DEPARTMENT OF ENVIRONMENT,
TECHNOLOGY AND TECHNOLOGY MANAGEMENT

# Split-Plot Experiments with Factor-Dependent Whole-Plot Sizes

**Eric Schoen, Bradley Jones & Peter Goos**


**UNIVERSITY OF ANTWERP**
**Faculty of Applied Economics**

Stadscampus
Prinsstraat 13, B.213
BE-2000 Antwerpen
Tel. +32 (0)3 265 40 32
Fax +32 (0)3 265 47 99
http://www.ua.ac.be/tew

# FACULTY OF APPLIED ECONOMICS

DEPARTMENT OF ENVIRONMENT,
TECHNOLOGY AND TECHNOLOGY MANAGEMENT

## Split-Plot Experiments with Factor-Dependent Whole-Plot Sizes

**Eric Schoen, Bradley Jones & Peter Goos**

RESEARCH PAPER 2010-001
JANUARY 2010

University of Antwerp, City Campus, Prinsstraat 13, B-2000 Antwerp, Belgium
Research Administration – room B.213
phone: (32) 3 265 40 32
fax: (32) 3 265 47 99
e-mail: joeri.nys@ua.ac.be



**D/2010/1169/001**

# Split-Plot Experiments with Factor-Dependent Whole-Plot Sizes

Eric Schoen, Bradley Jones, and Peter Goos

January 12, 2010

**Abstract**

In industrial split-plot experiments, the number of runs within each whole plot is usually determined independently from the factor settings. As a matter of fact, it is often equal to the number of runs that can be done within a given period of time or to the number of samples that can be processed in one oven run or with one batch. In such cases, the size of every whole plot in the experiment is fixed no matter what factor levels are actually used in the experiment. In this article, we discuss the design of a real-life experiment on the production of coffee cream where the number of runs within a whole plot is not fixed, but depends on the level of one of the whole-plot factors. We provide a detailed discussion of various ways to set up the experiment and discuss how existing algorithms to construct optimal split-plot designs can be modified for that purpose. We conclude the paper with a few general recommendations.

KEY WORDS: Coordinate-Exchange Algorithm; $D$-optimum Designs; Point-Exchange Algorithm; Restricted Randomization.

## 1 Introduction

In industrial experimentation, there are very often factors that are hard to change (HTC) and factors that are easy to change (ETC). As discussed by Letsinger, Myers and Lentner (1996) and Bisgaard (2000), among many others, this usually results in a restriction on the randomization of the runs. That is, the runs of the experiments are grouped into whole plots. These whole plots have a constant setting of the HTC factors. The ETC factors vary within a whole plot. The runs in this context are called sub-plots because they are nested within whole plots, and the resulting experiment is a split-plot experiment. The randomization is performed separately for the whole plots and the sub-plots. Commonly, the HTC factors are named whole-plot factors, while the ETC factors are referred to as sub-plot factors.

One example of a split-plot experiment was recently carried out at Friesland-Campina, a dairy company in the Netherlands. The purpose of the experiment, which is the motivating example for this paper, was to model the effect of nine factors on the properties of coffee cream. Two of these were HTC factors. The first one, number of process loops, has settings $1\times$ and $2\times$. A change in this factor from one setting to the other required disconnection of the piping of the production system. The second HTC factor, processing speed, controls the speed with which the runs were performed, and thereby the number of runs per unit of time. The remaining factors were all ETC, because they were material-related or required adjustment of individual components of the system.

The investigators imposed two constraints on the way the experiment was performed. First, at most six whole plots could be used. Second, the total number of sub-plots had to be about 65. Given these requirements, it was natural to design a split-plot experiment with two whole-plot factors, six whole plots, seven sub-plot factors, and 10 or 11 sub-plots per whole plot. The experimental factors had different numbers of levels, and two of them had constraints on their level-combinations. Therefore, neither the combinatorial construction methods for split-plot designs (see, e.g., Bingham, Schoen and Sitter (2004)) nor the equivalent-estimation designs proposed by Vining, Kowalski and Montgomery (2005) and Parker, Kowalski and Vining (2007) could be used. Instead, the investigators utilized the optimal experimental design approach which offers more flexibility when setting up split-plot experiments. In this paper, we use the optimal experimental design approaches of Goos and Vandebroek (2003), who presented a point-exchange algorithm to set up split-plot experiments, Jones and Goos (2007), who developed a coordinate-exchange algorithm for the same purpose, and Trinca and Gilmour (2001), who suggested a general method for the design of split-plot and other multi-stratum experiments, to construct various alternative designs for the coffee cream experiment.

