CALCUTTA

STATISTICAL ASSOCIATION

BULLETIN

Vol. 52 (SPECIAL VOLUME), 2002, Nos. 205 - 208

PROCEEDINGS
OF THE
FOURTH INTERNATIONAL TRIENNIAL
CALCUTTA SYMPOSIUM
ON
PROBABILITY AND STATISTICS
December 26 – 28, 2000

Editors
Arijit Chaudhuri and Malay Ghosh

Managing Editor
Manisha Pal

Calcutta Statistical Association Bulletin PP. 47-69
Vol 52 (Special 4-th Triennial Proceedings Volume), 2002, Nos. 205-208

RECENT ADVANCES IN INDUSTRIAL EXPERIMENTATIONS

DENNIS K.J. LIN
Pennsylvania State University, U.S.A.

ABSTRACT: Statistical experimental design has played an important role in industrial investigations. In this paper, we first discuss the major differences between running industrial and agricultural experiments. Thus, classical statistical experimental designs which are suitable for agricultural experimentation may not be appropriate for industrial experimentation. Some recent research, specifically for designing industrial experimentations, are then introduced. This includes computer experiment, impact of the dispersion effect in unreplicated fractional factorial design and optimal foldover plans.

Key Words and Phrases: Computer experiment, Dispersion effect, Foldover design, Rotated factorial.

AMS Subject Classification: Primary 62K15; Secondary 62K20

1. Industrial and Agricultural Experiments

Industrial management is becoming increasingly aware of the benefits of running statistically designed experiments. Statistical experimental designs, developed by Sir R.A. Fisher in the 1920's, largely originated from agricultural problems. Designing experiments for industrial problems and for agricultural problems are similar in the basic concerns. There are, however, many differences. The differences listed in Table 1 are based upon the overall characteristics of all problems. Exceptions can be found in some particular cases, of course.

- Industrial problems tend to contain a much larger number of factors under investigation and usually involve a much smaller number of runs in total.

- Industrial results are more reproducible; that is, industrial problems contain a much smaller replicated variation (pure error) than that of agricultural problems.
- Industrial experimenters are obliged to run their experimental points in sequence and naturally plan their follow-up experiments guided by previous results; in contrast, agricultural problems harvest all results at one time. Doubts and complications can be resolved in industry by immediate follow-up experiments. Confirmatory experimentation is readily available for industrial problems and become a routine procedure to resolve assumptions.
- The concept of blocking arose naturally in agriculture, but often is not obvious for industrial problems. Usually, industrial practitioners need certain specialized training to recognize and handle blocking variables.
- Missing values seem to occur more often in agriculture (mainly due to natural losses) than industry. Usually, such problems can be avoided for industrial problems by well-designed experiments.

Classical designs will remain important to solve daily problems. However, new problems in this IT (*Information Technology*) era require new designs. Given in this paper are some of my recent research results, specifically for industrial experimentation. This may not be an appropriate proposal for some agricultural problems. There are certainly many other important work from other researchers. Here, I will only focus on my own work. Three subjects to be discussed here are: computer experiment, impact of the dispersion effect in unreplicated fractional factorial design, and optimal foldover plans.

Table 1: *Differences Between Agricultural and Industrial Experiments*

Subject	Agriculture	Industry
Number of Factors	Small	Large
Number of Runs	Large	Small
Reproducibility	Less Likely	More Likely
Time Taken	Long	Short
Blocking	Nature	Not Obvious
Missing Values	Often	Seldom

2. Computer Experiment

Computer models are often used to describe complicated physical phenomena encountered in science and engineering. These phenomena are often governed by a set of equations, including linear, nonlinear, ordinary, and partial differential equations. The equations are often too difficult to be solved simultaneously by any person, but can be by a computer modeling program. These programs, due to the number and complexity of the equations, may have long running times, making their use difficult for comprehensive scientific investigation.

One goal in this setting is to build an approximating program which, although not as precise as the computer model, would run fast enough to study the phenomenon in detail. Construction of an adequate approximating function (or program) to the computer model requires the selection of design points (a designed experiment) at which to approximate. Because the computer models are mostly deterministic, these computer experiments require special designs. Standard factorial designs are inadequate here; in the absence of certain main effects, replication cannot be used to estimate random error, but instead produces redundancy. That is, they are hindered by their non-unique projections to lower dimensions. This section presents a new and simple strategy for designs for computer experiments, developed from the rotation of the standard factorial design to yield a Latin hypercube.

2.1 Previous Work

Selection of an appropriate designed experiment depends to an extent on the experimental region, the model to be fit, and the method of analysis. In order to assess design criteria for computer experiments, it is valuable to study the progression of proposed designs. Koehler and Owen (1996) provide an overview of past and current approaches. The two main geometric designs are the standard (full or fractional) factorial designs and the Latin hypercube designs, but also include other traditional designs for physical experiments, such as central composite designs. Easterling (1989) points out that standard factorial designs have many attractive properties for physical experiments: balance (factor levels used an equal number of times), symmetry (permutation of design matrix columns yields same design), orthogonality (separability of main effects), collapsibility (projects to lower subspace as factorial design, sometimes redundantly), equally-spaced projections to each dimension, and straightforward measurability of main effects.

McKay, Beckman, and Conover (1979) introduced the use of the

Latin hypercube (LH) in computer experiments. A n-point LH design matrix is constructed by randomly permuting the integers $\{1, 2, \ldots, n\}$ for each factor and rescaling to the experimental region, so that the points project uniquely and equally-spaced to each dimension. The unique projections of LHs allow for great flexibility in model fitting. Box and Draper (1959) showed that when the true model is a polynomial of unknown degree, the best design places its points evenly spaced over the design region. Thus, equally-spaced projections are also of value. For these reasons, the LH has become the standard for computer experiments. However, random LHs are susceptible to high correlations between factors, even complete confounding, and to omitting regions of the design space.

