Quality Engineering, 19:1–15, 2007

ISSN: 0898-2112 print/1532-4222 online
DOI: 10.1080/08982110601057179

Tutorial: Industrial Split-plot Experiments

Scott M. Kowalski
Minitab Inc. State College, PA

Peter A. Parker
NASA Langley Research Center, Hampton, VA

G. Geoffrey Vining
Virginia Polytechnic Institute and State University, Blacksburg, VA

ABSTRACT Many industrial experiments involve two types of factors: those that are hard-to-change and those that are easy-to-change (ETC). Hard-to-change (HTC) factors have levels that are difficult and/or expensive to change. As a result, the experimenter would prefer to run the experiment in such a manner as to minimize the number of times that he/she must change the levels of these factors. Unfortunately, it is precisely the changing of these levels that provides the information about the effects of the HTC factors. Consequently, when we minimize the number of times we change the levels of these factors, we also minimize the relevant information about their effects.

This paper summarizes the structure and the analysis of industrial split-plot experiments. The purpose of this article is to teach practitioners how to identify split-plot experimental conditions, how to run the experiment efficiently, and then how to analyze the results. The article illustrates both first-order and second-order experiments. The first four sections provide a basic background on experimental design and an introduction to first-order split-plot experiments. The remainder of this article contains more advanced topics dealing with second-order, split-plot experiments.

KEYWORDS design of experiments, response surface methodology, split-plot experiments

1. INTRODUCTION

In an effort to continually improve their products, many organizations engage in some form of quality initiative, such as Six Sigma. A commonly used approach involves a sequence of designed experiments, where the results from the initial experiments guide the design of future experiments. Response surface methodology (RSM) is an example of such a strategy for product or process optimization.

This article seeks to introduce industrial split-plot experiments to practitioners with various levels of background in experimental design and analysis. The first section gives an overview of the fundamentals of designed experiments. Section 1 is an introduction that is intended to serve as both a refresher for those familiar with experimental design and as a primer for those less familiar. The next three sections cover the necessary information for designing and analyzing a basic split-plot experiment. Section 2 is an overview of the structure of a basic split-plot experiment. The third section

Address correspondence to Scott M. Kowalski, 1829 Pine Hall Road, Minitab Inc., State College, PA. E-mail: skowalski@minitab.com

presents some simple examples of split-plot experiments. Section 4 discusses the design and analysis of first-order split-plot experiments. The fifth section provides the general conditions on designs that make the ordinary least squares (OLS) estimates of the model parameters equivalent to the generalized least squares (GLS) estimates. Section 6 addresses second-order split-plot experiments. This article concludes with some suggested reading for those who would like more information.

1.1. Design of Experiments

Design of experiments involves making simultaneous, intentional changes to process input variables, called factors, in order to observe changes in the output response. Before conducting an experiment, the experimenter must identify the potentially important factors and then define the appropriate levels for these factors. A treatment or treatment combination is simply a specific combination of the levels of all the factors used in the experiment. If all of the k factors are at two levels, then a full factorial involves running all of the 2^k treatment combinations.

Experimental protocol leads to two types of units: experimental units and observational units. The smallest unit of experimental material to which a treatment can be applied *independently* of all other units is known as the experimental unit. By definition experimental units must be able to receive different treatments. Each experimental unit contains one or more observational units on which the measurements are taken. It is crucial to note that by their nature observational units for the same experimental unit must receive the same treatment. Consequently, the treatments are not assigned independently to the observational units.

In most experiments, the experimental unit and the observational unit are the same. For example, consider an experiment that applies a treatment combination to a cylindrical object and then measures the object's breaking strength. Because the smallest unit to which a treatment is applied is an individual cylinder and the single measurement of breaking strength is also taken on the cylinder, then both the experimental and observational units are the same cylindrical object. However, suppose the moistness of a cake is of interest. This experiment mixes up the process factors, which are poured into a pan and then baked. After the cake is baked, suppose it is cut into four slices, and the moistness is measured for each slice. The experimental unit is the cake because it is the smallest unit to which the factor combinations were applied independently of any other combination. The observational units are the smaller sections from each cake because the moistness measurements are taken on them. This distinction is important for all experimental designs because the variation among the experimental units is the basis for the experimental error used in the analysis of the data. In the cake example, it is common to average the four moistness responses for each cake and use that average as the response. We will return to this important topic later for split-plot experiments.

People often speak of three basic principles for the conduct of experimental design:

- randomization,
- replication, and
- local control of error

Randomization means randomly assigning the treatments to the experimental units. This typically involves using a computer or random number table to mix the treatments up into some random order. Randomization ensures that every experimental unit has an equal chance of receiving any specific treatment. Besides helping to satisfy certain statistical properties of the analysis, randomization helps to eliminate the effect of any extraneous factors on the results. Replication means assigning at least one treatment to more than one experimental unit. The main purpose of replication is to create an estimate of experimental error which can be used to construct statistical tests for factor significance. Local control of error seeks to minimize the impact of effects of possible "nuisance" influences not of direct concern for the specific experiment. Blocking, which is a classic example of local control of error, groups similar experimental conditions into "blocks." For example, suppose an experiment requires more than one batch of raw material. The experimenter recognizes that batches of raw material differ, but the differences among these batches are not of interest to the experimenter. Essentially, the batches of raw materials represent a source of nuisance variation. Blocking refers to running one set of treatments using one batch of raw material and then another set of treatments using a different batch. It is important to note that blocking affects randomization. In the first block, the treatments are run in some random order. In the second block, the treatments are run according to a second randomization scheme. It is important to note that while the batch defines a block, it is not a factor in the experiment. The experimenter does not control the batch or make intentional changes to the batch levels (the levels are simply one and two). The experimenter simply is acknowledging that there could be batch differences and wants to make sure the factor comparisons are as unaffected as possible by these block differences.

There are many types of experimental designs including the split-plot design. The completely randomized design (CRD) and the randomized complete block design (RCBD) are the two most commonly used industrial experimental designs. In a CRD, the t treatments are replicated r times, and all of the $N = rt$ runs in the experiment are performed in random order. In other words, suppose $t = 4$ and $r = 3$. Think of putting 12 slips of paper in a hat (3 from each of treatments 1–4). Select one piece of paper and run that treatment first. Then, select another piece of paper and run that treatment next. Continue until all 12 pieces have been selected (see Table 1). In a RCBD, the t treatments are replicated r times in b blocks with $r = b$. Using the same scenario as above, place just one piece of paper for each treatment in the hat (so there are only four pieces of paper in the hat). For block one, select a piece of paper and assign that treatment to the first unit. Select another piece of paper and assign that treatment to the second unit. Once all four pieces of paper have been selected, place them back in the hat and repeat the process for blocks two and three. Notice now that in the first four runs of block one, it is guaranteed that all four treatments are run. So essentially a block design restricts how the randomization is done, see Table 1. Hopefully, it is clear that the experimental protocol accounts for any possible differences among the blocks, which then minimizes the effect of the blocks on the treatment analysis. In essence, the proper design of the experiment allows the analyst to remove the nuisance variation created by the blocking variable.

