QUALITY AND RELIABILITY ENGINEERING INTERNATIONAL
Qual. Reliab. Engng. Int. 2005; **21**:539–551 (DOI: 10.1002/qre.740)

Special Issue

Analysis of Split-plot Designs with Mirror Image Pairs as Sub-plots

John Tyssedal[1,*,†] and Murat Kulahci[2]

[1] *Department of Mathematical Sciences, The Norwegian University of Science and Technology, 7491 Trondheim, Norway*
[2] *Department of Industrial Engineering, Arizona State University, Tempe, AZ 85287-5906, U.S.A.*

In this article we present a procedure to analyze split-plot experiments with mirror image pairs as sub-plots when third- and higher-order interactions can be assumed negligible. Although performing a design in a split-plot manner induces correlation among observations, we show that with such designs the essential search for potentially active factors can be done in two steps using ordinary least squares. The suggested procedure is tested out on a real example and on two simulated screening examples; one with a split-plot design based on a geometric design and one with a split-plot design based on a non-geometric Plackett and Burman design. The examples also illustrate the advantage of using non-geometric designs where the effects are partially aliased instead of being fully aliased as in highly fractionated fractional factorials. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS: two-level designs; restricted randomization; alias structure; least squares; generalized least squares

1. INTRODUCTION

Split-plot experiments are common in industry when experiments are used for process and product optimization. The need for such experimental plans typically arises when some factors are harder to change, so that a complete randomization of the experimental runs is difficult or sometimes impossible to perform. Split-plotting may also occur when (1) two or more process steps are involved, (2) in robust product design experimentation and (3) it is of interest to estimate some factors with higher precision than others where the latter are then handled as sub-plot factors.

In split-plot experiments the factors that are easy to change (the *sub-plot* factors) are changed according to a design matrix called sub-plot design and the factors that are hard to change (the *whole-plot* factors) are changed less frequently according to a second design matrix called whole-plot design. For each whole-plot, the same or different sub-plot design may be employed. So for a randomly chosen level combination of the whole-plot factors, a design in the sub-plot factors is run in random order. If the same number of sub-plot runs is used for each whole-plot, the total number of runs will then be equal to the number of whole-plot level combinations times the number of runs in the design for the sub-plot factors.

When there are many factors, a two-level experimental plan will often help to keep the number of runs at an acceptable level. Examples of such split-plot designs can be found in Addelman[1], Huang *et al.*[2], Bingham and Sitter[3], Bisgaard[4] and Kulahci[5]. Tyssedal *et al.*[6] further explored two-level split-plot designs where for

*Correspondence to: John Tyssedal, Department of Mathematical Sciences, The Norwegian University of Science and Technology, 7491 Trondheim, Norway.
†E-mail: tyssedal@math.ntnu.no

Copyright © 2005 John Wiley & Sons, Ltd.

each level combination of the whole-plot factors, two runs as mirror image pairs are used at the sub-plot leve These designs are referred to as SPMIP designs and can be constructed using both fractional factorials and/ non-geometric designs such as Plackett and Burman (PB) designs (Plackett and Burman[7]). Tyssedal *et al.*[6] als showed how different columns from the PB designs can be used for whole-plot and sub-plot level combinatio to avoid full aliasing among interaction effects.

One of the most appealing properties of SPMIP designs is that they divide the estimated effects into tw orthogonal subspaces, separating sub-plot main effects and sub-plot by whole-plot interactions from the re In addition their projective properties can be quite attractive compared to their run sizes making them ve appealing for screening purposes (Tyssedal *et al.*[6]).

In this article we propose a procedure of analyzing SPMIP designs in situations when a small subset the factors investigated is of importance, i.e. a typical screening situation. The method proposed is a two-st procedure where active effects are identified and estimated using ordinary least squares (OLS).

We start this paper by presenting some designs that fit into the description of SPMIP designs. The rationale f the estimation method together with its algorithmic description is given in Section 3. In Section 4 the algorithm procedure is illustrated on three examples and its performance is evaluated by simulation. Concluding remar are given in Section 5.

2. SPLIT-PLOT DESIGNS WITH MIRROR IMAGE PAIRS AS SUB-PLOTS

SPMIP designs have a design matrix that can be written as

$$\begin{bmatrix} \mathbf{W} & \mathbf{S} \\ \mathbf{W} & -\mathbf{S} \end{bmatrix} \qquad (1)$$

where $\begin{bmatrix} \mathbf{W} \\ \mathbf{W} \end{bmatrix}$ contains the whole-plot factors and $\begin{bmatrix} \mathbf{S} \\ -\mathbf{S} \end{bmatrix}$ the sub-plot factors. An example of such a design can b found in Bisgaard *et al.*[8]. They presented a split-plot experiment where the surface characteristics of a securi paper are modified via plasma treatment to make it more susceptible to ink. Four factors: A, pressure; B, powe C, gas flow rate and D, type of gas (oxygen or $SiCl_4$) that are suspected to affect plasma creation in addition the paper type (factor E) are tested. The response is the 'wettability' of the paper measured as the contact ang between the paper and a water droplet placed on the paper right after the plasma treatment.

A caveat about this experiment is that the creation of plasma requires vacuum. For this, a sample of securi paper is placed in a reactor and vacuum is generated by pumping the air out. This process takes approximate half an hour. Thus, a completely randomized two-level design would require $2^5 = 32$ runs which will take mo than 16 hours to run. To save labor and time, it was decided to place two types of papers in the reactor at t same time. The design array and the responses are given in Table I. Note that with $\mathbf{W} = (A, B, C, D)$ a $\mathbf{S} = E$, this design lay out may be written in the format given in (1). We return to the analysis of this experime in Section 4.

