_1$ and $\mathcal{K}_2$ can also be verified through the correlation matrices of the interactions for designs in $\mathcal{K}_1$ and $\mathcal{K}_2$ (see Table 2, where $x_A, \ldots, x_D$ are coded as $(-1, 0, 1)$ for levels $(0, 1, 2)$ respectively). The latter has a more serious collinearity among its linear-by-linear interactions than the former.

Table 2. Correlation matrix of interactions for designs in $\mathcal{K}_1$ and $\mathcal{K}_2$

$\mathcal{K}_1$							$\mathcal{K}_2$						
	x_Ax_B	x_Ax_C	x_Ax_D	x_Bx_C	x_Bx_D	x_Cx_D		x_Ax_B	x_Ax_C	x_Ax_D	x_Bx_C	x_Bx_D	x_Cx_D
x_Ax_B	1	0	0	0	0	0.25	x_Ax_B	1	0	0	0	0	0.5
x_Ax_C	0	1	0	0	0.25	0	x_Ax_C	0	1	0	0	0.5	0
x_Ax_D	0	0	1	0.25	0	0	x_Ax_D	0	0	1	0.5	0	0
x_Bx_C	0	0	0.25	1	0	0	x_Bx_C	0	0	0.5	1	0	0
x_Bx_D	0	0.25	0	0	1	0	x_Bx_D	0	0.5	0	0	1	0
x_Cx_D	0.25	0	0	0	0	1	x_Cx_D	0.5	0	0	0	0	1

In Example 2 (and throughout the paper), for consistency, we use the coding scheme $(0, 1, 2) \rightarrow (-1, 0, 1)$ for linear effects. A different coding scheme would affect the numerical results and classification of designs.

4. Projection-Efficiency Properties for 3^{n-k} Designs with 27 Runs

In this section, we study the performance of 27-run 3^{n-k} designs under the projection-efficiency criteria. Because $n - k = 3$ for 27 runs, we use $3^{n-(n-3)}$ to denote these designs. Only designs with resolution III or higher are considered.

4.1. Eligible projections and estimation efficiencies

Suppose the projection of a 3^{n-k} design onto three factors is a 3^{3-1} design with the defining relation **I**=ABC. This projection results in an ineligible design because there are only nine distinctive runs in the projected design while there are ten parameters in a second-order model with three factors. With the help of the following theorem, it can be further concluded that any 3^{n-k} design with resolution III is ineligible because it contains one nine-run three-factor projected design.

Theorem 2. *If a p-factor projected design of a design with n factors is ineligible, then the n-factor design is also ineligible.*

Proof. Let $\mathbf{X_p}$ and $\mathbf{X}$ be the model matrix of the p-factor projected design and the original n-factor design, respectively. Then $\mathbf{X_p}$ is a submatrix of $\mathbf{X}$ with the same number of rows. Because the p-factor design is ineligible, $\mathbf{X_p}'\mathbf{X_p}$ is singular, which implies that $\mathbf{X}'\mathbf{X}$ is singular. Therefore, the original n-factor design is ineligible.

Noting that a three-letter word in the defining contrast subgroup causes the design to be ineligible, we refer to this as the *curse of three-letter words*.

Let p denote the number of factors identified in factor screening. For p=1, 2, and 3, there is only one type of eligible projected design. For $p = 1$, it is a nine-replicate of the 3^1 design, labeled as 27-1; for $p = 2$, it is a three-replicate of the 3^2 design, labeled as 27-2; and for $p = 3$, it is the 3^3 design, labeled as 27-3. For $p = 4$, in terms of combinatorial isomorphism, there is only one design, i.e., the 3^{4-1} design with $\mathbf{I}$ = ABCD. In terms of model isomorphism, there are two types of eligible projected designs, 27-4.1 and 27-4.2 noted in Example 2. For $p \geq 5$, any projected design has resolution III and therefore is ineligible.

Consider next the number of eligible projections. There are $\binom{n}{p}$ projections of a $3^{n-(n-3)}$ design $\mathfrak{D}$ onto p factors. All projections are eligible when $p = 1$ or 2. For $p = 3$, by Theorem 1, each three-letter defining word corresponds to a set of three factors onto which the projection is ineligible. Therefore, the number of three-letter words in the defining contrast subgroup equals the number of ineligible three-factor projected designs. For $p = 4$, any projected design must be a 3^{4-1} design. For this design to be eligible, its defining word cannot be a three-letter word. It cannot have five letters as there are only four factors. Therefore, the defining word for an eligible projected design must have four letters and belong to either $\mathcal{K}_1$ or $\mathcal{K}_2$. This proves that the number of eligible projected designs for $p = 4$ equals the number of four-letter words in the defining contrast subgroup. The results are summarized in the following theorem.

Theorem 3. *For* $p = 3$*, the total number of eligible projected designs equals* $\binom{n}{3} - A_3(\mathfrak{D})$*, where* $A_3(\mathfrak{D})$ *is the number of three-letter words in the defining contrast subgroup* $\mathcal{G}$ *for the original* $3^{n-(n-3)}$ *design. For* $p = 4$*, the total number of eligible projected designs is* $A_4(\mathfrak{D})$*, which is the number of four-letter words in* $\mathcal{G}$*. Furthermore, the total numbers of projected designs with types* 27-4.1 *and* 27-4.2 *are the numbers of words in* $\mathcal{G}$ *that belong to sets* $\mathcal{K}_1$ *and* $\mathcal{K}_2$*, respectively.*

Finally, we consider the efficiencies of eligible projected designs for second-order model fitting. Their D-, G-efficiencies and relative D-, G-efficiencies to the CCD are given in Table 3. Overall, they are quite efficient for second-order model fitting: at least 92.2% as efficient as the CCD in D-efficiency, and at least 71.3% as efficient as the CCD (except for the design 27-4.2) in G-efficiency.

Table 3. Efficiencies of eligible projected designs of $3^{n-(n-3)}$ designs.

Design	D_{eff}	G_{eff}	rel. D_{eff}	rel. G_{eff}
27-1	1	1		
27-2	.974	.828	1	1
27-3	.932	.727	.989	.713
27-4.1	.878	.556	.964	.713
27-4.2	.840	.417	.922	.535

4.2. Optimal $3^{n-(n-3)}$ designs

With the help of Theorem 3, we can find the $3^{n-(n-3)}$ designs with the best projection properties. In order to increase the number of eligible three-factor projected designs, choose a design with the smallest $A_3(\mathfrak{D})$. If several designs have the same smallest value, then choose one with the largest $A_4(\mathfrak{D})$ to maximize the total number of eligible four-factor projected designs. (Note that the first step is the same as the minimum aberration criterion while the second step is the *reverse* of the same criterion.) From Table 6 of Chen, Sun and Wu (1993), it can be verified that the minimum aberration $3^{n-(n-3)}$ designs happen to have the smallest $A_3(\mathfrak{D})$ and the largest $A_4(\mathfrak{D})$, but this is not generally true. Hence, for $p = 3$ and 4, the minimum aberration $3^{n-(n-3)}$ designs perform the best in terms of the eligible projection criterion.

For the minimum aberration $3^{n-(n-3)}$ designs, the total numbers of the corresponding types of projected designs and the frequencies (= total/$\binom{n}{p}$) are given in Table 4. The frequency can be interpreted as the chance of obtaining the corresponding type of projected designs when factors are *randomly* assigned to the columns of the design matrix. Sometimes, prior information or experience may suggest that some factors are more important than others. It is then desirable to assign factors to columns in such a way that the projection onto the more important factors is eligible. For example, suppose three factors are regarded as more important. By Theorem 1, in order to make the projection onto the three factors eligible, they should be assigned to columns that do not form any three-letter defining words in the original design. For the design in Example 1, there are two three-letter defining words: ABD^2 and CEF^2. Therefore, the three factors cannot be assigned to columns {A, B, D} or {C, E, F}. For four factors, suppose we want the projected designs to be of the type 27-4.1 (which has better *D*- and *G*-efficiencies than 27-4.2). By Theorem 1, the four factors should be assigned to columns that form a four-letter word of the type $\mathcal{K}_1$. There are six such four-letter words: $AB^2C^2F^2$, AB^2EF, AC^2DE, ACDF, $BCDE^2$, and BDEF. Therefore, the four factors can be assigned to columns {A, B, C, F}, {A, B, E, F}, {A, C, D, E}, {A, C, D, F}, {B, C, D, E}, or {B, D, E, F} so that the projected designs are eligible and have high efficiency.

Table 4. Total numbers and frequencies of projected designs of minimum aberration $3^{n-(n-3)}$ designs.

	number of factors, n									
	4		5		6		7		8	
	Total	Freq.	Total	Freq.	Total	Freq.	Total	Freq.	Total	Freq.
27-3	4	1.00	9	.90	18	.90	30	.86	48	.86
ineligible	0	.00	1	.10	2	.10	5	.14	8	.14
27-4.1	1	1.00	2	.40	6	.40	10	.29	19	.27
27-4.2	0	.00	1	.20	3	.20	5	.14	11	.16
ineligible	0	.00	2	.40	6	.40	20	.57	40	.57

	number of factors, n									
	9		10		11		12		13	
	Total	Freq.	Total	Freq.	Total	Freq.	Total	Freq.	Total	Freq.
27-3	72	.86	99	.83	135	.82	180	.82	234	.82
ineligible	12	.14	21	.18	30	.18	40	.18	52	.18
27-4.1	33	.26	36	.17	60	.18	90	.18	126	.18
27-4.2	21	.17	36	.17	48	.15	72	.15	108	.15
ineligible	72	.57	138	.66	222	.67	333	.67	481	.67

4.3. A historical note

In their seminal paper, Box and Wilson (1951) gave two reasons for not using 3^{n-k} designs as second-order designs:

> "For example, for the 3^n designs[1] , no useful fractional factorial exists when $n = 3$, so that it would be necessary to carry out all the 27 experiments of the complete factorial design in order to determine the 10 effects of second-order or less. Similarly when $n = 4$, 81 experiments are needed to determine 15 effects. When $n = 5$ a one-third replicate may be used to determine the 21 effects of order 2 and less; but even this involves the carrying out of 81 experiments,..."

Their argument is correct only for $n = 3$ and $n = 5$. For $n = 3$, a 3^3 design (27 runs) is necessary for fitting a second-order model with 10 parameters. For $n = 5$, a 3^{5-1} design (81 runs) is needed for fitting a second-order model with 21 parameters. On the other hand, a CCD only needs 15 runs for $n = 3$ and 27 runs (using $\mathbf{I}$ = ABCDE for cube points) for $n = 5$. Therefore, for $n = 3$ and $n = 5$, the run size of 3^{n-k} designs is too large for fitting a second-order model. Hence, 3^{n-k} designs become unattractive because too many degrees of freedom are allocated for the estimation of higher-order interactions, or alternately for error variance estimation. For $n = 4$, they stated that a 3^4 design (81 runs) is necessary for fitting a second-order model with 15 parameters. This is *not true* as any 27-run 3^{4-1} design with resolution IV, such as 27-4.1 and 27-4.2, can be used for this purpose. A CCD needs 25 runs, which is close to 27. Their statement was apparently based on the traditional approach that classified 3^{n-k} designs in terms of combinatorial isomorphism and insisted that designs need to have resolution at least V.[2] However, this approach only holds for 2^{n-k} designs. For 3^{n-k} designs, although a defining word like $\mathrm{ABC^2D^2}$ causes the component AB (of the A×B interaction) to be aliased with the component CD (each component having two degrees of freedom), the two linear-by-linear interactions $x_A x_B$ and $x_C x_D$ (each having one degree of freedom) are not aliased. Further discussion along these lines can be found in Wu and Hamada (2000, Section 5.6).

The main credible reason for not using 3^{n-k} designs in the Box and Wilson argument is their inefficient run size. However, this argument is only relevant when 3^{n-k} designs are used for fitting quadratic response surfaces. With the two-stage analysis strategy, it no longer applied. In the screening step, the degrees of freedom of 3^{n-k} designs can be efficiently utilized to screen a large number of factors. For example with 27 runs, up to 13 factors can be studied. Furthermore, when the number of important factors is not too large, we can fit a quadratic model for these factors with good efficiency.

[1] Box and Wilson (1951) use k to denote the number of factors. For consistency with the symbols in the paper, we replace k by n.

[2] Designs in recent work (e.g., Draper and Lin (1990)) do not require resolution V by using Plackett-Burman designs for the cube points.

5. Projection-Efficiency Properties for Nonregular Orthogonal Designs

Many nonregular designs (see Chapter 7 of Wu and Hamada (2000)) have economic run size and are more flexible in accommodating factors with different levels. In this section, two popular nonregular designs, $OA(18, 3^7)$ (Table 7C.2, Wu and Hamada (2000)) and $OA(36, 3^{12})$ (Table 7C.6, Wu and Hamada (2000)) are studied, where OA stands for orthogonal array. For these two orthogonal arrays, the classification of projected designs is quite complicated. Compared with 3^{n-k} designs, there is no theoretical tool like the defining contrast subgroup to describe the structure of designs. Therefore, no theoretical results are available for the classification of projected designs. Computer search is used instead.

5.1. The $OA(18, 3^7)$ design

With 27 runs, the $3^{n-(n-3)}$ designs can screen at most 13 factors. For fewer factors, a smaller orthogonal design like the $OA(18, 3^7)$ in Table 11 can be used to study up to seven factors.