TABLE 1 **One Possible Randomization for a Completely Randomized Design (CRD) with Four Treatments and One Possible Randomization for a Randomized Complete Block Design (RCBD) with Four Treatments in Three Blocks**

CRD	RCBD	
Treatment	Block	Treatment
1	1	2
2	1	3
2	1	1
3	1	4
4	2	3
1	2	4
3	2	1
2	2	2
4	3	4
1	3	2
3	3	1
4	3	3

1.2. Analysis of Variance

Consider a 12 run CRD involving a single factor with four levels. In this case $N = 12$, $t = 4$ treatments, and $r = 3$ replicates for each treatment. In this case, the total number of experimental runs, N, is $N = rt$. The analysis of data from this experiment is referred to as a one-way analysis of variance (ANOVA). The *one* here refers to having only one factor. It is useful to describe the experimental situation using a model. For this example, an appropriate model is

$$Y_{ij} = \mu_i + \epsilon_{ij} \quad \text{for } i = 1, 2, 3, 4 = t \text{ and } j = 1, 2, 3 = r$$

where Y_{ij} is the response for the jth observation from treatment i, μ_i is mean of the ith treatment, and ε_{ij} is the random error associated with the jth replicate of the ith treatment. Let

$$Y_{i.} = \sum_{j=1}^{r} Y_{ij}, \quad \bar{Y}_{i.} = \frac{\sum_{j=1}^{r} Y_{ij}}{r}$$

$$Y_{..} = \sum_{i=1}^{t}\sum_{j=1}^{r} Y_{ij}, \quad \text{and} \quad \bar{Y}_{..} = \frac{\sum_{i=1}^{t}\sum_{j=1}^{r} Y_{ij}}{N}$$

The dot notation simply means the summation over the particular subscript.

The question of interest from this experiment is whether there are any differences among the treatment means. Statistics addresses this question through a hypothesis test. The appropriate hypotheses are

$$\text{null hypothesis} : H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$
$$\text{alternative hypothesis} : H_1 : \text{at least one} \neq$$

TABLE 2 Analysis of Variance Table for One Factor with Four Levels

Source of variation	Degrees of freedom	Sums of squares	Mean square	F-statistic
Treatment	$t-1=3$	$SS_{Treat} = r\sum_{i=1}^{t}(\bar{Y}_{i.} - \bar{Y}_{..})^2$	$MS_{Treat} = \frac{SS_{Treat}}{t-1}$	$F = \frac{MS_{Treat}}{MS_{Error}}$
Error	$N-t=8$	$SS_{Error} = \sum_{i=1}^{t}\sum_{j=1}^{r}(Y_{ij} - \bar{Y}_{i.})^2$	$MS_{Error} = \frac{SS_{Error}}{N-t}$	
Total	$N-1=11$	$SS_{Total} = \sum_{i=1}^{t}\sum_{j=1}^{r}(Y_{ij} - \bar{Y}_{..})^2$		

An ANOVA table is used to make a decision about the two hypotheses by breaking down the variation in the data. Table 2 shows the appropriate ANOVA table for this example. The sums of squares column breaks down the total variation in the data into the important pieces for the experiment. The degrees of freedom are the number of independent elements in the sums of squares. The mean squares are standardized versions of the sums of squares. If the F-ratio is large, there is statistical evidence that at least one treatment mean is different from the others (i.e., there is support for the alternative hypothesis).

Consider a variable called Rep that indicates how many of each treatment are run. In our example, Rep would look like

Treatment	1	1	1	2	2	2	3	3	3	4	4	4
Rep	1	2	3	1	2	3	1	2	3	1	2	3

It is important to note that the sums of squares error from the CRD in Table 2 is identical to a more complicated statistical modeling term involving nesting (because what is called Rep 1 in treatment 1 is not the same as what is called Rep 1 in treatments 2 and 3). Hence, the term is written as Rep(treatment) and indicates that Rep = 1 in treatment one is not the same as Rep = 1 in treatment two. In other words,

$$SS_{\text{Rep(Treatment)}} = \sum_{i=1}^{t}\sum_{j=1}^{r}(Y_{ij} - \bar{Y}_{i.})^2 = SS_{\text{Error(CRD)}}$$

and the degrees of freedom are $t(r-1) = 4(3-1) = 8$, which is the same as the degrees of freedom from the CRD error term in Table 2. This nested idea will play an important role in the analysis of the split-plot design shown in later sections. For more details on design of experiments and their analysis, see Montgomery (2004).

We will now build on this basic introduction to design of experiments and analysis of variance to understand how to recognize and analyze a split-plot experiment.

2. INTRODUCTION TO SPLIT-PLOT EXPERIMENTS

Industrial experimenters often encounter situations where some experimental factors are hard to change or where there is significant discrepancy in the size of some experimental units. In particular, it may be difficult to change the level for one of the factors. In this case, practitioners typically fix the level of the difficult-to-change factor and run all the combinations of the other factors, which leads to a split-plot design. Indeed the mixture of hard-to-change (HTC) and easy-to-change (ETC) factors is probably the norm rather than the exception in industrial practice. Too frequently, practitioners analyze a split-plot experiment as a CRD, which often leads to very misleading conclusions, especially with regard to the HTC factors.

Split-plot designs have their origins in agricultural experiments. In these experiments, a factor such as irrigation method is randomly applied to large sections of land (called whole plots). These sections then are *split* into smaller *plots* (called subplots) and another factor, such as different fertilizers, is applied in random order (see Table 3). In split-plot experiments, the experimental units for the whole-plot factors (irrigation method) are large. The smaller plots represent observational units (or repeats) for the whole-plot factor, but they are the experimental units for the subplot factors (fertilizer type). This is extremely important because, as discussed earlier, experimental error comes from replication of experimental units. Therefore, the analysis of a split-plot design will consist of two error terms because there are two types of experimental units.

Table 3 is similar to the RCBD discussed in the first section. However, as pointed out in Section 1, a block variable is not intentionally changed to see its effect on the response (i.e., it is not a factor in the experiment). Clearly, irrigation method is a factor and is intentionally changed. Section 4 discusses in more detail the differences between a split-plot design and a block design.

TABLE 3 An Agricultural Split-plot Design

Apply irrigation method 1	Apply irrigation method 2	Apply irrigation method 2	Apply irrigation method 1
Apply fertilizer 1	Apply fertilizer 2	Apply fertilizer 3	Apply fertilizer 1
Apply fertilizer 3	Apply fertilizer 1	Apply fertilizer 2	Apply fertilizer 4
Apply fertilizer 4	Apply fertilizer 3	Apply fertilizer 1	Apply fertilizer 2
Apply fertilizer 2	Apply fertilizer 4	Apply fertilizer 4	Apply fertilizer 3

In recent years, industrial experimenters have recognized the importance of split-plot designs (see Huang et al., 1998; Bingham and Sitter, 1999, 2001; Bisgaard, 2000; Bisgaard and Kulahci, 2001). Most of this work focuses on first-order models and the design of screening experiments. In addition, some work has been done on split plots in the robust parameter design setting (Box and Jones, 1992, 2001; Bisgaard, 2000; Kowalski, 2002).

