Split-Plot Designs: What, Why, and How

BRADLEY JONES

SAS Institute, Cary, NC 27513

CHRISTOPHER J. NACHTSHEIM

Carlson School of Management, University of Minnesota, Minneapolis, MN 55455

The past decade has seen rapid advances in the development of new methods for the design and analysis of split-plot experiments. Unfortunately, the value of these designs for industrial experimentation has not been fully appreciated. In this paper, we review recent developments and provide guidelines for the use of split-plot designs in industrial applications.

Key Words: Block Designs; Completely Randomized Designs; Factorial Designs; Fractional Factorial Designs; Optimal Designs; Random-Effects Model; Random-Run-Order Designs; Response Surface Designs; Robust Parameter Designs.

Introduction

All industrial experiments are split-plot experiments.

This provocative remark has been attributed to the famous industrial statistician, Cuthbert Daniel, by Box et al. (2005) in their well-known text on the design of experiments. Split-Plot experiments were invented by Fisher (1925) and their importance in industrial experimentation has been long recognized (Yates (1936)). It is also well known that many industrial experiments are fielded as split-plot experiments and yet erroneously analyzed as if they were completely randomized designs. This is frequently the case when hard-to-change factors exist and economic constraints preclude the use of complete randomization. Recent work, most notably by Lucas and his coworkers (Anbari and Lucas (1994), Ganju and Lucas (1997, 1999, 2005), Ju and Lucas (2002), Webb et al. (2004)) has demonstrated that many experiments previously thought to be completely randomized experiments also exhibit split-plot structure. This surprising result adds creedence to the Daniel proclamation and has motivated a great deal of pioneering work in the design and analysis of split-plot experiments. We note in particular the identification of minimum-aberration split-plot fractional factorial designs by Bingham and Sitter and coworkers (Bingham and Sitter (1999, 2001, 2003), Bingham et al. (2004), Loeppky and Sitter (2002)), the development of equivalent-estimation split-plot designs for response surface and mixture designs (Kowalski et al. (2002), Vining et al. (2005), Vining and Kowalski (2008)), and the development of optimal split-plot designs by Goos, Vandebroek, and Jones and their coworkers (Goos (2002), Goos and Vandebroek (2001, 2003, 2004), Jones and Goos (2007 and 2009)).

Our goal here is to review these recent developments and to provide guidelines for the practicing statistician on their use.

What Is a Split-Plot Design?

In simple terms, a split-plot experiment is a blocked experiment, where the blocks themselves serve as experimental units for a subset of the factors. Thus, there are two levels of experimental units. The blocks are referred to as whole plots, while the experi-

Dr. Jones is Director of Research and Development for the JMP Division of SAS Institute, Inc. His email address is bradley.jones@sas.com.

Dr. Nachtsheim is the Curtis L. Carlson Professor and Chair of the Operations and Management Science Department at the Carlson School of Management of the University of Minnesota. His email address is nacht001@umn.edu.

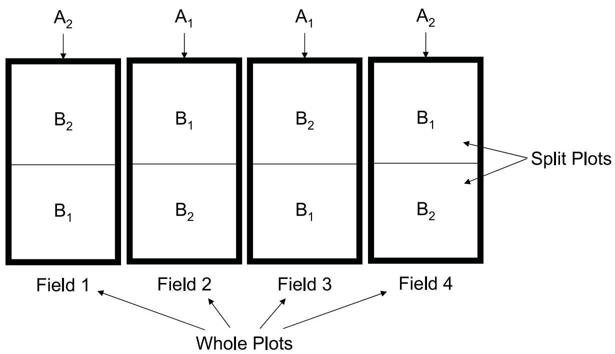

FIGURE 1. Split Plot Agricultural Layout. (Factor A is the whole-plot factor and factor B is the split-plot factor.)

mental units within blocks are called split plots, split units, or subplots. Corresponding to the two levels of experimental units are two levels of randomization. One randomization is conducted to determine the assignment of block-level treatments to whole plots. Then, as always in a blocked experiment, a randomization of treatments to split-plot experimental units occurs within each block or whole plot.

Split-plot designs were originally developed by Fisher (1925) for use in agricultural experiments. As a simple illustration, consider a study of the effects of two irrigation methods (factor A) and two fertilizers (factor B) on yield of a crop, using four available fields as experimental units. In this investigation, it is not possible to apply different irrigation methods (factor A) in areas smaller than a field, although different fertilizer types (factor B) could be applied in relatively small areas. For example, if we subdivide each whole plot (field) into two split plots, each of the two fertilizer types can be applied once within each whole plot, as shown in Figure 1. In this split-plot design, a first randomization assigns the two irrigation types to the four fields (whole plots); then within each field, a separate randomization is conducted to assign the two fertilizer types to the two split plots within each field.

In industrial experiments, factors are often differentiated with respect to the ease with which they can be changed from experimental run to experimental run. This may be due to the fact that a particular treatment is expensive or time-consuming to change, or it may be due to the fact that the experiment is to be run in large batches and the batches can be subdivided later for additional treatments. Box et al. (2005) describe a prototypical split-plot experiment with one easy-to-change factor and one hard-to-change factor. The experiment was designed to study the corrosion resistance of steel bars treated with four coatings, C_1, C_2, C_3, and C_4, at three furnace temperatures, 360°C, 370°C, and 380°C. Furnace temperature is the hard-to-change factor because of the time it takes to reset the furnace and reach a new equilibrium temperature. Once the equilibrium temperature is reached, four steel bars with randomly assigned coatings C_1, C_2, C_3, and C_4 are randomly positioned in the furnace and heated. The layout of the experiment as performed is given in Table 1. Notice that each whole-plot treatment (temperature) is replicated twice and that there is just one complete replicate of the split-plot treatments (coatings) within each whole plot. Thus, we have six whole plots and four subplots within each whole plot.

TABLE 1. Split-Plot Design and Data for Studying the Corrosion Resistance of Steel Bars (Box et al. (2005))

Whole-plot	Temperature (°C)	Coating (randomized order)			
1.	360	C_2	C_3	C_1	C_4
		73	83	67	89
2.	370	C_1	C_3	C_4	C_2
		65	87	86	91
3.	380	C_3	C_1	C_2	C_4
		147	155	127	212
4.	380	C_4	C_3	C_2	C_1
		153	90	100	108
5.	370	C_4	C_1	C_3	C_2
		150	140	121	142
6.	360	C_1	C_4	C_2	C_3
		33	54	8	46

Other illustrative examples of industrial experiments having both easy-to-change and hard-to-change factors include

1. Box and Jones (1992) describe an experiment in which a packaged-foods manufacturer wished to develop an optimal formulation of a cake mix. Because cake mixes are made in large batches, the ingredient factors (flour, shortening, and egg powder) are hard to change. Many

packages of cake mix are produced from each batch, and the individual packages of cake mix can be baked using different baking times and baking temperatures. Here, the easy-to-change, split-plot factors are time and temperature.

2. Another example involving large batches and the subsequent processing of subbatches was discussed by Bingham and Sitter (2001). In this instance, a company wished to study the effect of various factors on the swelling of a wood product after it has been saturated with water and allowed to dry. A batch is characterized by the type and size of wood being tested, the amounts of two additives used, and the level of moisture saturation. Batches can later be subdivided for processing, which is characterized by three easy-to-change factors, i.e., process time, pressure, and material density.

3. Bisgaard (2000) discusses the use of split-plot arrangements in parameter design experiments. In one example, he describes an experiment involving four prototype combustion engine designs and two grades of gasoline. The engineers preferred to make up the four motor designs and test both of the gasoline types while a particular motor was on a test stand. Here the hard-to-change factor is motor type and the easy-to-change factor is gasoline grade.

The analysis of a split-plot experiment is more complex than that for a completely randomized experiment due to the presence of both split-plot and whole-plot random errors. In the Box et al. corrosion-resistance example, a whole-plot effect is introduced with each setting or re-setting of the furnace. This may be due, e.g., to operator error in setting the temperature, to calibration error in the temperature controls, or to changes in ambient conditions. Split-plot errors might arise due to lack of repeatability of the measurement system, to variation in the distribution of heat within the furnace, to variation in the thickness of the coatings from steel bar to steel bar, and so on. We will assume that the whole-plot errors are independent and identically distributed as $\mathrm{N}(0, \sigma_w^2)$, that the split-plot errors are independent and identically distributed as $\mathrm{N}(0, \sigma^2)$, and that whole-plot errors and the split-plot errors are mutually independent. Randomization guarantees that the whole-plot errors are mutually independent and that the split-plot errors are mutually independent within whole plots.

When the split-plot experiment is balanced and the ANOVA sums of squares are orthogonal, a standard, mixed-model, ANOVA-based approach to the analysis is possible. This is the approach that is taken by most DOE textbooks that treat split-plot designs (see, e.g., Kutner et al. (2005), Montgomery (2008), Box et al. (2005)). With this approach, all experimental factor effects are assumed to be fixed. The standard ANOVA model for the balanced two-factor split-plot design, where there are a levels of the whole-plot factor A and b levels of the split-plot treatment B, where each level of whole-plot factor A is applied to c whole plots (so that the whole plot effect is nested within factor A), and where the number of runs is $n = abc$, is given by

$$Y_{ijk} = \mu_{...} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_{k(i)} + \varepsilon_{ijk}, \quad (1)$$

where: $\mu_{...}$ is a constant; α_i, the a whole-plot treatment effects, are constants subject to $\sum \alpha_i = 0$; β_j, the b split-plot treatment effects, are constants subject to $\sum \beta_j = 0$; $(\alpha\beta)_{ij}$, the ab interaction effects, are constants subject to $\sum_i (\alpha\beta)_{ij} = 0$ for all j and $\sum_j (\alpha\beta)_{ij} = 0$ for all i; $\gamma_{k(i)}$, the $n_w = ac$ whole-plot errors, are independent $N(0, \sigma_w^2)$; ε_{ijk} are independent $N(0, \sigma^2)$; $i = 1, \ldots, a$, $j = 1, \ldots, b$, $k = 1, \ldots, c$. The ANOVA model (1) is appropriate for balanced factorial and fractional factorial designs. When balance is not present due to, e.g., missing observations or a more complex design structure, such as a response surface experiment, more general methods are required for analysis. For that reason, we prefer a more general, regression-based modeling approach.

The regression model for a split-plot experiment is a straightforward extension of the standard multiple regression model, with a term added to account for the whole-plot error. Let $\mathbf{w}_i' = (w_{i1}, w_{i2}, \ldots, w_{im_w})$ and $\mathbf{s}_i' = (s_{i1}, s_{i2}, \ldots, s_{im_s})$ denote, respectively, the settings of the m_w whole-plot factors and the settings of the m_s split plot factors for the ith run. The linear statistical model for the ith run is

$$y_i = \mathbf{f}'(\mathbf{w}_i, \mathbf{s}_i)\boldsymbol{\beta} + \mathbf{z_i}'\boldsymbol{\gamma} + \varepsilon_i, \quad (2)$$

where $\mathbf{f}'(\mathbf{w}_i, \mathbf{s}_i)$ gives the $(1 \times p)$ polynomial model expansion of the factor settings, $\boldsymbol{\beta}$ is the $p \times 1$ parameter vector of factor effects, $\mathbf{z_i}$ is an indicator vector whose kth element is one when the ith run was assigned to the kth whole plot and zero if not, $\boldsymbol{\gamma}$ is a $b \times 1$ vector of whole-plot random effects, and ε_i is the split-plot error. We prefer the regression approach for its simplicity and flexibility. It permits the use of polynomial terms for continuous or quantitative treatments, as well as the use of indicators for qualitative predictors.