By computer search, different p-factor projected designs are classified in terms of combinatorial and model isomorphisms. For $p = 1$ and 2, there is only one type of projected design. For $p = 3$ and 4 there are three and four types, respectively, in terms of combinatorial isomorphism, and six and seven types, respectively, in terms of model isomorphism. A detailed description of these projected designs is given in Appendix 1. (The labels for the projected designs in Tables 5 and 6 are taken from Appendices 1 and 2.) Among the projected designs, only one three-factor and one four-factor projected design are ineligible. For five or more factors, a second-order model has at least 21 parameters, which is larger than 18. Therefore, all projected designs of five or more factors in the $OA(18, 3^7)$ are ineligible.

From Table 5, the frequency of eligible projections is 100%, 100%, 97%, and 89% for p= 1, 2, 3, and 4, respectively. In terms of eligible projection, the $OA(18, 3^7)$ is much better than the $3^{n-(n-3)}$ designs, even though the latter has nine more runs. Suppose seven factors are under study. For the $3^{7-(7-3)}$ designs, 86% of the three-factor projected designs and 43% of the four-factor projected designs are eligible, while for the $OA(18, 3^7)$, the corresponding frequencies are 97% and 89%. For n factors, $n \leq 6$, an n-column submatrix of the $OA(18, 3^7)$ can be chosen for use and have 100% eligible three- and four-factor projected designs, while for $3^{n-(n-3)}$ designs with n= 5 and 6, only 90% are eligible for $p = 3$ and 60% for $p = 4$. The reason that a nonregular design like the $OA(18, 3^7)$ with a smaller size can outperform regular designs in terms of the eligible projection property will be explained in the next section.

The D- and G-efficiencies of eligible projected designs are given in Table 5. The one- and two-factor projected designs have the same D- and G-efficiencies as the 27-1 design and the 27-2 design, respectively, because the 18-1 (18-2) design has the same support and weights as the 27-1 (27-2) design. For $p = 3$ and 4, the estimation efficiencies of eligible projected designs of the $OA(18, 3^7)$ are not as good as those of the $3^{n-(n-3)}$ designs. For p=3, when compared with the CCD, the eligible projected designs are 82.59% – 97.03% as efficient in D-efficiency and 22.85% – 72.49% as efficient in G-efficiency. For $p = 4$, the projected designs are 68.17% – 82.77% as efficient as the CCD in D-efficiency and only 9.10% – 28.97% as efficient in G-efficiency.

Table 5. Total numbers, frequencies, and efficiencies of projected designs of the $OA(18, 3^7)$.

Design	total	freq.	D_{eff}	G_{eff}	rel. D_{eff}	rel. G_{eff}
18-1	7	1.00	1	1		
18-2	21	1.00	.974	.828	1	1
18-3.1.1	24	.69	.890	.476	.945	.569
18-3.1.2	4	.11	.865	.318	.918	.380
18-3.2.1	2	.06	.914	.606	.970	.725
18-3.2.3	2	.06	.788	.191	.837	.228
18-3.2.4	2	.06	.778	.215	.826	.257
18-3.3	1	.03	ineligible			
18-4.1	15	.43	.736	.226	.808	.290
18-4.2.1	4	.11	.754	.154	.828	.197
18-4.2.2	4	.11	.664	.100	.729	.128
18-4.2.3	2	.06	.663	.092	.728	.118
18-4.2.4	2	.06	.655	.128	.719	.164
18-4.3	4	.11	ineligible			
18-4.4	4	.11	.621	.071	.682	.091

Although the $OA(18, 3^7)$ does not perform well in terms of the D- and G-criteria, it has other virtues. First, and a noteworthy one, is its run size economy. With 18 runs, it can screen seven factors and, after projection, perform as a second-order design for four factors. It has only three more runs than the three-factor CCD and seven fewer runs than the four-factor CCD. For four factors, its run size is close to being saturated (because the second-order model has 15 parameters).

5.2. The $OA(36, 3^{12})$ design

Suppose that five or more factors are expected to be important in factor screening. Neither the 27-run 3^{n-k} designs nor the $OA(18, 3^7)$ are suitable because they do not contain any eligible projected designs of five or more factors. In this situation, the $OA(36, 3^{12})$ in Table 12 is a viable option.

Table 6. Total numbers, frequencies and efficiencies of projected designs of the $OA(36, 3^{12})$

Design	total	freq	D_{eff}	G_{eff}	rel. D_{eff}	rel. G_{eff}
36-1	12	1.00	1	1		
36-2	66	1.00	.974	.828	1	1
36-3.1.1	96	.44	.922	.559	.979	.669
36-3.1.2	96	.44	.919	.663	.976	.793
36-3.2	12	.05	.890	.476	.945	.569
36-3.3.1	4	.01	.864	.362	.917	.433
36-3.3.2	12	.05	.829	.403	.880	.482
36-4.1	72	.15	.855	.417	.939	.535
36-4.2	72	.15	.850	.444	.933	.569
36-4.3	24	.05	.848	.365	.931	.468
36-4.4	48	.10	.837	.411	.919	.527
36-4.5	6	.01	.834	.452	.915	.579
36-4.6	72	.15	.830	.289	.911	.371
36-4.7	48	.10	.829	.342	.910	.438
36-4.8	36	.07	.803	.229	.881	.294
36-4.9	24	.05	.798	.362	.876	.464
36-4.10	12	.02	.796	.375	.873	.481
36-4.11	6	.01	.791	.273	.868	.350
36-4.12	72	.15	.786	.270	.863	.346
36-4.13	3	.01	.736	.226	.808	.290
36-5 eligible	792	1.00	.631∼.770	.102∼.365	.750∼.916	.136∼.487
36-6 eligible	895	.97	.488∼.627	.060∼.125	.573∼.736	.016∼.234
36-6 ineligible	29	.03				
36-7 eligible	348	.44	.327∼.416	.001∼.011	.387∼.492	.000∼.025
36-7 ineligible	444	.56				

The classification of p-factor projected designs is done by computer search. For $p = 1$ and 2, there is only one type of projected design. For $p = 3$, in terms of combinatorial and model isomorphisms, there are three and five types, respectively. For $p \geq 4$, there is a greater number of designs. By an exhaustive search, different types of eligible projected designs (in terms of model isomorphism) are presented. For p= 4, 5, 6, and 7, there are 13, 36, 45, and 15 types, respectively. A detailed description of the projected design with $p \leq 4$ is given in Appendix

2. For $p \geq 8$, because of insufficient degrees of freedom, all projected designs are ineligible. The frequencies of having an eligible projected design are given in Table 6. For $p \leq 5$, all projected designs are eligible. For p= 6 and 7, the frequencies are 97% and 44%, respectively. Compared with the $3^{n-(n-3)}$ design and the $OA(18, 3^7)$, the $OA(36, 3^{12})$ can offer eligible projections of higher dimensions (up to seven). This would be useful if the system under study is complicated and has a larger number of important factors.

The estimation efficiency of eligible projected designs in the $OA(36, 3^{12})$ is compared with the CCD (see Table 6). For $p = 1$ and 2, the projected designs have the same supports and weights as the design 27-1 and the design 27-2, respectively. Therefore, the comparisons for $p = 1$ and $p = 2$ can be found in Section 4. For $p = 3$, 4, and 5, the eligible projected designs are nearly as efficient as the CCDs in D-efficiency. However, for $p = 6$ and 7, the eligible projected designs are less efficient than the CCDs (38.7% – 73.6%). In G-efficiency, the eligible projected designs are far less efficient than the CCDs, especially for $p = 5$, 6, and 7. It can be seen that, as p increases, the performance of eligible projected designs becomes worse. This is not surprising in view of the run sizes of the designs. As p increases, the run size of CCD also increases, while the run size of the projected designs is fixed at 36 for any p.

It is surprising to find that the $3^{n-(n-3)}$ designs perform better than the $OA(36, 3^{12})$ in terms of D- and G-efficiencies of their eligible projected designs. For example, the design 27-3 (resp. 27-4.1) has higher D- and G-efficiencies than any three-factor (resp. four-factor) projected design of the $OA(36, 3^{12})$. This indicates that the nine additional runs in the $OA(36, 3^{12})$ are not well allocated in the experimental region to allow the extraction of maximum information and to attain highest efficiencies.

6. A New Nonregular 27-Run Design and Its Projection-Efficiency Properties

As pointed out in Section 5, with nine more runs, the $3^{n-(n-3)}$ designs do not perform significantly better than the $OA(18, 3^7)$. In terms of the frequencies of eligible projections, their performance is even worse. What causes this poor performance? The explanation lies in the *curse of three-letter words*. All ineligible projected designs of the $3^{n-(n-3)}$ designs contain three-letter defining words that entail insufficient degrees of freedom for fitting a second-order model. It happens that, in the $OA(18, 3^7)$, the ineligible three- and four-factor projected designs also suffer from the curse. This observation leads us to consider alternative nonregular designs that are free from the curse and hence have a better eligible projection property. An example is given as follows.

Consider the minimum aberration $3^{8-(8-3)}$ design. Its defining generators are D = AB, E = ABC, F = AB^{2}C, G = AC^2, and H = BC^2. Because the design has 27 runs and independent factors A, B, and C, its design matrix can be given an alternative description. First, fix the 3^2 design given by A and B. Then form the 27-run design by considering the product between this design and C = $\mathbf{I}$, C = u, and C = u^2, where $\mathbf{I} = (0, 0, \ldots, 0)^T$, $u = (1, 1, \ldots, 1)^T$, and $u^2 = (2, 2, \ldots, 2)^T$, respectively. Among the remaining generators, D = AB is unaffected by the choice among C = $\mathbf{I}$, u, or u^2; E = ABC equals AB, ABu, or ABu^2, and G = AC^2 equals A, Au^2, or Au, depending on C = $\mathbf{I}$, u, or u^2. A similar construction is applied to the other factors. Then the particular $3_{III}^{8-(8-3)}$ design can be regarded as a *parallel-flat design* (Srivastava, Anderson and Mardekian (1984)) which is a combination of three 3_I^{8-6} designs. Based on this observation, we propose to construct a nonregular $3^{8-(8-3)}$ design by combining three different 3^{8-6} designs in such a way that the design has no defining words of length three and still retains the orthogonality property.

A design that satisfies these properties is given in Table 7 and is denoted by 3_{NR}^{8-5}, where NR stands for "nonregular". A key idea in the construction is to break the aliasing caused by a three-letter word by exchanging or modifying *cells*. (Each cell in this case is a 9×1 vector, such as AB, A^{2}B$^2u^2$, ... in Table 7.) For example, suppose all the u letters in the columns G and H are dropped, and denote the resulting columns by G′ and H′, respectively. Then G′ is aliased with AB^2 and H′ aliased with A^{2}B. These aliasing relations result in three-letter words. However, by replacing the second cell of G′ by AB^2u and the third cell of H′ by A^{2}Bu^2, as in Table 7, the aliasing relations G′ = AB^2 and H′ = A^{2}B are broken while orthogonality is retained in columns G and H. Another example is the columns E and F. If the two middle cells A^{2}B$^2u^2$ and AB are interchanged and the resulting columns are denoted by E′ and F′ respectively, E′ is then aliased with AB. The aliasing relation brings the curse of three-letter word, which can be broken by exchanging the two middle cells in the columns E′ and F′. A general construction of these nonregular designs can be found in Cheng (1999).

Table 7. An example of 27-run nonregular design.

A B	C	D	E	F	G	H
3^2	A	B	AB	A^{2}B^2	AB^2	A^{2}B
3^2	Au	Bu^2	A^{2}B$^2u^2$	AB	AB^2u	A^{2}B
3^2	Au^2	Bu	AB	A^{2}B^2u	AB^2	A^{2}Bu^2

For the 3_{NR}^{8-5} design in Table 7, all the linear and quadratic main effects are orthogonal. Surprisingly, all its three- and four-factor projected designs are eligible. When projected onto five factors, only *one* out of 56 projected designs is ineligible. By contrast, for the minimum aberration $3^{8-(8-3)}$ design, only 86% of the three-factor projected designs and 43% of the four-factor projected designs are eligible (see Table 4), and all five-factor projected designs are ineligible. It is clear that the eligible projection property is improved significantly in the 3_{NR}^{8-5} design. For $p = 6$, the degrees of freedom (27) is insufficient to fit a second-order model (which has 28 parameters). Therefore, no eligible projected design exists. However, by adding *one* more run to the 3_{NR}^{8-5} design, some six-factor projected designs become eligible.

A brief summary of efficiency values for the eligible projected designs in the 3_{NR}^{8-5} is given in Table 8. It includes the ranges of D- and G-efficiencies of eligible projected designs and the ranges of their relative efficiencies to the CCDs. The corresponding values for the $3^{n-(n-3)}$ designs are also given in Table 8 for comparison. For p=3 and 4, the ineligible projected designs of the $3^{n-(n-3)}$ designs are also included in the table and indicated by 0 efficiency value. In D-efficiency, the projected designs of the 3_{NR}^{8-5} design are nearly as efficient as the CCDs for $p = 3$ and 4, and less efficient for $p = 5$. Their performance in G-efficiency is less satisfactory. Compared with the $3^{n-(n-3)}$ designs, the D- and G-efficiencies of the projected designs of the 3_{NR}^{8-5} design is significantly better. This is because the $3^{n-(n-3)}$ designs have many ineligible projected designs that hurt performance. Only one particular 27-run nonregular design is studied in this section. There are many other nonregular designs with similar properties. They will be reported elsewhere.