Computer-generated designs include those of Sacks, Schiller, and Welch (1989) that try to minimize the integrated mean square error (IMSE) of prediction when prediction errors are taken as a realization of a spatial stochastic process. Johnson, Moore, and Ylvisaker (1990) proposed similar designs to minimize the correlations between observations when responses are taken as a realization of a spatial stochastic process. The latter authors' design D^* they call a maximin distance design if

$$\min_{x_1, x_2 \in D^*} d(x_1, x_2) = \max_{D} \min_{x_1, x_2 \in D} d(x_1, x_2), \tag{2.1}$$

where d is a distance measure and $\min_{x_1, x_2 \in D} d(x_1, x_2)$ is the minimum interpoint distance (MID) of design D; that is, its points are moved as far apart from one another as possible. Attempts have been made to bridge the gap between geometric designs and computer-generated designs. However, being themselves computer-generated designs leaves many susceptible to the aforementioned problems. With this in mind, we seek a new design for computer experiments with these properties: the unique and equally-spaced projections to each dimension and flexibility in model selection provided by Latin hypercube design and the orthogonality and ease of construction provided by standard factorial designs. In addition, these new designs should perform reasonably well in terms of other criteria mentioned, such as MID correlation and coverage of the design space.

2.2 Rotated Factorial Designs in Two Dimensions

The strategy taken here is to modify the standard factorial design by rotation so as to yield a Latin hypercube. To see how this is done, first consider the standard 3^2 factorial design, represented by the 3×3 square of points in Figure 1, and how it can be rotated to yield equally-spaced projections. The key to finding all such rotations is in the relationship

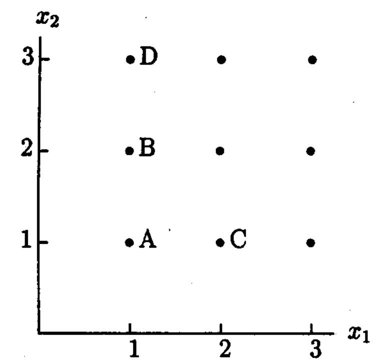

Figure 1: Standard 3^2 factorial design before rotation

between points A-D. We focus on nontrivial angles between 0 and 45 degrees clockwise due to the symmetry of the rotation problem.

The matrix equation to rotate a set of points clockwise by an angle w about the origin is

$$\begin{bmatrix} x_1 & x_2 \end{bmatrix} \times \begin{bmatrix} \cos(w) & -\sin(w) \\ \sin(w) & \cos(w) \end{bmatrix}$$

so that if (x_1, x_2) are the coordinates of a design point in the standard factorial design, then the rotation moves the point to $(x_1 \cos(w) + x_2 \sin(w), -x_1 \sin(w) + x_2 \cos(w))$. Notice first that as the points are rotated clockwise about the origin that A will have the smallest x_1-coordinate for any angle between 0° and 45°. (A 45° rotation will place A directly on the x_1-axis and A is the closest point to the origin.) Also notice that the x_1-projections of points with the same initial x_1-coordinate (like A, B, and D) will be equally spaced, by $\sin(w)$, regardless of the rotation angle. Likewise, the x_1-projections of points with the same initial x_2-coordinate (like A and C) will be equally spaced, by $\cos(w)$, regardless of the rotation angle. It suffices to find all angles that make the x_1-projections of points A-D equally spaced. For the x_1-coordinates of A-D, see the table below.

point	x_1-coordinate
A	$\cos(w) + \sin(w)$
B	$\cos(w) + 2\sin(w)$
C	$2\cos(w) + \sin(w)$
D	$\cos(w) + 3\sin(w)$

Between $0°$ and $45°$, $\sin(w) \leq \cos(w)$, so the point with the next smallest x_1-coordinate will always be B (although C will tie B when $w = 45°$) and the distance between the smallest two x_1-projections will always be $\sin(w)$. To achieve equally-spaced x_1-projections, the distance between all x_1-projections must equal $\sin(w)$. We've already seen that this is true when $w = 45°$ (equivalently, $\tan^{-1}(1)$) and both C and B have the second smallest x_1-coordinate.

Another possibility is that C will have the third smallest x_1-coordinate, and that the "x_1-distance" between B and C will be $\sin(w)$. However, the "x_1-distance" between B and D is always $\sin(w)$. In this case, C and D will have the same x_1-coordinate, hence $\cos(w) = 2\sin(w) \implies w = \tan^{-1}(1/2)$. Continuing in this manner, consider the case where C has the fourth smallest x_1-coordinate – after A, B, and D – and the "x_1-distance" between D and C is $\sin(w)$. Then $\cos(w) - 2\sin(w) = \sin(w) \implies w = \tan^{-1}(1/3)$. Point C cannot have the fifth smallest x_1-coordinate, so these three rotations are the only ones (again, among nontrivial angles between $0°$ and $45°$) that yield equally-spaced x_1-projections from the 3^2 design. It is easily verified that these also yield equally-spaced x_2-projections.

Figure 2 displays the standard 3^2 factorial design, shown in open circles, and the designs that result from these rotations, shown in solid circles. Boxes are drawn around the rotated designs to identify the design regions. In practice, one would then scale this design (by subtraction and division) to the experimental region of interest. Along each axis, we have provided dot plots of the projections from which it is plain to see the equally-spaced property.

Among the rotated standard p^2 factorial designs with equally-spaced projections, only those obtained from rotation angles of $\tan^{-1}(1/p)$ contain no redundant projections. Therefore, we define a p^2-point rotated full factorial design to be a rotated standard p^2 factorial design with unique, equally-spaced projections to each dimension (which is a Latin hypercube). Following the argument above, a general result for factorial designs can be stated.

Theorem 1: *For nontrivial rotations between* $0°$ *and* $45°$, *a rotated standard* p^2 *factorial design will produce equally-spaced projections to each dimension if and only if the rotation angle is* $\tan^{-1}(1/k)$, *where* $k \in \{1, \ldots, p\}$.

2.3 High-Dimensional Rotation Theory

Consider a standard full factorial design consisting of d factors, each with p levels. The goal is to rotate this design to convert it into a LH design, so that the p^d points create unique and equally-spaced projections to each individual factor. For certain values of d (notably when d is a power of 2) such a rotation exists, but not for general d. The following proof proceeds in three parts: identification of the required form of the rotation matrix, construction of the power-of-2 rotation matrix, and failure of the transformation matrix to be a rotation matrix when d is not a power of two.