3. EXAMPLES OF INDUSTRIAL SPLIT-PLOT EXPERIMENTS

HTC factors occur often in real world applications. Many times the water chilling temperature used in injection molding processes can be considered hard-to-change because of the time to stabilize. Wind tunnel experiments often require the tunnel to be shut down in order to change certain environmental factors, which again takes a significant amount of time to restabilize the tunnel conditions.

Consider an experiment that examines the image quality of a printing process. The four factors, all at two-levels, identified as potentially important are: (a) Blanket type (in terms of thickness), (b) Cylinder gap, (c) Ink flow, and (d) Press speed. Changing the cylinder gap, ink flow, and press speed is a very simple procedure and merely consists of making an adjustment on a control panel while the printing press is still running. Therefore, these three are the ETC factors in the experiment. Changing the blanket type, on the other hand, requires the press to be stopped and a manual replacement of the blanket type. Thus, blanket type is a HTC factor. A CRD would run the 16 treatment combinations in a random order, requiring frequent stopping of the press so that the blanket type could be changed. Instead, a split-plot design could be used with one whole-plot variable and three subplot variables. A clever experimenter can run the same 16 combinations only changing the blanket type twice.

Kowalski and Potcner (2003) give a nice overview of how to recognize a split-plot design in industry. One of their examples involves the water resistance property of wood, which can be related to the type of wood pretreatment and the stain type used. Consider an experiment involving two types of pretreatment and four types of stains. It is difficult to apply the pretreatment to a small piece of wood. Therefore, the experiment is conducted by applying a pretreatment to a an entire board, then the board is cut into four pieces and the different stains are applied. As a result, the experimental unit for the pretreatment is an entire board while the experimental unit for the stain is one of the smaller cut pieces.

An example of robust parameter design involves finding the settings that optimize the moisture content of a cake. Consider flour, oil, and egg powder as the design factors. The noise factors are oven temperature and baking time. Fixing the temperature and time allows eight cakes to be baked together. Designating the noise factors as the whole plots and the design factors as the subplots creates an efficient split-plot design.

Simpson et al. (2004) study the lift and drag of a racecar in a wind tunnel experiment. The front and rear end heights of the car were hard-to-change factors because the wind tunnel had to be shut down and it took about 45 minutes to change either factor. In addition, the yaw angle and grille cover were two ETC factors. They showed how using a split-plot design reduced the experimental time by about 1/3 over a CRD.

4. FIRST-ORDER SPLIT-PLOT EXAMPLE

Many designed experiments in industrial environments result in split-plot applications. For every split-plot experiment there are two randomizations. The hard-to-change factors, often called whole-plot treatments, are randomly assigned to whole plots based on the whole-plot design. Within each whole-plot, the ETC factors, often called subplot treatments, are randomly assigned to subplots with a separate randomization within each whole-plot. As previously mentioned, this strategy leads to a split-plot design where the experimental unit for the HTC factors is subdivided into experimental units for the ETC factors. As previously mentioned, a consequence of this is that there are two error terms: one for the whole-plot treatments and one for subplot treatments. Typically, the interaction between whole-plot treatments and subplot treatments is a subplot effect as well.

We use an example to illustrate the design and analysis of a split-plot design for fitting a first-order model with interactions. Consider an experiment involving the strength of plastic. The four factors identified as potentially important are: (a) baking temperature, (b) additive percentage, (c) agitation rate, and (d) processing time. Each factor has two levels: low = –1 and high = 1. The levels are left in coded units for proprietary reasons. The experiment first makes a mold from the additive percentage, the agitation rate, and the processing time. Then the mold is baked at one of the temperatures. Finally, the strength of the plastic is measured.

Consider two replicates in order to estimate the experimental error. To conduct this experiment as a CRD, all $2^4 = 16$ treatment combinations would be run twice in random order, yielding 32 observations. This involves making 32 molds, which represent four for each of the eight different combinations of additive, rate, and time. Then, these 32 molds would be baked each individually at one of two temperatures. If the experiment is conducted in this manner, it would involve frequent changes to the baking oven's temperature, which takes some time to stabilize. A CRD also implies each run of the oven is a true experimental run, which means 32 separate runs of the oven. Such an approach adds considerable time and expense to the experiment.

It is important to note that the oven can accommodate more than one mold at a time. A more efficient approach bakes all $2^3 = 8$ mold combinations involving additive percentage, agitation rate, and processing time for one temperature setting at the same time. This would be followed by another run of the oven at the other temperature level, repeating the process until all the desired number of replicates for the temperature factor (two in our example) are run.

We note that such an approach is no longer a CRD but a split-plot design because

- the experimental unit for temperature is the oven while for the other three factors (additive percentage, agitation rate and processing time), the experimental unit is one mold of plastic. Therefore, the observational units for the temperature factor are the experimental units for the other three factors.
- Temperature is a HTC factor, and the three ETC factors are varied within a level of the HTC factor.
- The temperature factor uses a different randomization scheme from the other factors. First, a temperature is chosen at random, then the eight combinations of the ETC factors are randomly assigned within the oven.

TABLE 4 Layout of Split-plot Design for Plastic Example (Additive is Abbreviated as Add)

High temperature			Low temperature			Low temperature			High temperature		
Add	Rate	Time	Add	Rate	Time	Add	Rate	Time	Add	Rate	Time
–1	1	1	1	–1	–1	–1	–1	–1	–1	–1	–1
1	–1	1	1	–1	1	–1	–1	1	1	1	–1
–1	–1	–1	–1	1	–1	1	1	–1	–1	1	1
1	1	1	1	1	1	1	1	1	1	–1	–1
–1	1	–1	–1	1	1	–1	1	–1	–1	1	–1
1	–1	–1	1	1	–1	1	–1	1	1	1	1
–1	–1	1	–1	–1	1	–1	1	1	1	–1	1
1	1	–1	–1	–1	–1	1	–1	–1	–1	–1	1

The resulting design (see Table 4) looks similar to a block design. However, there are several differences:

- A block is a collection of relatively homogenous experimental units. It is not a factor that is applied to the experimental units. For example, consider an experiment replicated twice with each replicate run on a different day. Thus, the design is blocked by day. The experimenter does not apply the day (say Wednesday) to the units. Therefore, it is not under the control of the experimenter. Instead, day represents a collection of experimental units.
- In some experiments, the HTC by ETC factor interactions are important to estimate. For example, consider a robust parameter design situation. Suppose the HTC factor represents a noise factor, and the ETC factors represent the design factors. The main focus of the experiment is to understand the noise by design interactions. In a randomized complete block design, we cannot estimate the interaction between blocks and treatments since this interaction represents the error term.
- There is only one error term in a block design because there is only one type of experimental unit.

The main purpose of using blocks is to get groups of units that are more homogenous and to remove the variation between blocks from the estimate of experimental error.

The split-plot design for this experiment is just an extension of a two-level factorial design. Therefore, the design supports the fitting of a first-order model with interactions. In general, we could write the model as

$$Y = \text{WP factors} + \text{WP error} + \text{SP factors} + \text{WP} \times \text{SP interactions} + \text{SP error}$$

where WP represents whole-plot and SP represents subplot. For our particular example, temperature is the whole-plot factor. Additive percentage, agitation rate and processing time are the subplot factors. It is sometimes easier to think of the analysis of a split-plot experiment as two separate experiments corresponding to the two levels of the split-plot experiment: the whole-plot level and the subplot level.