It is often assumed that only a few factors out of many considered for experimentation are active. This particularly valid for screening experiments. For these cases the projective properties of the original design a very important. The concept of projectivity for two-level designs is introduced by Box and Tyssedal[9]. A $N \times$ design with N runs and k factors each at two levels is said to be of projectivity P if every subset of P facto out of possible k contains a complete 2^P factorial design, possibly with some points replicated. It is convenie to describe such designs as (N, k, P) screens. This concept is further extended to SPMIP designs by Tyssed *et al.*[6]. They pointed out that for geometric designs, in order to construct a $P = 3$ SPMIP design with the tot number of N runs, there can be at most $N/4$ factors at each of the whole-plot and sub-plot levels. One w to obtain such designs is to let $\mathbf{W}$ consist of main effects and odd-order interaction columns from a geometr design in $N/2$ runs and let $\mathbf{S}$ consist of even-order interactions and a column of $+$ or -1. For example a 3 run SPMIP design of projectivity $P = 3$ can be constructed by considering the 2^4 full factorial design in fo factors, say A, B, C and D, fully expanded with main effects and interaction columns. The largest possib $P = 3$ SPMIP design is the one obtained by allocating up to eight whole-plot factors and eight sub-plot facto

Copyright © 2005 John Wiley & Sons, Ltd.

Table I. Split plot design for the plasma experiment

				E	
A	B	C	D	−	+
−	−	−	−	48.6	57.0
+	−	−	−	41.2	38.2
−	+	−	−	55.8	62.9
+	+	−	−	53.5	51.3
−	−	+	−	37.6	43.5
+	−	+	−	47.2	44.8
−	+	+	−	47.2	54.6
+	+	+	−	48.7	44.4
−	−	−	+	5.0	18.1
+	−	−	+	56.8	56.2
−	+	−	+	25.6	33.0
+	+	−	+	41.8	37.8
−	−	+	+	13.3	23.7
+	−	+	+	47.5	43.2
−	+	+	+	11.3	23.9
+	+	+	+	49.5	48.2

to the columns of **W** and **S** defined as

$$\mathbf{W} = [A \quad B \quad C \quad D \quad ABC \quad ABD \quad ACD \quad BCD] \quad \text{and} \quad \mathbf{S} = [AB \quad AC \quad AD \quad BC \quad BD \quad CD \quad ABCD \quad I]$$

where I is a column of $+1$.

Tyssedal *et al.*[6] also showed how to construct SPMIP designs from non-geometric designs such as PB designs where the number of runs, n, is a multiple of 4 and not necessarily a power of 2. Their projective properties make PB designs very appealing alternatives to fractional factorials in screening experiments (Box and Bisgaard[10], Lin and Draper[11], Samset and Tyssedal[12]). Tyssedal *et al.*[6] showed that this is also the case when these designs are used in split-plot experiments.

An example of such a SPMIP design with five whole-plot factors and six sub-plot factors constructed from the 12-run PB design (PB12) is shown in Table II. Note that the first 11 rows constitute the 12-run PB design.

SPMIP designs constructed from non-geometric designs as described above can accommodate a total number of $N/2$ whole-plot and sub-plot factors if a column of $+1$ are added in [**S**]. Apart from that, there is no restriction on the number of whole-plot and sub-plot factors in order to have a final design of projectivity $P = 3$. Another advantage of these designs compared with those generated from the fractional factorials is that full aliasing among two-factor interactions may be avoided. For a more detailed description of these designs and their properties, see Tyssedal *et al.*[6].

3. ANALYSIS OF SPMIP DESIGNS

In the following w and s are used to denote whole-plot and sub-plot main effects and ww, ws and ss are used to denote whole-plot-by-whole-plot, whole-plot-by-sub-plot and sub-plot-by-sub-plot interactions, respectively.

Due to the way split-plot experiments are executed, one has to deal with two variance regimes. The form of the covariance matrix of the responses is given as

$$\mathbf{V} = \begin{bmatrix} \mathbf{C} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{C} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{C} \end{bmatrix} \tag{2}$$

Copyright © 2005 John Wiley & Sons, Ltd.

Table II. A design layout where **W** and **S** consists of different columns from a PB12 design

	Whole plot					Sub-plot					
Run	*A*	*B*	*C*	*D*	*E*	*P*	*Q*	*R*	*S*	*T*	*U*
1	1	−1	1	−1	−1	−1	1	1	1	−1	1
2	1	1	−1	1	−1	−1	−1	1	1	1	−1
3	−1	1	1	−1	1	−1	−1	−1	1	1	1
4	1	−1	1	1	−1	1	−1	−1	−1	1	1
5	1	1	−1	1	1	−1	1	−1	−1	−1	1
6	1	1	1	−1	1	1	−1	1	−1	−1	−1
7	−1	1	1	1	−1	1	1	−1	1	−1	−1
8	−1	−1	1	1	1	−1	1	1	−1	1	−1
9	−1	−1	−1	1	1	1	−1	1	1	−1	1
10	1	−1	−1	−1	1	1	1	−1	1	1	−1
11	−1	1	−1	−1	−1	1	1	1	−1	1	1
12	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1
13	1	−1	1	−1	−1	1	−1	−1	−1	1	−1
14	1	1	−1	1	−1	1	1	−1	−1	−1	1
15	−1	1	1	−1	1	1	1	1	−1	−1	−1
16	1	−1	1	1	−1	−1	1	1	1	−1	−1
17	1	1	−1	1	1	1	−1	1	1	1	−1
18	1	1	1	−1	1	−1	1	−1	1	1	1
19	−1	1	1	1	−1	−1	−1	1	−1	1	1
20	−1	−1	1	1	1	1	−1	−1	1	−1	1
21	−1	−1	−1	1	1	−1	1	−1	−1	1	−1
22	1	−1	−1	−1	1	−1	−1	1	−1	−1	1
23	−1	1	−1	−1	−1	−1	−1	−1	1	−1	−1
24	−1	−1	−1	−1	−1	1	1	1	1	1	1