A p-level, d-factor standard full factorial design can be represented by a $p^d \times d$ matrix, D, with entries from $\{1, 2, \ldots, p\}$ and all p^d combinations represented.

$$D= \left[\begin{array}{ccccccccccccccccccc} 1 & 1 & & & \cdots & & & & 1 & \cdots & p & p & & & \cdots & & & & p \\ \vdots & & & & & & & & & \vdots & & & & & & & & & \vdots \\ 1 & 1 & \cdots & 1 & \cdots & p & p & \cdots & p & \cdots & 1 & 1 & \cdots & 1 & \cdots & p & p & \cdots & p \\ 1 & 2 & \cdots & p & \cdots & 1 & 2 & \cdots & p & \cdots & 1 & 2 & \cdots & p & \cdots & 1 & 2 & \cdots & p \end{array}\right]^T \quad (2.2)$$

A rotation of this matrix is accomplished by post-multiplication by a $d \times d$ matrix R with the property that $R^T R = I_d$ where I_d is the $d \times d$ identity matrix. (In this section, we relax the definition of rotation to be a matrix R that satisfies $R^T R = kI_d$ for some scalar k, since the true rotation can be obtained as $(1/\sqrt{k})R$.) Let the multiplication matrix R have entries denoted as $r_{[i,j]}$, which is the entry from the ith row and jth column. Lemma 1 below will not be concerned with whether the multiplication matrix is indeed a rotation matrix, but with how such a matrix would yield unique and equally-spaced projections to each dimension.

Lemma 1: The entries of each column of the transformation matrix R must be unique from the set $\{p^t | t = 0, 1, \ldots, d-1\}$ in order to yield unique and equally-spaced projections.

The previous lemma shows that every column of the transformation matrix must be a permutation of the set $\{1, p, \ldots, p^{d-1}\}$ (allowing sign changes to elements and multiplication of entire columns by a constant).

However, every rotation matrix R satisfies $R^T R = kI_d$, so that the sum of squares for all columns of R must be equal. Then, WLOG, every column of the transformation matrix must be a permutation of the set $\{1, p, \ldots, p^{d-1}\}$ (allowing only sign changes to elements). It is obvious that the columns of the transformation matrix cannot be identical, for otherwise the columns of the transformed matrix would be identical. The following lemma shows that the ith entries for the d columns must be unique in magnitude in order for the transformation to be a rotation.

Lemma 2: For a rotation matrix R, the ith entries of the d columns are unique in magnitude for all i.

Lemmas 1 and 2 proved that all the rows and columns of the transformation matrix must be permutations of the set $\{1, p, \ldots, p^{d-1}\}$ (up to sign changes). However, this is not sufficient to guarantee that the matrix will also be a rotation. Another requirement implied by the rotation condition $R^{\mathrm{T}} R = kI_d$ is that the columns of R must be orthogonal. Any matrix satisfying the requirements of the lemmas and this last condition will rotate factorial designs into Latin hypercubes. The remainder of this chapter shows how to create these matrices for d that are powers of two and illustrates why other choices of d, in general, have no such rotation matrix.

Let d be a power of 2. Let $c = \log_2 d$. Let

$$V_1 = [v_1 \quad v_2] = \begin{bmatrix} +1 & +p \\ +p & -1 \end{bmatrix}. \tag{2.3}$$

Now, for $c > 1$, let V_c be defined inductively from V_{c-1} as follows:

$$V_c = \begin{bmatrix} V_{c-1} & -(p^{2^{c-1}} V_{c-1})^* \\ p^{2^{c-1}} V_{c-1} & (V_{c-1})^* \end{bmatrix} \tag{2.4}$$

where the operator $(\cdot)^*$ works on any matrix with an even number of rows by multiplying the entries in the top half of the matrix by -1 and leaving those in the bottom half unchanged.

Theorem 2: *The matrix V_c is a rotation of the d-factor ($d = 2^c$), p-level standard full factorial design which yields unique and equally-spaced projections to each dimension.*

Reviewing the two-dimensional result from section 2.2, when $d = 2 = 2^1$, with $w = \tan^{-1}(1/p)$ we have

$$V_1 = \begin{bmatrix} \cos(\tan^{-1}(1/p)) & -\sin(\tan^{-1}(1/p)) \\ \sin(\tan^{-1}(1/p)) & \cos(\tan^{-1}(1/p)) \end{bmatrix} = \frac{1}{\sqrt{1+p^2}} \begin{bmatrix} +1 & -p \\ +p & +1 \end{bmatrix}, \tag{2.5}$$

which is the correctly scaled rotation matrix V_1 given in equation (2.3).

Other scaled rotation matrices for cases of interest ($d = 4, 8$ corresponding to $c = 2, 3$) are

$$V_2 = \sqrt{\frac{p^2-1}{p^8-1}} \begin{bmatrix} +1 & -p & +p^2 & -p^3 \\ +p & +1 & -p^3 & -p^2 \\ +p^2 & -p^3 & -1 & +p \\ +p^3 & +p^2 & +p & +1 \end{bmatrix} \tag{2.6}$$

and

$$V_3 = \sqrt{\frac{p^2-1}{p^{16}-1}} \begin{bmatrix} +1 & -p & +p^2 & -p^3 & +p^4 & -p^5 & +p^6 & -p^7 \\ +p & +1 & -p^3 & -p^2 & +p^5 & +p^4 & -p^7 & -p^6 \\ +p^2 & -p^3 & -1 & +p & -p^6 & +p^7 & +p^4 & -p^5 \\ +p^3 & +p^2 & +p & +1 & -p^7 & -p^6 & -p^5 & -p^4 \\ +p^4 & -p^5 & +p^6 & -p^7 & -1 & +p & -p^2 & +p^3 \\ +p^5 & +p^4 & -p^7 & -p^6 & -p & -1 & +p^3 & +p^2 \\ +p^6 & -p^7 & -p^4 & +p^5 & +p^2 & -p^3 & -1 & +p \\ +p^7 & +p^6 & +p^5 & +p^4 & +p^3 & +p^2 & +p & +1 \end{bmatrix} \tag{2.7}$$

respectively.