Table 8. Summary of efficiency values for the 3_{NR}^{8-5} design.

	D_{eff}		G_{eff}		rel. D_{eff}		rel. G_{eff}	
p	3_{NR}^{8-5}	$3^{n-(n-3)}$	3_{NR}^{8-5}	$3^{n-(n-3)}$	3_{NR}^{8-5}	$3^{n-(n-3)}$	3_{NR}^{8-5}	$3^{n-(n-3)}$
3	.762-.932	.000-.932	.145-.727	.000-.727	81%-99%	00%-99%	17%-87%	00%-87%
4	.592-.878	.000-.878	.065-.556	.000-.556	65%-96%	00%-96%	8% -71%	00%-71%
5	.506-.684	none	.015-.149	none	60%-81%	none	2% -20%	none

7. An Illustrative Example

The proposed design and analysis strategy is illustrated with a 27-run experiment (Taguchi (1987, p.423)) to study the PVC insulation for electric wire. The objective of the experiment is to understand the compounding method of plasticizer, stabilizer, and filler for avoiding embrittlement of PVC insulation, and to find the most suitable process conditions. All nine factors are continuous and their levels are chosen to be equally spaced. Among the factors, two are about plasticizer: DOA (denoted by A) and n-DOP (B); two about stabilizer:

Tribase (C) and Dyphos (D); three about filler: Clay (E), Titanium white (F), and Carbon (G); the remaining two about process condition: number of revolutions of screw (H) and cylinder temperature (J). The measured response is the embrittlement temperature. The design matrix and data are given in Table 9.

Table 9. Design matrix and response data, PVC insulation data.

run	A	B	C	D	E	F	G	H	J	response
1	0	0	0	0	0	0	0	0	0	5
2	0	0	0	0	1	1	1	1	1	2
3	0	0	0	0	2	2	2	2	2	8
4	0	1	1	1	0	0	0	2	2	−15
5	0	1	1	1	1	1	1	0	0	−6
6	0	1	1	1	2	2	2	1	1	−10
7	0	2	2	2	0	0	0	1	1	−28
8	0	2	2	2	1	1	1	2	2	−19
9	0	2	2	2	2	2	2	0	0	−23
10	1	0	1	2	0	1	2	0	1	−13
11	1	0	1	2	1	2	0	1	2	−17
12	1	0	1	2	2	0	1	2	0	−7
13	1	1	2	0	0	1	2	2	0	−23
14	1	1	2	0	1	2	0	0	1	−31
15	1	1	2	0	2	0	1	1	2	−23
16	1	2	0	1	0	1	2	1	2	−34
17	1	2	0	1	1	2	0	2	0	−37
18	1	2	0	1	2	0	1	0	1	−29
19	2	0	2	1	0	2	1	0	2	−27
20	2	0	2	1	1	0	2	1	0	−27
21	2	0	2	1	2	1	0	2	1	−30
22	2	1	0	2	0	2	1	2	1	−35
23	2	1	0	2	1	0	2	0	2	−35
24	2	1	0	2	2	1	0	1	0	−38
25	2	2	1	0	0	2	1	1	0	−39
26	2	2	1	0	1	0	2	2	1	−40
27	2	2	1	0	2	1	0	0	2	−41

This design is a regular 3^{9-6} design with $C = AB$, $D = A^2B$, $F = AE$, $G = A^2E$, $H = B^2E$, and $J = AB^2E$ (with A, B, and E as generators). There are 15 three-letter words and 42 four-letter words in its defining contrast subgroup. By Theorem 3, 69 ($= \binom{9}{3} - 15$) three-factor projected designs and 42 four-factor projected designs are eligible. Based on the projection-efficiency criteria, this design is not optimal. A better choice would be to use the minimum aberration 3^{9-6} design (see Table 5A.2 of Wu and Hamada (2000)), in which 72 three-factor and 54 four-factor projected designs are eligible.

In Taguchi's analysis, only the factor main effects were of concern; in other words, the analysis was performed for factor screening. In the ANOVA analysis, factors A, B, C, D, and G are identified as significant, with respective p-values 0.00000, 0.00000, 0.00178, 0.00638, and 0.00113. Among them, A and B are more significant than C, D, and G.

In the two-stage analysis, a second-order response surface is fitted after factor screening. Based on the discussion in Section 4, there is no eligible projected design of five factors in the 3^{9-6} design. Therefore, it is impossible to fit a second-order model for A, B, C, D, and G. Factors A and B are the first two factors that should be considered because they are most significant. After A and B are chosen, neither C nor D can be included because the defining relations $C = AB$ and $D = A^2B$ make the projected designs on $\{A, B, C\}$ or $\{A, B, D\}$ ineligible. Because there is no three-letter defining word formed by A, B, and G, the projected design on A, B, and G can be used to fit a second-order model, yielding the results in Table 10 with $R^2 = 0.9811$. (In the analysis, the three levels 0, 1, and 2 in Table 9 are converted into -1, 0, and 1, respectively.) The significance of x_A, x_B, x_G, and x_G^2 is consistent with the finding from factor screening. An interesting discovery is that the linear-by-linear interaction x_Ax_B between A and B is significant. Because $C = AB$ and $D = A^2B$, x_Ax_B is partially aliased (i.e., correlation not equal 0 or ±1) with the linear and quadratic main effects of C and D, and thus can be expressed as a linear combination of these four effects. The conclusion in factor screening on C and D may be due to the significance of x_Ax_B. This is confirmed by further analysis, in which neither C or D is significant once x_Ax_B is included in the model. This finding is further supported by Taguchi's comment (1987, p.428). After identifying the five factors as important, Taguchi commented that: "It is only natural that A and B are significant since both were originally compounded as plasticizers by which to improve the embrittlement temperature, ... It was not expected that stabilizers C and D would influence the embrittlement temperature ... That filler G would influence the embrittlement temperature had been anticipated..." It seems more reasonable to conclude, either based on statistical analysis or specialist's opinion, that x_Ax_B is significant and the significance of C and D is a faulty conclusion caused by the omission of interactions in model fitting. A second-order fitted model can be obtained by using the estimates in Table 10. Based on this model, further studies of the response surface (e.g., canonical analysis, identification of stationary point, or contour plots) can be performed.

Table 10. Least squares estimates, t Statistics and p-values, PVC insulation data.

Effect	Estimate	Standard Error	t	p-value
intercept	−22.78	1.22	−18.62	0.0000
x_A	−12.56	0.57	−22.17	0.0000
x_A^2	1.67	0.98	1.70	0.1076
x_B	−10.22	0.57	−18.05	0.0000
x_B^2	2.00	0.98	2.04	0.0573
x_G	1.94	0.57	3.43	0.0032
x_G^2	−3.50	0.98	−3.57	0.0024
x_Ax_B	4.08	0.69	5.89	0.0000
x_Ax_G	−0.50	0.69	−0.72	0.4808
x_Bx_G	−0.08	0.69	−0.12	0.9058

Comparing the results of Taguchi's analysis and the two-stage analysis, the latter goes further to explore the interactions and reveals some important information. Suppose that the experimenter used the conclusion of the former analysis and performed a central composite design to study the response surface of the five identified factors. It would require 27 runs and the experimenter might find that no effects of C and D are significant. Without adding more runs, the two-stage analysis identifies an important interaction x_Ax_B, explains why the significance of C and D may be caused by their partial aliasing with x_Ax_B, and provides a response surface equation for A, B, and G. It is interesting to note that, even though the projected design on A, B, C, D, and G is not eligible, the projection onto A, B, and G still provides valuable information. It shows that an eligible projection on some of the important factors may suffice for building an appropriate model, especially when the main effects and interactions are partially aliased.

8. Conclusions and Remarks

By reducing the experimentation from two stages to a single stage, the proposed strategy has the advantage of saving time and possibly run size. The D-efficiencies of the projected designs are generally quite good and the G-efficiencies are worse. Since the proposed use of designs is multi-objective, one cannot expect its efficiency for fitting second-order models to be as good as the central composite designs. The CCDs cannot be used for factor screening, though.

Traditionally the research for factor screening and for response surface exploration proceed on separate lines and rarely interact. The former involves concepts like resolution, minimum aberration, and the number of clear effects, and the latter involves concepts like rotatability, D-optimality, and prediction variance. Considering these two lines of work in a unified framework has led to the development of new concepts like eligible projection, model isomorphism and projection-efficiency criteria, and may inspire further advances in design theory and experimentation strategy. In the paper the projection-efficiency properties are studied only for designs with 27 runs (one regular and one nonregular), 18 runs and 36 runs. The same study can be extended to other designs with larger runs and/or more levels. The superiority of the 27-run nonregular design in terms of eligible projections is encouraging. It suggests a new direction for research, namely, to find nonregular three-level designs with 18, 27, 36, 45 or 54 runs. Recently the minimum aberration criterion has been extended to nonregular designs (Deng and Tang (1999); Tang and Deng (1999); Xu and Wu (2001)). The proposed projection-efficiency criteria provide a new performance measure for nonregular designs. While the extensions of minimum aberration are single-valued criteria, the projection-efficiency criteria are multi-valued.

An alternative to the proposed two-stage analysis is *Bayesian model search and inference*. Box and Meyer (1993), Chipman (1996), and Chipman, Hamada and Wu (1997) have proposed Bayesian approaches for related problems. By exploiting the power in Bayesian computation, the Bayesian alternative can conceivably provide a more comprehensive search of the model space. On the other hand, it is much less user-friendly and may not be widely used by experimenters (who, alas, are not statisticians with Bayesian expertise.) These Bayesian methods do not focus on the inference for the parameters in the second-order models. Another difference is that they do not treat screening and surface exploration in separate steps. This may be an advantage for computation. It is, however, conceptually easier to treat these two as separate steps. The proposed approach also provides some new criteria for evaluating the performance of designs, which may stimulate corresponding work from the Bayesian perspective. It would be interesting to see further development of Bayesian analysis and design for the present problem.

Acknowledgements

The research is supported by NSF DMS-0072489 and NSA MSPR-00G-091. The authors would like to thank the referees, Derek Bingham and Huaiqing Wu for helpful comments.

Appendix 1

The columns of the $OA(18, 3^7)$ in Table 11 are denoted by **1**, **2**, **3**, **4**, **5**, **6**, **7**. The classification of its projected designs was also studied in Wang and Wu (1995), which is mainly based on combinatorial isomorphism. When the

$OA(18, 3^7)$ is projected onto three factors, any projected design consists of two 3^{3-1} designs. This is used to classify the projected designs into three types.

Table 11. 18-Run orthogonal array, $OA(18, 3^7)$.

run	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1	0	0	0	0	0	0	0
2	0	1	1	1	1	1	1
3	0	2	2	2	2	2	2
4	1	0	0	1	1	2	2
5	1	1	1	2	2	0	0
6	1	2	2	0	0	1	1
7	2	0	1	0	2	1	2
8	2	1	2	1	0	2	0
9	2	2	0	2	1	0	1
10	0	0	2	2	1	1	0
11	0	1	0	0	2	2	1
12	0	2	1	1	0	0	2
13	1	0	1	2	0	2	1
14	1	1	2	0	1	0	2
15	1	2	0	1	2	1	0
16	2	0	2	1	2	0	1
17	2	1	0	2	0	1	2
18	2	2	1	0	1	2	0

Design 18-3.1. It has two 3^{3-1} designs with no point in common, with 18 degrees of freedom (df).

Design 18-3.2. It has two 3^{3-1} designs with three points in common, with 15 df.

Design 18-3.3. It has two identical copies of a 3^{3-1} design, with 9 df and is hence ineligible. Only the set $\{\mathbf{1}, \mathbf{3}, \mathbf{4}\}$ is of this type.

There are more types of projected designs in terms of model isomorphism. In the following, instead of studying only the types of three-factor projected designs in the $OA(18, 3^7)$, we explore all types of non-isomorphic combinations of two 3^{3-1} designs. Because any three-factor projected design in the $OA(18, 3^7)$ consists of two 3^{3-1} designs, the resulting conclusion is more general and is potentially useful for other three-level orthogonal arrays. When a design consists of two 3^{3-1} designs, they can be defined by choosing two defining relations from the set:

{ABC = $\mathbf{I}$, ABC = u, ABC = u^2, ABC^2 = $\mathbf{I}$, ABC^2 = u, ABC^2 = u^2,
AB^{2}C = $\mathbf{I}$, AB^{2}C = u, AB^{2}C = u^2, AB^{2}C^2 = $\mathbf{I}$, AB^{2}C^2 = u, AB^{2}C^2 = u^2},
where $u = (1, 1, ..., 1)^T$ and $u^2 = (2, 2, ..., 2)^T$. We call each element in the set a *defining relation*, and the left expression of a defining relation a *defining word*. For example, the three defining relations ABC = $\mathbf{I}$, ABC = u and ABC = u^2 have the same defining word ABC. Note that the defining relations involving u and u^2 are not the regular defining relations in fractional factorial designs. First, we separate these 12 defining relations into the three sets: S_1 = {ABC = u, ABC^2 = $\mathbf{I}$, AB^{2}C = $\mathbf{I}$, AB^{2}C^2 = $\mathbf{I}$}, S_2 = {ABC = $\mathbf{I}$, ABC^2 = u, AB^{2}C = u, AB^{2}C^2 = u^2}, S_3 = {ABC = u^2, ABC^2 = u^2, AB^{2}C = u^2, AB^{2}C^2 = u}. In terms of model isomorphism, there are two types for design 18-3.1, labeled as 18-3.1.1 and 18-3.1.2, and four types for design 18-3.2, labeled as 18-3.2.1, 18-3.2.2, 18-3.2.3 and 18-3.2.4. These designs are arranged in descending order of D-efficiency.