In matrix form, model (2) can be expressed as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \tag{3}$$

where $\mathbf{Y}$ is the $n \times 1$ vector of responses, $\mathbf{X}$ is the $n \times p$ model matrix with ith row given by $\mathbf{f}'(\mathbf{w}_i, \mathbf{s}_i)$, $\mathbf{Z}$ is an $n \times b$ matrix with ith row equal to $\mathbf{z}'_i$, and $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)'$. We assume that $\boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}_n, \sigma^2\mathbf{I}_n)$, $\gamma \sim \mathrm{N}(\mathbf{0}_c, \sigma_w^2\mathbf{I}_c)$, and $\mathrm{Cov}(\boldsymbol{\varepsilon}, \boldsymbol{\gamma}) = \mathbf{0}_{c\times n}$, where $\mathbf{I}_n$ denotes an $n \times n$ identity matrix.

Our model assumptions imply that the variance–covariance matrix of the response vector $\mathbf{Y}$ can be written as

$$\mathbf{V} = \sigma^2\mathbf{I}_n + \sigma_w^2\mathbf{Z}\mathbf{Z}', \tag{4}$$

where $\mathbf{I}_n$ denotes the $n\times n$ identity matrix. If the runs are grouped by whole plots, the variance–covariance matrix is block-diagonal,

$$\mathbf{V} = \begin{bmatrix} \mathbf{V_1} & 0 & \cdots & 0 \\ 0 & \mathbf{V_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{V_b} \end{bmatrix}.$$

The ith matrix on the diagonal is given by

$$\mathbf{V}_i = \sigma^2\mathbf{I}_{s_i} + \sigma_w^2\mathbf{1}_{s_i}\mathbf{1}'_{s_i},$$

where s_i is the number of observations in the ith whole plot and $\mathbf{1}_{s_i}$ is an $s_i \times 1$ vector of ones.

For statistical inference, we are interested primarily in the estimation of the fixed effects parameter vector $\boldsymbol{\beta}$ and the variance components σ^2 and σ_w^2. There are several alternative methods available for estimation of model parameters for this general approach. Recommendations for estimation and testing are taken up in the later section "How to analyze a split-plot design."

Why Use Split-Plot Designs?

In this paper, we advocate greater consideration of split-plot experiments for three reasons: cost, efficiency, and validity. In this section, we discuss each of these advantages in turn.

Cost

The cost of running a set of treatments in split-plot order is generally less than the cost of the same experiment when completely randomized. As Ganju and Lucas (1997, 1998, 2005) have noted, a properly implemented completely randomized design (CRD) requires that all factors must be independently reset with each run. If a factor level does not change from one run to the next and the factor level is not reset (e.g., by changing its level and then changing it back), *inadvertent split plotting* occurs. Such designs, termed *random run order* (RRO) designs by Ganju and Lucas (1997) are widely used and always analyzed inappropriately. We consider this issue in greater detail, under "Validity" later, and discuss here the relative cost of split-plot experiments and true CRDs. Simple accounting tells us that the cost of a split-plot experiment can be significantly less because much of the cost (or time required) of running a split-plot experiment is tied to changes in the hard-to-change factors. If the cost of setting a hard-to-change factor is C_w and the cost of setting an easy-to-change factor is C_s, and if there are n_w whole plots and n_s split plots, the CRD cost is $n_s(C_w + C_s)$, whereas the SPD cost is $n_wC_w + n_sC_s$, so that the additional cost for the CRD is $(n_s - n_w)C_w$. Split-plot designs are useful precisely because this additional cost can be substantial. Anbari and Lucas (1994, 2008) contrast the costs of CRDs and SPDs for various design scenarios. Bisgaard (2000) also discussed cost functions.

Efficiency

Split-plot experiments are not just less expensive to run than completely randomized experiments; they are often more efficient statistically. Ju and Lucas (2002) show that, with a single hard-to-change factor or a single easy-to-change factor, use of a split-plot layout leads to increased precision in the estimates for all factor effects except for whole-plot main effects. Overall measures of design efficiency have been considered by Anbari and Lucas (1994) and Goos and Vandebroek (2004). Considering two-level, full factorial designs only, Anbari and Lucas (1994) showed that the maximum variance of prediction over a cuboidal design region for a completely randomized design in some instances is greater than that for a blocked split-plot design involving one hard-to-change factor. In other words, the G-efficiency of the split-plot design can, at times, be higher than the corresponding CRD. Similarly, Goos and Vandebroek (2001, 2004) demonstrated that the determinant of a D-optimal split-plot design (discussed later) frequently exceeds that of the corresponding D-optimal completely randomized design. These results indicate that running split-plot designs can represent a win–win proposition: less cost with greater overall precision. As a result, many authors recommend the routine use of split-plot layouts, even when a CRD is a feasible alternative.

Validity

Completely randomized designs are prescribed frequently in industry, but are typically not run as such in the presence of hard-to-change factors. One of two shortcuts may be taken by the experimenter in the interest of saving time or money, i.e.,

1. A *random-run-order* (RRO) design—alternately referred to as a *random-not-reset* (RNR) design—is utilized. As noted above, in this case, the order of treatments is randomized, but the factor levels are not reset with each run of the experiment, or perhaps all factors, *except* the hard-to-change factors are reset. Not resetting the levels of the hard-to-change factor leads to a correlation between adjacent runs. Statistical tests that do not consider these correlations will be biased (Ganju and Lucas (1997)).
2. Although a random run order is provided to the experimenter, the experimenter may decide to resort the treatment order to minimize the number of changes of the hard-to-change factor. Although this leads to a split-plot design, the results are typically analyzed as if the design had been fielded as a completely randomized experiment.

Either scenario leads to an incorrect analysis. The Type I error rate for whole plot factors increases, as does the Type II error rate for split-plot factors and whole-plot by split-plot interactions (see, e.g., Goos et al. (2006)). The corrosion-resistance example of Table 1 is a good example. The table below provides p-values for tests of effects in the corrosion-resistance example for the standard (incorrect) analysis and the split-plot analysis. At the $\alpha = .05$ level of significance, the p-values for the completely randomized analysis lead to the conclusion that the temperature main effect is statistically significant, whereas the coating and temperature-by-coating interaction effects are not statistically significant. The conclusions based on the split-plot analysis are exactly the opposite.

p-Values for corrosion-resistance example

Effect	Split-plot analysis	Standard Analysis
Temperature (A)	.209	.003 ✓
Coating (B)	.002 ✓	.386
Temp × Coating ($A \times B$)	.024 ✓	.852

One way to avoid these mistakes is to plan the experiment as a split-plot design in the first place. Not only does this avoid mistakes, it also leverages the economic and statistical efficiencies already described. Interestingly, it is not necessarily the case that an RRO experiment is to be universally avoided. Webb et al. (2004) discuss comparisons of prediction properties of experiments that are completely randomized versus experiments that have one or more factors that are not reset. They argue that, in process-improvement experiments, the cost savings that accrues from *not* resetting each factor on every run may at times justify the use of an RRO design, even though prediction variances will be inflated and tests will be biased.

In general, we believe that considerations about the presence of hard-to-change factors and their potential impact on the randomization employed should be a part of the planning of every industrial experiment. That is, asking questions about necessary groupings of experimental runs is something that a consultant should do regularly. This is in contrast with the usual standard advice dispensed by consultants and short courses, that designs should always be randomized. Finally, to ensure the validity of the experiment and the ensuing analysis, the experiment should be carefully monitored to verify that whole-plot and split-plot factor settings are reset as required by the design.

How to Choose a Split-Plot Design

Choosing a "best" split-plot design for a given design scenario can be a daunting task, even for a professional statistician. Facilities for construction of split-plot designs are not as yet generally available in software packages (with SAS/JMP being one exception). Moreover, introductory textbooks frequently consider only very simple, restrictive situations. Unfortunately, in many cases, obtaining an appropriate design requires reading (and understanding) a relevant recent research paper or it requires access to software for generating the design. To make matters even more difficult, construction of split-plot experiments is an active area of research and differences in opinion about the value of the various recent advances exist. In what follows, we discuss the choice of split-plot designs in five areas: (1) two-level full factorial designs; (2) two-level fractional factorial designs;

(3) mixture and response surface designs; (4) split-plot designs for robust product experiments; and (5) optimal designs.

Full Factorial Split-Plot Designs

Construction of single replicates of a two-level full factorial designs is straightforward. In this case, a completely randomized design in the whole-plot factors is conducted and, within each whole plot, a completely randomized design in the split-plot factors is also conducted. Note that there are $n_w + 1$ separate randomizations. The disadvantage of running the split-plot in a single replicate is the same as that for a completely randomized design: there are no degrees of freedom for error, either at the whole-plot level or the split-plot level, and so tests for various effects cannot be obtained without assuming that various higher order interactions are negligible. One design-based solution to this is to fully replicate the experiment, although this may be expensive. Intermediate alternatives include replicating the split-plots within a given whole plot (if the size of the whole plot permits), giving an estimate of the split-plot error variance, or the use of classical split-plot blocking (as advocated by Ju and Lucas (2002)), where the split-plot treatments are run in blocks and the whole-plot factor level is reset for each block. This increases the number of whole plots (while decreasing the number of split-plots per whole plot), providing an estimate of whole-plot error.

An example of a blocked full-factorial split-plot design is given in Table 2. For this example, there are two hard-to-change factors, A and B, and three easy-to-change factors p, q, and r. Run as an unblocked full-factorial split-plot design, there would be $2^2 = 4$ whole plots with each whole plot consisting of the $2^3 = 8$ treatment combinations involving p, q, and r. Here we use $b = pqr$ as the block generator to divide the eight split-plot treatment combinations in each whole plot into two blocks of size four. Each block is now treated as a separate whole-plot, thereby increasing the number of whole plots from four to eight. An abbreviated ANOVA table is shown in Table 3. (Note that the four degrees of freedom for whole-plot error are obtained by pooling each of the four whole-plot ($= pqr$) effects that are nested within the levels of A and B.) The fact that the number of whole plots has doubled will increase the cost of the experiment. It is also important to reset the whole-plot factor level between whole plots even if level does not change.