The choice of rotation matrices for higher dimensions ($d > 2$) is not unique. Other inductive definitions for V_c in equation (2.4) are possible, namely

$$\begin{bmatrix} V_{c-1} & -p^{2^{c-1}} V_{c-1} \\ p^{2^{c-1}} V_{c-1} & V_{c-1} \end{bmatrix} \tag{2.8}$$

However, the point is still clear, such rotations do exist.

Theorem 3 below proves that all designs obtained by rotation of standard factorial designs, specifically rotated full factorial designs, will also be orthogonal. Let k be the sum of squares of the first column of X. As X is an orthogonal matrix, $X^{\mathrm{T}} X = kI_d$. So $(XR)^{\mathrm{T}}(XR) = R^{\mathrm{T}} X^{\mathrm{T}} XR = R^{\mathrm{T}} kI_d R = kR^{\mathrm{T}} R = kI_d$, a diagonal matrix. Therefore, the rotated design matrix XR is an orthogonal design.

Theorem 3: *Let X be an orthogonal design matrix of n rows and d columns in which the sums of squares for columns are equal. Let R be a $d \times d$ rotation matrix. The design resulting from the matrix product XR is also an orthogonal design.*

Recall that Johnson et al. (1990) introduced the use of minimum interpoint distance (MID) as an important design criterion (see equation (2.1)). It can be shown that the MID using Euclidean distance for a p^d-point rotated factorial design scaled to the unit hypercube, $[0,1]^d$, is $\sqrt{1+p^2+\cdots,+p^{d+1}/(p-1)} = \sqrt{(p^{2d}-1)/((p^2-1)(p-1)^2)}$. Additionally, it can be shown this is the maximal MID for $d=2$. We are unable to obtain a formal proof for higher dimensions, however. For other theoretical properties, see Beattie and Lin (1998).

3. Impact of Dispersion Effects

When studying both location and dispersion effects in unreplicated fractional factorial designs, a "standard" procedure is to identify location effects using ordinary least squares analysis, fit a model, then identify dispersion effects by analyzing the residuals. Traditionally, the primary use of these designs has been in detecting and modeling location effects (changes in the mean response). An assumption of constant variance is usually made. In this section, we show that if the model in the above procedure does not include all active location effects, then null dispersion effects may be mistakenly be identified as active. We also derive an exact relationship between location and dispersion effects.

3.1 An Illustrative Example

Montgomery (1990) analyzed data from an injection molding experiment where the response to be optimized was shrinkage. The factors studied were mold temperature (A), screw speed (B), holding time (C), gate size (D), cycle time (E), moisture content (F), and holding pressure (G). The design is a 2^{7-3}_{IV} fractional factorial, meaning it is a resolution IV, $1/2^3$ fraction of a 2^7 design. (See Box, Hunter, and Hunter (1978).) The generators of this design are E=ABC, F=BCD, and G=ACD. The data are shown in Table 2.

The least squares regression coefficients were obtained from fitting a saturated model. In 2^{k-p} experiments, "effects" are calculated as the average difference in the response at the +1 and −1 levels of the column. Here, $effect_j = 2\beta_j$. Montgomery used a normal probability plot of the estimated effects and determined that columns 1, 2, and 5 (A, B, and AB) produce active location effects. He fit this location model, which we denote M1.

$$(M1) \qquad \hat{y} = 27.3125 + 6.9375A + 17.8125B + 5.9375AB$$

Table 2: Design Matrix and Response for Injection Molding Experiment

i \ j	0	A 1	B 2	C 3	D 4	5	6	7	8	9	10	E 11	12	G 13	F 14	15	y_M
1	1	−	−	−	−	+	+	+	+	+	+	−	−	−	−	+	6
2	1	+	−	−	−	−	−	−	+	+	+	+	+	+	−	−	10
3	1	−	+	−	−	−	+	+	−	−	+	+	+	−	+	−	32
4	1	+	+	−	−	+	−	−	−	−	+	−	−	+	+	+	60
5	1	−	−	+	−	+	−	+	−	+	−	+	−	+	+	−	4
6	1	+	−	+	−	−	+	−	−	+	−	−	+	−	+	+	15
7	1	−	+	+	−	−	−	+	+	−	−	−	+	+	−	+	26
8	1	+	+	+	−	+	+	−	+	−	−	+	−	−	−	−	60
9	1	−	−	−	+	+	+	−	+	−	−	−	+	+	+	−	8
10	1	+	−	−	+	−	−	+	+	−	−	+	−	−	+	+	12
11	1	−	+	−	+	−	+	−	−	+	−	+	−	+	−	+	34
12	1	+	+	−	+	+	−	+	−	+	−	−	+	−	−	−	60
13	1	−	−	+	+	+	−	−	−	−	+	+	+	−	−	+	16
14	1	+	−	+	+	−	+	+	−	−	+	−	−	+	−	−	5
15	1	−	+	+	+	−	−	−	+	+	+	−	−	−	+	−	37
16	1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	52

The estimated residuals under M1 are (-2.50, -0.50, -0.25, 2.00, -4.50, 4.50, -6.25, 2.00, -0.50, 1.50, 1.75, 2.00, 7.50, -5.50, 4.75, -6.00). As a measure of the dispersion effect magnitude for column j, Montgomery calculates the statistic $F_j^* = ln\frac{s^2_{j+}}{s^2_{j-}}$ which is the natural logarithm of the ratio of the sample variances of the residuals at the +1 and −1 levels of column j. Note that Box and Meyer (1986b) point out this statistic is approximately normally distributed with mean 0 and variance 1. Montgomery compared these statistics to an appropriate normal quantile to determine significance. He also used a normal plot of these statistics. Using either the normal quantile or the probability plot, it is evident that column 3 (C) has a dispersion effect with

$$F^*_{3|M1} = ln\frac{s^2_{3+|M1}}{s^2_{3-|M1}} = ln\frac{32.44}{2.66} = 2.50.$$

Thus, Montgomery (1990) concludes that factors A (mold temperature) and B (screw speed) impact the mean shrinkage of the mold and that factor C (holding time) impacts the variation in shrinkage. By studying the interaction between mold temperature and screw speed, it is apparent that the low screw speed is better for reducing mean shrinkage and that the setting of mold temperature is not crucial at this speed. To reduce the variation in shrinkage, holding time should be set at its low level.