4.1. Whole-plot Level Only

Again, suppose the experiment uses two replicates of the temperature factor with the data shown in Table 5. This approach involves four uses of the oven: two at the high temperature setting and two at the low temperature setting. For now, focus on only these four oven uses, ignoring the other three factors. Because the molds from the other three factors represent observational units for temperature, the 8 repeats in each oven will be averaged, which results in only four observations, each being the mean of one of the whole plots. Because these four oven cycles are randomly assigned a temperature, the whole-plot design can be thought of as a CRD with two replicates. Because there is only one factor, temperature, the analysis is a one-way analysis of variance with temperature at two levels. Hence, there are $4-1=3$ total degrees of freedom (df) for the whole-plot level of the experiment. Temperature has $2-1=1$ df since it has

TABLE 5 **Data for the Plastic Example**

Whole-plot	Temp	Additive	Rate	Time	Strength	Whole-plot mean
1	1	−1	1	1	68.5	
1	1	1	−1	1	66.8	
1	1	−1	−1	−1	58.5	
1	1	1	1	1	70.8	62.94
1	1	−1	1	−1	61.3	
1	1	1	−1	−1	51.9	
1	1	−1	−1	1	59.5	
1	1	1	1	−1	66.2	
2	−1	1	−1	−1	57.4	
2	−1	1	−1	1	57.5	
2	−1	−1	1	−1	56.5	
2	−1	1	1	1	63.9	57.81
2	−1	−1	1	1	56.4	
2	−1	1	1	−1	58.1	
2	−1	−1	−1	1	53.2	
2	−1	−1	−1	−1	59.5	
3	−1	−1	−1	−1	66.6	
3	−1	−1	−1	1	63.9	
3	−1	1	1	−1	62.6	
3	−1	1	1	1	63.2	62.93
3	−1	−1	1	−1	56.1	
3	−1	1	−1	1	63.3	
3	−1	−1	1	1	62.7	
3	−1	1	−1	−1	65.0	
4	1	−1	−1	−1	59.5	
4	1	1	1	−1	64.0	
4	1	−1	1	1	68.0	
4	1	1	−1	−1	65.6	64.34
4	1	−1	1	−1	58.6	
4	1	1	1	1	73.3	
4	1	1	−1	1	61.5	
4	1	−1	−1	1	64.2	

TABLE 6 Whole-plot Analysis of the means using MINITAB

Source	df	SS	MS	F	p
Temperature	1	10.68	10.68	1.52	.343
Error	2	14.05	7.02		
Total	3	24.73			

two levels and there are 2 df for error coming from the two replicates (1 df from running the high level twice and 1 df from running the low level twice).

Table 6 shows the whole-plot analysis using the means from each whole-plot. This analysis gives the F-statistic $= MS_{Temp}/MS_{\mathrm{Error}}$ and the p-value. Please note that the sums of squares in Table 6 are reduced by a factor of 8 (the number of subplots in each whole-plot) from what will be produced from the actual split-plot analysis later. Recall from Section 1.2 that the error term in a CRD is a nested term. In other words, if we number the four uses of the oven temperature 1–4 and call it WP (similar to Rep in Section 1.2), then the error term (MS_{Error}) from Table 6 is the same as putting the nested term WP (temperature) in the model and indicating that WP is random. We point this fact out because in some popular software packages, such as MINITAB and Proc GLM in SAS, the user must build the whole-plot error term by entering this type of nested term (see the Appendix). It is not necessary to do this when using the whole-plot means as we have done above, but it is necessary when doing the combined analysis presented next.

4.2. Split-plot Analysis

The first-order model including two-factor interactions for our split-plot design is

$$\begin{aligned}\text{Strength} = \beta_0 + \beta_1 Z + \text{whole-plot error} + \beta_2 X_2 \\ + \beta_3 X_3 + \beta_4 X_4 + \beta_{23} X_2 X_3 + \beta_{24} X_2 X_4 \\ + \beta_{34} X_3 X_4 + \beta_{12} Z X_2 + \beta_{13} Z X_3 \\ + \beta_{14} Z X_4 + \text{subplot error}\end{aligned}$$

where

- β_0 is the intercept
- the other β's are the corresponding coefficient for the main effects and two-factor interactions
- Z represents temperature (the HTC factor)
- X_2 is additive, X_3 is rate, X_4 is time (the ETC factors)

and there are two error terms: one for the temperature effect (whole-plot error) and one for all the other terms (subplot error), see Figure 1. It is important to note that the whole-plot by subplot interactions are generally subplot terms, see Bisgaard (2000). An analysis of variance approach can be used to determine which effects are significant. Table 7 gives the split-plot analysis. Notice that the F-statistic and the p-value for the temperature term are identical to the analysis of the whole-plot means from Table 6. However, as pointed out earlier, the sums of squares are eight times larger in Table 7 because of not averaging the eight subplots in each whole-plot.

Split-plot experiments sacrifice information and precision on the whole-plot factors to gain precision on the subplot factors and the whole-plot by subplot interactions. This approach typically leads to the whole-plot error being much larger than the subplot error, which is the case for most split-plot experiments, but it is not always true. We can see from Table 7 that the whole-plot error for our example is almost six times the size of the subplot error. Using a significance level of 0.05, the following terms are significant: additive, time, rate × time interaction, temperature × rate interaction, and temperature × time interaction.

Many times in practice the split-plot nature of the design is overlooked, and the data are analyzed as if they came from a CRD, see Potcner and Kowalski (2004). Because there is only one error term for testing all of the effects in a completely randomized analysis, it should be intuitive that this error term lies somewhere between the whole-plot error and the subplot

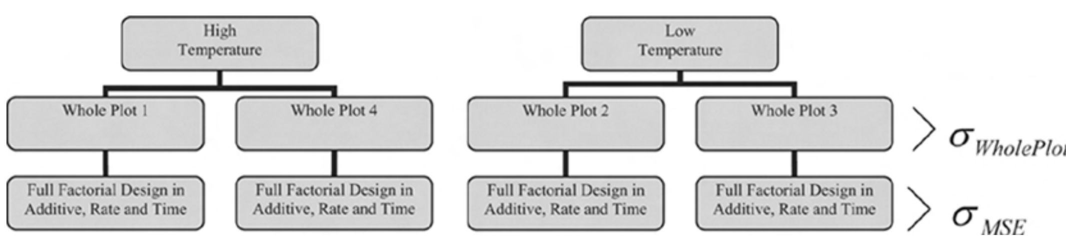

FIGURE 1 Understanding the two errors in the example.