Design 18-3.1.1. It is formed by choosing one defining relation from S_2 and one from S_1 with the same defining word (e.g., ABC = $\mathbf{I}$ and ABC = u), or by choosing one defining relation from S_2 and one from S_3 with the same defining word (e.g., ABC = $\mathbf{I}$ and ABC = u^2).

Design 18-3.1.2. It is formed by choosing one defining relation from S_1 and one from S_3 with the same defining word (e.g., ABC = u and ABC = u^2).

Design 18-3.2.1. It is formed by choosing one defining relation from S_1 and one from S_3 with different defining words (e.g., ABC = u and ABC^2 = u^2).

Design 18-3.2.2. It is formed by choosing two defining relations from S_2.

Design 18-3.2.3. It is formed by choosing one defining relation from S_2 and one from S_1 with different defining words (e.g., ABC = $\mathbf{I}$ and ABC^2 = $\mathbf{I}$) or by choosing one defining relation from S_2 and one from S_3 with different defining words (e.g., ABC = $\mathbf{I}$ and ABC^2 = u^2).

Design 18-3.2.4. It is formed by choosing two defining relations from S_1 or from S_3.

Except for 18-3.2.2, all other types of designs appear among the three-factor projected designs of the $OA(18, 3^7)$. The columns {**1**, **2**, **3**}, {**1**, **3**, **7**}, {**1**, **4**, **5**}, {**1**, **4**, **6**}, and any three columns not containing **1** are of type 18-3.1.1. The columns {**1**, **2**, **4**}, {**1**, **3**, **5**}, {**1**, **3**, **6**}, and {**1**, **4**, **7**} are of type 18-3.1.2, {**1**, **2**, **6**} and {**1**, **5**, **7**} of type 18-3.2.1, {**1**, **2**, **5**} and {**1**, **6**, **7**} of type 18-3.2.3, and {**1**, **2**, **7**} and {**1**, **5**, **6**} of type 18-3.2.4.

Each four-factor projected design is classified in terms of its four sets of three columns. For example, the projected design {**1**, **2**, **4**, **5**} contains the four sets of columns {**1**, **2**, **4**}, {**1**, **4**, **5**}, {**2**, **4**, **5**}, and {**1**, **2**, **5**}. The first three are of type 18-3.1 and the last of type 18-3.2. Hence {**1**, **2**, **4**, **5**} is classified as a design of type 18-4.2 (to be defined below). Based on the different combinations of designs 18-3.1, 18-3.2, and 18-3.3, four types of projected designs are obtained, each having 18 df.

Design 18-4.1. Any set of its three columns is of type 18-3.1. Any four columns not containing **1** is of this type.

Design 18-4.2. Three sets of its three columns are of type 18-3.1 and the other set of type 18-3.2.

Design 18-4.3. Three sets of its three columns are of type 18-3.1 and the other set of type 18-3.3. Since a type 18-3.3 design is ineligible, by Theorem 2, 18-4.3 designs are ineligible. Exclusive list of the type: {**1**, **2**, **3**, **4**}, {**1**, **3**, **4**, **5**}, {**1**, **3**, **4**, **6**}, and {**1**, **3**, **4**, **7**}.

Design 18-4.4. One set of its three columns are of type 18-3.1 and the other sets of type 18-3.2. Exclusive list of the type: {**1**, **2**, **5**, **6**}, {**1**, **2**, **5**, **7**}, {**1**, **2**, **6**, **7**}, and {**1**, **5**, **6**, **7**}.

Consider the further classification in terms of model isomorphism. The 18-4.2 designs can be classified into four types, labeled as 18-4.2.1, 18-4.2.2, 18-4.2.3 and 18-4.2.4. They are arranged in descending order of D-efficiency.

Design 18-4.2.1. Two sets of its three columns are of type 18-3.1.1 and the other two contain one 18-3.2.1 design and one 18-3.1.2 design. Exclusive list of the type: {**1**, **2**, **3**, **6**}, {**1**, **2**, **4**, **6**}, {**1**, **3**, **5**, **7**}, and {**1**, **4**, **5**, **7**}.

Design 18-4.2.2. Two sets of its three columns are of type 18-3.1.1 and the other two contain one 18-3.1.2 design and one 18-3.2.3 design. Exclusive list of the type: {**1**, **2**, **3**, **5**}, {**1**, **2**, **4**, **5**}, {**1**, **3**, **6**, **7**}, and {**1**, **4**, **6**, **7**}.

Design 18-4.2.3. Three sets of its three columns are of type 18-3.1.1 and the other set of type 18-3.2.4. Exclusive list of the type: {**1**, **2**, **3**, **7**} and {**1**, **4**, **5**, **6**}.

Design 18-4.2.4. Two sets of its three columns are of type 18-3.1.2 and the other two contain one 18-3.1.1 design and one 18-3.2.4 design. Exclusive list of the type: {**1**, **2**, **4**, **7**} and {**1**, **3**, **5**, **6**}.

Appendix 2

When the $OA(36, 3^{12})$ in Table 12 is projected onto three factors, there are three types of projected designs in terms of combinatorial isomorphism.

Table 12. 36-Run orthogonal array, $OA(36, 3^{12})$

run	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1	0	0	0	0	0	0	0	0	0	0	0	0
2	1	1	1	1	1	1	1	1	1	1	1	1
3	2	2	2	2	2	2	2	2	2	2	2	2
4	0	0	0	0	1	1	1	1	2	2	2	2
5	1	1	1	1	2	2	2	2	0	0	0	0
6	2	2	2	2	0	0	0	0	1	1	1	1
7	0	0	1	2	0	1	2	2	0	1	1	2
8	1	1	2	0	1	2	0	0	1	2	2	0
9	2	2	0	1	2	0	1	1	2	0	0	1
10	0	0	2	1	0	2	1	2	1	0	2	1
11	1	1	0	2	1	0	2	0	2	1	0	2
12	2	2	1	0	2	1	0	1	0	2	1	0
13	0	1	2	0	2	1	0	2	2	1	0	1
14	1	2	0	1	0	2	1	0	0	2	1	2
15	2	0	1	2	1	0	2	1	1	0	2	0
16	0	1	2	1	0	0	2	1	2	2	1	0
17	1	2	0	2	1	1	0	2	0	0	2	1
18	2	0	1	0	2	2	1	0	1	1	0	2
19	0	1	0	2	2	2	0	1	1	0	1	2
20	1	2	1	0	0	0	1	2	2	1	2	0
21	2	0	2	1	1	1	2	0	0	2	0	1
22	0	1	1	2	2	0	1	0	0	2	2	1
23	1	2	2	0	0	1	2	1	1	0	0	2
24	2	0	0	1	1	2	0	2	2	1	1	0
25	0	2	1	0	1	2	2	0	2	0	1	1
26	1	0	2	1	2	0	0	1	0	1	2	2
27	2	1	0	2	0	1	1	2	1	2	0	0
28	0	2	1	1	1	0	0	2	1	2	0	2
29	1	0	2	2	2	1	1	0	2	0	1	0
30	2	1	0	0	0	2	2	1	0	1	2	1
31	0	2	2	2	1	2	1	1	0	1	0	0
32	1	0	0	0	2	0	2	2	1	2	1	1
33	2	1	1	1	0	1	0	0	2	0	2	2
34	0	2	0	1	2	1	2	0	1	1	2	0
35	1	0	1	2	0	2	0	1	2	2	0	1
36	2	1	2	0	1	0	1	2	0	0	1	2

Design 36-3.1. It contains a complete 3^3 design and a 3^{3-1} design, with 27 df.

Design 36-3.2. It is a two-replicate of a 18-3.1.1 design, with 18 df. Example: {**1**, **3**, **8**}

Design 36-3.3. It contains a three-replicate of a 3^{3-1} design and a single replicate of another 3^{3-1} design, with 18 df.

In terms of model isomorphism, type 36-3.1 can be classified into two types, labeled as 36-3.1.1 and 36-3.1.2, and type 36-3.3 can be classified into two types, labeled as 36-3.3.1 and 36-3.3.2. These designs are arranged in descending order of D-efficiency.

Design 36-3.1.1. It consists of a complete 3^3 design and a 3^{3-1} design from S_1 or S_3. Example: {**1**, **2**, **6**}.

Design 36-3.1.2. It consists of a complete 3^3 design and a 3^{3-1} design from S_2. Example: {**1**, **2**, **3**}.

Design 36-3.3.1. It consists of a three-replicate of a 3^{3-1} design from S_1 or S_3, and a single replicate of a 3^{3-1} design from S_1 with the same defining word as the one chosen from S_1 or S_3. Example: {**1**, **4**, **9**}.

Design 36-3.3.2. It consists of a three-replicate of a 3^{3-1} design from S_2, and a single replicate of a 3^{3-1} design from S_1 or S_3 with the same defining word as the one chosen from S_2. Example: {**1**, **2**, **5**}.

For projection onto four factors, by an exhaustive search, we found 13 types of projected designs in terms of model isomorphism. They are labeled as 36-4.1 to 36-4.13 and arranged in descending order of D-efficiency.

Design 36-4.1 Three sets of its three columns are of type 36-3.1.2 and the other one of type 36-3.1.1, with 36 df. Example: {**1**, **2**, **3**, **7**}.

Design 36-4.2 Two sets of its three columns are of type 36-3.1.1 and the other two of type 36-3.1.2, with 36 df. Example: {**1**, **2**, **3**, **6**}.

Design 36-4.3 Two sets of its three columns are of type 36-3.1.1 and the other two of type 36-3.1.2, with 33 df. Example: {**1**, **2**, **6**, **12**}.

Design 36-4.4 One set of its three columns is of type 36-3.1.1, two of type 36-3.1.2, and the remaining one of type 36-3.2, with 36 df. Example: {**1**, **2**, **3**, **12**}.

Design 36-4.5 Any set of its three columns is of type 36-3.1.2, with 33 df. Example: {**1**, **4**, **8**, **10**}.

Design 36-4.6 Three sets of its three columns are of type 36-3.1.1 and the other one of type 36-3.1.2, with 36 df. Example: {**1**, **2**, **3**, **11**}.

Design 36-4.7 Two sets of its three columns are of type 36-3.1.1, one of type 36-3.1.2, and the remaining one of type 36-3.2, with 36 df. Example: {**1**, **2**, **3**, **8**}.

Design 36-4.8 One set of its three columns is of type 36-3.1.1, two of type 36-3.1.2, and the remaining one of type 36-3.3.1, with 33 df. Example: {**1**, **2**, **4**, **9**}.

Design 36-4.9 Three sets of its three columns are of type 36-3.1.2 and the other of type 36-3.3.2, with 33 df. Example: {**2**, **8**, **9**, **11**}.

Design 36-4.10 Three sets of its three columns are of type 36-3.1.1 and the other of type 36-3.3.2, with 33 df. Example: {**1**, **2**, **5**, **8**}.

Design 36-4.11 All four sets of its three columns are of type 36-3.1.1, with 33 df. Example: {**1**, **2**, **7**, **8**}.

Design 36-4.12 Two sets of its three columns are of type 36-3.1.1, one of type 36-3.1.2, and the remaining one of type 36-3.3.2, with 33 df. Example: {**1**, **2**, **3**, **5**}.

Design 36-4.13 One set of its three columns is of type 36-3.1.1 and the other three of type 36-3.2, with 18 df. Example: {**1**, **3**, **8**, **12**}.

References

Box, G. E. P. and Draper, N. R. (1987). *Empirical Model-Building and Response Surfaces.* John Wiley, New York.

Box, G. E. P. and Hunter, J. S. (1957). Multifactor experimental designs for exploring response surfaces. *Ann. Math. Statist.* **28**, 195-241.

Box, G. E. P., Hunter, W. G. and Hunter, J. S. (1978). *Statistics for Experimenters.* Wiley, New York.

Box, G. E. P. and Meyer, R. D. (1993). Finding the active factors in fractionated screening experiments. *J. Qual. Tech.* **25**, 94-105.

Box, G. E. P. and Wilson, K. B. (1951). On the experimental attainment of optimum conditions (with disscussion). *J. Roy. Statist. Soc. Ser. B* **13**, 1-45.

Chen, H. (1998). Some projective properties of fractional factorial designs. *Statist. Probab. Lett.* **40**, 185-188.

Chen, J., Sun, D. X. and Wu, C. F. J. (1993). A catalogue of two-level and three-level fractional factorial designs with small runs. *Internat. Statist. Rev.* **61**, 131-145.

Cheng, S.-W. (1999). Optimal blocking schemes and projection properties for fractional factorial designs. Ph.D. dissertation, Department of Statistics, University of Michigan.

Chipman, H. (1996). Bayesian variable selection with related predictors. *Canad. J. Statist.* **24**, 17-36.