This logical procedure has been used by many and has become a standard practice. However, the identification of dispersion effects is quite sensitive to the location model that is fit. To illustrate, note that another reasonable interpretation from the normal plot is that columns 7 and 13 have active location effects in addition to columns 1, 2, and

5 (see McGrath and Lin (2001b) for details). Due to the confounding associated with this design, column 13 represents not just the factor G effect alone, but also the ACD interaction and other effects. The AD interaction effect appears in column 7 and the interaction of columns 7 and 13 appears in column 3.

We denote this model with five location effects (columns 1, 2, 5, 7, and 13) as M2.

$$(M2) \quad \hat{y} = 27.3125 + 6.9375A + 17.8125B + 5.9375AB - 2.6875AD - 2.4375ACD.$$

The residuals from model M2 are (-2.250, -0.750, 0.000, 1.750, 0.625, -0.625, -1.125, -3.125, -0.750, 1.750, 1.500, 2.250, 2.375, -0.375, -0.375, -0.875). From this model we have the F_j^* statistic for column 3,

$$F^*_{3|M2} = ln\frac{s^2_{3+|M2}}{s^2_{3-|M2}} = ln\frac{2.42}{2.58} = -0.06.$$

Here, it is apparent there is no dispersion effect associated with column 3 (factor C) as the sample variance of residuals is quite similar at the -1 and $+1$ levels of column 3.

So we have two feasible models for mold shrinkage, M1 and M2. M1 shows two factors important for determining the location (mean) of the response, and also includes another factor that is important for controlling the variation in the response. M2 includes four factors that affect the mean response and no dispersion factors. Which model is more appropriate? Is one model better than the other? Some additional information may be helpful. The experiment actually included four center points (25, 29, 24, 27) in addition to the fractional factorial. From these center points, we have an estimate of the variance of the response, σ^2, of $\hat{\sigma}^2 = 4.92$. M1 produces $\hat{\sigma}^2_{M1} = 20.73$ and M2 produces $\hat{\sigma}^2_{M2} = 3.81$. The M2 estimate is in much better agreement with the center point estimate.

Therefore, a reasonable conclusion based on model M2 is that there are four important factors: mold temperature, screw speed, holding time and gate size (D). If this experiment is truly a screening experiment, then fitting M1 would have eliminated a potentially important factor, gate size. So we have two distinctly different possibilities: (1) failing to include a pair of location effects created a spurious dispersion effect, or (2) failing to account for a dispersion effect created two location effects. These spurious dispersion effects are not uncommon.

It can be shown that the exclusion of a pair of active location effects will create an apparent (spurious) dispersion effect in the interaction of these two columns. Box and Meyer (1986a) and Bergman and Hynèn (1997) both noted a relationship between location and dispersion effects. We next provide a theoretical explanation showing that failure to include two location effects in a model before calculating residuals can produce a spurious dispersion effect.

3.2 Spurious Dispersion Effects

Assume some method is used to identify m active location effects in an unreplicated fractional factorial design. A model is fit and residuals are estimated, but assume there are two active location effects that are excluded from this model. Let the excluded active location effects be in columns $\mathbf{x_j}$ and $\mathbf{x_{j'}}$ and let $\mathbf{x_d}$ be the column associated with the interaction of $\mathbf{x_j}$ and $\mathbf{x_{j'}}$. Then $x_{ij}x_{ij'} = x_{id}$. Let $\hat{\beta}_j$ and $\hat{\beta}_{j'}$ be the usual least squares estimators of β_j and $\beta_{j'}$, the regression coefficients associated with $\mathbf{x_j}$ and $\mathbf{x_{j'}}$ respectively. We will show that failure to include β_j and $\beta_{j'}$ in the regression model will create a difference in the expected value of the sample variances at the $+1$ and -1 levels of $\mathbf{x_d}$.

Define the following sets of rows using the convention P for 'plus' and M for 'minus': $M = \{i : x_{id} = -1\}$, $P = \{i : x_{id} = +1\}$. A dispersion effect occurs when the variance of the response, independent of the location effects, (or equivalently, the variance of the residuals from a known location model) is higher at one level of a column than the other. We can compare sample variances of the residuals at the plus and minus levels of a column to determine if it has a dispersion effect. Let

$$s^2_{d+} = \frac{2}{n-2}\sum_{i\in P}(e_i - \bar{e}_p)^2 \text{ and } s^2_{d-} = \frac{2}{n-2}\sum_{i\in M}(e_i - \bar{e}_m)^2,$$

where, $\bar{e}_m = \frac{2}{n}\sum_{i\in M} e_i$ and $\bar{e}_p = \frac{2}{n}\sum_{i\in P} e_i$. It can be shown that the expected sample variance of the residuals when $x_{id} = -1$ $(i \in M)$ is

$$E\left[s^2_{d-}\right] = E\left[\frac{2}{n-2}\sum_{i\in M}(e_i - \bar{e}_m)^2\right] = \frac{n-1-m}{n-1}\sigma^2 + \frac{n}{n-2}(\beta_j - \beta_{j'})^2 \tag{3.1}$$

and when $x_{id} = +1$ $(i \in P)$,

$$E\left[s^2_{d+}\right] = E\left[\frac{2}{n-2}\sum_{i\in P}(e_i - \bar{e}_p)^2\right] = \frac{n-1-m}{n-1}\sigma^2 + \frac{n}{n-2}(\beta_j + \beta_{j'})^2. \qquad (3.2)$$

From (3.1) and (3.2) we have

$$E\left[s^2_{d+}\right] - E\left[s^2_{d-}\right] = \frac{4n}{n-2}\beta_j\beta_{j'}. \qquad (3.3)$$

Thus, consider the following three scenarios involving β_j and $\beta_{j'}$:

- If $\beta_j = \beta_{j'} = 0$, then these two location effects are not active and $E\left[s^2_{d-}\right] = E\left[s^2_{d+}\right] = \frac{n-1-m}{n-1}\sigma^2$ and $E\left[s^2_{d+}\right] - E\left[s^2_{d-}\right] = 0$. Thus, any difference is just random error so there will should be no spurious dispersion effect.
- If only one of the coefficients is nonzero, then (3.3) is still zero as mentioned in Bergman and Hynèn (1997), although both are biased upwards as estimates of σ^2.
- If β_j and $\beta_{j'} \neq 0$, the residuals will have different expected variance at the -1 and $+1$ levels of $\mathbf{x_d}$. Thus, excluding two location effects from a model and then studying residuals can create a spurious dispersion effect.