TABLE 7 Split-plot Analysis for Plastic Experiment using MINITAB

Source	df	SS	MS	F	p
Temp	1	85.478	85.478	1.52	0.343
WP Error	2	112.391	56.195	*	*
Add	1	45.363	45.363	4.64	0.044
Rate	1	41.178	41.178	4.21	0.054
Time	1	75.953	75.953	7.76	0.012
Add*Rate	1	27.938	27.938	2.86	0.107
Add*Time	1	2.940	2.940	0.30	0.590
Rate*Time	1	43.945	43.945	4.49	0.047
Temp*Add	1	1.088	1.088	0.11	0.742
Temp*Rate	1	78.438	78.438	8.02	0.011
Temp* Time	1	62.440	62.440	6.38	0.021
SP Error	19	185.858	9.782		
Total	31	763.010			

*Not necessary to test.

error. In addition, the completely randomized error term tends to be closer to the subplot error because there are typically more degrees of freedom at the subplot level. Using this one error term can lead to overstating significance of the whole-plot effects because the actual error term used is smaller than the correct whole-plot error term. Also, using this one error term can lead to understating significance of the subplot effects (including the whole-plot $\times$ subplot interactions) because the actual error term used is larger than the correct subplot error term. It is easy to calculate the completely randomized error term from the split-plot analysis. It is simply

$$\frac{\text{WP error SS} + \text{SP error SS}}{\text{WP error df} + \text{SP error df}}$$

Note, this is the incorrect completely randomized error term from carrying out the experiment as a split-plot. It is **not** the error term that would have resulted from running the experiment as a CRD (randomly running all 16 combinations twice).

For the plastic experiment, the incorrect completely randomized error term is

$$\frac{112.391 + 185.858}{2 + 19} = 14.203.$$

Using this error term for all effects and a significance level of 0.05, the following terms are significant: temperature, time, temperature $\times$ rate interaction, and temperature $\times$ time interaction. Notice that the temperature main effect is claimed to be significant and the main effect for additive and the rate $\times$ time interaction are thought to be insignificant.

The correct analysis of a two-level split-plot design is fairly straightforward with available software. The Appendix gives the SAS and MINITAB code for generating the correct split-plot analysis for the plastic example shown in Table 7. Some extensions to the basic first-order split-plot design and analysis are:

- more than one HTC factor
- having more than one level of split-plotting, typically referred to as a split-split-plot design
- using a fractional factorial in each whole-plot
- no replication of the hard-to-change factor
- response surface designs with HTC factors.

We began this article with some basic experimental design philosophy. Then, we introduced an important extension called the split-plot design, which is very common in industry because of the time and/or cost constraints on certain factors. A thorough presentation on the design and analysis of the basic first-order split-plot case has been given. These first four sections are meant to serve as a tutorial for the everyday practitioner. The next two sections are more advanced. They focus on the last bullet above and are presented for those practitioners who need to optimize a process that involves HTC factors.

5. AN EQUIVALENCE RESULT

Vining et al. (2005) (VKM) study the conditions under which split-plot experimental designs produce OLS estimates of the model that are equivalent to the GLS estimates. It is easy to produce experimental designs for the first-order model and for the first-order model plus interactions. However, it is more difficult to construct experimental designs that achieve this equivalence for the second-order model. Parker, Kowalski, and Vining (2006a, b) provide a catalog of balanced designs (the same number of subplots in each whole-plot) that achieve this equivalence.

In some sense, the OLS-GLS equivalence can be thought of as an optimality criteria. Equivalent estimation designs possess many important attributes:

- Model parameter estimates are independent of the variance components. This is an attractive feature because the usual estimation technique involves GLS and requires estimation of the variance components.
- Model parameter estimates are BLUE (best linear unbiased estimate).
- Equivalent estimation designs can be constructed such that the equivalence property holds for any projected model consisting of a subset of terms from a complete second-order model. In this way, prior assumptions of a model do not influence the choice or performance of the design; only that the design supports the model.
- Model estimation is straightforward using OLS, which can be easily performed with software capable of matrix manipulations and any statistical software package. Furthermore, OLS estimation is approachable to practitioners with modest statistical training. For a discussion of the inferential advantages of equivalent estimation designs see Vining and Kowalski (2006).
- The designs can be augmented by replication, which enables pure-error estimates of the variance components. From pure-error estimates, we can construct a model-independent estimate of the variance-covariance matrix of the parameter estimates.

These features make OLS-GLS equivalent designs attractive to practitioners.

6. SECOND-ORDER SPLIT-PLOT EXAMPLE

Frequently scientific research employs second-order response surface designs to reveal factor levels that improve system performance. As an example, consider an aerodynamic configuration study to identify the best airfoil configuration that maximizes the coefficient of lift (response) during take-off and landing of a passenger aircraft, as discussed by Payne et al. (2000). A wind tunnel experiment using a scaled aircraft model allows two parameters of the outboard control surface on the trailing edge of the wing, known as the flap, to be varied in its offset gap (z_1) and deflection angle (z_2), see Figure 2. Changing the levels of either configuration variable, z_1 or z_2, requires a costly and time consuming personnel access into an environmentally controlled test section to make the necessary configuration adjustments. Therefore, these two factors present a practical restriction on randomization. Furthermore, we desire a design strategy that minimizes the number of levels and settings of these HTC factors in order to achieve high experimental efficiency from both a cost and statistical perspective.

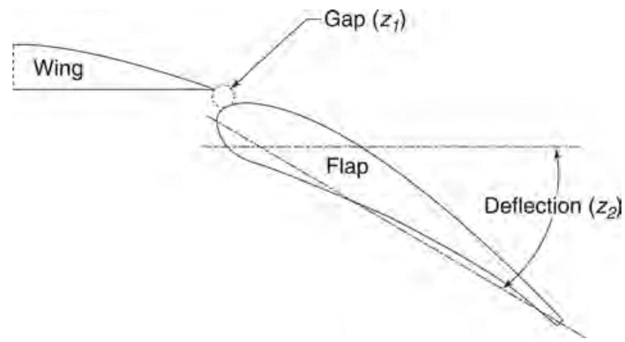

FIGURE 2 **Diagram of main wing element denoting the flap deflection and gap for the aerodynamic configuration experiment.**

In addition to these HTC factors, the effect of model attitude relative to the airflow, known as the angle-of-attack (x_1), and Reynolds number (x_2), a flow field parameter enabling full-scale flight correlation, are studied. Specific levels of the Reynolds number are achieved by setting the temperature and pressure of test environment. The setting of the angle-of-attack and Reynolds number can be performed remotely from outside of the test section; making them relatively ETC factors. Therefore, we require a second-order split-plot design for two HTC and two ETC factors.

Payne et al. (2000) noted the presence of HTC factors in this experimental investigation, however they did not employ a split-plot design. In this section, we explore the execution of this experiment as a second-order split-plot design. Since the original researchers did not conduct this experiment as a split-plot design, we have simulated the data based on the original results. The design factors are provided in coded units and the response has been multiplied a factor of 10,000 to covert the lift coefficient to counts. In addition, the intercept has been set equal to zero in the simulation, thereby expressing the response as the relative change in the coefficient of lift.