TABLE 2. Full-Factorial Split-Plot Design with Split-Plot Blocking (in standard order)

	Whole plot		Split plot		
Whole plot number	A	B	p	q	r
1	−1	−1	−1	−1	−1
	−1	−1	1	1	−1
	−1	−1	1	−1	1
	−1	−1	−1	1	1
2	−1	−1	1	−1	−1
	−1	−1	−1	1	−1
	−1	−1	−1	−1	1
	−1	−1	1	1	1
3	1	−1	−1	−1	−1
	1	−1	1	1	−1
	1	−1	1	−1	1
	1	−1	−1	1	1
4	1	−1	1	−1	−1
	1	−1	−1	1	−1
	1	−1	−1	−1	1
	1	−1	1	1	1
5	−1	1	−1	−1	−1
	−1	1	1	1	−1
	−1	1	1	−1	1
	−1	1	−1	1	1
6	−1	1	1	−1	−1
	−1	1	−1	1	−1
	−1	1	−1	−1	1
	−1	1	1	1	1
7	1	1	−1	−1	−1
	1	1	1	1	−1
	1	1	1	−1	1
	1	1	−1	1	1
8	1	1	1	−1	−1
	1	1	−1	1	−1
	1	1	−1	−1	1
	1	1	1	1	1

Fractional Factorial Split-Plot Designs

If a full factorial split-plot design is too large, consideration of a fractional factorial split-plot (FFSP)

TABLE 3. ANOVA (Source and DF) for Design in Table 1

Source	DF
A	1
B	1
AB	1
Whole plot error	4
p	1
q	1
r	1
pq	1
pr	1
qr	1
Ap	1
Aq	1
Ar	1
Apq	1
Apr	1
Aqr	1
Bp	1
Bq	1
Br	1
Split-plot error	9
Bpq	
Bpr	
Bqr	
ABp	
ABq	
ABr	
$ABpq$	
$ABpr$	
$ABqr$	
Total	31

design may be warranted. As noted by Bisgaard (2000), there are generally two approaches to constructing FFSPs. The first is to (1) select a $2^{p_1-k_1}$ fractional factorial design in the p_1 whole-plot factors of resolution r_1 (or a full factorial if $k_1 = 0$); (2) select a $2^{p_2-k_2}$ fractional factorial design of resolution r_2 in the p_2 split-plot factors (or a full factorial if $k_2 = 0$); and then (3) cross these designs to obtain a $2^{(p_1+p_2)-(k_1+k_2)}$ fractional factorial design. (This design will be a fractional factorial of resolution $\min(r_1, r_2)$, as long as $k_1 + k_2 > 0$). By *crossing*, we simply mean that every whole-plot treatment combination is run in combination with every split-plot treatment combination. Bisgaard (2000) refers to designs that are obtained by crossing as *Cartesian-product* designs, and we adopt this terminology in what follows. The use of Cartesian-product designs is a simple strategy that is frequently used, but it is not guaranteed to give the FFSP design of maximum resolution. An alternative approach is to use the technique of *split-plot confounding*.

To illustrate the concept of split-plot confounding, we borrow from an example provided by Bingham and Sitter (1999). Assume that there are four whole-plot factors, A, B, C, and D, and three split-plot factors, p, q and r, resources permit at most 16 runs, and we require a design of resolution III or greater. Employing a Cartesian-product design, the highest resolution possible is obtained by crossing a resolution IV 2^{4-1} design in the whole plots (based on the generator $D = ABC$) with a 2^{3-2} design in the split plots (based, for example, on the split-plot generators $q = p$ and $r = p$). This leads to a $2^{(4+3)-(1+2)}$ FFSP with the following defining contrast subgroup:

$$I = ABCD = pq = pr = ABCDpq = ABCDpr$$
$$= qr = ABCDqr,$$

which is resolution II. A better design can be constructed using split-plot confounding (Kempthorne (1952)) by including whole-plot factors in the split-plot factorial generators. In the current example, if we use $D = ABC$ as the whole-plot design generator and $q = BCp$ and $r = ACp$ as the split-plot design generators, the defining contrast subgroup is

$$I = ABCD = BCpq = ACpr = CDpq = BDpr$$
$$= BCqr = ADqr,$$

which is resolution IV.

Huang et al. (1998, pp. 319-322) give tables of minimum aberration (MA) FFSP designs having five through fourteen factors for 16, 32, and 64 runs and these designs were extended by Bingham and Sitter (1999).

To illustrate the construction of an FFSP design using the tables of Huang et al., consider again the previous example involving four whole-plot factors and three split-plot factors, to be fielded in 16 runs. In this case, we have $n_1 = 4$, $n_2 = 3$, $k_1 = 1$, and $k_2 = 2$. The "word-length pattern/generators" for this case are provided on line 9 of Table 1 (Huang et al. (1998), p. 319) and are $ABCD$, $BCpq$, and $ACpr$. Hence, we have $D = ABC$, $q = BCp$, and $r = ACp$. The independent columns are A, B, C, and p. Most

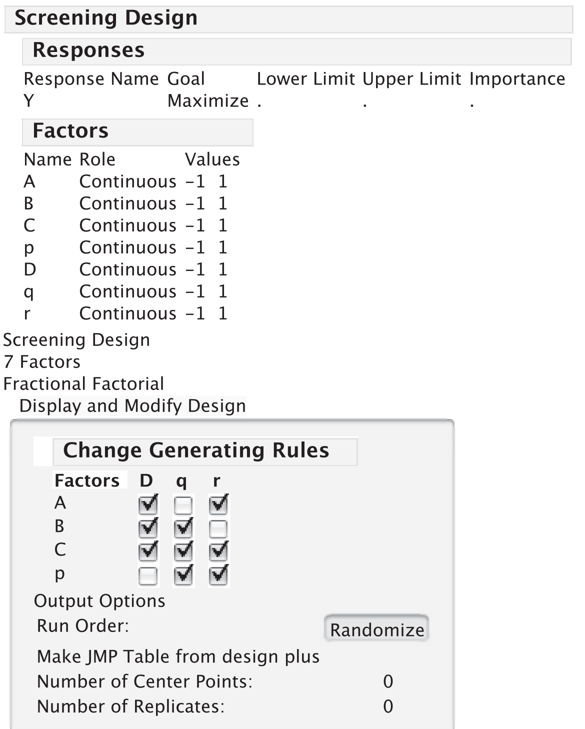

FIGURE 2. Selecting Generators for Minimum Aberration Split-Plot Fractional Factorial Design.

TABLE 4. Minimum Aberration SPFF Design

Whole plot	A	B	C	D	p	q	r
1	−1	−1	−1	−1	−1	−1	−1
	−1	−1	−1	−1	1	1	1
2	−1	−1	1	1	−1	1	1
	−1	−1	1	1	1	−1	−1
3	−1	1	−1	1	1	−1	1
	−1	1	−1	1	−1	1	−1
4	−1	1	1	−1	1	1	−1
	−1	1	1	−1	−1	−1	1
5	1	−1	−1	1	−1	−1	1
	1	−1	−1	1	1	1	−1
6	1	−1	1	−1	−1	1	−1
	1	−1	1	−1	1	−1	1
7	1	1	−1	−1	1	−1	−1
	1	1	−1	−1	−1	1	1
8	1	1	1	1	−1	−1	−1
	1	1	1	1	1	1	1

standard software packages for the design of experiments permit the user to specify generators or, equivalently, a preferred defining contrast subgroup. Figure 2 illustrates this step for SAS/JMP. The resulting design is displayed in standard order in Table 4.

While the tables in Huang et al. (1998) and Bingham and Sitter (1999) are useful, they are by no means exhaustive, and a number of papers have augmented these results. One limitation of these designs is that the whole-plot designs were not replicated. Bingham et al. (2004) developed methods for splitting the whole plots (as discussed above in the case of a full factorial split-plot design), thereby increasing the number of whole plots while decreasing the number of split plots per whole plot. McLeod and Brewster (2004) developed three methods for constructing blocked minimum aberration 2-level fractional factorial split-plot designs, for $n = 32$ runs. McLeod and Brewster (2008) also developed foldover plans for two-level fractional factorial designs. Relatedly, Almimi et al. (2008) develop follow-up designs to resolve confounding in split-plot experiments. Kowalski (2002) provides additional SPFF designs with 24 runs, in an effort to fill the gap between the 16- and 32-run designs provided by Huang et al. (1998). He starts with a 16-run design that uses fewer whole plots than those in the minimum aberration tables, and then adds 8 runs to this design using semi-folding to break various alias chains. He shows that the 24-run design is a good compromise between the 16- and 32-run designs. We consider these designs in a bit more detail in our discussion of optimal split-plot designs below.

Split-Plot Designs for Robust Product Experiments

The purpose of robust product experiments is to identify settings of product-design factors, sometimes termed *product-design parameters* that lead to products or processes that are insensitive to the effects of uncontrollable environmental factors. As illustrated by Box and Jones (1992), product-design parameters for a cake mix might be the amount of flour, shortening, and egg powder contained in the mix. Uncontrollable environmental factors include the baking time and baking temperature used by the consumer to produce the cake. These ideas were introduced by Taguchi (see, e.g., Taguchi (1987)), who recommended the use of Cartesian product experiments to determine the optimal settings of the design parameters. Specifically, Taguchi employed a designed experiment in the design factors—the inner array, a second designed experiment for the environmental factors—the outer array, and these designs were crossed to form the robust product experiment. A "best" setting of the product design factors is one that leads to an average response closest to the product designer's target, and that also exhibits little variation as the environmental factors are changed.

Unfortunately, a barrier to the widespread use of crossed-array parameter-design experiments has been cost. Clearly, if n_1 is the number of runs in the design-factors array and n_2 is the number of runs in the environmental-factors array, then the required number of runs—and therefore the number of factor resettings—is $n = n_1 \times n_2$. Thus, the size of the the required robust parameter CRD will be prohibitive unless both n_1 and n_2 are small. Recently, various authors, beginning with Box and Jones (1992), have demonstrated how split-plot designs can be used effectively to reduce the cost of robust product experiments when there are hard-to-change factors. Box and Jones (1992) discuss three ways in which split-plot designs might arise in robust product experiments:

1. The environmental factors are hard to change while the product-design factors are easy to change. In the context of the cake-mix example, suppose that the cake mixes can be easily formulated as individual packages, but the ovens are large and it takes a good deal of time to reset the temperature. In this case, the time for the experiment may be minimized by using time and temperature as whole-plot factors.
2. The product-design factors are hard to change and the environmental factors are easy to change. In the context of the cake-mix example, it may be that the cake mixes must be made in large batches, which can be subdivided for baking at the different conditions indicated by the environmental-factor (outer) array. Thus, we want to minimize the number of required batches (whole plots).
3. In this alternative, Box and Jones (1992) considered the use of strip-plot experiments, which can further reduce the cost of the experiment for either of alternatives (1) and (2). We refer the reader to their paper for further details.

Of course, while alternatives (1) and (2) arise frequently, environmental- and product-design factors need not be of one type or the other. The case may arise in which the set of hard-to-change factors is comprised of both environmental- and product-design factors. Generally, in robust product experiments, however, interest centers on the design-factor effects and the design-factor × environmental-factor interactions. Main effects of the environmental factors are of secondary interest, as they have no bearing on determination of best settings for the design factors. For this reason, Box and Jones (1992) conclude that environmental factors will be more appropriately used as the whole-plot treatment.

A number of authors have explored the question of how to choose a "best" split-plot design for a given set of design and environmental factors. Bisgaard (2000) considered the use of $2^{p_1-k_1} \times 2^{p_2-k_2}$ fractional factorial split-plot experiments, where there are p_1 whole-plot factors and p_2 split-plot factors. He began by considering Cartesian-product designs, as advocated by Taguchi, and showed how to determine the defining contrast subgroup for the combined array as a function of the defining contrast subgroups for the inner and outer arrays. He then noted that split-plot confounding can lead to better designs and gave rules for implementing split-plot confounding.