where **C** corresponds to each whole-plot and is a symmetric matrix of the following form

$$\mathbf{C} = \begin{bmatrix} \sigma_w^2 + \sigma_s^2 & \sigma_w^2 & \cdots & \sigma_w^2 \\ \sigma_w^2 & \sigma_w^2 + \sigma_s^2 & \cdots & \sigma_w^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_w^2 & \sigma_w^2 & \cdots & \sigma_w^2 + \sigma_s^2 \end{bmatrix} \tag{3}$$

where σ_w^2 and σ_s^2 are the between whole-plots and within whole-plot variances respectively. This implies the us of the generalized least squares to estimate the effects. However, as shown in Kulahci[5], for split-plot design with orthogonal whole-plot and sub-plot arrays, generalized least squares (GLS) estimates are the same a OLS estimates. Now let $\mathbf{X} = [\mathbf{W}_*, \mathbf{S}_*]$, where $\mathbf{W}_*$ contains the columns for the w, ww and ss effects and **S** contains the columns for the s and ws effects to be estimated. The estimated effects are then given by 2l where $\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$. In Appendix A it is shown that for SPMIP designs, the covariance matrix for th corresponding OLS estimators for the coefficients, $\hat{\boldsymbol{\beta}}$, and thereby also the GLS estimators is given by

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}) = \sigma_s^2 \begin{bmatrix} (\mathbf{W}_*'\mathbf{W}_*)^{-1} & 0 \\ 0 & (\mathbf{S}_*'\mathbf{S}_*)^{-1} \end{bmatrix} + 2\sigma_w^2 \begin{bmatrix} (\mathbf{W}_*'\mathbf{W}_*)^{-1} & 0 \\ 0 & 0 \end{bmatrix} \tag{4}$$

which implies that the inference about w, ww and ss effects and s and ws effects can in fact be done separately

In order to get a better understanding of the properties that follow from the design structure let Y_i and Y_{i2} be the two sub-plot responses obtained at the whole-plot level combination i. Y_{i1} and Y_{i2} have common part that consist of a possible constant, w main effects, ww and ss interactions and a whole-plo error, ε_i^w. The part that differs consists of s main effects and ws interactions, which occur with opposit

Copyright © 2005 John Wiley & Sons, Ltd.

signs in the two sub-plot responses, and the sub-plot errors, ε_{i1}^{s} and ε_{i2}^{s}. Now define $Y_i^{W} = (Y_{i1} + Y_{i2})/2$ and $Y_i^{S} = (Y_{i1} - Y_{i2})/2$. Assuming the expected values of the errors are zero, $E(Y_i^{W})$ consists of a possible constant and the corresponding terms for w main effects, ww and ss interactions, while $E(Y_i^{S})$ consists of the corresponding terms for s main effects and ws interactions. The error part in Y_i^{W} equals $\varepsilon_i^{w} + (\varepsilon_{i1}^{s} + \varepsilon_{i2}^{s})/2$ and the error part in Y_i^{S} equals $(\varepsilon_{i1}^{s} - \varepsilon_{i2}^{s})/2$. Further Y_i^{W}, $i = 1, 2, \ldots, N/2$, are all independent with the same variance. The same applies to Y_i^{S}. This explains the rationale for a two-step algorithm where in each step standard OLS screening procedure can be used on both Y_i^{W} and Y_i^{S} in order to identify the active factors. The use of OLS enables a search procedure as proposed below to run reasonably fast and breaking the analysis into two parts reduces one large search into two smaller searches, thereby experiments with more runs and a larger number of factors can be handled with relative ease. The proposed algorithm goes as follows.

1. Calculate $Y_i^{S} = (Y_{i1} - Y_{i2})/2$, $i = 1, \ldots, N/2$. Identify the most important s main effects and ws interactions by using a search procedure based on OLS. The mean squared error of the final model estimates $(\frac{1}{2})\hat{\sigma}_s^2$.
2. Calculate $Y_i^{W} = (Y_{i1} + Y_{i2})/2$, $i = 1, \ldots, N/2$. Use a search procedure based on OLS to identify the most important w main effects, ww and ss interactions. The mean squared error of the final model estimates $\hat{\sigma}_w^2 + (\frac{1}{2})\hat{\sigma}_s^2$.
3. Evaluate the significance of the model parameters and if necessary re-estimate the parameters, $\mathbf{V}$ and $\mathrm{Cov}(\hat{\boldsymbol{\beta}})$.

(In case of ambiguities, redo steps 1 and 2.)

The degrees of freedom will be equal to $N/2 - n_2$ for s main effects and ws interactions and $N/2 - n_1$ for the w main effects, ww and ss interactions where n_2 is the total number of s main effects and ws interactions, and n_1 is the total number of w main effects, ww and ss interactions plus 1 if a constant term is included in the model. Various search procedures may be applied in steps 1 and 2. Both factor-based and effect-based search procedures have been proposed. Box and Meyer[13] showed how to perform a factor-based search where the goal is to identify a few active factors among many that are being considered. The analysis method considers each possible subset of factors as defining a different model, which includes main effects and all interactions among those factors. A projectivity P design ensures that all of the effects in a model with P factors can be estimated. Hence, the performance of such a search depends heavily on the projection properties of the design used. In the effect-based searches the goal is to identify significant effects. The use of OLS to identify active main effects and two-factor interactions in non-geometric designs was first proposed by Hamada and Wu[14]. Their method is based on the precepts of effect sparsity and weak heredity. In order for the method to work well, the number of two-factor interactions should be few and their sizes should be small. The weak heredity principle requires excluding an interaction to be in a model unless at least one of the main effects associated with the interaction is also included in the model. However, the heredity principle is not necessarily valid in every experimental situation (Box *et al.*[14], Box and Tyssedal[15]). For SPMIP designs, one can even have a final model that obeys the weak heredity principle without obeying it in each of the two algorithmic steps described earlier. Also in step 2 of the algorithm, even if the model has few terms, the numbers of factors in the model may easily become larger than the projectivity of the design. Based on these considerations and the number of runs in our examples we have found that it is appropriate to use a search procedure where in each step we evaluate all candidate models with up to three main effects and three two-factor interactions without imposing any heredity or size restrictions on two-factor interaction terms. The choice of the number of main effects and two-factor interactions in each step is guided by the projectivity of the designs used. Since in our examples we mainly use $P = 3$ designs, we in turn search through models with up to three main effects and three two-factor interactions. It should also be noted that the procedure proposed in this article includes a factor-based search for three factors if three-factor interactions can be assumed negligible. In screening situations it is usually assumed that only a small number of factors are responsible for most of the variation in the data. Since our final model can include six main effects and six two-factor interactions, our procedure should be able to handle a wide variety of screening cases. If in an experimental situation it could be afforded to increase the projectivity of the final design, a more complete