To build the design, we follow the construction strategy proposed by Vining, Kowalski, and Montgomery (2005) and begin with a popular three-level spherical design in four factors, known

TABLE 8 Data for Aerodynamic Configuration Example Based on a Split-plot Box-Behnken Design for D(2, 2) (z_1 and z_2 are HTC Factors, x_1 and x_2 are ETC Factors)

Whole-plot	z_1	z_2	x_1	x_2	y	Whole-plot	z_1	z_2	x_1	x_2	y
1	−1	−1	0	0	−779.08	7	0	−1	−1	0	−2531.95
1	−1	−1	0	0	−787.94	7	0	−1	1	0	2281.86
1	−1	−1	0	0	−777.95	7	0	−1	0	−1	−440.14
1	−1	−1	0	0	−798.51	7	0	−1	0	1	83.73
2	−1	1	0	0	472.72	8	0	1	−1	0	−2433.91
2	−1	1	0	0	486.24	8	0	1	1	0	2393.37
2	−1	1	0	0	470.04	8	0	1	0	−1	−379.46
2	−1	1	0	0	471.14	8	0	1	0	1	210.95
3	1	−1	0	0	−157.34	9	0	0	−1	−1	−2680.81
3	1	−1	0	0	−152.63	9	0	0	−1	1	−2558.52
3	1	−1	0	0	−145.39	9	0	0	1	−1	1683.26
3	1	−1	0	0	−161.69	9	0	0	1	1	2669.83
4	1	1	0	0	−1246.41	10	0	0	0	0	15.28
4	1	1	0	0	−1250.66	10	0	0	0	0	12.98
4	1	1	0	0	−1255.47	10	0	0	0	0	7.01
4	1	1	0	0	−1262.32	10	0	0	0	0	−1.22
5	−1	0	−1	0	−2653.34	11	0	0	0	0	1.70
5	−1	0	1	0	2208.63	11	0	0	0	0	−6.66
5	−1	0	0	−1	−667.63	11	0	0	0	0	−17.70
5	−1	0	0	1	94.61	11	0	0	0	0	−7.04
6	1	0	−1	0	−3167.77	12	0	0	0	0	−14.34
6	1	0	1	0	1571.00	12	0	0	0	0	−21.35
6	1	0	0	−1	−1010.47	12	0	0	0	0	−17.58
6	1	0	0	1	−689.23	12	0	0	0	0	2.19

as a Box-Behnken (BBD; Box and Behnken, 1960). Rearranging the completely randomized version of the design, and replicating subplot center runs to balance the whole-plot size, results in the split-plot design shown in Table 8. This design features $m = 12$ whole plots with $n = 4$ subplot runs per whole-plot. We have augmented the base design with $m_r = 3$ replicates of the whole-plot containing all centers points (whole plots 10, 11, and 12) to enable a pure-error estimate of the whole-plot variance.

Note that the design is shown in standard order. In practice, the whole plots and the subplot runs within each whole-plot would be executed in a randomized order and require the resetting of the factor levels between every experimental run. Failing to reset the factors between consecutive runs will influence the estimated variance components and could lead to inappropriate inference, as previously illustrated in the first-order case.

To visualize a split-plot design structure, we find it useful to use a nested diagram; plotting the HTC and ETC factor combinations separately. Figure 3 provides a graphical representation of the design in Table 8 and illustrates the distribution of the design points within the whole-plot and subplot design spaces. Note that there are two whole-plot designs superimposed at the center of the design space representing the structure of whole-plot number 9 and 10.

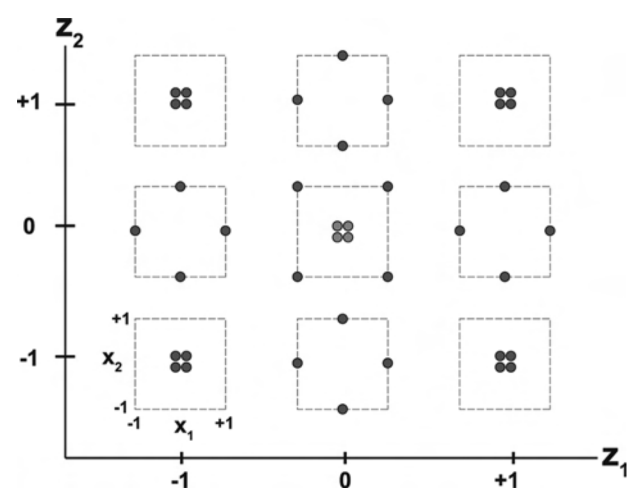

FIGURE 3 D(2,2) Split-plot Box-Behnken design for the aerodynamic configuration experiment (z_1 and z_2 are hard-to-change and x_1 and x_2 are easy-to-change factors).

We can numerically verify that this design possesses the property of equivalent estimation by testing the following condition defined by Parker, Kowalski, and Vining (2006),

$$\mathbf{XK} = \mathbf{JX},$$

where $\mathbf{X}$ is the $N \times p$ model matrix for a complete second-order model including the intercept, N is the total number of runs, p is the number of terms in the model, $\mathbf{K} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{JX}$, and $\mathbf{J}_{N\times N} = \mathbf{I}_m \otimes \mathbf{J}_n$. The block diagonal matrix $\mathbf{J}$ for the design in Table 8, where $N = 48$, is given by

$$\mathbf{J} = \begin{bmatrix} \mathbf{1}_4\mathbf{1}_4' & 0 & \cdots & 0 \\ 0 & \mathbf{1}_4\mathbf{1}_4' & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{1}_4\mathbf{1}_4' \end{bmatrix}$$

For this particular design, the $\mathbf{X}$ matrix is

$$\mathbf{X} = [1\, z_1\, z_2\, z_1z_2\, z_1^2\, z_2^2\, x_1\, x_2\, x_1\, x_2\, z_1x_1\, z_1x_2\, z_2x_1\, z_2x_2\, x_1^2\, x_2^2]$$

where the elements in $\mathbf{X}$ are vectors of length N representing each model term.

A feature of this design is that it supports pure-error based estimates of the variance components. First, we obtain a pooled estimate of the subplot error variance as

$$S_{SP}^2 = \frac{\sum_{i=1}^{m_s} S_i^2}{m_s},$$

where S_i^2 is the sample variance for the i^{th} whole-plot containing subplot replicates, and m_s is the number of whole plots of this type. The $\mathrm{E}(S_{SP}^2) = \sigma_\varepsilon^2$, the subplot variance. Therefore, $\hat{\sigma}_\varepsilon^2 = S_{SP}^2$ is an unbiased estimate of σ_ε^2 with $m_s(n-1)$ degrees of freedom. Note that in general the degrees of freedom for this estimate is equal to the sum of the degrees of freedom for each sample variance, S_i^2.

Now for the whole-plot variance estimate, we let $\bar{\boldsymbol{Y}}_{i.}$ be the sample mean of the responses in the ith whole-plot replicate, and $\bar{\boldsymbol{Y}}_{..}$ the overall mean of the responses in the WP replicates and define S_{WP}^2 as

$$S_{WP}^2 = \frac{\sum_{i=1}^{m_r} (\bar{Y}_{i.} - \bar{Y}_{..}^2)}{m_r - 1}.$$

The $\mathrm{E}(S_W^2) = \sigma_\delta^2 + \frac{1}{n}\sigma^2$, where σ_δ^2 is the whole-plot error variance. Therefore, using the method of moments we obtain an unbiased estimate of σ_δ^2 as

$$\hat{\sigma}_\delta^2 = S_{WP}^2 - \frac{1}{n}\hat{\sigma}_\epsilon^2,$$

with $m_r - 1$ degrees of freedom. If of $\hat{\sigma}_\delta^2 < 0$, then we recommend biasing the estimate by setting it equal to zero.