Chipman, H., Hamada, M. and Wu, C. F. J. (1997). A Bayesian variable selection approach for analyzing designed experiments with complex aliasing. *Technometrics* **39**, 372-381.

Deng, L. Y. and Tang, B. (1999). Generalized resolution and minimum aberration criteria for Plackett-Burman and other nonregular factorial designs. *Statist. Sinica* **9**, 1071-1082.

Draper, N. R. and Lin, D. K. J. (1990). Small response surface designs. *Technometrics* **32**, 187-194.

Khuri, A. I. and Cornell, J. A. (1996). *Response Surface: Design and Analyses.* 2nd edition. Marcel Dekker, New York.

Lin, D. K. J. and Draper, N. R. (1992). Projection properties of Plackett and Burman designs. *Technometrics* **34**, 423-428.

Myers, R. H. and Montgomery, D. C. (1995). *Response Surface Methodology: Process and Product Optimization Using Designed Experiments.* John Wiley, New York.

Srivastava, J., Anderson, D. and Mardekian, J. (1984). Theory of factorial designs of parallel flats type. I: the coefficient matrix. *J. Statist. Plann. Inference* **9**, 333-345.

Taguchi, G. (1987). *System of Experimental Design.* Unipub/Kraus International Publications, White Plains, New York.

Tang, B. and Deng, L. Y. (1999). Minimum G_2-aberration for nonregular fractional factorial designs. *Ann. Statist.* **27**, 1914–1926.

Wang, J. C. and Wu, C. F. J. (1995). A hidden projection property of Plackett-Burman and related designs. *Statist. Sinica* **5**, 235-250.

Wu, C. F. J. and Hamada, M. (2000). *Experiments: Planning, Analysis and Parameter Design Optimization.* John Wiley, New York.

Xu, H. and Wu, C. F. J. (2001). Generalized minimum aberration for asymmetrical fractional factorial designs. *Ann. Statist.* To appear.

Institute of Statistical Science, Academia Sinica, Taipei 115, Taiwan.

E-mail: swcheng@stat.sinica.edu.tw

Department of Statistics, The University of Michigan, 4062 Frieze Building, 105 South State Street, Ann Arbor, MI 48109-1285, U.S.A.

E-mail: jeffwu@umich.edu

(Received December 2000; accepted June 2001)

COMMENTS

Derek Bingham

University of Michigan

The authors must be congratulated on taking an innovative look at design and analysis procedures for response surface methodology. Indeed they have challenged us to re-visit some of the "well-known" facts regarding the use of 3-level designs and the way *sequential assembly* problems are addressed. One is left to wonder how many common assumptions regarding other designs should be reconsidered.

There are two related concepts I would like to emphasize for possible improvement. Firstly, the analysis procedure performs factor screening and then recycles the data to fit a second order model on a projected design space. While this approach makes the analysis quite simple, it has important implications on the models being considered. Secondly, the proposed designs might fail to estimate some quite reasonable models. These issues will be discussed in the next section.

In addition to making a valuable contribution to this response surface application, it is worth noting that the authors make the important observation that combinatorially isomorphic designs do not necessarily estimate the same set of models (model isomorphism). This highlights the crucial connection between the design and analysis of an experiment. When some effects are assumed negligible, seemingly isomorphic designs are non-isomorphic. Interactions with quadratic terms are assumed negligible in the second-order model and thus the combinatorially isomorphic designs in Example 2 are not model isomorphic. If one considered interactions with the quadratic terms, there would be no distinction between the two designs. Because this assumption is made in many applications of 3-level designs, it would seem that a deeper look into what constitutes the best design in other applications should be undertaken.

1. Design and Analysis Approach

My main concern with the approach taken by the authors is that the model space that is explored is unrealistically small. As a consequence, the proposed optimality criterion may not be suitable. I discuss both these issues below.

The analysis approach takes place in two-stages: (i) factor screening, and (ii) fitting a second order model to the significant factors from stage 1. This methodology implies some unstated, but important assumptions with respect to the underlying structure of the model. Because the first step considers main effects only, an underlying assumption is that every factor affecting the response has a significant main effect. Interactions are entertained only on the projected design space, thus one is assuming that all factors in significant interaction effects also have significant main effects. The assumption that a two-factor interaction (2fi) can only be active if both main effects are active is called *strong heredity* (Chipman (1996)). This is a very restrictive assumption and may cause interactions to be missed and also cause the misspecification of the response surface.

The strong heredity assumption can be relaxed if the analysis considers some interactions in the first stage. This is important later for fitting a response surface where efficient estimation of the regression effects is the goal. To achieve this, one instead might use Bayesian variable selection for designs with complex aliasing

(Chipman, Hamada and Wu (1997)). A distinguishing feature of this approach is that hierarchical priors are used to specify the prior probability of an effect being included in the model. In this setting a model is denoted by a vector, δ, of ones and zeros, indicating whether an effect is active or inactive. The probability that an interaction is active, and thus in the model, is dependent on whether the corresponding main effects are active. For instance, for the AB-interaction,

$$p(\delta_{AB}|\delta_A, \delta_B) = \begin{cases} .00 & \text{if } (\delta_A, \delta_B) = (0,0) \\ .10 & \text{if } (\delta_A, \delta_B) = (1,0) \\ .10 & \text{if } (\delta_A, \delta_B) = (0,1) \\ .25 & \text{if } (\delta_A, \delta_B) = (1,1). \end{cases}$$

This specification of the prior (weak heredity) implies that interactions where both parents are active are more likely to be significant than interactions with only one active parent. Similarly, interactions where both parents are inactive are assumed to be negligible. While it is convenient to perform the first stage of the analysis ignoring interactions, it seems unrealistic to expect that the strong heredity assumption will apply in most applications. The Bayesian variable selection procedure will consider a richer and more realistic class of models than the one proposed by the authors.

The analysis procedure impacts the authors' choice of designs. The strong heredity assumption leads the authors to consider optimal projections as a criterion. Indeed, it is the projection approach that forces only strong heredity models to be entertained. Again, it is my feeling that this is too restrictive because interactions with at least one inactive parent will be missed. An alternate approach is to design the experiment so that parsimonious models may be identified. Questions facing the experimenter are (i) what are the likely models, and (ii) how many effects are likely to be significant? A criterion which attempts to address these questions is estimation capacity (Cheng, Steinberg and Sun (1999)). Roughly stated for 2-level designs, estimation capacity is the proportion of models containing all main effects and a pre-specified number (g) of 2fi's that a design is capable of estimating. To compare designs, the estimation capacity sequence (Cheng, Steinberg and Sun (1999)) has been proposed where the proportion of estimable models with $g = 1, 2, \ldots$, etc. 2fi's is computed.

Because one is interested in identifying more parsimonious models, Bingham and Li (2001) adjusted the estimation capacity criterion to measure the proportion of weak heredity models with g_1 main effects and g_2 2fi's (weak heredity maximum estimation capacity).

A criterion for this application, more in line with the possible model space, is to adjust the weak heredity maximum estimation capacity criterion to entertain the second order models of interest. For example, one could first specify the number of significant linear and quadratic main effects and interactions with at least one active parent that are likely to be present. The design criterion would then compute the proportion of such models that are estimable. For designs that are tied based on this criterion, the best design would have the best average D-efficiency. By specifying the likely number of significant effects, the experimenter is forced to consider explicitly what they might expect from the experiment, thereby running experiments that are neither too big nor too small. In any case, this procedure is meant as a suggestion to improve the potential deficiencies associated with designing and analyzing the experiment for the restrictive strong heredity models.

Department of Statistics, University of Michigan, Ann Arbor, MI, 48109-1027, U.S.A.
E-mail: dbingham@umich.edu

COMMENTS

Ching-Shui Cheng

University of California, Berkeley

The interesting discovery of good projection properties of nonregular designs was made by Lin and Draper (1992, 1993), Box and Bisgaard (1993) and Wang and Wu (1995). An important implication is that designs with complex aliasing, which have traditionally been used for screening main effects only, can be used to entertain and estimate certain interactions when these designs are projected onto a small subset of factors. Indeed, Hamada and Wu (1992) proposed a data analysis strategy for entertaining and estimating interactions from experiments with complex aliasing. The success of their strategy was attributed to the hidden projection properties of such designs.

Except for an 18-run design studied in Wang and Wu (1995), investigations of projection properties of factorial designs have been restricted to the two-level case. Now Cheng and Wu have carried out a detailed study of some three-level designs of small sizes and, more importantly, have proposed the idea of performing factor screening and response surface exploration on the same experiment, taking advantage of the hidden projection properties of three-level designs. The authors are to be congratulated for proposing this innovative alternative to central composite designs.

For three-level designs, the presence of a three-letter defining word causes the projection onto the three factors that appear in the defining word to be ineligible for estimating the second-order model. The presence of such defining words prevents the projected design from having enough degrees of freedom for estimating all parameters in the model, and is termed "curse of three-letter words" by the authors. In general, the absence of certain defining words is necessary for the projected design to be eligible, but there is no guarantee that it is sufficient. For a two-level regular design, a trivial necessary and sufficient condition for all its projections to be eligible for estimating all the main effects and two-factor interactions is that there is no defining word of length three or four; in other words, the design is of resolution at least five. But results for other designs are limited. Cheng (1995) showed that the absence of defining words of length three or four is also necessary and sufficient for all four-factor projections of a nonregular two-level orthogonal array with strength two to be eligible for estimating all the main effects and two-factor interactions. For a two-level orthogonal array with strength three, a necessary and sufficient condition for all its five-factor projections to be eligible is that there is no defining word of length four (Cheng (1998)). For three-level regular designs, Cheng and Wu showed that the absence of three-letter words is enough for the projection onto any three or four factors to be eligible for the second-order model (The eligibility is not affected by defining words of length four since the second-order model contains only the linear by linear components of two-factor interactions.) It is not clear whether similar results can be established for projections of nonregular designs or projections of regular designs onto more than four factors. An additional complication for three-level designs is that the usual defining words do not provide direct information on the aliasing of factorial effects defined by orthogonal polynomials (such as the linear by linear effect in a second-order model).

Nevertheless, designs with few defining words of short lengths are expected to have good projection properties. Cheng and Wu successfully used this idea to construct a 27-run design with 8 factors such that all its projections onto four factors are eligible for the second-order model, while among projections onto five factors only one projection is ineligible. It is an interesting problem to investigate the general construction of three-level designs with good projection properties. For example, a 27-run design with 8 factors such that all its projections onto five factors are eligible for the second-order model can be found. If one does not insist on using an orthogonal array, then an even smaller design with the aforementioned projection property can be constructed. These and other designs will be reported in Bulutoglu and Cheng (2001).

As the authors noted, the projection-efficiency criteria proposed in this article are multi-valued. Even though the generalized minimum aberration criterion defined by Deng and Tang (1999), an extension of minimum aberration to nonregular designs, is a single-valued criterion, it does serve well the dual goal of model robustness and efficiency: entertaining many models containing lower order factorial effects and estimating these effects efficiently (see Cheng, Deng and Tang (2000)). I wonder if the generalized minimum aberration criterion or a suitable modification could be a satisfactory single-valued surrogate for the projection-efficiency criteria.

Department of Statistics, University of California, Berkeley, CA 94270-3860, U.S.A.
E-mail: cheng@stat.berkeley.edu

COMMENTS

Mong-Na Lo Huang

National Sun Yat-sen University

The paper by Cheng and Wu brings interesting perspective to the issues of designing efficient experiments which combine the tasks of factor screening and response surface exploration in one design. In order to achieve the two objectives in one design, an approach based on a two-stage analysis is proposed where factor screening is performed in the first stage and, after the important factors are identified, a second-order model is fitted to the identified factors.

Two-stage statistical analyses are quite common in practice. The procedure of stepwise regression can be thought of as a multi-stage model building procedure, for example, although the steps of selecting parameters and models are different. The interesting part of this paper is that it discusses the problem from the design of experiments point of view. The proposed approach is quite natural and cleverly combines considerations of factor screening, projection and response surface exploration with orthogonality, projection eligibility and estimation efficiency. Projection-efficiency criteria are proposed where, for projection eligibility, designs with larger number of eligible projected designs are favor so; giving more weight to lower-dimension than higher-dimension projection is preferred; for estimation efficiency, designs with high D- and G-efficiency are preferred. After setting up the is objectives, these criteria arise quite naturally. Although the criteria are not new, the way the authors combine them is very nice.

At the factor screening stage, each factor should have at least three levels since one intends to fit a second-order model at the second stage. Projection-efficiency properties of $3^{n-(n-3)}$ regular orthogonal designs with 27 runs are studied for their projection eligibility and estimation efficiencies. The distinction between design isomorphism and model isomorphism is interesting, perhaps some studies along this line for other types of regular orthogonal designs would also be of interest. Although based on the criteria of projection-efficiency the $3^{n-(n-3)}$ designs seem to be doing well only for cases when the number of important factors is not too large, less than five factors can be fitted with a second-order model, and the frequency of eligible projections is not as good as the nonregular design $OA(18,3^7)$. As for the three nonregular designs $OA(18,3^7)$, $OA(36,3^{12})$, and 3^{8-5}_{NR}, the advantages lie in economic run size, the estimation efficiencies seem to be low in many cases, especially for the G-criterion.