Copyright © 2005 John Wiley & Sons, Ltd.

Table III. Subplot analysis for Example 1

Model	Effects				MSE	AIC	BIC	CV
1	*E*	*AE*	*DE*		1.815	0.971	1.116	1.117
2	*E*	*AE*	*BE*	*DE*	1.906	1.145	1.338	1.271
3	*E*	*AE*	*CE*	*DE*	1.954	1.170	1.363	1.302
4	*E*	*AE*			2.286	1.077	1.173	1.306
5	*E*	*AE*	*BE*		2.406	1.253	1.398	1.481
6	*E*	*AE*	*CE*		2.450	1.271	1.416	1.508
7	*E*	*AE*	*BE*	*CE*	2.594	1.453	1.646	1.729
8	*E*				20.699	3.155	3.203	11.039
9	*E*	*DE*			21.577	3.322	3.418	12.330
10	*E*	*BE*			22.126	3.347	3.443	12.643
11	*E*	*CE*			22.166	3.349	3.445	12.666
12	*E*	*BE*	*DE*		23.181	3.518	3.663	14.265
13	*E*	*CE*	*DE*		23.225	3.520	3.665	14.292
14	*E*	*BE*	*CE*		23.816	3.545	3.690	14.656
15	*E*	*BE*	*CE*	*DE*	25.100	3.723	3.916	16.734

factor-based search, including interactions of higher order than two, would be appealing. For other OLS-base search algorithms, we refer to Samset[17].

When a geometric SPMIP design is used, $N - 1$ different contrasts can be estimated. These contrasts can b divided into two sets according to which error regime they belong to. If there are sufficiently many in eac regime, half-normal or normal plots can be used on each of the sets of contrasts separately to judge thei influence on the response (Daniel[18]). More objective methods are given in Lenth[19] and Langsrud *et al.*[20] However, in fractionated designs there is a lot of uncertainty associated with the process of allocating th influential contrasts to the individual effects. Another option is to search through all models of a certain typ as proposed in this article. The candidate models can be chosen based on some simple criterion such as mea squared errors. More elaborate criteria that take into account the number of terms used in the model such a Akaike information criterion (AIC), Bayesian information criterion (BIC) or the 'leave one out' cross-validatio (CV) criterion can also be employed. It should be noted that the final outcome of this procedure is a list o candidate models most likely containing the 'right' model. In the following section, we illustrate the use o the proposed procedure for the analysis of SPMIP experiments. We explore the possibility of using fractiona factorial designs in such experiments as well as PB designs.

4. EVALUATION OF THE PROCEDURE

Example 1

Let us return to the design given in Table I. To analyze this experiment, Bisgaard *et al.*[8] used two separat normal plots; one for the whole-plot effects, i.e. the effects that are affected by the whole-plot and sub-plo error variances and another one for the sub-plots effects, i.e. the effects that are affected by only the sub-plo error variance. They concluded that A, D, AD, E and AE are the significant effects and rough estimates for th whole-plot and sub-plot standard deviations are 7 and 1, respectively.

We now apply the procedure discussed in Section 3 to this experiment. In step 1, we perform the sub-plo analysis. The results are given in Table III.

The 'best' model includes the main effect of E, and the AE and DE interactions. Although this conclude step 1, we strongly recommend not to pick the 'best' model as the direct interpretation of the results but t study the 'nearly best' models in order to check whether they reveal any additional information about the data A quick look at the decision criteria reveals that there is a large jump in the magnitude of each of these criteri going from models including AE to those without it. Moreover, there are reasons to argue that the DE interactio

Copyright © 2005 John Wiley & Sons, Ltd.

Table IV. Whole-plot analysis for Example 1

Model	Effects					MSE	AIC	BIC	CV
1	*A*	*D*		*AD*		44.095	4.161	4.306	63.692
2	*A*	*B*	*D*	*AD*		41.819	4.296	4.513	66.358
3	*A*	*D*		*AB*	*AD*	41.854	4.109	4.254	66.413
4	*A*	*B*	*D*	*AB*	*AD*	39.168	4.230	4.448	68.936
5	*A*	*C*	*D*	*AD*		43.944	4.345	4.563	69.730
6	*A*	*D*		*AD*	*BD*	44.112	4.162	4.307	69.996
7	*A*	*D*		*AC*	*AD*	44.819	4.178	4.322	71.118
8	*A*	*B*	*D*	*AD*	*BD*	41.631	4.291	4.509	73.270
9	*A*	*D*		*AD*	*CD*	46.834	4.222	4.366	74.315
10	*A*	*C*	*D*	*AC*	*AD*	44.721	4.363	4.580	78.708
⋮									

should also be considered in the model. Note that MSE values given in Table III and in the subsequent tables for the sub-plot analysis are multiplied by 2 to directly provide the estimate of the sub-plot error variance.