The problem of finding a "best" split-plot fractional factorial design for robust product design was taken up later by Bingham and Sitter (2003). They began by noting that their own tables of MA SPFF designs (Bingham and Sitter (1999)) and those developed previously by Huang et al. (1998) were not directly applicable to product-design experiments. Let C denote any (controllable) product-design factor and let N denote any environmental (noise) factor. If we ignore the robust design aspect of the design problem, the importance of the various factors (of third order or less) to the investigator will typically be ranked in descending order of likely significance—main effects, followed by second-order effects, followed by third-order effects—as shown in the top panel of Table 5. However, in a product-design experiment, we are less interested in main effects of the noise factors, and Bingham and Sitter (2003) argue that a more relevant ranking of factor effects is as shown in the bottom panel of Table 5. The key point is that the importance ranking of the control-by-noise interactions has been increased

TABLE 5. Ranking of Factor Effects in Standard vs. Parameter Design Experiments

Importance ranking	Effect type
	Standard experiment
1.	C, N
2.	$C \times C$, $N \times N$, $C \times N$
3.	$C \times C \times C$, $C \times C \times N$, $C \times N \times N$, $N \times N \times N$
	Robust parameter experiment
1.	C, N
2.	$C \times N$
3.	$C \times C$, $N \times N$
4.	$C \times C \times N$, $C \times N \times N$
5.	$C \times C \times C$, $N \times N \times N$

relative to other effects of the same order, consistent with the experimenter's objectives in robust product experiments. As noted by Bingham and Sitter (2003), the low-order noise main effects must be retained because they are likely to be significant, and it is important to keep other effects from becoming aliased with them. They then modify the definition of aberration to reflect the new importance rankings and use their search algorithm (Bingham and Sitter (1999)) to produce two tables of MA FFSP designs for robust product design experiments. The first table is for use when the product-design factors are hard to change, the second is for applications when the environmental factors are hard to change. These tables are useful for 16- and 32-run experiments and for total numbers of factors ranging from 5 to 10.

Split-Plot Designs for Response Surface Experiments

Relatively little research has been conducted with regard to the construction of split-plot designs for response surface experiments. Recall that, in a response surface experiment, we are typically interested in estimating a response model comprised of all first- and second-order polynomial terms.

Letsinger et al. (1996) first discussed the notion of optimality criteria in connection with split-plot response surface experiments. They noted that a best design will maximize, according to the determinant criterion C_D,

$$C_D(\mathbf{X}, \mathbf{Z}, \mathbf{V}) = \left|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right|. \tag{5}$$

Notice that, from Equation (4), we have

$$\begin{aligned}\mathbf{V} &= \sigma^2\mathbf{I}_n + \sigma_w^2\mathbf{ZZ}' \\ &= \sigma^2(\mathbf{I}_n + d\mathbf{ZZ}'),\end{aligned} \tag{6}$$

where $d = \sigma_w^2/\sigma^2$ is the *variance ratio*—the ratio of the whole-plot variance component to the split-plot variance component. From Equations (5) and (6), it follows that

$$\begin{aligned}C_D(\mathbf{X}, \mathbf{Z}, \mathbf{V}) &= (\sigma^2)^{-p}\left|\mathbf{X}'(\mathbf{I}_n + d\mathbf{ZZ}')^{-1}\mathbf{X}\right| \\ &\propto \left|\mathbf{X}'(\mathbf{I}_n + d\mathbf{ZZ}')^{-1}\mathbf{X}\right|,\end{aligned} \tag{7}$$

which indicates that the determinant criterion depends on the unknown variance components only through their ratio, d. The absolute magnitudes are not required for comparing designs. Two alternative designs, $[\mathbf{X}_1, \mathbf{Z}_1]$ and $[\mathbf{X}_2, \mathbf{Z}_2]$, can be compared by computing their relative efficiencies. The relative D-efficiency of design $[\mathbf{X}_1, \mathbf{Z}_1]$, relative to the design $[\mathbf{X}_2, \mathbf{Z}_2]$ is given by

$$RE_D([\mathbf{X}_1, \mathbf{Z}_1], [\mathbf{X}_2, \mathbf{Z}_2]) = \left(\frac{\left|\mathbf{X_1}'(\mathbf{I}_n + d\mathbf{Z_1}\mathbf{Z_1}')^{-1}\mathbf{X_1}\right|}{\left|\mathbf{X_2}'(\mathbf{I}_n + d\mathbf{Z_2}\mathbf{Z_2}')^{-1}\mathbf{X_2}\right|}\right)^{1/p}.$$

The relative efficiency gives the number times the design $[\mathbf{X}_2, \mathbf{Z}_2]$ must be replicated in order to achieve the overall precision of design $[\mathbf{X}_1, \mathbf{Z}_1]$.

Letsinger et al. (1996) also gave the integrated variance of prediction criterion (IV) for split-plot designs as

$$\begin{aligned}C_{IV}(\mathbf{X}, \mathbf{Z}, \mathbf{V}) &= \frac{\sigma^2}{\int_R d\mathbf{x}}\int_{\mathbf{R}} \mathbf{f}'(\mathbf{x})[\mathbf{X}'(\mathbf{I}_n + d\mathbf{ZZ}')^{-1}\mathbf{X}]^{-1}\mathbf{f}'(\mathbf{x})d\mathbf{x} \\ &\propto \text{Trace}\left\{[\mathbf{X}'(\mathbf{I}_n + d\mathbf{ZZ}')^{-1}\mathbf{X}]^{-1} \times \left[\int_{\mathbf{R}} \mathbf{f}(\mathbf{x})\mathbf{f}'(\mathbf{x})d\mathbf{x}\right]\right\} \\ &= \text{Trace}\{[\mathbf{X}'(\mathbf{I}_n + d\mathbf{ZZ}')^{-1}\mathbf{X}]^{-1}[\mathbf{M}_R]\},\end{aligned} \tag{8}$$

where $\mathbf{x}' = (\mathbf{w}', \mathbf{s}')$ denotes the vector of factor settings, $\mathbf{R}$ is the region of interest, $\mathbf{M}_R$ is the moment matrix, and we again note the dependence of the criterion on the unknown variance ratio d. For the integrated variance criterion, the efficiency of design $[\{\mathbf{X}_1, \mathbf{Z}_1\}]$, relative to design $[\{\mathbf{X}_2, \mathbf{Z}_2\}]$ is given by

$$RE_{IV}([\mathbf{X}_1, \mathbf{Z}_1], [\mathbf{X}_2, \mathbf{Z}_2]) = \frac{\text{Trace}\{[\mathbf{X}_2'(\mathbf{I}_n + d\mathbf{Z}_2\mathbf{Z}_2')^{-1}\mathbf{X}_2]^{-1}[\mathbf{M}_R]\}}{\text{Trace}\{[\mathbf{X}_1'(\mathbf{I}_n + d\mathbf{Z}_1\mathbf{Z}_1')^{-1}\mathbf{X}_1]^{-1}[\mathbf{M}_R]\}}.$$

Letsinger et al. examined, for the case of four factors, the efficiencies of the Box–Behnken design, the uniform precision design, the rotatable central composite design, the small composite design for the case of one or two whole-plot factors. They conclude that the central composite design, run as a split-plot, seemed to give the best overall performance.

Other recent developments in this area include Draper and John (1998), who discuss how to choose factor levels for their CUBE and STAR designs in split-plot situations. They use an index of rotatability as their criterion for choosing the factor levels to create optimal designs. Split-plot mixture response surface experiments have only very recently been considered (Kowalski et al. (2002), Goos and Donev (2006)).

Recently, Vining et al. (2005) discussed the design and analysis of response surface experiments when run in the form of a split plot. Their paper established conditions under which ordinary least squares (OLS) estimates, which are usually inappropriate for analysis of split-plot experiments, and the preferred generalized least squares (GLS) estimates are equivalent. Their result builds on the Letsinger et al. (1996) paper just discussed, which demonstrated that OLS and GLS estimates are equivalent for Cartesian-product split-plot designs. Vining et al. (2005) then show how various points in central composite designs and Box–Behnken designs can be replicated in such a way as to induce the equivalence of OLS and GLS. These designs, named *equivalent estimation designs* (EEDs) by the authors, have two useful features. First, the required replication structure leads to designs with pure replicates of whole plots and split plots, so that model-free estimates of the variance components σ_w^2 and σ^2 are provided. Second, OLS software, which is widely available, can be used to obtain parameter estimates.

An equivalent-estimation design for a split-plot response surface scenario, developed by Vining et al. (2005), is shown in Figure 3, and the design points are listed in Table 6. In this example, there are two hard-to-change factors, Z_1 and Z_2, and there are two easy-to-change factors, X_1 and X_2, and the experimenter is interested in estimating all first- and second-order terms. The design consists of a Box–Behnken design with 12 whole plots corresponding to the four corner points, the four axial points, and four center points. In the four axial whole plots and in three of the four center-point whole plots, the center points for the split-plot treatments (i.e., $X_1 = X_2 =$ 0) are replicated four times. In one of the four center-point whole plots, the split-plot treatments occupy the four corners. This leads to $7 \times (4 - 1) = 21$ degrees of freedom for split-plot pure error. Also, because three of the four whole-plot center points have the identical split-plot treatment structure (i.e., four center points), they are true replicates, and we obtain three degrees of freedom for whole-plot pure error. In general, if the center points of the split-plot factors are replicated n_{spc} times within each of the $2k$ whole-plot axial points and r whole-plot center points, there will be $(2k + r)(n_{spc} - 1)$ degrees of freedom for split-plot pure error and $r - 1$ degrees of freedom for whole-plot pure error. An unbiased estimate of the split-plot pure-error variance is given by

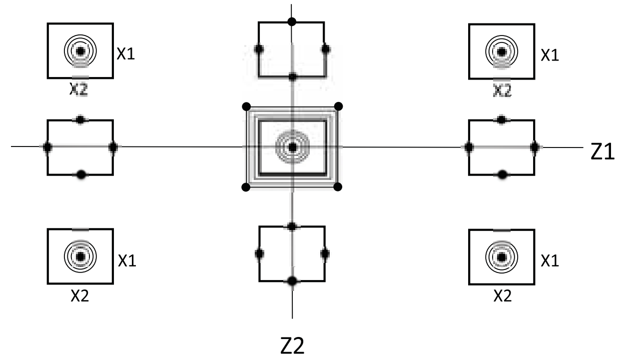

FIGURE 3. Equivalent Estimation Response Surface Design.

$$s_{sp}^2 = \frac{\sum_{l=1}^{2k} s_{A_l}^2}{2k} + \frac{\sum_{i=1}^{r} s_{cp0_i}^2}{r},$$

where $s_{A_l}^2$ and $s_{cp0_i}^2$ are the usual sample variances of the responses from the lth whole-point axial point and the ith whole-plot center point containing split-plot center points. An estimate of the whole-plot variance component is also easy to compute. Let $\bar{Y}_{i.}$ denote the mean of the values in the ith replicated whole plot and let s_{wp}^2 denote the usual sample variance of the r replicated whole-plot means, $\bar{Y}_{1.}, \ldots, \bar{Y}_{r.}$. An unbiased estimate of σ_w^2 is given by

$$s_w^2 = s_{wp}^2 - \frac{s_{sp}^2}{n_{spc}}.$$

As was the case with the standard ANOVA estimator (9), this estimator can, at times, be negative.