However, none of these three approaches could be used without modification because the sizes of the whole plots in the coffee cream experiment depend on the setting of the factor process speed. The low level of the factor process speed results in material for seven batches of coffee cream which allows for seven runs of the experiment, while the factor's high level results in material for 14 batches and thus 14 runs. So a change in level of one of the whole-plot factors also changes the number of runs in a whole plot. This feature makes the coffee cream experiment different from any other split-plot experiment in the literature and motivated us to investigate how the existing optimal design approaches could be modified to deal with this complication.

The rest of the paper is organized as follows. First, we discuss the coffee cream experiment more comprehensively. Next, we briefly review important issues in the statistical design and analysis of split-plot experiments. Then, we present design options for the coffee cream experiments with three different run sizes, and we discuss three possible construction methods adapted from existing algorithms. The paper ends with a discussion of the computational results and recommendations.

## 2 Motivating example

In a project carried out in Autumn, Winter and early Spring of 2007-2008, the production process of coffee cream of FrieslandCampina was redesigned. We suppress the details of the redesign to protect the intellectual property of the company. The next two subsections give a sanitized summary of the factors and of the challenges to construct the experimental design, respectively.

## 2.1 Factors

The first factor bears on the content of ingredient A in the final product. The content should either be 16, or 17, or 18 % (factor 1). This can be achieved by mixing milk with ingredient B. Ingredient B can be applied in various concentrations as characterized by a three-level continuous factor with equidistant levels 5, 10, and 15 % (factor 2). Second, the mix is to be conducted through one or two process loops (factor 3) using one of four possible settings of the process parameter coded with PP (factor 4). The treatment applied in process step 2 can be extended over some time; this time can be varied (factor 5). In addition, the temperature maintained in step 2 can be set at several levels (factor 6). Process step 3 can be applied at different rates (factor 7). Next, there is a process step 4 where the material is condensed. The temperature used and the time spent during this process can be set at several possible values (factors 8 and 9).

A key goal in the redesign of the coffee cream's production process was to derive a simple polynomial model that relates the settings of controllable factors to viscosity and other properties of the coffee cream. As there was no theoretical approximation available, the project team decided to conduct a series of experiments with different settings of the product factors 1 and 2, and the process factors 3-9. The factors and relevant settings are all shown in Table 1. The middle column shows whether the factors are easy or hard to change. Factor 3, process loop, is a hard-to-change factor, because adding a process loop requires disconnecting the piping of the production system. Another hard-to-change factor is factor 7, the step-3 processing speed. The remaining factors are all easy to change, as they are material-related or require adjustment of individual components of the system only.

The factors 3, 5, 7, and 9 were two-level factors. Factor 3 is categorical with just two options. Factors 5, 7, and 9 are numeric, but intermediate settings were hard to attain. The investigators chose to vary factors 1, 2, 6, and 8 at three levels, because they were concerned about possible curvature of the factor-response relationship over the range of these factors. They had similar concerns with factor 4, process parameter PP. There is an additional complication with this factor, because the possible settings for it depend on whole-plot factor 3;

Table 1: Experimental factors and their settings

| | Factor | Type† | Settings |
|---|---|---|---|
| 1 | Ingredient A | ETC | 16, 17, 18 % |
| 2 | Ingredient B | ETC | 5, 10, 15 % |
| 3 | Process loop | HTC | 1×, 2× |
| 4 | Process parameter (PP) | ETC | 1, 1.5, 2, 2.5 $\log_{10}$ units |
| 5 | Step 2 processing time | ETC | 60, 120 sec |
| 6 | Step 2 temperature | ETC | 40, 50, 60 °C |
| 7 | Step 3 processing speed | HTC | 0.5, 1 |
| 8 | Step 4 processing time | ETC | 60, 150, 240 sec |
| 9 | Step 4 temperature | ETC | 90, 95 °C |

†ETC: easy to change; HTC: hard to change

see Table 2. The value of 2.5 for the process parameter PP could not be attained with a single process loop, and the value of 1 was too low to be attained with two process loops. This led the investigators to employ four settings of this factor.