In step 2, we consider the average response for each whole-plot and correspondingly include only the whole-plot main effects and their two-factor interactions. It should also be noted that since there is only one sub-plot factor, there are no ss interactions. The results are given in Table IV for the top 10 models ranked using CV. The 'best' model based on CV is also embedded in the best models based on other criteria. To keep the final model parsimonious, the model with A, D and AD is considered. At this point, we also have the estimates for σ_w^2 and σ_s^2 as $44.095 - 1.815/2 = 43.188$ and 1.815, respectively, which give us the following estimate of the matrix $\mathbf{C}$ given in (3)

$$\hat{\mathbf{C}} = \begin{bmatrix} 45.002 & 43.188 \\ 43.188 & 45.002 \end{bmatrix}$$

The final model with estimated standard errors given in parentheses under the coefficient estimates is

$$\hat{y} = \underset{}{40.98} + \underset{(1.66)}{5.91}\,A - \underset{(1.66)}{7.55}\,D + \underset{(1.66)}{1.57}\,E + \underset{(0.24)}{8.28}\,AD - \underset{(1.66)}{2.95}\,AE + \underset{(0.24)}{0.51}\,DE \qquad (5)$$

This suggests that all of the terms in the final model are indeed significant with the exception of the DE interaction. The results from our procedure are very similar to those from Bisgaard *et al.*[8]. It should be noted that this example is included mainly for illustration purposes since in this case there are enough degrees of freedom to estimate all main effects and two-factor interactions directly by means of OLS without first having to perform the screening.

Example 2

Let us consider a more complicated case where the same example is studied as if it were a screening experiment. For this, to the original design we add six dummy factors: two whole-plot (f and g) and four sub-plot factors (p, q, r and s) which are denoted with lower case letters. That is, instead of the $2^{5-1} \times 2^1$ design, we now have a $2^{6-2} \times 2^{5-4}$ design. To construct this design we use the approach given in Tyssedal *et al.*[6], which guarantees a $P = 3$ design. For the whole-plot array, we use a standard 2^{6-2} fractional factorial with $f = ABC$ and $g = ABD$ as generators. The generators for the subplot array are given as $p = -ABE$, $q = -ACE$, $r = -ADE$ and $s = -BCE$. The final design has the form given in Table V. Coincidentally this design is also on the list of the minimum aberration designs provided by Huang *et al.*[2], but unfortunately not all $P = 3$ designs are on this list.

Consider now the first step where we try to identify important s main effects and ws interactions. In our search, we allowed for up to three s main effects and three ws interactions in each model. Table VI shows the

Copyright © 2005 John Wiley & Sons, Ltd.

Table V. The alternative design layout for Example 2

Whole-plot						Sub-plot				
A	B	C	D	f	g	E	p	q	r	s
		2^{6-2}				-1				
						-1				
						⋮	AB	AC	AD	BC
						-1				
		2^{6-2}				$+1$				
						$+1$				
						⋮	$-AB$	$-AC$	$-AD$	$-BC$
						$+1$				

Table VI. Sub-plot analysis for Example 2 (screening case)

Model	Effects						MSE	AIC	BIC	CV
1	E	r	s	Ar	Dr	Ds	0.553	0.158	0.448	0.443
2	E	r	s	Ar	Dr	fr	0.553	0.158	0.448	0.443
3	E	r	s	Ar	Ds	fs	0.553	0.158	0.448	0.443
4	E	r	s	Ar	fr	fs	0.553	0.158	0.448	0.443
5	E	r	s	DE	Dr	Ds	0.553	0.158	0.448	0.443
6	E	r	s	DE	Dr	fr	0.553	0.158	0.448	0.443
⋮										
12	E	r	s	AE	DE	fr	0.553	0.158	0.448	0.443
13	E	s		AE	Ar	Ds	0.983	0.608	0.849	0.715
14	E	s		AE	Ar	fr	0.983	0.608	0.849	0.715
⋮										

top models ranked using CV. The values of the other criteria such as MSE, AIC and BIC are also provided One interesting feature in Table VI is that the top 12 best models give exactly the same mean squared erro AIC, BIC and CV values. Due to model aliasing these models give exactly the same fit and the design used i this case cannot distinguish between these models.

We can now move on to the second step. The analysis for this step is summarized in Table VII, which indicate that the models that contain the main effects of A, B and D with certain combinations of ww and ss interaction performed better than the rest. Due to aliasing, the top three models give exactly the same fit. Let us now, fo illustration purposes, represent all of the terms in the model in terms of the original factors. For example, qs i aliased with AB, and since both A and B main effects will be in the model, it may be better to use AB in the fina model instead of qs. The same argument can be made favoring the inclusion of DE instead of Ar. Consequentl we include the following effects in the final model: A, B, D, E, AB, AD, AE ($= -Dr$), DE, r, s, rs and Ds We also have the estimates for σ_s^2 and σ_w^2 as 0.553 and 24.04 ($= 24.316 - 0.553/2$), respectively, and

$$\hat{\mathbf{C}} = \begin{bmatrix} 24.59 & 24.04 \\ 24.04 & 24.59 \end{bmatrix}$$

The final model is then given as

$$\begin{aligned} \hat{y} = {} & \underset{(1.23)}{40.98} + \underset{(1.23)}{5.91}\,A + \underset{(1.23)}{2.11}\,B - \underset{(1.23)}{7.55}\,D + \underset{(0.13)}{1.57}\,E + \underset{(0.13)}{0.41}\,r - \underset{(0.13)}{0.45}\,s \\ & - \underset{(1.23)}{2.11}\,AB + \underset{(1.23)}{8.28}\,AD - \underset{(0.13)}{2.95}\,AE + \underset{(0.13)}{0.51}\,DE + \underset{(1.23)}{3.42}\,rs - \underset{(0.13)}{0.44}\,Ds \end{aligned} \tag{6}$$

Copyright © 2005 John Wiley & Sons, Ltd.