In summary, equivalent estimation designs for response-surface experiments designs possess two notable properties. First, they permit the use of OLS

TABLE 6. Equivalent Estimation Split-Plot Response Surface Design

Whole plot	Z_1	Z_2	X_1	X_2	Whole plot	Z_1	Z_2	X_1	X_2
1	−1	−1	0	0	7	0	−1	−1	0
(Corner)	−1	−1	0	0	(Axial)	0	−1	1	0
	−1	−1	0	0		0	−1	0	−1
	−1	−1	0	0		0	−1	0	1
2	1	−1	0	0	8	0	1	−1	0
(Corner)	1	−1	0	0	(Axial)	0	1	1	0
	1	−1	0	0		0	1	0	−1
	1	−1	0	0		0	1	0	1
3	−1	1	0	0	9	0	0	−1	−1
(Corner)	−1	1	0	0	(Center)	0	0	1	−1
	−1	1	0	0		0	0	−1	1
	−1	1	0	0		0	0	1	1
4	1	1	0	0	10	0	0	0	0
(Corner)	1	1	0	0	(Center)	0	0	0	0
	1	1	0	0		0	0	0	0
	1	1	0	0		0	0	0	0
5	−1	0	−1	0	11	0	0	0	0
(Axial)	−1	0	1	0	(Center)	0	0	0	0
	−1	0	0	−1		0	0	0	0
	−1	0	0	1		0	0	0	0
6	1	0	−1	0	12	0	0	0	0
(Axial)	1	0	1	0	(Center)	0	0	0	0
	1	0	0	−1		0	0	0	0
	1	0	0	1		0	0	0	0

for obtaining factor-effects estimates. Second, the authors' recommended replication structure leads to model-free estimates of the whole-plot and split-plot variance components. However, these designs are not without drawbacks. Inadvertent changes to the design during implementation, e.g., the occurrence of a missing observation, will destroy the OLS–GLS equivalence. Moreover, Goos (2006) noted that the relatively large proportion of runs devoted to pure replication significantly reduces the overall efficiency of the design. Goos advocates the use of optimal experimental designs, which we take up next.

Optimal Split-Plot Designs

A general approach to the problem of constructing split-plot designs is provided by the optimal-design approach. Optimal designs for split-plot experiments were first proposed by Goos and Vandebroek (2001). The general approach is as follows:

1. Specify the number of whole plots, n_w.
2. Specify the number of split plots per whole plot, n_s.
3. Specify the response model, $\mathbf{f}'(\mathbf{w}, \mathbf{s})$.
4. Specify a prior estimate of the variance ratio, d.
5. Use computer software to construct the design for steps 1 through 4 that maximizes the D-optimality criterion of Equation (5).
6. Study the sensitivity of the optimal design to small changes in d, n_w, and n_s.

The algorithm given by Goos and Vandebroek (2001) also requires that the user specify a candidate set of possible design points. This is usually the 2^p vertices of the hypercube ($[-1,1]^p$) for main-effects models or for main-effects models plus interactions. It may consist of points from the 3^p factorial when pure quadratic effects are present. An alternative algorithm based on the coordinate exchange algorithm (Meyer and Nachtsheim (1995)), which does not require candidate sets, was later developed by Jones and Goos (2007). Note that the specified numbers of whole plots and split plots must both be large enough to enable estimation of all of the desired model effects and that the designs produced are optimal for the model as specified in item (3) above.

The approach is general in the sense that it can be applied to virtually any split-plot design problem and, unlike the minimum-aberration designs described earlier, there is no need to limit run sizes to powers of two. We demonstrate the use of the optimal split-plot design approach in connection with three of the examples already discussed.

Example 1: Screening

Consider the screening design that was constructed earlier involving four hard-to-change factors and three easy-to-change factors. The minimum-aberration design, shown in Table 4, was based on eight whole plots with two split plots per whole plot. Thus, we specify $n_w = 8$, $n_s = 2$, and the model of interest is the seven-factor first-order model. The D-optimal design (produced by SAS/JMP) is shown in Table 7. Note that the design is balanced overall and that every whole plot is separately balanced. The design, in fact, is orthogonal, although it is not a *regular* orthogonal design. For nonregular orthogonal designs, effects are not directly aliased with other effects, as they are for MA fractional factorial designs. This example illustrates advantages and disadvantages of the optimal design approach. The simplicity of this approach is undeniable, and whereas the MA design is a D-optimal design for the first-order model, a D-optimal design need not be an MA design. One might question the marginal value of a resolution IV design in this instance, when most of the two-factor interactions are aliased with two or more other two-factor interactions. Correctly identifying the active two-factor interactions will be difficult, at best, without substantial followup testing.

TABLE 7. D-Optimal Split-Plot Screening Design

Whole plot	A	B	C	D	p	q	r
1	−1	−1	−1	1	−1	−1	1
1	−1	−1	−1	1	1	1	−1
2	−1	−1	1	−1	−1	−1	−1
2	−1	−1	1	−1	1	1	1
3	−1	1	−1	−1	−1	1	1
3	−1	1	−1	−1	1	−1	−1
4	−1	1	1	1	−1	−1	−1
4	−1	1	1	1	1	1	1
5	1	−1	−1	−1	−1	1	−1
5	1	−1	−1	−1	1	−1	1
6	1	−1	1	1	−1	1	−1
6	1	−1	1	1	1	−1	1
7	1	1	−1	1	−1	1	1
7	1	1	−1	1	1	−1	−1
8	1	1	1	−1	−1	−1	1
8	1	1	1	−1	1	1	−1

Example 2: Main Effects Plus Interactions

We noted that Kowalski (2002) provided approaches for constructing 24-run split-plot designs by using semi-folding of 16-run designs or balanced incomplete blocks. In Kowalski's case 1, based on balanced incomplete block designs, there were two whole-plot factors, four split-plot factors, and the objective was to run the experiment in four whole-plots with six split plots per whole plot in order to estimate all main effects and two-factor interactions. The design recommended by Kowalski is displayed in Table 8. The optimal design approach produced the design shown in Table 9. In both cases, all main effects and second-order interactions are estimable, and in both cases, the designs are balanced both overall and within whole plots. The D-efficiency of the combinatoric design, relative to the D-optimal design, is 88.9%.

Example 3: Response Surface Designs

Vining et al. (2005) developed an EED based on 2 hard-to-change and 2 easy-to-change factors, in 12 whole plots having 4 split plots per whole plot. The design, shown previously in Figure 3, is a Box–Behnken design with the split-plot and whole-plot

TABLE 8. 24 Run Design for Estimating Main Effects and Two-Factor Interactions (Kowalski, 2002)

Whole plot	A	B	p	q	r	s
1	1	−1	1	−1	−1	−1
1	1	−1	−1	1	−1	−1
1	1	−1	−1	−1	1	−1
1	1	−1	1	1	−1	1
1	1	−1	1	−1	1	1
1	1	−1	−1	1	1	1
2	−1	1	−1	1	−1	−1
2	−1	1	−1	−1	1	−1
2	−1	1	−1	−1	−1	1
2	−1	1	1	1	1	−1
2	−1	1	1	1	−1	1
2	−1	1	1	−1	1	1
3	1	1	1	1	−1	−1
3	1	1	1	−1	1	−1
3	1	1	1	−1	−1	1
3	1	1	−1	1	1	−1
3	1	1	−1	1	−1	1
3	1	1	−1	−1	1	1
4	−1	−1	1	−1	1	−1
4	−1	−1	1	−1	−1	1
4	−1	−1	−1	1	1	−1
4	−1	−1	−1	1	−1	1
4	−1	−1	−1	−1	−1	−1
4	−1	−1	1	1	1	1

replication structure chosen in such a way that an EED results. The D-optimal design for the same second-order model is listed in Table 10 and displayed in Figure 4. It is clear that the D-optimal design has traded split-plot and whole-plot center-point replicates for treatment combinations at the boundaries of the design space. Notice, for example, that there are no center-point whole plots and that all whole plots include corner points. Consistent with its criterion, the optimal-design approach is sacrificing information on variance components for information on factor effects.

These results indicate that there are currently a variety of approaches for obtaining split-plot designs. We have discussed the use of MA fractional factorial designs, various combinatoric modifications to the MA designs, EED response surface designs, and optimal split-plot designs. The results show that each approach emphasizes a different characteristic of the design and the resulting designs can be quite different. Our view is that all of these designs are *excellent*, especially when viewed against the alternatives of running a CRD incorrectly or simply not running a designed experiment. The particular choice depends on the goals of the experiment.

TABLE 9. 24 Run D-Optimal Design for Estimating Main Effects and Two-Factor Interactions

Whole plot	A	B	p	q	r	s
1	−1	−1	−1	−1	−1	1
1	−1	−1	−1	−1	1	−1
1	−1	−1	−1	1	−1	−1
1	−1	−1	1	−1	−1	−1
1	−1	−1	1	1	−1	1
1	−1	−1	1	1	1	−1
2	−1	1	−1	−1	−1	−1
2	−1	1	−1	1	−1	1
2	−1	1	−1	1	1	−1
2	−1	1	−1	1	1	1
2	−1	1	1	−1	1	1
2	−1	1	1	1	−1	−1
3	1	−1	−1	−1	−1	−1
3	1	−1	−1	1	−1	1
3	1	−1	−1	1	1	−1
3	1	−1	−1	1	1	1
3	1	−1	1	−1	1	1
3	1	−1	1	1	−1	−1
4	1	1	−1	−1	1	1
4	1	1	−1	1	−1	−1
4	1	1	1	−1	−1	−1
4	1	1	1	−1	−1	1
4	1	1	1	−1	1	−1
4	1	1	1	1	1	1

In this subsection, we have emphasized the use of the D-optimality criterion, although alternative criteria and computer-based approaches do exist. For example, the ability to construct split-plot designs that are optimal by the average variance criterion (8) may be constructed, although general-purpose software for doing so is not yet available. Such designs would be particularly appropriate for estimating response surfaces. In an alternative computer-based

TABLE 10. D-Optimal Split-Plot Response Surface Design

Whole plot	Z_1	Z_2	X_1	X_2	Whole plot	Z_1	Z_2	X_1	X_2
1	−1	−1	−1	−1	7	1	−1	−1	−1
1	−1	−1	−1	1	7	1	−1	−1	1
1	−1	−1	1	−1	7	1	−1	0	1
1	−1	−1	1	1	7	1	−1	1	0
2	−1	1	−1	−1	8	1	1	−1	−1
2	−1	1	−1	1	8	1	1	−1	1
2	−1	1	1	−1	8	1	1	0	−1
2	−1	1	1	1	8	1	1	1	0
3	1	−1	−1	−1	9	−1	0	−1	−1
3	1	−1	−1	1	9	−1	0	−1	1
3	1	−1	1	−1	9	−1	0	0	0
3	1	−1	1	1	9	−1	0	1	1
4	1	1	−1	−1	10	1	0	−1	0
4	1	1	−1	1	10	1	0	0	−1
4	1	1	1	−1	10	1	0	1	−1
4	1	1	1	1	10	1	0	1	1
5	−1	−1	−1	1	11	0	−1	−1	−1
5	−1	−1	0	−1	11	0	−1	−1	0
5	−1	−1	1	0	11	0	−1	0	1
5	−1	−1	1	1	11	0	−1	1	−1
6	−1	1	−1	−1	12	0	1	−1	0
6	−1	1	−1	1	12	0	1	0	1
6	−1	1	0	0	12	0	1	1	−1
6	−1	1	1	−1	12	0	1	1	1

approach, Trinca and Gilmour (2001) presented a strategy for constructing multistratum response surface designs (split plots are a special case of these). This is a multistage approach that moves from the highest stratum to the lowest, where, at each stage, designs are constructed to ensure near orthogonality between strata. Weighted trace-optimal designs—not generally available—would be useful for robust-product designs, because the rankings of effects utilized by Bingham and Sitter (2003) could be easily reflected in the design-optimality criterion.

How to Analyze a Split-Plot Design

In this section, we review methods for the analysis of split-plot designs. We begin with the simplest case—the ANOVA approach for balanced, replicated split-plot experiments—and move on to more complex cases, including the analysis of unreplicated split-plot designs and a completely general approach, based on restricted maximimum likelihood estimation (REML).