Returning to the injection molding example, if we assume columns 7 and 13 produce active location effects but were left out of the model, then we have

$$\begin{aligned} E\left[s^2_{3+|M1}\right] - E\left[s^2_{3-|M1}\right] &= \frac{4n}{n-2}\hat{\beta}_7\hat{\beta}_{13} \\ &= \frac{(4)(16)}{14}(-2.6875)(-2.4375) \doteq 29.95 \end{aligned}$$

Recalling that $s^2_{3-|M1} = 2.66$ and $s^2_{3+|M1} = 32.44$, we have

$$s^2_{3+|M1} - s^2_{3-|M1} = 29.79.$$

So the observed difference in sample variances is almost the same as that caused by not including β_7 and β_{13} in the model. This indicates the dispersion effect detected by fitting model M1 is spurious.

3.3 Theoretical Summary

- McGrath and Lin (2001a) shows that (1) failing to include a pair of location effects creates a spurious dispersion effect in its interaction column; and (2) two dispersion effects create a dispersion effect in their interaction column. They also provide a way to simultaneously analyze the location and dispersion effects.
- Let $\hat{\beta}_j$ and $\hat{\beta}_{j'}$ be the OLS estimates for columns $\mathbf{x_j}$ and $\mathbf{x_{j'}}$ respectively in a 2^{k-p} experiment. If the interaction of $\mathbf{x_j}$ and $\mathbf{x_{j'}}$ is in column $\mathbf{x_d}$, $Var(\epsilon_i|x_{id} = -1) = \sigma^2_{d-}$ and $Var(\epsilon_i|x_{id} = 1) = \sigma^2_{d+}$, then the correlation of $\hat{\beta}_j$ and $\hat{\beta}_{j'}$ is

$$\rho_{j,j'|d} = \frac{\sigma^2_{d+} - \sigma^2_{d-}}{\sigma^2_{d+} + \sigma^2_{d-}}. \qquad (3.4)$$

- Let m be the number of active location effects in the model fit from a 2^{k-p} experiment. Let $g =$ the number of alias pairs $(\mathbf{x_j}, \mathbf{x_{j'}})$ not in the model such that $x_{ij}x_{ij'} = x_{id}$ for $i = 1, \ldots, n$. Then s^2_{d+} and s^2_{d-} are independent if and only if $g = (n-1-m)/2$ and $\mathbf{x_d}$ is in the effect matrix for the fitted model.

4. Foldover Plan

A standard follow-up strategy discussed in many textbooks involves adding a second fraction, which is called a foldover design (or simply foldover), by reversing the signs of one or more columns of the initial design (e.g., Box, et al., 1978; Montgomery, 2001; Neter, et al., 1996; Wu and Hamada, 2000). Here, we develop optimal foldover plans for commonly used fractional factorial designs. The criterion we use is the aberration of the combined design. (A combined design refers to the combination of the initial design and its foldover.) Note that foldovers may be constructed for various reasons. If the analysis of the initial design reveals a particular set of main and interaction effects that are significant, then the foldover design should be chosen to resolve confounding problems with these significant effects. For example, if one particular factor is very important and should not be confounded with other factors, then a foldover based on reversing the sign of this factor is appropriate. (or as many as possible) main effects from two-factor interactions, and (b) to de-alias as many as possible two-factor interactions from each other, then the aberration criterion appears to be a good choice.

Note that the aberration criterion has been used (sometimes implicitly) among the existing foldover strategies. For example, a commonly used foldover strategy for a resolution III design involves reversing the signs of all factors. This is usually considered to be a good strategy because the resulting combined design has resolution IV—which is higher than the resolution of the initial design. This section demonstrates that the use of the aberration criterion can lead to further improvement. The combined design may have a higher resolution or the same resolution with fewer numbers of two-factor interactions which are confounded with each other.

4.1 Existing Work

Let w_i denote the number of words of length i in the defining relation of a design $\mathbf{d}$. The vector $W(\mathbf{d}) = (w_1, w_2, w_3, \ldots, w_k)$ is called the word length pattern (WLP) of the design. (For simplicity, only $(w_3, \ldots, w_7)$ of WLP's are displayed in this article.) The resolution of $\mathbf{d}$ is defined as the smallest r such that $w_r \geq 1$. For any two designs $\mathbf{d}_1$ and $\mathbf{d}_2$, let s be the smallest integer such that $w_s(\mathbf{d}_1) \neq w_s(\mathbf{d}_2)$. Then $\mathbf{d}_1$ is said to have less aberration than $\mathbf{d}_2$, denoted by $W(\mathbf{d}_1) < W(\mathbf{d}_2)$, if $w_s(\mathbf{d}_1) < w_s(\mathbf{d}_2)$. When there is no design with less aberration than $\mathbf{d}_1$, $\mathbf{d}_1$ has minimum aberration.