Table VII. Whole-plot analysis for Example 2 (screening case)

Model	Effects						MSE	AIC	BIC	CV
1	*A*	*B*	*D*	*AD*	*qs*	*rs*	24.316	3.754	3.971	48.031
2	*A*	*B*	*D*	*AB*	*Er*	*rs*	24.316	3.754	3.971	48.031
3	*A*	*B*	*D*	*AB*	*AD*	*rs*	24.316	3.754	3.971	48.031
4	*A*	*B*	*D*	*Er*	*rs*		28.558	3.914	4.132	50.262
5	*A*	*D*		*AD*	*qs*	*rs*	28.596	3.728	3.873	50.330
6	*A*	*D*		*AD*	*Df*	*Ep*	28.596	3.728	3.873	50.330
7	*A*	*D*		*AD*	*Df*	*qs*	28.596	3.728	3.873	50.330
8	*A*	*D*		*AD*	*Dg*	*rs*	28.596	3.728	3.873	50.330
9	*A*	*D*		*AB*	*Er*	*rs*	28.596	3.728	3.873	50.330
10	*A*	*D*		*Ag*	*Er*	*rs*	28.596	3.728	3.873	50.330
⋮										

From (6), we see that the main effect of B and the AB interaction are not significant. Apart from the effects that were significant in (5), we now have r, s, rs and Ds effects significant. However, a closer look reveals that rs is in fact aliased with $ABCD$ and Ds with $BCDE$. Indeed $ABCD$ and $BCDE$ interactions, although not significant in the original analysis, had the biggest effects after the significant ones. Thus, in this case we would have mistakenly included these extra effects in the model. This is, in fact, a very good example of the effects of heavy aliasing in fractional factorial split-plot designs. Similarly other simple techniques of analysis of these designs such as half-normal and normal plots will most likely fail to produce the right answer. One way to work around this problem is to use PB designs instead of fractional factorials. Even though the alias pattern of PB designs is considered to be complex, this represents only minor problems if only a few factors are active. Half-normal plots and normal-plots may be heavily affected and there may be some loss in efficiency when PB designs are used for estimating models, since interaction columns may not be orthogonal to main effect columns. However, if only a few factors are active this loss in efficiency is normally small. The fact that the aliasing in these designs is partial rather than full makes it a lot easier to distinguish between models. Hence when a search procedure is used to evaluate all models of a certain type we are less likely to end up in the same ambiguity as in Example 2. As shown in Kulahci and Bisgaard[21] and Tyssedal *et al.*[6], the applications of PB designs in split-plot experiments also offer more choices than the competing fractional factorials. We now consider a simulated example where we use a PB design in a split-plot situation as described in Tyssedal *et al.*[6].

Example 3

In Table II we have a SPMIP design with 11 factors in 24 runs constructed from the 12-run PB design. To construct this design, the runs of the first five columns from a 12-run PB design are replicated to represent the whole-plot factors (A through E) and the last six columns are folded over to represent the sub-plot factors (P through U).

The following model is used to generate the response values

$$y_{ij} = 5 + 2A + 1.5C + 2P + 2AD + 1.5AP + 1.5PR + \varepsilon_i + \varepsilon_{ij} \tag{7}$$

where y_{ij} is the response value for the ith whole-plot and jth sub-plot, ε_i is the whole-plot error and ε_{ij} is the sub-plot error. In Tables VIII and IX, as an example we summarize the results from one simulation run for which we used $\varepsilon_i \sim N(0, 0.3)$ and $\varepsilon_{ij} \sim N(0, 0.3)$. We then consider the top model from each stage and form the final model. The analysis of that model gives

$$\hat{y} = \underset{(0.14)}{4.74} + \underset{(0.19)}{1.68}\,A + \underset{(0.21)}{1.26}\,C - \underset{(0.19)}{0.40}\,D + \underset{(0.11)}{1.88}\,P - \underset{(0.14)}{0.10}\,R + \underset{(0.19)}{1.71}\,AD + \underset{(0.12)}{1.43}\,AP - \underset{(0.12)}{0.10}\,CP + \underset{(0.12)}{0.13}\,DP + \underset{(0.20)}{1.49}\,PR - \underset{(0.24)}{0.31}\,RU \tag{8}$$

Copyright © 2005 John Wiley & Sons, Ltd.

Table VIII. Sub-plot analysis for Example 3

Model	Effects						MSE	AIC	BIC	CV
1	*P*	*R*		*AP*	*CP*	*DP*	0.302	−0.364	−0.162	0.196
2	*P*	*S*		*AP*	*BP*	*CS*	0.277	−0.450	−0.248	0.207
3	*P*	*Q*	*S*	*AP*	*BS*		0.249	−0.558	−0.356	0.211
4	*P*	*U*		*AP*	*BP*	*CP*	0.279	−0.444	−0.241	0.216
5	*P*	*U*		*AP*	*CP*		0.308	−0.511	−0.350	0.224
6	*P*	*U*		*AP*			0.344	−0.568	−0.447	0.230
7	*P*	*T*	*U*	*AP*	*DP*		0.305	−0.355	−0.153	0.236
8	*P*	*U*		*AP*	*CP*	*DP*	0.303	−0.362	−0.160	0.240
9	*P*	*S*	*U*	*AP*	*BP*	*AU*	0.290	−0.239	0.004	0.244
10	*P*	*R*	*S*	*AP*	*BP*	*CP*	0.333	−0.101	0.142	0.245
⋮										