Balanced, Replicated Split-Plot Designs

When full factorial designs are balanced and replicated, ANOVA model (1) is fully applicable. The ANOVA table for the corrosion-resistance experiment is given in Table 11. Note that there are two replicates of each of the three whole-plot factor settings, leading to three degrees of freedom for pure whole-plot error. In contrast, the split-plot treatments are not replicated within the whole plots and, for this reason, there are no degrees of freedom for

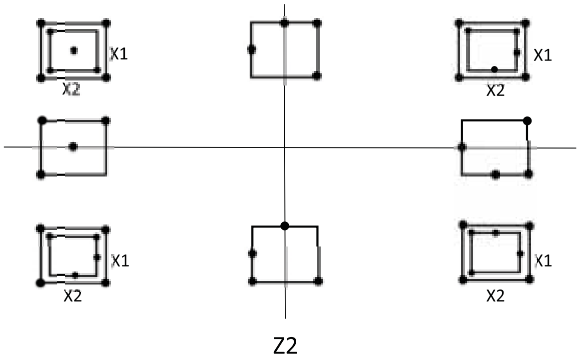

FIGURE 4. D-optimal Split-Plot Response Surface Design Based on 12 Whole Plots. (Note: Figure shows layouts for two whole plots at each corner.

pure split-plot error. An assumption frequently made is that there is no whole-plot error by split-plot factor (i.e., block-by-treatment) interaction and, from this, we obtain $3 \times 3 = 9$ degrees of freedom for split-plot error. From an examination of the EMS column, we discern (by setting all fixed effects to zero in the expressions) that the whole-plot error mean square is the appropriate term for the denominator of the F-test for the effect of temperature, and the subplot error mean square is the appropriate denominator for F-tests for the effects of coatings and for the temperature-by-coating interaction. The general rule for testing factor effects is simple: whole-plot main effects are tested based on the whole-plot mean square error, while all other effects—split-plot main effects and split-plot by whole-plot interactions—are tested using the split-plot mean square error.

From the expected mean square column of the ANOVA table (Table 11), we note that

$$\begin{aligned} \text{E[MS(Whole-Plot Error)]} &= \sigma^2 + b\sigma_w^2 \\ \text{E[MS(Split-Plot Error)]} &= \sigma^2. \end{aligned}$$

It follows that

$$\sigma_w^2 = \frac{\text{E[MS(Whole-Plot Error)]}}{b} - \frac{\text{E[MS(Split-Plot Error)]}}{b}.$$

We obtain unbiased estimators of σ^2 and σ_w^2 by substituting observed for expected mean squares as

$$\begin{aligned} s^2 &= \text{MS(Split-Plot Error)} \\ s_w^2 &= \frac{\text{MS(Whole-Plot Error)}}{b} - \frac{\text{MS(Split-Plot Error)}}{b}. \end{aligned} \tag{9}$$

From the corrosion-resistance ANOVA in Table (11), we obtain $s^2 = 124.5$ and $s_w^2 = (4813.2 - 124.5)/4 = 1172.2$. We note that, if MS(split-plot error) is greater than MS(whole-plot error), the ANOVA-based estimated variance component for whole-plot effects, s_w^2, will be negative. Procedures for constructing confidence intervals for the variance components are detailed in standard texts (see, e.g., Kutner et al. (2005)).

As noted previously, in the corrosion-resistance example, whole plots are replicated, but split plots are not. In this case, a pure-error estimate of whole-plot error is available, whereas the estimate of the split-plot error is based on the whole-plot error by split-plot factor interaction. In choosing a design, the experimenter must weigh the importance of true replicates at both levels of the experiment and their implications for the power of the factor-effects tests to be conducted. In industrial experiments, the cost of whole plots frequently precludes the use of replicates at the whole-plot level and methods for the analysis of unreplicated experiments are required. We turn now to the analysis of unreplicated split-plot designs.

Unreplicated Two-Level Factorial Split-Plot Designs

As is the case with completely randomized unreplicated two-level factorial experiments, an experimenter has a number of options for the analysis of unreplicated split-plot two-level designs. If the design permits the estimation of higher order interactions and if it can be safely assumed that these higher order interactions are not active, the estimates can be pooled to provide unbiased estimates of experimental error at both levels of the experiment. Factor-effects tests can then be conducted in standard fashion. Another approach recommended by Daniel (1959) was the use of two half-normal (or normal) plots of effects, one for whole-plot factor main effects and interactions and one for split-plot factor effects and split-plot-factor by whole-plot-factor interactions.

To illustrate these approaches, we use data from a robust product design reported by Lewis et al. (1997). This experiment has also been discussed by Bingham and Sitter (2003) and Loeppky and Sitter (2002). The experiment studied the effects of customer usage factors on the time between failures of a wafer handling subsystem in an automated memory that relied on optical-pattern recognition to position a wafer for repair. The response was a

TABLE 11. ANOVA for Corrosion Resistance Data

Source	df	SS	MS	F	p	EMS
Temperature (A)	2	26519.3	13259.6	$F_{2,3} = \dfrac{13259.6}{4813.2} = 2.75$	0.209	$\sigma^2 + b\sigma_w^2 + bc\dfrac{\sum \alpha_j^2}{a-1}$
Coatings (B)	3	4289.1	1429.7	$F_{3,9} = \dfrac{1429.7}{124.5} = 11.48$	0.002	$\sigma^2 + ac\dfrac{\sum \beta_j^2}{b-1}$
Temp × Coating (AB)	6	3269.8	545.0	$F_{6,9} = \dfrac{545.0}{124.5} = 4.38$	0.024	$\sigma^2 + c\dfrac{\sum(\alpha\beta)_{ij}^2}{(a-1)(b-1)}$
Whole-plot error	3	14439.6	4813.2	$F_{3,9} = \dfrac{4813.2}{124.5} = 38.65$	0.000	$\sigma^2 + b\sigma_w^2$
Split-plot error	9	1120.9	124.5			σ^2
Corrected total	23	49638.6				

measure of the correlation between an image repair system provided by the pattern-recognition system and the expected image. The experiment had eight control factors, each at two levels: A = training box size, B = corner orientation, C = binary threshold, D = illumination level, E = illumination angle, F = illumination uniformity, G = teach scene angle, and H = train condition. There were three noise factors, each at two levels: p = ambient light intensity, q = initial wafer displacement, and r = initial wafer orientation. The design implemented was a Cartesian-product design, based on a 2^{8-4} resolution IV fractional factorial in the control factors and a 2^{3-1} resolution III fractional factorial in the environmental factors, where the design generators for the whole-plot fractional factors were $E = ABD$, $F = ABC$, $G = BCD$, and $H = ADC$. The single generator for the split-plot design was $r = pq$. Thus, there were 64 runs, with 16 whole plots and four runs in each whole plot.

A JMP half-normal plot of the estimates of the 15 independent whole-plot contrasts is provided in Figure 5. The plot clearly suggests that the effects

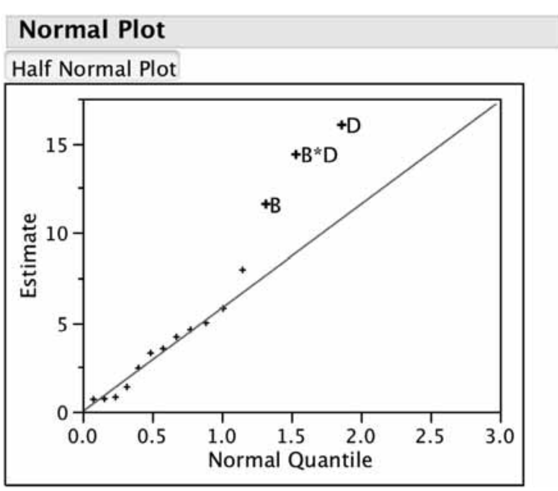

FIGURE 5. Half Normal Plot of Whole-Plot Effects for Lewis et al. (1997) Example.

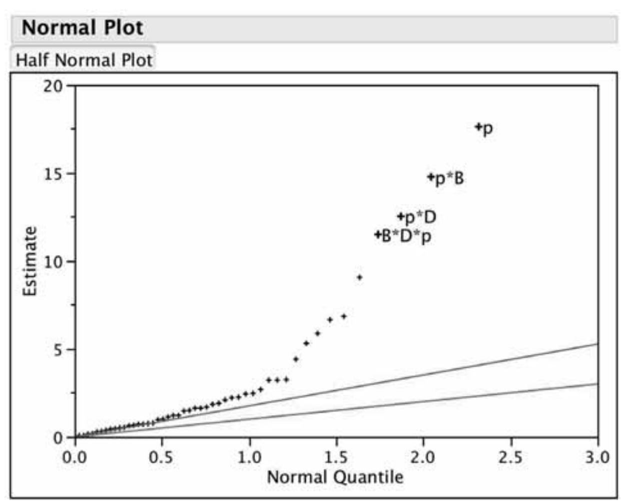

FIGURE 6. Half Normal Plot of Split-Plot Effects for Lewis et al. (1997) Example.

B, D, and BD tend to "fall off" the line and may therefore be active. A half-normal plot of the 48 independent contrasts involving split-plot effects and split-plot by whole-plot interactions is shown in Figure 6. Here, the plot suggests the presence of 10 active effects, p, Bp, Dp, BDp, ABp, Ep, Ap, ADp, BCp, and Cp. Bingham and Sitter (2003) show how the control-by-noise interactions could be used in this application to identify robust settings of the control factors.

Thus far, we have discussed the use of pooling of higher order effects and the use of half-normal plots for the analysis of two-level unreplicated split-plot designs. For completely randomized designs, a number of other popular options exist, including the use of Pareto plots, Lenth's (1989) method, and Bayes plots of estimated effects (Box and Meyer (1986)). Loughlin and Noble (1997) devised a permutation test for effects in unreplicated factorial experiments. Pareto plots can be easily implemented for the analysis of two-level unreplicated factorial designs. As with the half-normal plots, it is only necessary to separate the effects into separate plots for whole-plot factor effects and split-plot factor effects. Lenth's method and the permutation test approach of Loughlin and Noble were recently adapted for use in split-plot experiments by Loeppky and Sitter (2002). To our knowledge, the use of Bayes plots have not yet been discussed in connection with split-plot designs.

The General Case: Restricted Maximum Likelihood Estimation

In the balanced cases discussed thus far, OLS is used for fixed-effects estimates, and ANOVA-based estimators are used to obtain unbiased estimates of the variance components. For the general case, more complex estimation procedures are required. Letsinger et al. (1996) examined three alternatives—OLS, restricted maximum likelihood (REML), and iteratively reweighted least squares (IRLS), where the authors provided special estimating equations for OLS and IRLS. On the basis of simulation studies, they conclude, as do Goos et al. (2006), that the REML approach is preferred. In what follows, we briefly review REML and illustrate its use with two examples.