Denote a foldover plan γ as the collection of columns whose signs are to be reversed in the foldover design, then each foldover is generated by a foldover plan. For example, $\gamma = 456$ produces a foldover design by reversing the signs of factors 4, 5, and 6. A classic approach to constructing a foldover design is to reverse the signs of all k factors. We call this type of foldover plan a full-foldover plan and denote it by $\gamma^f = 1 \cdots k$. The corresponding foldover design $\mathbf{d}'(\gamma^f)$ is called a full-foldover. Most popular statistical software packages (e.g., SAS) take this approach. The combined design generated by this foldover plan, however, may not be optimal with respect to its WLP. Consider a fractional factorial 2^{7-2}_{IV} design generated by: $6 = 1234$ and $7 = 1245$. When the signs of all 7 factors are reversed, the combined design has $W = (0, 1, 0, 0, 0)$. This is a resolution IV design that has a pair of two-factor interactions that are fully aliased. A quick search reveals that $\gamma = 6$ produces a resolution V design with $W = (0, 0, 1, 0, 0)$, namely, all two-factor interactions are clear.

Other foldover plans have also been proposed in the literature. Sign-reversal of one factor was considered in, for example, Box et al. (1978) and Wu and Hamada (2000). Montgomery and Runger (1996) considered foldovers generated by reversing the signs of one or two factors. For resolution IV designs, the rule of reversing signs of all factors is not directly applicable because the resulting combined design will have the same number of length-4 words. Some software packages consider different foldover strategies for these designs. For example, in Design-Expert V6, it is suggested that the sign of a single column be reversed. Another software package, called RS/Discover, suggested reversing the sign of the factor if the generator in which this factor is involved is an odd-length word. It is not clear, however, whether any of these previously given foldovers is optimal (with respect to WLP of the combined design). To our knowledge, the optimality of foldover designs has not been addressed in the literature.

4.2 Construction of Optimal Foldovers

Denote the initial design, the foldover, and the combined design by $\mathbf{d}$, $\mathbf{d}'$, and $\mathbf{D}$, respectively. The optimal foldover plan γ^* is the one such that $W(\mathbf{D}(\gamma^*)) = \min_{\gamma \in \Gamma} W(\mathbf{D}(\gamma))$, where $\Gamma = \{\gamma_1, \ldots, \gamma_q\}$ is the foldover plan space and q is the total number of possible foldover plans. The resulting foldover $\mathbf{d}'(\gamma^*)$ and combined design $\mathbf{D}(\gamma^*)$ are called optimal foldover and optimal combined design, respectively. Given a 2^{k-p} design, the optimal foldover plan γ^* can be found by searching all $q = 2^k$ possible foldover plans. Note, however, that many of these q foldover plans produce the same foldover design. We call them equivalent foldover plans. Consider, for example, a 2^{5-1}_V design defined by a generating relation 5=1234. Obviously, the foldover plans $\gamma_i = i$ $(i = 1, 2, 3, 4)$ are equivalent to each other. And they are all equivalent to a foldover plan $\gamma_c = 5$, which only involves the generated factor of the design—factor 5. (Without loss of generality, we use $1, \ldots, k-p$ to denote the basic factors and $k-p+1, \ldots, k$ to denote the generated factors.) In general, if a foldover plan γ_c consists only of the generated factors, we call it a core foldover plan. An important property is that every foldover plan is equivalent to a specific core foldover plan (the proof is given in Li and Lin, 2000):

Theorem 4: *For a 2^{k-p} design with p generators $G_1, \ldots, G_p$, any foldover plan is equivalent to a core foldover plan. Moreover, for every core foldover plan, there are 2^{k-p} foldover plans that are equivalent to it.*

Based on Theorem 4, we present an algorithm to search for optimal foldover plans. The algorithm is an exhaustive search method based on a specific set of the p generated factors. Note that the number of candidate foldover plans is only 2^p—a fraction of the total number of candidate foldover plans 2^k. The computer program consists of the following steps:

1. Input n, k, and p of the initial design $\mathbf{d}$.
2. Generate all $2^p - 1$ defining words of $\mathbf{d}$ for a given set of $G_1, \ldots, G_p$.
3. For each core foldover plan γ_i, $(i = 1, \ldots, 2^p)$.
   (a) Consider all defining words of $\mathbf{d}$. For each word, if there is an even number of factors whose signs are reversed by γ_i, this word is retained in the defining relation of the combined design $\mathbf{D}(\gamma_i)$; otherwise, the word is deleted.
   (b) Compare $W(\mathbf{D}(\gamma_i))$ with $W(\mathbf{D}(\gamma^*))$ where γ^* is the best core foldover plan among those that are considered before γ_i. Update γ^* when $W(\mathbf{D}(\gamma_i)) < W(\mathbf{D}(\gamma^*))$.
4. Output γ^* and $W(\mathbf{D}(\gamma^*))$.

We use this algorithm to construct the optimal foldovers for 16- and 32-run designs. Although fractional factorial designs with the minimum aberration are commonly used in practice, in some situations other designs can meet practical needs better. Chen et al. (1993) presented a catalog of complete 16-run designs and selected 32-run designs. Finding optimal foldovers of these designs would be important and useful for practitioners. Thus, by using the computer search method described in this section, we constructed optimal foldovers of all these designs. The methodology described in this article is applicable to any 2^{k-p} fractional factorial design. We have focused here on 16-run and 32-run designs with $k \leq 11$ because most standard textbooks give designs of up to 11 factors and foldovers of designs with $n \geq 64$ are rarely used in practice. Foldovers of other (larger) designs, however, can be constructed in a straightforward manner. These optimal foldover plans for all 16-run and 32-run designs are given in Li and Lin (2000).

4.3 Major Findings

For most designs, there exist better foldover plans than the classic full-foldover plans. In 52 out of 77 cases we have found better foldover plans than the corresponding full-foldover plans. While some of them may be obtained by previously reported methods in the literature, most are new. It shows the numbers of optimal foldover designs that are better than the full-foldovers for each given set of (k, p).

For resolution III designs, the full-foldover plans produce combined resolution IV designs. An optimal foldover plan can further improve this desirable property by two means. Firstly, it may further increase the resolution of the combined design $\mathbf{D}$. One example is Design 7–2.5, for which the WLP of the combined design from full-foldover is $W(\mathbf{D}(\gamma^f))$=(0,1,0,0,0), whereas the optimal combined design has $W(\mathbf{D}(\gamma^*))$=(0,0,0,0,1). Secondly, it may lead to a resolution IV design with fewer length-4 words. Thus, the optimal combined design can de-alias more two-factor interactions from each other. Consider, for example, Design 7–3.2, for which the full-foldover plan produces a combined design with three length-4 words $I = 2356 = 2347 = 4567$. But the corresponding optimal combined design has only one length-4 word $I = 2356$.