Table IX. Whole-plot analysis for Example 3

Model	Effects						MSE	AIC	BIC	CV
1	*A*	*C*	*D*	*AD*	*PR*	*RU*	0.244	−0.661	−0.479	0.691
2	*C*	*D*		*AD*	*PQ*	*QR*	0.304	−0.690	−0.569	0.694
3	*A*	*C*		*AD*	*CD*	*PR*	0.275	−0.791	−0.670	0.761
4	*A*	*C*		*AD*	*PQ*	*PR*	0.295	−0.721	−0.600	0.786
5	*A*	*C*		*AC*	*AD*	*PR*	0.335	−0.593	−0.472	0.789
6	*A*	*C*		*AD*	*PR*	*RU*	0.363	−0.514	−0.393	0.792
7	*A*	*C*		*AD*	*PR*	*PS*	0.360	−0.521	−0.399	0.796
8	*A*	*C*		*AD*	*PR*	*QR*	0.353	−0.543	−0.422	0.801
9	*A*	*C*	*D*	*AD*	*PR*	*PU*	0.305	−0.439	−0.257	0.827
10	*A*	*C*	*D*	*AD*	*CD*	*PR*	0.258	−0.605	−0.423	0.850
⋮										

After eliminating the effects that are not significant such as *D*, *R*, *CP*, *DP* and *RU*, the right model i obtained.

Allowing for many terms in the final model may of course lead to wrong conclusions. However, there are als several options to construct a more parsimonious final model. Since each stage can be considered independently one can consider the top model in each stage separately and use backward elimination methods to get rid of th terms that are not significant. The new variance estimates can then be calculated using the mean squared error of the reduced models. Another option is to consider the top five models with two, four and six terms (equall divided between main-effects and two-factor interactions) at each stage of the algorithm. Those models ca then be compared to obtain a more parsimonious final model. As indicated earlier, the model presented abov was from one simulation run only. In Table X we summarize the results for different error variances obtaine from 200 simulations for each case. The right model at the sub-plot level has been detected in almost all cases However, the problem occurs at the whole-plot level when the magnitude of the error against which the effect are compared with gets larger. Considering the fact that at the whole-plot level we are searching for a model tha contains five active factors using only 12 runs, the results for Scenario 3 in Table X are still quite promising They also underline the importance of a screening procedure to seek a set of candidate models for furthe investigation and not only take the top candidate model ranked by an arbitrary criterion.

5. *DISCUSSION*

Split-plot experiments with mirror image pairs as sub-plots offer the possibility to estimate s main effects an ws interactions independently from the w main effects, ww and ss interactions. For these designs, the tw

Copyright © 2005 John Wiley & Sons, Ltd.

Table X. Simulation results for Example 3 (with noise)

		$\sigma_w^2 + (1/2)\sigma_\varsigma^2 = 0.1$			$\sigma_w^2 + (1/2)\sigma_\varsigma^2 = 0.5$			$\sigma_w^2 + (1/2)\sigma_\varsigma^2 = 1.0$		
		$\sigma_w^2 = \sigma_s^2$	$\sigma_w^2 = 2\sigma_s^2$	$2\sigma_w^2 = \sigma_s^2$	$\sigma_w^2 = \sigma_s^2$	$\sigma_w^2 = 2\sigma_s^2$	$2\sigma_w^2 = \sigma_s^2$	$\sigma_w^2 = \sigma_s^2$	$\sigma_w^2 = 2\sigma_\varsigma^2$	$2\sigma_w^2 = \sigma_\varsigma^2$
Sub-plot analysis	Scenario 1*	2**	8	6	3	1	3	5	2	5
	Scenario 2	200	200	200	200	200	200	200	200	199
	Scenario 3	200	200	200	200	200	200	200	200	200
Whole-plot analysis	Scenario 1	0	0	0	0	0	0	0	0	0
	Scenario 2	178	185	182	134	153	137	84	87	80
	Scenario 3	200	200	196	192	195	195	159	168	163
Overall***	Scenario 4	178	185	182	134	153	137	84	87	79

*Scenario 1: First model is the right model. Scenario 2: right model embedded in first model. Scenario 3: right model embedded in the top 10 models. Scenario 4: reduced first model is the right model.

**Each cell represents the number simulations out of 200 in which the corresponding scenario holds.

***First models from both sub-plot and whole-plot analyses.

Copyright © 2005 John Wiley & Sons, Ltd.

independent estimation steps can be performed using OLS. Hence, for screening situations where a searcl through possible candidate models is necessary, fewer candidate models need to be investigated in each ste and this can be done by means of rather simple calculations. We used a search procedure where in each step w investigated all possible models with up to three main effects and three two-factor interactions. This is a quit exhaustive search. Considering the fact that the important task for a screening procedure is to provide a set o candidate models, where the correct model is included or imbedded, the results from the proposed procedur look quite promising. The somehow poorer performance for the search procedure at the whole-plot level shoul be expected since the variance at that level is larger. Finally we point out that the procedure proposed in thi article, can be used for any design that can be written in the $\begin{bmatrix} \mathbf{W} & \mathbf{S} \\ \mathbf{W} & -\mathbf{S} \end{bmatrix}$ form whether the columns in $\mathbf{W}$ and $\mathbf{S}$ ar orthogonal or not. However, the performance of the procedure will of course depend on the design's properties In particular, full aliasing among effects, as may be the case when the columns in $\mathbf{W}$ and $\mathbf{S}$ are from geometri designs and many factors are screened, complicates the identification of active factors.