### 2.2 Challenges

One production run took a week, and the total experimental effort was not to exceed six weeks. The presence of two hard-to-change factors clearly calls for a split-plot experiment, where the two hard-to-change factors, process loop and processing speed, are fixed to one of their levels at the start of every week. As a result, different weeks could have different settings of the hard-to-change factors. However, within a week, several batches of coffee cream, involving various settings of the easy to change factors, could be produced. In this setup, the weeks correspond to the whole plots of the design, yielding six whole plots in total.

There were two special challenges that made the design of the split-plot experiment a non-standard problem. First, the number of different coffee cream batches within a week depended on the setting of the step-3 processing speed. At low speed, 7 batches could be made, while the high speed permitted 14 batches to be made. So the size of the whole plots depended on the setting of one of the whole-plot factors, processing speed: some whole plots involve 7 sub-plots, while others have 14 sub-plots. The second challenge in the design is the dependence of the feasible levels of one sub-plot factor on the level of a whole-plot factor (see Table 2, as discussed earlier).

Both challenges enforce restrictions on the randomization additional to the split-plot restrictions.

## 3 $D$-optimum split-plot designs

The restriction in the randomization of a split-plot experiment results in the presence of two random errors in the statistical model:

$$Y = X\beta + Z\delta + \epsilon \tag{1}$$

In this expression, $Y$ is an $N \times 1$ vector of responses, and the $p$ columns of the $N \times p$ matrix $X$ are vectors with the settings of the explanatory variables. The model terms included in $X$ are the intercept, nine linear effects, five quadratic effects, and 36 linear-by-linear product terms. Further, $\beta$ is a $p \times 1$ column vector containing the effects of the explanatory variables, $\epsilon$ is an $N \times 1$ vector of random contributions of individual sub-plots, $Z$ is an $N \times b$ indicator matrix

Table 2: Feasible settings for the process parameter PP

| Process | PP | | | |
|---|---|---|---|---|
| Loop | 1 | 1.5 | 2 | 2.5 |
| 1× | ✓ | ✓ | ✓ | |
| 2× | | ✓ | ✓ | ✓ |

denoting from which whole plot an observation was taken, and $\delta$ is a $b \times 1$ column vector with the random contributions of the whole plots.

We adopt the usual assumptions that the random contributions in $\delta$ and $\epsilon$ are independent and normally distributed with mean 0 and variance $\sigma_1^2$ and $\sigma_0^2$, respectively. The former variance is the whole-plot error variance, while the latter is the sub-plot error variance. It can be shown that the variances and covariances of the responses in $Y$ are given by the matrix

$$V = \sigma_1^2 ZZ' + \sigma_0^2 I_N. \tag{2}$$

The best linear unbiased estimator of $\beta$ is the generalized least squares estimator $\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}Y$. The information matrix about $\beta$ is then $X'V^{-1}X$. The determinant $D = |X'V^{-1}X|$ of this matrix is inversely related to the volume of a joint confidence region about the parameters in $\beta$. So the larger the determinant, the more precisely the joint estimation of the parameters in $\beta$. For this reason, the determinant is often used as a quality measure of a split-plot design. It is usually referred to as the $D$-criterion. Another criterion used as a quality measure for a design is the $A$-criterion, which is the trace of the variance-covariance matrix of $\hat{\beta}$, $(X'V^{-1}X)^{-1}$. Thus, the $A$-criterion sums the variances with which the individual parameters in $\beta$ are estimated.