Although foldovers of resolution III designs are more common in practice, augmenting resolution IV designs can sometimes be important as well. Such examples were discussed in Montgomery and Runger (1996). The objective here is to de-alias two-factor interactions from each other. We find that the improvement over the full-foldover plans from the optimal foldover plans is usually substantial. For example, optimal foldovers of the minimum-aberration designs 8–3.1, 9–4.1, 10–5.1, and 11–6.1 de-alias 2, 4, 6, and 15 out of 3, 6, 10, and 25 pairs of two-factor interactions, respectively. The percentages of de-aliased pairs of two-factor interactions by optimal foldovers of non-minimum-aberration designs are also in the range of 60%–80%. Notable exceptions are Designs 7–2.1 and 7–2.2, for which the optimal combined designs have resolution V and VI, respectively. This demonstrates that augmenting resolution IV designs may also produce designs with a higher resolution.

We conclude this section by giving two remarks. Firstly, one disadvantage of the foldover design is that the run size may become large in some situations. In these cases, partial foldovers proposed by Mee and Peralta (2000) can be considered. Optimal partial foldover plans are currently under study. Mee and Peralta (2000) pointed out that foldover designs are sometimes inefficient. Such an argument, however, is valid for conventional full-foldover, but may not apply to the optimal foldover plan given here. Secondly, there are various reasons for using a foldover design. Thus, the aberration criterion should not be considered as the only criterion. However, in the situation where the use of aberration criterion is justified, the optimal foldover plans presented in Li and Lin (2000) are recommended. We also note that the proposed approach can be applied to other design criteria in a straightforward manner.

5. Discussion

This paper introduced some recent developments in industrial experimentation. Section 2 discusses the popular computer experiment and its related issues. A class of design suitable for computer experiments is proposed. Section 3 discusses the impact of dispersion effect in analyzing the screening designs, using 2^{k-p} design as an example. Section 4 discussed the foldover plans. Most designs discussed here provide somewhat unique features to the experimenters. These are some of my undergoing research topics, many of them deserves further investigation. I hope that this paper will be useful for those who are interested in research problems in design area.

REFERENCES

Beattie, S. D. and Lin, D. K. J. (1998): Some Theorems on High-Dimensional Rotated Factorial Designs for Computer Experiments. *Technical Report*, **98-02** Department of Statistics, The Pennsylvania State University.

Bergman, B. and Hynén, A. (1997): Dispersion Effects From Unreplicated Designs in the 2^{k-p} Series. *Technometrics*, **39** 191-198.

Box, G. E.P. and Draper, N.R. (1959): A Basis for The Selection of A Response Surface Design. *J. Amer. Statist. Assoc.* **54** 622-654.

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. (1986a): An Analysis for Unreplicated Fractional Factorials. *Technometrics* **28** 11-18.

Box, G. E. P. and Meyer, R. D. (1986b): Dispersion Effects From Fractional Designs. *Technometrics* **28** 19-27.

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. *Int. Statist. Rev.* **61** 131–145.

Daly, B.J. and Torrey, M.D. (1984): A Transient, Three-dimensional Algorithm for Fluid-thermal Mixing And Wall Heat Transfer in Complex Geometries. Technical Report LA-10132-MS, Los Alamos National Laboratory.

Easterling, R.G. (1989): Comment: Design and analysis of computer experiments. *Statistical Science* **4(4)** 425–427.

Haaland, P., McMillan, N., Nychka, D. and Welch, D. (1994): Analysis of Space-filling Designs. *Computationally Intensive Statistical Methods* **26** 111-120.

Johnson, M.E., Moore, L.M. and Ylvisaker, D. (1990): Minimax And Maximin Distance Designs. *J. Statis. Plann.Infer.* **26** 131-148.

Koehler, J.R. and Owen, A.B. (1996): Computer Experiments. In S. Ghosh and C. R. Rao (Eds.) *Handbook of Statistics*, **Vol. 13** Chapter 9, 261-308.

Li, W.W. and Lin, D.K.J. (2000): Optimal Foldover Plans for Two-Level Fractional Factorial Designs. *Technical Report*, **MS&IS 01-01**, The Pennsylvania State University.

McGrath, Richard N. and Lin, Dennis K.J. (2001a): Testing Multiple Dispersion Effects in Unreplicated Fractional Factorial Designs. *Technometrics* **43** 406-414.

McGrath, Richard N. and Lin, Dennis K.J. (2001b): Confounding of Location and Dispersion Effects in Unreplicated Fractional Factorials. *J. Qual. Tech.* **33** 129-139.

McKay, M.D., Beckman, R.J. and Conover, W.J. (1979): A Comparison of Three Methods for Selecting Values of Input Variables in The Analysis of Output From A Computer Code. *Technometrics* **21** 239-245.

Mee, R. W. and Peralta, M. (2000): Semifolding 2^{k-p} Designs. *Technometrics* **42** 122–134.

Montgomery, D. C. (1990): Using Fractional Factorial Designs for Robust Process Development. *Qual. Engg.* **3** 193-205.

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

Montgomery, D. C. and Runger, G. C. (1996): Foldovers of 2^{k-p} Resolution IV Experimental Designs *J. Qual. Tech.* **28** 446–450.

Morris, M.D. and Mitchell, T.J. (1992): Exploratory Designs for Computational Experiments. Technical Report TM-12045, Oak Ridge National Laboratory.

Neter, J., Kutner, M. H., Nachtsheim, C. J. and Wasserman, W. (1996): *Applied Statistical Models*, 4th de., Richard D. Irwin, Burr Ridge, IL.

Sacks, J., Schiller, S.B. and Welch, W.J. (1989): Designs for Computer Experiments. *Technometrics* **31** 41-47.

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