Apparently, when we try to achieve two objectives in one design, we cannot help losing some of the good properties that optimal designs have for each objective separately. This paper has made a nice contribution in finding good designs to achieve both objectives when the important factors are not too large. This can be very useful in growth curve models, when more than one observation is obtained under each factor combination in different times or periods, or in other models where a density function or regression curve under each factor combination needs to be compared to see the effects of the factors. Economic run size for factor combination as well as high estimation efficiency in the second-order model proposed here are certainly very helpful in reducing the total number of experimental runs.

But if the important factors are large, the advantanges of trying to achieve two objectives in one design diminish very quickly. Also as the number of important factors increases, the run size also increases very quickly. Then it becomes harder to attain good performances in both objectives with only one design, not to mention that when the run size becomes large, we might encounter the problem of blocking the experiments. In cases like that, do we still want to stay with a one-stage design and worry whether it can lead to efficient estimation on the important factors? Or should we simply adopt a two-stage design strategy to achieve our goals? Of course to know whether there are many important factors or not, we need to have some prior information to determine what would be a better approach.

Department of Applied Mathematics, National Sun Yat-sen University, Kaohsiung, Taiwan.
E-mail: lomn@math.nsysu.edu.tw

COMMENTS

André I. Khuri

University of Florida

The article by Cheng and Wu offers a novel approach in response surface methodology (RSM) that combines factor screening and response surface exploration using a single design. The primary motivation for proposing such an approach is to develop an experimental strategy that can be carried out when it is either time consuming or too costly to perform new experiments. This represents an interesting departure from one of the most fundamental tenets of RSM.

Most applications of RSM are sequential in nature whereby information acquired from one set of experiments is effectively used to plan the strategy for a follow-up set of experiments. This sequential pattern of experimentation was suggested by Box and Youle (1955). Important factors, or variables, are identified at an early stage of experimentation. Subsequent stages of the experiment serve to consolidate information concerning the important variables and to weed out factors deemed unimportant. Further experimentation will then be confined to a region of the factor space where the actual exploration of the response surface will be carried out. The time interval between successive sets of experimental runs is usually short (often a matter of hours), as in many experiments in the chemical, physical, and engineering sciences. However, in some experiments, as in agricultural field experiments and clinical trials, much longer periods of time may be needed before a new set of experiments can be performed. This will obviously hinder the implementation of the sequential approach. The proposed experimental strategy by Cheng and Wu represents a step in the right direction in trying to resolve this problem.

There are certain issues an experimenter may be concerned about when presented with a single design, a summary of which follows.

1. If the entire experiment were to be designed at the outset, some prior information should be available concerning the importance of some of the variables and the presence or absence of some key interactions. For example, if one factor is expected to produce a large effect, appreciable interactions are practically certain to arise between it and the other factors. In this respect, it would be very helpful to include in the design process input from a technical assistant who is involved in the running of the experiment and is therefore quite familiar with its operational development.
2. The sequential approach has always been stressed in RSM. It has been suggested, with some irony, "that the best time to design an experiment is after

the experiment has been completed because one then has more knowledge of the process under study–what variables are important, over what ranges, in what metrics, and so on. By designing experiments sequentially, we can, in a sense, approximate this happy situation by peeking at the answer and modifying the design accordingly." (see Steinberg and Hunter (1984, p.88)). The converse of this is "that the worst time is at the beginning, when least is known." (see Box, Hunter, and Hunter (1978, p.303)). Thus a retraining of an RSM experimenter may be necessary in order to become acclimated to a nonsequential environment.

3. In the sequential approach, variables included in one stage may be dropped in later stages, and new variables may be introduced. In addition, some of the factors may be varied over new ranges and the factor space being explored can change. This produces lively mobility and an active exchange of variables, which appear to be lacking in a single-design experiment. How can an experimenter compensate for the loss of such information?
4. As is pointed out in the article, the main-effect ANOVA analysis may not be appropriate to identify important variables if the interactions are large enough to bias the main effect estimates. Some variables may erroneously be admitted in Stage 1 of the two-stage analysis. Fortunately, Stage 2 can provide some safeguards for weeding out variables admitted in Stage 1 if their linear and quadratic effects are aliased with a significant interaction, as was shown in the example. An inexperienced research worker, however, may not be able to discover this since the construction of three-level fractional factorial designs and their interpretation is more complicated than in two-level fractional factorials. Another issue of some concern is that estimates of the main effects may not be given with sufficient precision in Stage 1.
5. Some mechanism is needed for testing lack of fit once a model is chosen. The proposed single design does not provide such a mechanism. The success of any response surface exploration in Stage 2 depends to a large extent on the form of the model and whether or not it provides a satisfactory representation of the true mean response. In the sequential approach, testing for lack of fit is carried out in every stage.
6. The proposed design assumes that the error variance is constant. This assumption may not be valid. It would be desirable to have a design that provides a check on the constancy of variance assumption. This is one of the design criteria listed in Box and Draper (1975).
7. The paper considers only orthogonal designs because they are suited for factor screening. Tukey once suggested the use of designs that are not orthogonal, but in which the correlations among the model's parameter estimates are quite small. "By sacrificing some orthogonality, it may be possible to gain much

in terms of the number of factors that can be studied." (see Steinberg and Hunter (1984, p.87)).

8. In an unrelated paper, Sitter and Wu (1999) proposed a two-stage design for binary response data, which takes advantage of the information from the initial stage of an experiment to design a follow-up study and still not unduly prolong the experiment's duration. Can a similar approach work here? It should be remembered that the central composite design introduced in Box and Wilson (1951) is really a two-stage design with a factorial portion chosen in the first stage followed by an axial portion in the second stage for the purpose of fitting a second-degree model. Replications at the design center provide additional design properties, such as orthogonality and uniform precision, and can be used to test for lack of fit.

In summary, Cheng and Wu are to be commended for proposing this novel approach in the design and analysis of response surface experiments. As they pointed out, further development of this approach would be interesting and, I may add, welcome.

Department of Statistics, University of Florida, Gainesville, FL 32611-8545, U.S.A.
E-mail: ufakhuri@stat.ufl.edu

COMMENTS

S. M. Lewis

University of Southampton

This interesting paper addresses the important problem of how to experiment economically in product and process development. I have two comments on where some caution may be needed in applying the approach.

The first concerns the number of factors investigated, and applies generally when screening is an aim of experimentation. At the outset of an investigation the list of factors that might possibly influence a response is often long perhaps consisting of twenty or more factors. Designs to screen a modest number of factors (as in the paper) are appropriate when this list can be reliably reduced by subject specialists through scientific understanding. or evidence from previous experiments. When such well-informed decisions are not possible, the exclusion of important factors at the early planning stage in order to achieve short-term economy can lead to the need for further experiments at a much later stage, and unscheduled delays in achieving the objectives for the product or process.

Larger numbers of factors could be included in the authors' one-experiment approach by using suitable fractions of $2^m \times 3^n$ designs. Here the two-level factors could be those whose likely importance was difficult for the subject specialists to assess. For any such factor found to be important in the first-stage analysis, however, the model used in the second-stage analysis could not include a quadratic main effect term.

My second concern is that, in screening only main effects in the first-stage analysis, one or more substantive interactions could be overlooked when one of the factors involved in the interaction does not have a significant main effect; see the figure for an illustration with factors at two levels. This is of especial concern when one factor is a control or design factor (A) and the other (B) is a noise factor. The detection and manipulation of control$\times$noise interactions is then an important method of reducing variation in a response, see Shoemaker, Tsui and Wu (1991) for example. Some design strategies that enable screening for interaction among large numbers of factors have been investigated by Lewis and Dean (2001) . These strategies use two stages of experiments with the factors investigated in groups at the first stage.

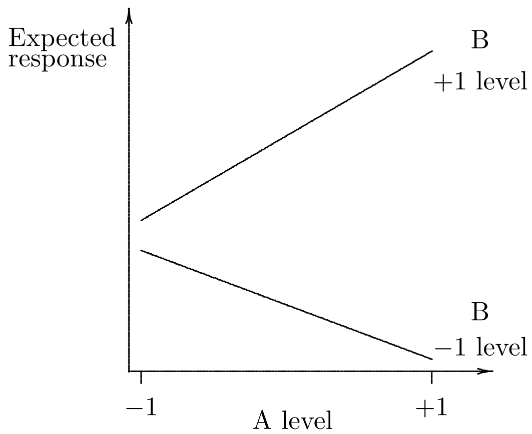

Figure 1. Positive interaction between factors A and B with zero A main effect.

When an interaction may be present without both main effects being significant, a conflict appears to arise in the authors' approach between the need for a design to have a large number of eligible projected designs of low dimension, and the ability of a design to offer some indication of such interactions in the second-stage analysis. As the authors point out, the inclusion of 3-letter words in the defining contrasts subgroup for a design is indeed a "curse" when many low-dimension projections is a design objective. However, it is through the inclusion of such 3-letter words in the defining contrasts of a design that information can be retrieved on interactions, through exploiting the partial aliasing of linear×linear interaction terms with linear and quadratic main effects, as the authors demonstrate in their illustrative example.

Department of Mathematics, University of Southampton, Southampton, SO17 1BJ, U.K.
E-mail: S.M.Lewis@maths.soton.ac.uk

COMMENTS

Douglas C. Montgomery and Connie M. Borror

Arizona State University

We thank Professors Cheng and Wu for an interesting and thought-provoking paper. There are some very original ideas in this work, and the authors have provided us with a novel and potentially quite useful way to employ the three-level fractional factorial design in response surface methodology (RSM). We will discuss the types of experimental problems where these designs can be useful and then focus on the prediction variance properties of the designs.

We traditionally think of RSM as consisting of three stages (see Myers and Montgomery (1995), Montgomery (2001)): (1) factor screening, (2) locating the region of the optimum, and (3) determining the optimum operating conditions. Myers and Montgomery refer to these as stages 0, 1, and 2, respectively, emphasizing the importance of factor screening before starting the "classical" RSM procedure. The reason for this is that many industrial processes have a relatively large region of operability and there is often little comfort in the assumption that the initial region of experimentation (even for a screening experiment) is large enough to contain the final optimum conditions. That is, the current operating conditions are very far from the optimum. This happens frequently in the chemical and process industries where the initial operating conditions are often determined through analysis of a pilot plant level process, and scale-up affects the results when they are translated to full-scale manufacturing. Furthermore, there may be variables in the full-scale process that were not present (or not adjustable) at the pilot plant level. Consequently, following factor screening it may be necessary to move to a new region of experimentation (via some technique such as steepest ascent) that is remote from the original one. Sometimes steepest ascent will be applied two or three times to reach this final region of experimentation (Myers and Montgomery (1995) refer to these as mid-course corrections). Consequently, the final region of interest will be far from the initial region of experimentation.

Now one possibility is to expand the initial region of experimentation so that it covers a much larger portion of the region of operability, thereby giving the experimenter a better chance that the original screening design actually covers the final optimum point(s). However, the risk here is that the region will be so large that the second-order model will be a poor fit to the true response surface. The true response surface is likely to be very nonlinear, and if the region of experimentation is very large this can lead to a situation where the Taylor-series argument that allows us to approximate the response surface with a low-order polynomial breaks down. The authors' designs would seem to be most useful in cases where (1) the region of operability is sufficiently small so that the experimenters can explore most of it with the original experiment, or (2) the process is sufficiently well-understood or mature so that the experimenters can be reasonably certain that the original region of experimentation includes the desired final operating conditions.

The authors give a very thorough discussion of the D- and G- efficiency properties of their designs. Myers and Montgomery (1995) observe that these single-number alphabetic optimality criteria often give a very incomplete summary of the prediction variance properties of a design. This is an important issue, since response surface designs are used primarily to fit models used for prediction. We focus on the scaled prediction variance

$$\frac{N\mathrm{Var}\,[\hat{y}(\boldsymbol{x})]}{\sigma^2} = N\boldsymbol{x}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x},$$

where $\boldsymbol{x}$ is the point of interest in the design space and N is the number of runs in the design. Giovannitti-Jensen and Myers (1989) recommend the use of variance dispersion graphs (VDGs) to display the scaled prediction variance over the design region. A VDG plots the minimum, maximum, and average scaled prediction variance for all points that are the same "distance" from the design center as a function of "distance". For designs on cubes, this "distance" is measured from the design center to the center of each cube face. VDGs have been used extensively to study the performance of response surface designs. For example, see Myers, Vining, Giovannitti-Jensen, and Myers (1992).

We chose to evaluate the proposed designs using scaled prediction variance by way of variance dispersion graphs. To illustrate, the design given in Table 11 for seven factors was projected into an 18-run design for factors x_1, x_2, and x_7. The projected design was compared to a face-centered cube (FCC) for the same set of factors (a design with $n = 17$ experimental runs). A full quadratic model in the three factors x_1, x_2, and x_7 is assumed. The VDGs for both designs are displayed in Figure 1. A horizontal line at 10 for the scaled prediction variance (since there are $p = 10$ terms in the full quadratic model) has been included as a frame of reference for G-optimality. Based on the VDG in Figure 1, the face-centered cube is a better design in terms of maximum scaled prediction variance. This is not surprising since the authors have pointed out that their designs generally do not enjoy high G-efficiencies. However, examining the average scaled prediction variance reveals that the proposed design is highly competitive with a standard FCC. It may be possible to improve the relatively low G-efficiency by selectively adding a few experimental runs to the proposed designs.