Now applying this to our example data set, we compute the subplot error term as follows.

Whole-plot	S_{ij}^2
1	**91.0097**
2	**56.9785**
3	**48.6740**
4	**46.6284**
10	**54.2897**
11	**63.1898**
12	**107.6388**
Total	**468.4090**

Because the design is balanced, the pooled estimate of σ^2 reduces to the unweighted average of the subplot variance estimates as

$$\hat{\sigma}^2 = S_{sp}^2 = \frac{468.4090}{7} = 66.9156.$$

Next, we compute the whole-plot error term.

Whole-plot	$\bar{y}_i$
10	**8.5128**
11	**−7.4259**
12	**−12.7698**
S_{wp}^2	**122.5913**

Therefore, our estimate of the whole-plot error variance $\hat{\sigma}_\delta^2$ is

$$\begin{aligned}\hat{\sigma}_\delta^2 &= S_{wp}^2 - \frac{1}{n}\hat{\sigma}^2 \\ &= 122.5913 - \frac{1}{4}(66.9156) \\ &= 105.8624.\end{aligned}$$

These pure-error estimates of the variance components are used to form the estimated variance-covariance matrix of the observations, $\widehat{\boldsymbol{\Sigma}}$ as

$$\widehat{\boldsymbol{\Sigma}} = \hat{\sigma}^2\mathbf{I} + \hat{\sigma}_\delta^2\mathbf{J},$$

where $\mathbf{I}$ is an $N \times N$ identity matrix. To estimate the model coefficients, we utilize the OLS estimates, which are equivalent to GLS for this design, as

$$\hat{\beta}_{OLS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{y}.$$

TABLE 9 Coefficients Using OLS and Standard Errors Using Pure Error Estimates of the Variance Components for the Aerodynamic Example Using SAS Proc Mixed

Term	Estimated coefficient	Standard error	df error	t-statistic	p	Comment
Intercept	−3.8943	6.3925	2	−0.61	0.6044	
z_1	−277.8047	4.5202	2	−61.46	0.0003	
z_2	43.4690	4.5202	2	9.62	0.0106	
z_1z_2	−590.0894	5.5360	2	−106.59	0.0001	
z_1^2	−431.0115	6.9047	2	−62.42	0.0003	
z_2^2	6.3182	6.9047	2	0.92	0.4568	
x_1	2402.8548	2.3614	21	1017.55	0.0000	
x_2	275.5522	2.3614	21	116.69	0.0000	
x_1x_2	216.0703	4.0901	21	52.83	0.0000	
z_1x_1	−30.8003	4.0901	21	−7.53	0.0000	
z_1x_2	−110.2498	4.0901	21	−26.96	0.0000	
z_2x_1	3.3666	4.0901	21	0.82	0.4197	
z_2x_2	16.6373	4.0901	21	4.07	0.0006	
x_1^2	−77.5064	5.6058	2	−13.83	0.0052	approx.
x_2^2	−135.6970	5.6058	2	−24.21	0.0017	approx.

The variance-covariance matrix for the OLS estimatesis

$$\widehat{\mathrm{Var}(\hat{\beta}_{OLS})} = (\mathbf{X}'\widehat{\Sigma}^{-1}\mathbf{X})^{-1},$$

which is the same as for GLS. The estimated standard errors for the coefficient estimates are the square root of the diagonal elements of $(\mathbf{X}'\widehat{\Sigma}^{-1}\mathbf{X})^{-1}$.

Table 9 summarizes the resulting coefficient estimates, standard errors, t-statistics, and p-values using OLS and estimated errors based on the pure-error estimates of the variance components (see the Appendix for the SAS code). The test for the subplot pure quadratics terms is approximate, as noted. Vining and Kowalski (2006) provide a thorough treatment on inference; providing a residual based analysis method for OLS-GLS designs. In addition, they establish design conditions for which exact tests are possible and show how to construct approximate tests when exact tests are not possible. While the pure-error estimates of the variance components are model independent, a residual based estimate increases the inferential power for the whole-plot terms by adding degrees of freedom.

7. CONCLUSIONS AND SUGGESTED READING

Many of the real world industrial experiments involve factors that are HTC. In these situations, experimenters realize that the most efficient way to run an experiment is by fixing a level of the hard-to-change factor and then running all or some of the combinations of the easy-to-change factors. This is then repeated a few times. As we have seen, this leads to a split-plot design. It is equally important to account for the split-plot nature of the design in the analysis of the data because the split-plot experiment has two error terms.

We began with the first-order split-plot case and introduced the basic analysis. The basic structure of this experiment produces two different size experimental units. This paper illustrated that understanding the nested structure of the error term for the HTC factor(s) leads to the correct split-plot analysis, which can be done in most statistical software packages.

We have demonstrated how the typical complex analysis requirements of split-plot designs can be greatly simplified by the use of an OLS-GLS equivalent design. In practice the conditions to build equivalent estimation designs are easily achieved. By finding the pure-error estimates and forming the appropriate matrices, all of the calculations in this example can be performed in any software package capable of matrix manipulations, including common spreadsheet applications. It is well-known that restrictions on randomization can be accommodated with a split-plot structure, however, these restrictions may not be properly accounted for in the analysis. We believe that the simplified approach to design and analysis of split-plot designs presented in this article will facilitate broader application among practitioners.

Montgomery (2004) is a good initial source to learn more about the basics on first-order split-plot designs. A very nice article to learn more about first-order industrial split-plot experiments is Bisgaard (2000). He gives a comprehensive discussion of running fractional factorials as split-plots. He uses the formal alias structure rather than minimum aberration as the basis for selecting designs. Although this approach requires more work on the part of the experimenter, it provides the most insight for the practitioner. Bisgaard's article represents an important development for the industrial use of fractional factorial experiments in a split-plot structure.

Cornell (1988) looks at running mixture component with process factor experiments within a split-plot structure. Essentially, he treats the process factors as "noise" in the sense of robust parameter design. He suggests crossing either a full factorial or a fractional factorial design in the process factors with standard mixture designs for the mixture components. This article shows in great detail how to perform the analysis of such experiments. In addition, Kowalski et al. (2002) discuss a pure-error approach for mixture component-process variable experiments run as split-plot designs.

Parker et al. (2006a,b) give a nice catalog of second-order split-plot designs. They propose balanced designs that possess the OLS-GLS equivalence. They also outline designs based on the central composite design that use a minimum number of whole plots. In addition, they are some D-optimal designs that achieve the OLS-GLS equivalence. Finally, Vining and Kowalski (2006) discuss in detail the testing issues for second-order split-plot designs. They also propose appropriate residual plots for evaluating the assumptions underlying the analysis.

ABOUT THE AUTHORS

Scott Kowalski is a Technical Training Specialist at Minitab Inc. He received his Ph.D. in Statistics from the University of Florida. He is an Associate Editor for *Quality Technology and Quantitative Management* and on the Editorial Board for *Quality Engineering* and the *Journal of Quality Technology*. He is a senior member of ASQ.

Peter Parker is a Research Scientist at the National Aeronautics and Space Administration's Langley Research Center. He received his Ph.D. in Statistics from Virginia Tech. He is on the Editorial Board for *Quality Engineering* and is a member of ASQ.