REFERENCES

1. Addelman S. Some two-level fractional factorial plans with split-plot confounding. *Technometrics* 1964; **6**:253–258.
2. Huang P, Chen D, Voelkel JO. Minimum-aberration two-level split-plot designs. *Technometrics* 1998; **40**(4):414–326
3. Bingham D, Sitter RS. Minimum aberration two-level fractional factorial split-plot designs. *Technometrics* 1999 **41**(1):62–70.
4. Bisgaard S. The design and analysis of $2^{k-p} \times 2^{q-r}$ split plot experiments. *Journal of Quality Technology* **32**:39–56.
5. Kulahci M. Applications of Plackett and Burman designs to split plot design. *PhD Thesis*, University of Wisconsin Madison, WI, 2001.
6. Tyssedal J, Kulahci M, Bisgaard S. *Split-plot Designs with Mirror Image Pairs as Subplot*, 2004; in review.
7. Plackett RL, Burman JP. The design of optimum multifactorial experiments. *Biometrika* 1946; **33**:305–325.
8. Bisgaard S, Fuller H, Barrios E. Two-level factorials run as split plot experiments. *Quality Engineering* 1996; **8**(4):705-708.
9. Box GEP, Tyssedal J. Projective properties of certain orthogonal arrays. *Biometrika* 1996; **83**(4):950–955.
10. Box GEP, Bisgaard S. What can you find out from 12 experimental runs. *Quality Engineering* 1993; **5**:663–668.
11. Lin DKJ, Draper NR. Projection properties of plackett and burman designs. *Technometrics* 1992; **34**:423–428.
12. Samset O, Tyssedal J. Repeat and mirror-image patterns and correlation structures of Plackett–Burman designs, thei foldovers and half-fractions. *Preprint, Statistics No. 13/1998*, Norwegian University of Science and Technology Trondheim, Norway. 1998.
13. Box GEP, Meyer RD. Finding the active factors in fractionated screening experiments. *Journal of Quality Technology* 1993; **25**:94–105.
14. Hamada H, Wu CFJ. Analysis of designed experiments with complex aliasing. *Journal of Quality Technology* 1992 **24**(3):130–137.
15. Box GEP, Meyer RD, Steinberg D. Follow-up designs to resolve confounding in multifactor experiments *Technometrics* 1996; **38**(4):303–332.
16. Box GEP, Tyssedal J. Sixteen run designs of high projectivity for screening. *Communications in Statistics-Simulation and Computation* 2001; **30**(2):217–228.
17. Samset O. Factor screening using non-geometric Plackett–Burman designs. *PhD Thesis*, Department of Mathematica Sciences, The Norwegian University of Science and Technology, 1999.
18. Daniel C. Use of half-normal plots in interpreting factorial two-level experiments. *Technometrics* 1959; **1**:311–340.
19. Lenth RV. Quick and easy analysis of unreplicated factorials. *Technometrics* 1989; **31**(4):469–473.
20. Langsrud Ø, Ellekjær MR, Næs T. Identifying significant effects in fractional factorial experiments. *Journal o Chemiometrics* 1994; **8**:205–219.
21. Kulahci M, Bisgaard S. The use of Plackett and Burman designs to construct split plot designs. *Technometrics* 2005 to appear.

APPENDIX A. DERIVATION OF THE COVARIANCE MATRIX FOR THE GLS ESTIMATOR $\hat{\beta}$

We only consider models with main effects and two-factor interactions.

Copyright © 2005 John Wiley & Sons, Ltd.

Result:

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}) = \sigma_s^2 \begin{bmatrix} (\mathbf{W}'_*\mathbf{W}_*)^{-1} & 0 \\ 0 & (\mathbf{S}'_*\mathbf{S}_*)^{-1} \end{bmatrix} + 2\sigma_w^2 \begin{bmatrix} (\mathbf{W}'_*\mathbf{W}_*)^{-1} & 0 \\ 0 & 0 \end{bmatrix}$$

where $\mathbf{W}_*$ *contains the columns for the* w, ww *and* ss *effects and* $\mathbf{S}_*$ *contains the columns for the* s *and* ws *effects.*

Proof. From Kulahci[5] the GLS estimator $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ and, hence, $\mathrm{Cov}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{V}\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}$. With $\rho = \sigma_w^2/(\sigma_w^2 + \sigma_s^2)$ the matrix $\mathbf{V}$ can be written as

$$\mathbf{V} = (\sigma_w^2 + \sigma_s^2)(1-\rho)\mathbf{I}_{N\times N} + (\sigma_w^2 + \sigma_s^2)\rho \begin{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} & 0 & \cdots & 0 \\ 0 & \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \end{bmatrix}$$

$$= \sigma_s^2\mathbf{I}_{N\times N} + \sigma_w^2 \begin{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} & 0 & \cdots & 0 \\ 0 & \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \end{bmatrix}$$

$$= \mathbf{V}_1 + \mathbf{V}_2$$

Hence, $\mathrm{Cov}(\hat{\boldsymbol{\beta}}) = \sigma_s^2(\mathbf{X}'\mathbf{X})^{-1} + (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{V}_2\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}$. Now let $\mathbf{X} = [\mathbf{W}_*, \mathbf{S}_*]$. Columns in $\mathbf{W}_*$ are orthogonal to columns in $\mathbf{S}_*$. Hence,

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} (\mathbf{W}'_*\mathbf{W}_*)^{-1} & 0 \\ 0 & (\mathbf{S}'_*\mathbf{S}_*)^{-1} \end{bmatrix}$$

Similarly it can be shown that $\mathbf{V}_2\mathbf{X} = 2\sigma_w^2[\mathbf{W}_*, \mathbf{0}]$, which leads to the covariance matrix given above.

Since the least squares estimator can be expressed as $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$, it follows that w, ww and ws effects can be calculated separately from s and ws effects and *vice versa* in the way explained by the two first steps of the algorithm in Section 3.

Authors' biographies

John Tyssedal is an Associate Professor in the Department of Mathematical Sciences at the Norwegian University of Science and Technology in Trondheim. His research interests include design of experiments and time series. He is a member of the American Statistical Association and the European Network of Business and Industrial Statistics (ENBIS).

Murat Kulahci is an Assistant Professor in the Department of Industrial Engineering at Arizona State University. His research interests include design of experiments, statistical process control and financial engineering. He is a member of the American Statistical Association, ENBIS and the Institute for Operations Research and the Management Sciences.

Copyright © 2005 John Wiley & Sons, Ltd.