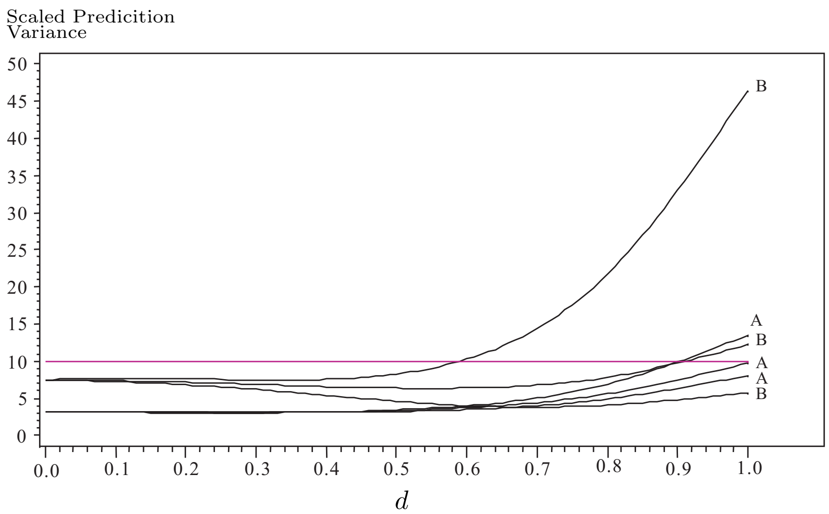

Figure 1. Variance Dispersion Graph for (A) Face-Centered Cube with 17 runs and (B) Cheng and Wu Design in three variables from Table 11.

As a second example, the only three-factor combination from Table 11 that will not project into a design that will support a full quadratic model in three factors is the combination of x_1, x_3, and x_4. This combination of factors yields a 3^{3-1} design in two replicates. This design can be augmented with additional runs to allow estimation of the full quadratic model by using a variety of augmentation criteria. We augmented the design with 2 experimental runs using the D-criteria for selecting the additional runs, with the resulting design shown in Table 1. The additional experimental runs are given in bold. The VDG for the augmented design is given in Figure 2. Note that the augmented design has reasonable scaled prediction variance properties.

Table 1. Augmented Cheng and Wu Design in three variables.

Run	x_1	x_3	x_4
1	−1	−1	−1
2	−1	0	0
3	−1	1	1
4	0	−1	0
5	0	0	1
6	0	1	−1
7	1	0	−1
8	1	1	0
9	1	−1	1
10	−1	−1	−1
11	−1	0	0
12	−1	1	1
13	0	−1	0
14	0	0	1
15	0	1	−1
16	1	0	−1
17	1	1	0
18	1	−1	1
19	**1**	**0**	**1**
20	**1**	**1**	**−1**

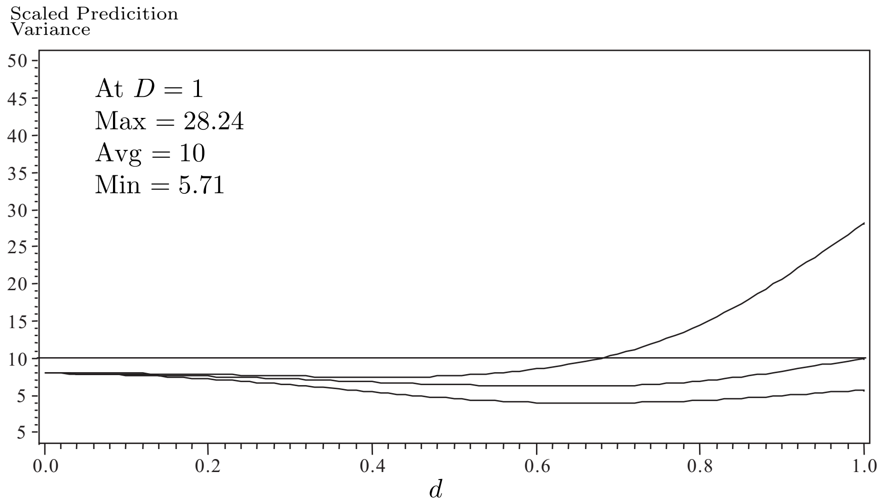

Figure 2. Cheng and Wu 3^{3-1} with two replicates in variables, x_1, x_3, x_4 augmented with 2 runs ($n = 20$).

In summary, we think that the designs proposed by Cheng and Wu are likely to be most useful when the experimenter is reasonably certain that the second-order model will be adequate over the region of experimentation, or when the region of operability is reasonably small. The authors have evaluated their designs in terms of D- and G-efficiencies and concluded that the designs have relatively low G-efficiencies. In our opinion, a single number criteria does not adequately describe the variance properties of the proposed designs and we suggest that the scaled prediction variance over the region of operability for these designs contributes valuable additional information. In particular, the average prediction variance based on our limited evaluation appears competitive with other designs. Furthermore, augmentation of these designs with a few additional runs could potentially improve G-efficiency in some situations.

Department of Industrial Engineering, Arizona State University, Tempe, AZ 85287-5906, U.S.A.
E-mail: doug.montgomery@asu.edu

Department of Industrial Engineering, Arizona State University, Tempe, A2 85287-5906, U.S.A.
E-mail: conni@asu.edu

COMMENTS

Rahul Mukerjee

Indian Institute of Management

I begin by congratulating the authors on this very scholarly and thought-provoking piece of work. In particular, the idea that non-regular fractions can have an edge over the regular ones in the present context is very promising and can open up whole new areas. I am sure that many other researchers will follow up this paper. For possible consideration in the future by the authors or others, I will indicate two issues that seem to emerge from the present work.

1. Non-regular fractions, including three-level ones, are often generated from difference matrices, and systematic constructions are known to exist for the latter - vide Wang and Wu (1991), Dey and Mukerjee (1999), and the references therein. Admittedly, these constructions can be much more complicated than the corresponding ones for regular fractions. Nevertheless, can there be any way of exploiting the systematic element in such constructions for analytical study of the performance of the resulting non-regular fractions, under at least some of the present criteria?

2. A salient feature of the two-stage analysis, discussed in Section 7, is that the active fac are identified in the first stage via a model that involves only

the main effects (MEs). As the authors note later in Section 7, the presence of interactions, say, two-factor interactions (2fis) can vitiate the identification of the active factors. Specifically, the trouble can arise in two ways:

(a) an inactive factor is declared active at the first stage;

(b) an active factor remains undetected at the first stage.

For example, in a regular fraction, if an insignificant ME is aliased with a significant 2fi then situation (a) can arise. Similarly, in such a fraction, if a significant ME is aliased with a 2fi which is also significant, with approximately the same magnitude but in the opposite direction, then situation (b) can arise.

As seen in Section 7, it may be possible to have some kind of protection against the situation in (a) under the present two-stage analysis. Is there any such protection against (b) within the frequentist paradigm? Is it possible to develop a systematic procedure for this purpose? Or does the Bayesian method of Box and Meyer (1993), hinted at in Section 2, represent the only way out?

Indian Institue of Management, Joka, Diamond Harbour Road, Post Box No. 16757, Alipore Post Office, Calcutta, 700 027, Indian.

E-mail: rmuk1@hotmail.com

COMMENTS

David M. Steinberg and Dizza Bursztyn

Tel Aviv University

Cheng and Wu present an interesting and useful framework for designing experiments that can accomplish the joint goals of factor screening and response modeling. It is essential to consider the scientific context of response modeling experiments to fully appreciate the ideas in their article and we begin our discussion with some brief comments on this topic. Then we describe an alternative class of designs that we have found efficient for joint screening and modeling. Our class is based on rotations of standard two-level factorials. Finally, we describe a design criterion that we have found useful in these settings and compare our rotation designs to the three-level fractions advocated by Cheng and Wu.

Scientists engaged in experimentation and empirical modeling are often faced with the problem of identifying the important factors and then modeling the response variables using those factors. Response surface methodology (RSM) evolved as a strategy for tackling such problems when rapid feedback is possible and a sequential approach can be taken. The spirit of the RSM strategy is captured with clarity in the very insightful articles by Box (1999a, 1999b) and Box and Liu (1999). One of the main points of those articles is that sequential investigation is an essential element of response surface methodology. At the very heart of the strategy is the flexibility to proceed in small, well-designed stages. The results at each stage lead to decisions in planning and analyzing the next stage, including possible changes in the experimental factors, their ranges, or the number of levels included, and in the equations used in modeling the response variables. Many RSM studies begin with factor screening experiments and eventually proceed to modeling experiments. However, the cursory characterization by Cheng and Wu that "RSM can typically be described as consisting of two parts" (screening and modeling) strikes us as stopping well short of the much broader strategic scope of RSM.

The fundamental motivation for this article by Cheng and Wu is that experimenters sometimes need to exploit data from a single experiment both to identify the important factors and to fit models that have more than linear main effects. We have also had experiences with experimenters who did not enjoy the flexibility required by the fully sequential nature of RSM. In one extreme case, we consulted with an engineer who had rented 6 hours of access time to an expensive facility. In that time, he had to make all the runs necessary to achieve both factor screening and modeling. Off-line data processing was necessary to summarize the output to each run, so bringing along a statistician to do "real time" data analysis was not feasible. We have consulted on other experiments in which the total time schedule was dominated by the need to specify and procure parts that matched the levels needed in the experiment. Each new order resulted in further delays in the project. Thus the sequential strategy of RSM simply was not practical.

One of the very interesting ideas in Cheng and Wu's article is their extension of the notion of projectivity. The original articles that examined projectivity were concerned with the ability of two-level fractional factorial designs to provide extra information on interactions when projected onto a subset of the factors. Cheng and Wu add the valuable step of looking at projectivity for fitting full second-order models and not just interactions.

Cheng and Wu propose the use of three-level fractional factorials for joint screening and modeling, pointing out the important difference between combinatorial and model isomorphism. This distinction was also made by D. Street with respect to 3^{3-1} designs in a presentation at the R. C. Bose Memorial Conference in 1995. Cheng and Wu consider the more interesting case of 3^{4-1} designs and show that there are two classes of model isomorphic designs, which they denote by $\mathcal{K}_1$ and $\mathcal{K}_2$. A simple geometric property distinguishes between these two classes of designs. The designs in $\mathcal{K}_2$ are precisely those which include the center point $(0, 0, 0, 0)$. The remaining 26 points in these designs form 13 foldover pairs.

There does not appear to be any useful 13 run design consisting of one member of each pair. So the foldover property does not offer any evident statistical benefits to the designs in $\mathcal{K}_2$. The design classes are specified by their defining word (as in the article) only if one chooses the fraction that sums to 0 modulo 3. If, instead, the sum is 1 or 2, different defining words are in each class. The geometric description is valid regardless of the sum.

One of the possible drawbacks to the three-level fractions proposed by Cheng and Wu is the relatively strong aliasing of main effects with two-factor interactions. These effects are only partially aliased, so many models with combinations of main effects and two-factor interactions can be fit. However, the aliasing might also lead to misidentification of the active factors. This problem is highlighted by the example in Section 7 of the article. The initial main effects analysis found significant effects for factors C and D, but the more detailed analysis that followed suggested that the real active effect might be the AB interaction, with the presumed effects for C and D a result of their aliasing with this interaction.

We have been exploring an alternative class of designs based on rotations of two-level factorials. These rotation designs are also useful for joint screening and modeling. They are orthogonal plans for first-order models and do not suffer from the aliasing of main effects with interactions noted above for the three-level fractional factorials. To construct a rotation design, let D denote an $n \times k$ design matrix for a two-level fractional factorial and let R be any $k \times k$ orthogonal matrix. Then the design $D_{Rot} = (1/c)DR$ is a rotation design. The constant c serves only to rescale the design to the unit cube. We have found that one useful class of rotation designs can be formed when $k = 2^t$ by taking orthogonal matrics of the form $R = HSH'$, where H is the standard Hadamard matrix for estimating all the effects in a 2^t factorial design (Bursztyn and Steinberg, in press). The matrix S is block diagonal, with each 2×2 block a simple planar rotation. Taking the initial design D to be the standard resolution IV design with $2k$ runs generates a rotated design with the same set of levels for each factor. The angles used in the planar rotations can be chosen to improve the statistical properties of the design. If the number of factors is not a power of two, one can use the above technique by adding "dummy" factors, doing the rotation, and then dropping the extra factors.

The rotation designs, like the three-level fractional factorials, are able to fit many extended models. Bursztyn and Steinberg (in press) present an example of an 8-factor, 16-run rotation design that can fit a full second-order model in every three-factor projection and can fit models with all linear effects and all pure quadratic main effects in 58 out of 70 four-factor projections.

We have found that a useful criterion for combined screening and modeling designs is to compute the sum of the squared elements of the alias matrix that results from fitting a first-order model in the presence of higher order (say up to third-order) terms (Bursztyn and Steinberg, in press). Small entries in the alias matrix imply that there is limited aliasing of main effects so that the misidentification problem pointed out earlier is minimized. Further, small entries in the alias matrix correlate well with second-order projectivity because they tend to indicate the ability to fit larger models that include some of the terms considered as "extra" regressors relative to the initial model. We make all of the polynomial terms in the alias matrix orthonormal with respect to uniform measure on the unit cube to ensure that our comparisons use polynomials that are defined independently of the designs.