In order to calculate the $D$- or $A$-criterion, we need the design matrix $X$, the split-plot structure $Z$ and the variance components $\sigma_1^2$ and $\sigma_0^2$. By writing $V$ as $V = \sigma_0^2(\gamma ZZ' + I_N)$, with $\gamma = \sigma_1^2/\sigma_0^2$, we see that the sub-plot error variance $\sigma_0^2$ can be worked out of the expressions for the $D$- and $A$-criteria. So the best designs in terms of these criteria do not depend on the size of the sub-plot error variance. Therefore, we set $\sigma_0^2 = 1$, without loss of generality. The best designs do, however, depend on the ratio $\gamma$. In this paper, we assume that this ratio is 1, which is in line with empirical values obtained by Letsinger, Myers and Lentner (1996), Littell *et al.* (1996), Kowalski, Cornell and Vining (2002) and Webb, Lucas and Borkowski (2004) for various published data sets from industrial experiments. If we fix the ratio $\gamma$, and if we assume a predetermined split-plot structure, then the $D$- and $A$-criteria depend only on the design matrix $X$. A $D$-optimal split-plot design is a design that maximizes the $D$-criterion, whereas an $A$-optimal design minimizes the $A$-criterion. In the next few sections we search for $D$-optimal designs for several feasible split-plot structures of the coffee cream experiment, and evaluate them in terms of the $D$- and $A$-criteria.

We refer to Goos (2002), Goos, Langhans and Vandebroek (2006) and Jones and Nachtsheim (2009) for a comprehensive account of the design issues peculiar to split-plot designs, and for a detailed discussion of the analysis of split-plot data.

# 4 Alternative lay-outs for the coffee cream experiment

In the coffee cream experiment, the investigators wanted to model the main effects and two-factor interaction effects of the whole-plot factors. This requires the use of the four different combinations of the levels of the whole-plot factors process loop and processing speed, as given by the $2^2$ factorial design. There are six whole plots in total. For practical application, the order in which these

Table 3: Schematic representation of the 56-run design options for the coffee cream experiment.

(a) Design 56 A

| Process speed | Process loop 1 | 2 |
|---|---|---|
| 0.5 | WP 1 | WP 3 |
| | WP 5 | WP 6 |
| 1.0 | WP 2 | WP 4 |

(b) Design 56 B1

| Process speed | Process loop 1 | 2 |
|---|---|---|
| 0.5 | WP 1 | WP 3 |
| | WP 5 | |
| | WP 6 | |
| 1.0 | WP 2 | WP 4 |

(c) Design 56 B2

| Process speed | Process loop 1 | 2 |
|---|---|---|
| 0.5 | WP 1 | WP 3 |
| | | WP 5 |
| | | WP 6 |
| 1.0 | WP 2 | WP 4 |

six whole plots are run should be randomized. For the clarity of our exposition, however, we allocate the first four whole plots to the $2^2$ different combinations of whole-plot factor levels. In the rest of the paper, these four whole plots are labeled WP 1 up to WP 4. We call the two whole plots that are not yet allocated to a combination of whole-plot factor levels WP 5 and WP 6, respectively.

There are ten possible scenarios for the allocation of WP 5 and WP 6. Three of these lead to a 56-run design, four lead to a 63-run design and the remaining three scenarios lead to a 70-run design. The three scenarios that lead to a 56-run design are shown in Table 3. In each of these scenarios both WP 5 and WP 6 are run at the low process speed. Design 56 A differs from designs 56 B1 and 56 B2 in the distribution of the whole plots over the process loops. For each level of process speed, design 56 A balances the number of whole plots at both levels of process loop. Designs 56 B1 and 56 B2 do not have such a balance. Statistically, scenarios 56 B1 and 56 B2 are equivalent, because switching the levels of the factor process loop and relabeling factor PP of any design of type 56 B1 results in a design of type 56 B2, or vice versa.

The four scenarios that lead to a 63-run design are shown in Table 4. In each of these scenarios, one of the whole plots WP 5 and WP 6 is allocated to the low level of the factor processing speed, while the other is allocated to the factor's high level. Here, there are two pairs of statistically equivalent types of designs. The first pair is labeled 63 A1 and 63 A2; the second pair is labeled 63 B1 and 63 B2. Each design of type 1 (either A1 or B1) can be turned into a design of type 2 (A2 or B2), and vice versa, by switching the levels of the factor process loop and relabeling factor PP. Pair A differs from pair B in the distribution of the whole plots over the process loops. For each level of the factor processing speed, the designs in pair A balance the numbers of whole plots at both levels of process loop, while this is not the case for the designs in pair B.

The three scenarios that lead to a 70-run design are shown in Table 5. This case is similar to the 56-run case, because WP 5 and WP 6 are run at the same level of the factor processing speed in these three scenarios.