REML Estimation

As indicated in Equation (4), for split-plot designs, the variance–covariance matrix of the response factor is given by

$$\mathbf{V} = \sigma^2 \mathbf{I}_n + \sigma_w^2 \mathbf{Z}\mathbf{Z}'. \qquad (10)$$

For known $\mathbf{V}$, the best linear unbiased estimator of the fixed-effects parameter vector is obtained using generalized least squares as

$$\hat{\boldsymbol{\beta}}_{GLS} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}. \qquad (12)$$

Because $\mathbf{V}$ is almost always unknown, we substitute estimates of the variance components σ^2 and σ_w^2 in the expression for $\mathbf{V}$ in Equation (10) to obtain the estimated variance–covariance matrix, $\hat{\mathbf{V}}$. Using $\hat{\mathbf{V}}$ in place of $\mathbf{V}$ in Equation (11) results in the feasible generalized least squares (FGLS) estimator,

$$\hat{\boldsymbol{\beta}}_{FGLS} = (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{Y}. \qquad (12)$$

In order to obtain estimates of σ^2 and σ_w^2, REML maximizes the partial log likelihood of the variance components as

$$\begin{aligned} &2\log \mathrm{L}(\sigma^2, \sigma_w^2) \\ &\quad = \text{const.} - \log|\mathbf{V}| - \log|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| \\ &\qquad - \mathbf{Y}'[\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}]\mathbf{Y}. \end{aligned}$$

The resulting estimates, $\hat{\sigma}^2$ and $\hat{\sigma}_w^2$, are used in Equation (12) to obtain the FGLS estimates of the fixed effects. Kackar and Harville (1984) showed that this procedure leads to an unbiased estimator of $\boldsymbol{\beta}$ in small samples.

Statistical Inference

Confidence intervals and tests of hypotheses concerning the fixed effects are based on the estimated variance–covariance matrix of the fixed effects, namely,

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}_{FGLS}) = \mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X},$$

and the asymptotic variance–covariance matrix, $\mathbf{A}$, of the augmented parameter vector $\boldsymbol{\beta}'_{\text{aug}} = (\hat{\boldsymbol{\beta}}', \hat{\sigma}^2, \hat{\sigma}_w^2)$. Suppose that we wish to test hypotheses of the form

$$\mathrm{H}_0 : \mathbf{c}'\boldsymbol{\beta} = 0 \quad \text{vs.} \quad \mathrm{H}_1 : \mathbf{c}'\boldsymbol{\beta} \neq 0,$$

where $\mathbf{c}'\boldsymbol{\beta}$ is an estimable linear combination of the elements of $\boldsymbol{\beta}$. Goos et al. (2006) studied five alternative approaches to conducting this test, as provided by the SAS Mixed procedure, and concluded that the Satterthwaite method (SATTERTH option) and the method of Kenward and Roger (1997) work well in split-plot studies. For the Satterthwaite method, the test statistic is

$$t = \frac{\mathbf{c}'\hat{\boldsymbol{\beta}}}{\sqrt{\mathbf{c}'(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})\mathbf{c}}}.$$

The approximate degrees of freedom for the t value above are

$$\nu = \frac{2[\mathbf{c}'(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})\mathbf{c}]^2}{\mathbf{g}'\mathbf{A}\mathbf{g}}, \tag{13}$$

where $\mathbf{g}$ is the derivative of $\mathbf{c}'(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})\mathbf{c}$ with respect to the augmented parameter vector $\boldsymbol{\beta}_{\text{aug}}$, evaluated at $\hat{\boldsymbol{\beta}}_{\text{aug}}$. The hypothesis-testing approach of Kenward and Roger (1997) is a bit more complex. It is based on an adjusted estimator of the covariance matrix and also employs an expression for approximate degrees of freedom. The method is available in both SAS and SAS/JMP.

Example 1

We consider again the robust product design reported by Lewis et al. (1997) and discussed at the beginning of this section. Because the design is unreplicated, there are no degrees of freedom for estimation of the split-plot and whole-plot error terms, and an initial REML run yields the same results as the OLS results. Examination of the normal probability plots suggests that at least half of the whole-plot and split-plot effects are inactive. In order to obtain p-values for the larger effects, we dropped the smallest 50% of whole-plot and split-plot effects from the model (pooling them into the respective error terms) and performed a second REML analysis. The results are displayed in Figure 7. The p-values will be biased low because of our pooling procedure and, for this reason, it is useful to employ a smaller α level when examining the results. At the $\alpha = 0.01$ level, four effects are identified as active at the whole-plot level (D, B, DB, and AB), and 18 effects (p through AHq in the figure) are significant at the split-plot level. The REML estimates of the two variance components are $\hat{\sigma}_w^2 = 64.825$ and $\hat{\sigma}_\varepsilon^2 = 39.833$. These lead to the estimated variance ratio $\hat{d} = 1.627$, suggesting that the whole-plot error is about 60% larger than the split-plot error.

Parameter Estimates

Term	Estimate	Std Error	DFDen	t Ratio	Prob>|t|
Intercept	936.4375	2.161936	7	433.15	<.0001*
D	16.03125	2.161936	7	7.42	0.0001*
B*D	14.375	2.161936	7	6.65	0.0003*
B	-11.59375	2.161936	7	-5.36	0.0010*
A*B	-7.9375	2.161936	7	-3.67	0.0079*
E	5.78125	2.161936	7	2.67	0.0318*
H	4.96875	2.161936	7	2.30	0.0551
A	4.59375	2.161936	7	2.12	0.0712
A*F	-4.1875	2.161936	7	-1.94	0.0940
p	17.625	0.788921	24	22.34	<.0001*
p*B	14.78125	0.788921	24	18.74	<.0001*
p*D	-12.53125	0.788921	24	-15.88	<.0001*
B*D*p	-11.5	0.788921	24	-14.58	<.0001*
A*B*p	9.0625	0.788921	24	11.49	<.0001*
p*E	-6.84375	0.788921	24	-8.67	<.0001*
p*A	6.65625	0.788921	24	8.44	<.0001*
A*D*p	-5.875	0.788921	24	-7.45	<.0001*
A*F*p	5.3125	0.788921	24	6.73	<.0001*
p*C	4.40625	0.788921	24	5.59	<.0001*
A*H*p	-3.25	0.788921	24	-4.12	0.0004*
p*G	-3.21875	0.788921	24	-4.08	0.0004*
A*D*q	-3.21875	0.788921	24	-4.08	0.0004*
A*G*p	-2.6875	0.788921	24	-3.41	0.0023*
p*H	-2.46875	0.788921	24	-3.13	0.0046*
r*H	2.4375	0.788921	24	3.09	0.0050*
q*H	-2.25	0.788921	24	-2.85	0.0088*
A*H*q	-2.21875	0.788921	24	-2.81	0.0096*
A*H*r	2.09375	0.788921	24	2.65	0.0139*
A*G*q	-1.90625	0.788921	24	-2.42	0.0237*
A*G*r	1.84375	0.788921	24	2.34	0.0281*
q*D	1.6875	0.788921	24	2.14	0.0428*
q*G	-1.625	0.788921	24	-2.06	0.0504
r*E	1.625	0.788921	24	2.06	0.0504

REML Variance Component Estimates

Random Effect	Var Ratio	Var Component	Std Error	95% Lower	95% Upper
WP	1.6274096	64.825149	40.076688	-13.72516	143.37546
Residual		39.833333	11.498893	24.286102	77.089623
Total		104.65848			
-2 LogLikelihood =		353.55646856			

FIGURE 7. REML Analysis of Lewis et al. (1997) Example.

Example 2

Letsinger et al. (1996) provided data on a 28-run, 5-factor central composite response surface design. The experimental design with the response are provided in Table 12. Details concerning the nature of the response could not be provided for reasons of confidentiality. Interestingly, this design was not fielded as either a split-plot design or a completely randomized design. However, it was reported that both Temp1 and Pres1 were difficult to change and so were held constant. From the reported layout, it appears that Temp1 was fully restricted and that inadvertent split-plotting may have occurred with both Pres1 and Temp2. Nonetheless, we will analyze the experiment treating Temp1 and Pres1 as whole-plot factors, as suggested by the authors.

The REML analysis of the second-order response surface model is provided in Figure 8. It identifies three active effects, namely, the quadratic main effect of Humid2, the Humid2 $\times$ Humid1 interaction, and the Humid2 $\times$ Pres1 interaction. The Humid2 $\times$ Temp2 and Temp1 $\times$ Humid1 interaction terms are also marginally significant, having p-values of .0706 and .0501, respectively. We note that the approximate degrees of freedom as given in the Kenward and Roger (1997) approach tend to be between 6.0 and 7.0 for split-plot effects and are smaller for the various whole-plot factors involving only Temp1 and Pres1. The estimate of the variance ratio is just 0.1023, and the point estimates of the whole-plot and split-plot variance components are 228.2 and 2230.8, respectively. (We note that Letsinger et al. (1996) provided both an OLS and an REML analysis of the

TABLE 12. 28 Run, Five-Factor Central Composite Design (Letsinger, et al., 1996)

Run	Whole plot	Temp1	Pres1	Humid1	Temp2	Humid2	Response
1	1	1	−1	−0.8	1	−1	1332
2	2	1	1	−0.73	1	−0.25	1296
3	2	1	1	−0.65	−1	−0.84	1413
4	3	1	−1	0.1	−1	−0.73	954
5	3	1	−1	−0.49	−1	−0.37	1089
6	3	1	−1	0.57	1	0.57	1044
7	4	1	0	0	0	0	1044
8	5	1	1	0.57	−1	0.57	1026
9	5	1	1	1	1	−0.96	1152
10	6	−1	0	−0.02	0	0.02	1026
11	7	−1	1	0.69	1	0.69	990
12	7	−1	1	−0.73	1	−1	1449
13	8	−1	−1	−0.41	1	−1	1170
14	8	−1	−1	1	1	−0.26	1197
15	8	−1	−1	0.41	−1	0.1	1062
16	8	−1	−1	−0.96	−1	−0.96	1017
17	9	−1	1	1	−1	−0.96	999
18	9	−1	1	−0.61	−1	1	882
19	10	0	0	0.09	0	−0.06	1080
20	11	0	1	−0.06	0	−0.06	1098
21	12	0	0	0.14	0	0	1089
22	13	0	−1	0.02	0	0.12	1071
23	14	0	0	0	1	−0.12	1008
24	14	0	0	0.12	0	−0.06	981
25	14	0	0	−0.33	0	−0.22	1035
26	14	0	0	1	0	−0.06	1134
27	14	0	0	0	0	0	1071
28	14	0	0	0	0	1	1260

full model for this experiment, although we have been unable to reproduce either analysis. The similarities in the results suggest that a typographical error in the published data table (Letsinger et al. (1996), p. 383) may have been the cause. Our analyses have been based on the data as published.)

Conclusions

We began this article with reference to the Daniel proclamation that "all industrial experiments are split-plot experiments." While many industrial experiments are, no doubt, run as split-plot experiments, few practitioners deliberately set out to design and analyze their investigations as split-plot experiments. Reasons for this include the widely held views that such designs are generally less powerful than completely randomized designs, that they can be more difficult to analyze, and the knee-jerk inclination by many investigators to treat the completely randomized design as the default option. In this paper, we have argued that split-plot designs should be intentionally implemented with far greater regularity than is currently the case. We have summarized the current state of the art for both the design and the analysis of split-plot experiments and have argued that split-plot designs are often superior to completely randomized designs in terms of cost, efficiency, and validity. In our view, the treatment of split plotting in standard textbooks and in nearly all industrial seminars on the design of experiments is completely inadequate.