We computed the alias matrix measure for several rotation designs and for some of the designs proposed by Cheng and Wu. The results for some 8-factor designs are described in Table 1. The rotation designs are much more efficient than the three-level fractions with respect to the alias matrix criterion. The rounded rotation design in Table 1 was derived from rotation design 2 by rounding each level to the nearest value from among ± 1, ± 0.5 and 0. This design overcomes the practical objection that rotation designs may include many levels for each factor and so may be difficult to implement. Some of the properties of the rotation design are lost by rounding, but the design here is still close to orthogonal and is quite efficient with respect to our alias matrix criterion. We note that the angles in our rotation designs were chosen via some very limited trial and error. No attempt was made to find optimal angles.

Cheng and Wu have opened an interesting new area for research in experimental designs for response modeling. They have also contributed some useful new designs. Our own research on rotation designs shares many common themes. We encourage further work in this area.

Table 1. The alias matrix criterion for several designs. The regular and special three-level fractional factorials are the designs in sections 4.2 and 6, respectively, of Cheng and Wu.

Design	Runs	Factors	Alias Criterion
Rotation 1	16	8	0.26
Rotation 2	32	8	2.50
Rounded Rotation	32	8	5.42
3^{8-5} Regular	27	8	25.75
3^{8-5} Special	27	8	27.25

Department of Statistics and Operations Research, Tel Aviv University, Tel Aviv, Israel.

E-mail: dms@post.tau.ac.il

E-mail: dizza@math.tau.ac.il

REJOINDER

Shao-Wei Cheng and C. F. J. Wu

The eight discussants have made insightful comments and asked penetrating questions on a variety of issues that our work has helped generate. We are thankful to them for their efforts in reading through a long paper, and their generosity in sharing ideas and unpublished results with the readers and us. Our reply is grouped into five categories.

1. Uni-stage vs. Sequential Experiments

The pros and cons of these two modes of experimentation are well documented in textbooks, our paper and the discussions. Because of its ability for adaptive learning, the sequential approach has been the mainstay in response surface literature. As the trend in technology favors a shorter product development cycle and a quicker reaction to market opportunities, shorter-duration non-sequential experiments will become more popular in engineering, and the physical and life sciences, as well as in agriculture and clinical trials.

Huang and Lewis raise concerns on the appropriateness of our approach to deal with many factors. Indeed, if the number of factors is too large, say more than 12 factors at three levels, it will become ineffective to perform screening and surface exploration with the same design. A preliminary stage of study /experimentation to screen and reduce the number of factors will be required, that is, a sequential strategy will be unavoidable. In the Myers-Montgomery terminology (per the Montgomery-Borror discussion), this is referred to as stage 0 in the RSM.

2. Analysis Methods for Factor Screening

Several discussants (Bingham, Khuri, Lewis, Mukerjee) correctly point out the limitations of using main effect analysis for factor screening. If a factor's significance is manifested through one or several of its interactions with other factors, but not through its main effect, it will be missed by the main effect analysis. Because of these limitations we discussed other analysis methods, such as the Bayesian approach of Box and Meyer (1993), and the more elaborate Bayesian approach of Chipman, Hamada and Wu (1997) that performs an intelligent search over the model space by using Gibbs sampling. A frequentist model search (see Section 8.4 of Wu and Hamada (2000) for three versions of this) that shares many features with the Bayesian approach of Chipman, Hamada and Wu is equally effective in many situations, and is easier to program. The first method (by Box and Meyer) is factor-based, while the last two are effect-based. (This terminology and distinction is due to Chipman and Hamada (1996)). Because these more elaborate methods can identify factors whose significance only appears through interactions, the models explored in the two-stage analysis strategy do not have to follow strong effect heredity (hinted at in Bingham's discussion.). In general, factor-based methods are more suitable for designs of small size, because such designs do not have enough degrees of freedom for model fitting. On the other hand, effect-based methods require more degrees of freedom and thus work better for larger designs. Therefore, an effect-based method (as suggested by Bingham) may not work for designs of medium size. By contrast, the two-stage analysis strategy can handle such situations as the model space is reduced after projection.

Because of its simplicity the main effect analysis method should be considered, especially when the investigators do not have access to software for more elaborate methods. In the majority of situations, main effect analysis coupled with second-stage analysis should perform well. This is especially true if a more lenient criterion is used in selecting factors.

If the aliasing between effects is too strong, no method can be effective in dealiasing them. In this case the misidentification of important factors, as pointed out by Mukerjee and Khuri, is unavoidable. An illustrative example can be found in Section 8.4.1 of Wu and Hamada (2000). The only recourse then is to find a better design or to sequentially add design points to ameliorate the strong aliasing. The related issue of choosing alternative designs is deferred to Point 4.

Lewis mentions the interesting alternative of using a two-stage group-screening to screen two-factor interactions (in addition to main effects). This approach requires sequential experimentation. She also expresses concern about "a conflict — between the need — to have a large number of eligible projected designs of low dimensions, and the ability — to offer some indication of such interactions in the second-stage analysis". We cannot see how this conflict will arise based on the theory and on our limited experience in data analysis. As shown in Section 6, the eligible projection is significantly improved (i.e., more second-order models can be fitted in the second-stage analysis) after the curse of three-letter words is removed. In addition, three-letter words should be avoided in the defining contrasts, a point echoed in Cheng's discussion.

3. Alternative Criteria

Montgomery and Borror correctly point out that G-efficiency is a pessimistic criterion for measuring the utility of designs. The VDG plot gives a more comprehensive picture of the utility of designs. Its usefulness is hampered by the fact that many plots need to be made, and a succinct summary from so many plots may pose a problem. For example, when the $OA(18, 3^7)$ is projected onto three factors, 35 VDG plots are needed.

Cheng raises the possibility of using the generalized minimum aberration criterion, or its modification, as a surrogate for the projection-efficiency criteria in the paper. This is an interesting idea. In our work with H. Xu (Cheng, Wu, and Xu (2001)), it is shown that eligible projection and estimation efficiency can be combined in one criterion through the frequencies of moment aberration of projected designs. The moment aberration criterion was introduced by Xu (2001). This result suggests that a criterion appropriate to Cheng's discussion can be obtained. Cheng and Mukerjee raise a related question of developing a theory to describe the projection properties found by computer search, which by itself is a challenging problem.

Two other criteria are mentioned: the sum of the squared elements of the alias matrix (Steinberg and Bursztyn) and the weak heredity maximum estimation capacity criterion (Bingham). We speculate that the latter is related to eligible projection, the former to the severity of complex aliasing. Both are interesting and deserve further study.

4. Alternative Designs and Other Design Strategies

Steinberg and Bursztyn suggest a novel class of rotation designs. The idea is ingenious and the designs look very promising. Cheng and Mukerjee raise the possibility of finding nonregular designs that are more efficient for the dual purpose of projection and efficiency. Mukerjee's suggestion of using difference matrix and systematic construction was independently conceived in our work with Xu (Cheng, Wu and Xu (2001)). By exploiting such ideas and using an intelligent algorithmic search driven by the minimum moment aberration criterion proposed by Xu (2001), we have constructed many 3-level and other combinatorial designs that are better than the ones in the paper. An example of such designs is a 27-run design with 8 factors such that all projections onto five factors are eligible. This was also found by Bulutoglu and Cheng. It will be interesting to compare the two approaches to construction.

Montgomery and Borror mention the limitations of assuming second-order models. For simplicity, this paper only considers second-order models and second-order designs but the proposed approach also covers more general designs and models. As mentioned by Huang, an interesting and challenging extension is to develop a comprehensive design theory and analysis strategy for more general models, spline regression models for example.

Khuri raises the possibility of using non-orthogonal designs for factor screening. In general this is not recommended. Non-orthogonal designs will be used only if they can accommodate more factors than orthogonal designs. In such situations the number of factors is too numerous for the dual task of screening and surface exploration to be effective, as discussed in a related comment under Point 1.

Khuri also expresses concern about the checking of model assumptions. Admittedly, sequential designs in the RSM literature are ideal for this purpose. But if a uni-stage experiment is mandated, one can still use a more comprehensive design than second-order design to accommodate a more flexible model.

5. Combinatorial Isomorphism vs. Model Isomorphism

We are glad that this distinction and its novelty were appreciated and recognized by several discussants. This distinction was extensively used in Chapter 5 of Wu and Hamada (2000) to motivate and justify a new data analysis strategy. Bingham speculates that "If one considered interactions with the quadratic terms, there would be no distinction between the two designs". What causes model non-isomorphism is the difference in the geometric structures of the designs. For example, Steinberg and Bursztyn observe that designs in $\mathcal{K}_1$ include a center point while designs in $\mathcal{K}_2$ do not. (More information about the geometric structure of the two designs can be found in Cheng (1999)). For quantitative factors, the issue of model isomorphism arises because different assignments of factor levels to 0, 1, 2 lead to different geometric structures of design points in the space. Therefore the geometric structure of design provides a more fundamental explanation for model non-isomorphism.

In conclusion, we are grateful and pleased that the ideas in our paper have helped generate further ideas, questions and results. These include a deeper understanding of data analysis methods, construction of new designs, and proposal of new design criteria and experimentation strategies. Judging by the enthusiasm and insights of the discussants, we can be hopeful that the modest start made in this paper will open up a new and fruitful direction of research in the design and analysis of experiments, and will eventually make some impact on its practice.

Additional References

Bingham, D. and Li, W. (2001). Resolution IV designs with weak heredity maximum estimation capacity. *Technical Report* **368**, Department of Statistics, University of Michigan.

Box, G. E. P. (1999a). Statistics as a catalyst to learning by scientific method part II – A discussion. *J. Qual. Tech.* **31**, 16-29.

Box, G. E. P. (1999b). Response surface methodology – Current status and future directions – Response. *J. Qual. Tech.* **31**, 67-72.

Box, G. E. P. and Bisgaard, S. (1993). What can you find out from 12 experimental runs? *Qual. Engrg.* **5**, 663-668.

Box, G. E. P. and Draper, N. R. (1975). Robust designs. *Biometrika* **62**, 347-352.

Box, G. E. P. and Liu, P. Y. T. (1999). Statistics as a catalyst to learning by scientific method part I – An example. *J. Qual. Tech.* **31**, 1-15.

Box, G. E. P. and Youle, P. V. (1955). The exploration and exploitation of response surfaces: an example of the link between the fitted surface and the basic mechanism of the system. *Biometrics* **11**, 287-323.

Bulutoglu, D. A. and Cheng, C. S. (2001). Hidden projection properties of some nonregular fractional factorial designs and their applications. In preparation.

Bursztyn, D. and Steinberg, D. M. (2001). Rotation designs for experiments in high bias situations. *J. Statist. Plann. Inference*, in press.

Cheng, C. S. (1995). Some projection properties of orthogonal arrays. *Ann. Statist.* **23**, 1223-1233.

Cheng, C. S. (1998). Some hidden projection properties of orthogonal arrays with strength three. *Biometrika* **85**, 491-495.

Cheng, C. S., Deng, L. Y. and Tang, B. (2000) Generalized minimum aberration and design efficiency for nonregular fractional factorial designs. Submitted for publication.

Cheng, C. S., Steinberg, D. M., and Sun, D. X. (1999) Minimum aberration and model robustness for two-level fractional factorial. *J. Roy. Statist. Soc. Ser. B* **61**, 85-93.

Cheng, S. W., Wu, C. F. J. and Xu, H. (2001). Optimal designs for factor screening and response surface exploration. Preprint.

Chipman, H. and Hamada, M. (1996). Factor-based or effect-based modeling? Implications for design. Discussion of "Follow-up designs to resolve confounding in multifactor experiments" by Meyer, Steinberg and Box. *Technometrics* **38**, 317-320.

Dey, A. and Mukerjee, R. (1999). *Fractional Factorial Plans*. Wiley, New York.

Giovannitti-Jensen, A. and Myers, R. H. (1989). Graphical assessment of the prediction capability of response surface designs. *Technometrics* **31**, 159-171.

Hamada, M. and Wu, C. F. J. (1992). Analysis of designed experiments with complex aliasing. *J. Qual. Technol.* **24**, 130-137.

Lewis, S. M. and Dean, A. M. (2001). Detection of interactions in experiments on large numbers of factors. *J. Roy. Statist. Soc. Ser. B* **63**, in press.

Lin, D. K. J. and Draper, N. R. (1993). Generating alias relationships for two-level Plackett and Burman designs. *Comput. Statist. Data Anal.* **15**, 147-157.

Montgomery, D. C. (2001). *Design and Analysis of Experiments*. 5th edition. John Wiley & Sons, New York.

Myers, R. H., Vining, G. G., Giovannitti-Jensen, A., and Myers, S. L. (1992). Variance dispersion properties of second-order response surface designs. *J. Qual. Tech.* **24**, 1-11.

Shoemaker, A. C., Tsui, K.-L. and Wu, C. F. J. (1991). Economical experimentation methods for robust design. *Technometrics* **33**, 415-427.

Sitter, R. R. and Wu, C. F. J. (1999). Two-stage design of quantal response studies. *Biometrics* **55**, 396-402.

Steinberg, D. M. and Hunter, W. G. (1984). Experimental design: review and comment (with discussion). *Technometrics* **26**, 71-130.

Wang, J. C. and Wu, C. F. J. (1991). An approach to the construction of asymmetrical orthogonal arrays. *J. Amer. Statist. Assoc.* **86**, 451-456.

Xu, H. (2001). Minimum moment aberration for nonregular designs and supersaturated designs. Preprint.