Geoff Vining is Professor of Statistics at Virginia Tech. He is on the Editorial Board of the *Journal of Quality Technology*. He is a Fellow of the ASQ.

REFERENCES

Bingham, D., Sitter, R. S. (1999). Minimum-Aberration two-level fractional factorial split-plot designs. *Technometrics*, 41:62–70.

Bingham, D., Sitter, R. S. (2001). Design issues in fractional factorial split-plot experiments. *Journal of Quality Technology*, 33:2–15.

Bisgaard, S. (2000). The design and analysis of $2^{k-p} \times 2^{q-r}$ split-plot experiments. *Journal of Quality Technology*, 32:39–56.

Bisgaard, S., Kulahci, M. (2001). Robust product design: Saving trials with split-plot confounding. *Quality Engineering*, 13:525–530.

Box, G. E. P., Behnken, D. W. (1960). Some new three-level designs for the study of quantitative variables. *Technometrics*, 2:455–475.

Box, G. E. P., Jones, S. (1992). Split-Plot designs for robust product experimentation. *Journal of Applied Statistics*, 19:3–26.

Box, G. E. P., Jones, S. (2001). Split-Plots for robust product and process experimentation. *Quality Engineering*, 13:127–134.

Cornell, J. A. (1988). Analyzing data from mixture experiments containing process variables: A split-plot approach. *Journal of Quality Technology*, 20:2–23.

Huang, P., Chen, D., Voelkel, J. O. (1998). Minimum-Aberration two-level split-plot designs. *Technometrics*, 40:314–326.

Kowalski, S. M. (2002). 24 run split-plot experiments for robust parameter design. *Journal of Quality Technology*, 34:399–410.

Kowalski, S. M., Cornell, J. A., Vining, G. G. (2002). Split-Plot designs and estimation methods for mixture experiments with process variables. *Technometrics*, 44:72–79.

Kowalski, S. M., Potcner, K. J. (2003). How to recognize a split-plot experiment. *Quality Progress*, 36(11):60–66.

Montgomery, D. C. (2004). *Design and Analysis of Experiments*, 6th ed. New York: John Wiley and Sons.

Parker, P. A., Kowalski, S. M., Vining, G. G. (2006a). Classes of split-plot response surface designs for equivalent estimation. *Quality and Reliability Engineering International*, 22:291–305.

Parker, P. A., Kowalski, S. M., Vining, G. G. (2006b). Construction of balanced equivalent estimation second-order split-plot designs. Submitted for publication.

Payne, F. M., Wyatt, G. W., Bogue, D. R., Stoner, R. C. (2000). High Reynolds number studies of a Boeing 777-200 high lift configuration in the NASA ARC 12 foot pressure tunnel and NASA LaRC National Transonic Facility. *American Institute of Aeronautics and Astronautics 18th Applied Aerodynamics Conference*, AIAA 2000-4220, Denver, CO.

Potcner, K. J., Kowalski, S. M. (2004). How to analyze a split-plot experiment. *Quality Progress*, 37(11):67–74.

Simpson, J. R., Kowalski, S. M., Landman, D. (2004). Experimentation with randomization restrictions: Targeting practical implementation. *Quality and Reliability Engineering International*, 20(5):481–495.

Vining, G. G., Kowalski, S. M. (2006). Exact inference for response surface designs within a split-plot structure. Submitted for publication.

Vining, G. G., Kowalski, S. M., Montgomery, D. C. (2005). Response surface designs within a split-plot structure. *Journal of Quality Technology*, 37:115–129.

8. APPENDIX: COMPUTER CODE FOR ANALYZING SPLIT-PLOT EXPERIMENTS

Analyzing split-plot designs can be done using many different statistical software packages. We have used two packages: SAS and MINITAB. The SAS code for the first-order case with the plastic example is given below.

```
data spd;
input WP Temp Add Rate Time Strength;
datalines;
Enter the data here.
proc glm;
  class WP Temp Add Rate Time;
  model strength = Temp WP(Temp) Add Rate
  Time Add*Rate Add*Time Rate Time
           Temp*Add Temp*Rate Temp*Time;
  test h = Temp e = WP(Temp);
run;
```

Notice that the nested term necessary for the whole-plot error is entered manually. In addition, it is necessary to specify the nested term as the error term for testing temperature. The code is very similar for SAS Proc Mixed, see below. The difference is that it is not necessary to specify the error term for the temperature factor, because SAS Proc Mixed recognizes the appropriate whole-plot error term.

```
data spd;
input WP Temp Add Rate Time Strength;
datalines;
Enter the data here.
proc mixed;
  class WP;
  model y = Temp Add Rate Time Add*Rate Add*
            Time Rate*Time
           Temp*Add  Temp*Rate  Temp*Time/
           solution ddfm =
           satterth;
  random WP;
run;
```

Using MINITAB is sort of a combination of the two SAS procedures above. In MINITAB, select Stat, ANOVA, General Linear Model, then enter Strength as the response and the same model as Proc GLM above: Temp WP(Temp) Add Rate Time Add*Rate Add*Time Rate*Time Temp*Add Temp*Rate Temp*-Time. Finally, enter WP as a Random Factor similar to Proc Mixed above and hit OK, see the screen shot in Figure 4.

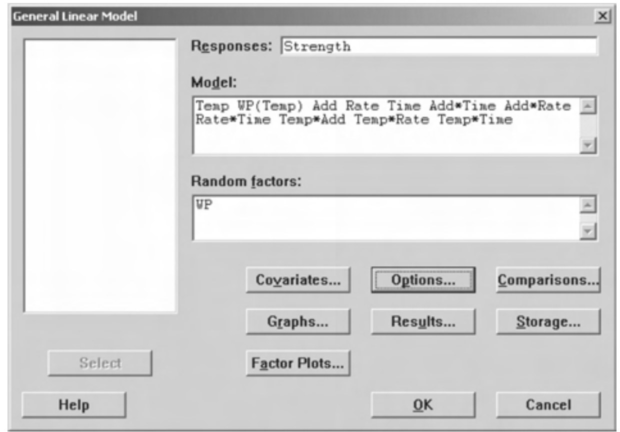

FIGURE 4 MINITAB screen shot for first-order split-plot plastic example.

Now consider the second-order Box-Behnken design for aerodynamic configuration experiment. The necessary SAS code is given below.

```
data spd;
input unique Rep z1 z2 x1 x2 y;
  z12 = z1*z2;
  z11 = z1*z1;
  z22 = z2*z2;
  x12 = x1*x2;
  z1x1 = z1*x1;
  z1x2 = z1*x2;
  z2x1 = z2*x1;
  z2x2 = z2*x2;
  x11 = x1*x1;
  x22 = x2*x2;
datalines;
Enter the data here.
proc mixed;
  class unique rep;
  model y = z1 z2 z12 z11 z22 x1 x2 x12 z1x1
  z1x2  z2x1  z2x2  x11  x22/solution
  ddfm = satterth;
  random Rep(unique);
run;
```

The unique variable enumerates the unique settings of the whole-plot, 10 in our example. The Rep variable enumerates the number of replicates of each unique setting; three replicates of whole-plot 10, 11, and 12 in our example.