While there has been significant methodological development in the past decade, much work remains. The minimum aberration fractional factorial split-

Parameter Estimates

Term	Estimate	Std Error	DFDen	t Ratio	Prob>|t|
Intercept	1059.1651	33.95721	1	31.19	0.0204*
Temp1	40.275617	21.92723	5.079	1.84	0.1248
Pres1	−16.03835	27.34588	5.324	−0.59	0.5815
Pres1*Temp1	35.212096	18.76278	2.441	1.88	0.1779
Temp1*Temp1	−29.68601	46.68923	2.642	−0.64	0.5756
Pres1*Pres1	3.2935849	47.23323	2.591	0.07	0.9494
Humid1	−19.25278	55.21444	6.31	−0.35	0.7387
Temp2	3.5909906	23.6646	6.693	0.15	0.8839
Humid2	−1.854362	45.99115	6.671	−0.04	0.9690
Temp2*Humid1	6.8647566	55.60722	6.803	0.12	0.9053
Humid2*Humid1	145.97299	51.56952	6.491	2.83	0.0275*
Humid2*Temp2	−96.5785	45.92918	6.828	−2.10	0.0746
Humid1*Humid1	88.864553	83.4841	6.946	1.06	0.3227
Temp2*Temp2	−75.248	48.94224	6.946	−1.54	0.1684
Humid2*Humid2	192.62544	50.35873	6.051	3.83	0.0086*
Temp1*Humid1	−121.1118	51.23504	6.996	−2.36	0.0501
Temp2*Temp1	−28.62106	30.7263	6.595	−0.93	0.3844
Humid2*Temp1	9.3102383	50.93679	6.565	0.18	0.8605
Humid1*Pres1	−55.29865	36.43087	6.981	−1.52	0.1729
Temp2*Pres1	25.929809	20.17673	6.996	1.29	0.2397
Humid2*Pres1	−116.9625	39.47101	6.921	−2.96	0.0213*

REML Variance Component Estimates

Random Effect	Var Ratio	Var Component	Std Error	95% Lower	95% Upper	Pct of Total
Whole Plot	0.1022923	228.19839	2106.8993	−3901.324	4357.721	9.280
Residual		2230.8455	1809.1055	719.82268	30152.848	90.720
Total		2459.0439				100.000

−2 LogLikelihood = 111.93225703

FIGURE 8. REML Analysis of Letsinger et al. (1996) Example.

plot designs developed by Huang et al. (1998) and Bingham and Sitter (1999, 2003) need to be incorporated into standard statistical packages. The optimal design algorithms (e.g., Goos and Vandebroek (2001), Jones and Goos (2007)) need to be more generally available. Integrated variance, trace-optimality, and other criteria need to be incorporated into the available design algorithms. The ability to generate equivalent estimation designs algorithmically or to combine classical optimality criteria with the generation of equivalent estimation designs would be a welcome development. Little work has been done in the areas of split-plot response surface and mixture experiments, and only recently have split-split-plot structures begun to receive renewed attention (Jones and Goos (2009)). On the analysis side, REML estimation routines need to be more accessible, and diagnostics for mixed model analysis of variance (e.g., Beckman et al. (1987)) are still not generally available. We look forward to these and other developments as the importance of these designs becomes more widely recognized.

References

Almimi, A. A.; Kulahci, M.; and Montgomery, D. C. (2008). "Follow-Up Designs to Resolve Confounding in Split-Plot Experiments". *Journal of Quality Technology* 40, pp. 154–166.

Anbari, F. T. and Lucas, J. M. (1994). "Super-Efficient Designs: How to Run Your Experiment for Higher Efficiencies and Lower Cost". *ASQC Technical Conference Transactions*, pp. 852–863.

Anbari, F. T. and Lucas, J. M. (2008). "Designing and Running Super-Efficient Experiments: Optimum Blocking with One Hard-to-Change Factor". *Journal of Quality Technology* 40, pp. 31–45.

Beckman, R. J.; Cook, R. D.; and Nachtsheim, C. J. (1987). "Diagnostics for Mixed-Model Analysis of Variance". *Technometrics* 29, pp. 413–426.

Bingham, D. and Sitter, R. R. (1999). "Minimum-Aberration Two-Level Fractional Factorial Split-Plot Designs". *Technometrics* 41, pp. 62–70.

Bingham, D. R. and Sitter, R. R. (2001). "Design Issues in Fractional Factorial Split-Plot Experiments". *Journal of Quality Technology* 33, pp. 2–15.

Bingham, D. and Sitter, R. R. (2003). "Fractional Factorial Split-Plot Designs for Robust Parameter Experiments". *Technometrics* 45, pp. 80–89.

Bingham, D. R.; Schoen, E. D.; and Sitter, R. R. (2004). "Designing Fractional Factorial Split-Plot Experiments with Few Whole-Plot Factors". *Journal of the Royal Statistical Society, Series C* 53, pp. 325–339.

Bisgaard, S. (2000). "The Design and Analysis of $2^{k-p} \times 2^{q-r}$ Split Plot Experiments". *Journal of Quality Technology* 32, pp. 39–56.

Box, G.; Hunter, W.; and Hunter, S. (2005). *Statistics for Experimenters: Design, Innovation, and Discovery*, 2nd edition. New York, NY: Wiley-Interscience.

Box, G. and Jones, S. (1992). "Split-Plot Designs for Robust Product Experimentation". *Journal of Applied Statistics* 19, pp. 3–26.

Box, G. and Meyer, R. D. (1986). "An Analysis for Unreplicated Fractional Factorials". *Technometrics* 28, pp. 11–18.

Daniel, C. (1959). "Use of Half-Normal Plots in Interpreting Factorial Two-Level Experiments". *Technometrics* 1, pp. 311–341.

Draper, N. R. and John, J. A. (1998). "Response Surface Designs Where Levels of Some Factors Are Difficult to Change". *Australian and New Zealand Journal of Statistics* 40, pp. 487–495.

Fisher, R. A. (1925). *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd.

Ganju, J. and Lucas, J. M. (1997). "Bias in Test Statistics when Restrictions in Randomization Are Caused by Factors". *Communications in Statistics: Theory and Methods* 26, pp. 47–63.

Ganju, J. and Lucas, J. M. (1999). "Detecting Randomization Restrictions Caused by Factors". *Journal of Statistical Planning and Inference* 81, pp. 129–140.

Ganju, J. and Lucas, J. M. (2005). "Randomized and Random Run Order Experiments". *Journal of Statistical Planning and Inference* 133, pp. 199–210.

Goos, P. (2002). *The Optimal Design of Blocked and Split-Plot Experiments*. New York, NY: Springer.

Goos, P. (2006). "Optimal Versus Orthogonal and Equivalent-Estimation Design of Blocked and Split-Plot Experiments". *Statistica Neerlandica* 60, pp. 361–378.

Goos, P. and Donev, A. (2007). "Tailor-Made Split-Plot Designs for Mixture and Process Variables". *Journal of Quality Technology* 39, pp. 326–339.

Goos, P.; Langhans, I.; and Vandebroek, M. (2006) "Practical Inference from Industrial Split-Plot Designs". *Journal of Quality Technology* 38, pp. 162–179.

Goos, P. and Vandebroek, M. (2001). "Optimal Split-Plot Designs". *Journal of Quality Technology* 33, pp. 436–450.

Goos, P. and Vandebroek, M. (2003). "D-Optimal Split-Plot Designs with Given Numbers and Sizes of Whole Plots". *Technometrics* 45, pp. 235–245.

Goos, P. and Vandebroek, M. (2004). "Outperforming Completely Randomized Designs". *Journal of Quality Technology* 36, pp. 12–26.

Goos, P.; Langhans, I.; and Vandebroek, M. (2006). "Practical Inference from Industrial Split-Plot Designs". *Journal of Quality Technology* 38, pp. 162-179.

Huang, P.; Dechang, C.; and Voelkel, J. O. (1998). "Minimum-Aberration Two-Level Split-Plot Designs". *Technometrics* 40, pp. 314–326.

Jones, B. and Goos, P. (2007). "A Candidate-Set-Free Algorithm for Generating D-optimal Split-Plot Designs". *Applied Statistics* 56, pp. 347–364.

Jones, B. and Goos, P. (2009). "D-Optimal Split-Split-Plot Designs". *Biometrika* 96, pp. 67–82.

Ju, H. L. and Lucas, J. M. (2002). "L^k Factorial Experiments with Hard-To-Change and Easy-To-Change Factors". *Journal of Quality Technology* 34, pp. 411–421.

Kackar, A. N. and Harville, D. A. (1984). "Approximations for Standard Errors of Estimators of Fixed and Random Effects in Mixed Linear Models". *Journal of the American Statistical Association* 79, pp. 963–974.

Kenward, M. G. and Roger, J. H. (1997) "Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood" *Biometrics* 53, pp. 983–997.

Kowalski, S. M. (2002). "24 Run Split-Plot Experiments for Robust Parameter Design". *Journal of Quality Technology* 34, pp. 399–410.

Kowalski, S. M.; Cornell, J. A.; and Vining, G. G. (2002). "Split-Plot Designs and Estimation Methods for Mixture Experiments with Process Variables". *Technometrics* 44, pp. 72–79.

Kowalski, S. and Vining, G. G. (2001). "Split-Plot Experimentation for Process and Quality Improvement". In *Frontiers in Statistical Quality Control 6*, H. Lenz (ed.), pp. 335–350. Heidelberg: Springer-Verlag.

Kutner, M.; Nachtsheim, C. J.; Neter, J.; and Li, W. (2005). *Applied Linear Statistical Models*. Chicago, IL: McGraw-Hill/Irwin.

Letsinger, J. D.; Myers, R. H.; and Lentner, M. (1996). "Response Surface Methods for Bi-Randomization Structures". *Journal of Quality Technology* 28, pp. 381–397.

Lewis, D. K.; Hutchens, C.; and Smith, J. M. (1997). "Experimentation for Equipment Reliability Improvement". In *Statistical Case Studies for Industrial Process Improvement*, V. Czitrom and P. D. Spagon (eds.), pp. 387–401. Philadelphia, PA: ASA and SIAM.

Loeppky, J. L. and Sitter, R. R. (2002). "Analyzing Unreplicated Blocked or Split-Plot Fractional Factorial Designs". *Journal of Quality Technology* 43, pp. 229–243.

Meyer, R. K. and Nachtsheim, C. J. (1995). "The Coordinate Exchange Algorithm for Constructing Exact Optimal Exponential Designs". *Technometrics* 37, pp. 60–69.

McLeod, R. G. and Brewster, J. F. (2004). "The Design of Blocked Fractional Factorial Split-Plot Experiments". *Technometrics* 46, pp. 135–146.

McLeod, R. G. and Brewster, J. F. (2008). "Optimal Foldover Plans for Two-Level Fractional Factorial Split-Plot Designs". *Journal of Quality Technology* 40, pp. 227–240.

Montgomery, D. C. (2008). *Design and Analysis of Experiments*, 7th ed. New York, NY: John Wiley & Sons.

Taguchi, G. (1987). *System of Experimental Design*. White Plains, NY: Kraus International Publications.

Trinca, L. A. and Gilmour, S. G. (2001). "Multistratum Response Surface Designs". *Technometrics* 43, pp. 25–33.

Vining, G. G.; Kowalski, S. M.; and Montgomery, D. C. (2005). "Response Surface Designs Within a Split-Plot Structure". *Journal of Quality Technology* 37, pp. 115–129.

Vining, G. G. and Kowalski, S. M. (2008). "Exact Inference for Response Surface Designs Within a Split-Plot Structure". *Journal of Quality Technology* 40, pp. 394–406.

Webb, D. F.; Lucas, J. M.; and Borkowski, J. J. (2004). "Factorial Experiments when Factor Levels Are Not Necessarily Reset". *Journal of Quality Technology* 36, pp. 1–11.

Yates, F. (1935) "Complex Experiments, with Discussion". *Journal of the Royal Statistical Society, Series B* 2, pp. 181–223.