Table 4: Schematic representation of the 63-run design options for the coffee cream experiment.

(a) Design 63 A1

| Process speed | Process loop 1 | Process loop 2 |
|---|---|---|
| 0.5 | WP 1<br>WP 5 | WP 3 |
| 1.0 | WP 2<br>WP 6 | WP 4 |

(b) Design 63 A2

| Process speed | Process loop 1 | Process loop 2 |
|---|---|---|
| 0.5 | WP 1 | WP 3<br>WP 5 |
| 1.0 | WP 2 | WP 4<br>WP 6 |

(c) Design 63 B1

| Process speed | Process loop 1 | Process loop 2 |
|---|---|---|
| 0.5 | WP 1<br>WP 5 | WP 3 |
| 1.0 | WP 2 | WP 4<br>WP 6 |

(d) Design 63 B2

| Process speed | Process loop 1 | Process loop 2 |
|---|---|---|
| 0.5 | WP 1 | WP 3<br>WP 5 |
| 1.0 | WP 2<br>WP 6 | WP 4 |

Table 5: Schematic representation of the 70-run design options for the coffee cream experiment.

(a) Design 70 A

| Process speed | Process loop 1 | Process loop 2 |
|---|---|---|
| 0.5 | WP 1 | WP 3 |
| 1.0 | WP 2<br>WP 5 | WP 4<br>WP 6 |

(b) Design 70 B1

| Process speed | Process loop 1 | Process loop 2 |
|---|---|---|
| 0.5 | WP 1 | WP 3 |
| 1.0 | WP 2<br>WP 5<br>WP 6 | WP 4 |

(c) Design 70 B2

| Process speed | Process loop 1 | Process loop 2 |
|---|---|---|
| 0.5 | WP 1 | WP 3 |
| 1.0 | WP 2 | WP 4<br>WP 5<br>WP 6 |

Obviously, the 70-run design options were statistically more attractive than the 56- and 63-run alternatives because they result in a higher precision of the parameter estimates and a larger power of the effect tests. Nevertheless, there was a practical reason why the alternatives with 56 and 63 runs could not be dismissed instantaneously: more runs imply more measurements, and thus an increase in the overall cost.

Below, we do not discuss the pairs of scenarios 56 B1 and 56 B2, 63 A1 and 63 A2, 63 B1 and 63 B2 and 70 B1 and 70 B2 separately, because they yield identical results, due to the fact that relabeling the levels of the factors process loop and PP suffices to convert a design from one type into a design of the other type. As a result, we report results on six (statistically) different scenarios, which we label 56 A, 56 B, 63 A, 63 B, 70 A and 70 B.

# 5 Adaptation of existing algorithms

In this section, we discuss adaptations of three state-of-the-art methods from the literature on the optimal design of split-plot experiments to design the coffee cream experiment. The original versions of the three methods have in common that a specification of both the number of whole plots and their sizes by the experimenter is required. In the coffee cream experiment, only the number of whole plots is fixed, so that an adaptation of the original methods is needed.

## 5.1 Three-step approach

Essentially, the design of a split-plot experiment requires the combination of two designs, one for the HTC or whole-plot factors and one for the ETC or sub-plot factors. Schoen (1999) demonstrated how different fractional factorial two-level designs can be merged to obtain a design with nested strata, such as a split-plot design. Trinca and Gilmour (2001) use a generalized version of this method to construct multi-stratum response surface designs, special cases of which are split-plot designs. The method proposed by Trinca and Gilmour (2001) would be an ideal method for the coffee cream experiment if the feasible levels of the process parameter PP had not been dependent on the level of the HTC factor process loop. Due to this complication, the point-wise interchange procedure employed by Trinca and Gilmour (2001) to optimize the merging of the whole-plot and sub-plot designs leads to infeasible combinations of levels for the process parameter PP and the process loop. Therefore, we used the following three-step procedure to set up suitable designs for the coffee cream experiment:

1. Construct a suitable whole-plot design for the two two-level HTC factors in six whole plots.

2. Construct a sub-plot design for the five three-level ETC factors and the two two-level ETC factors in six blocks of predetermined size matching the whole-plot sizes in the scenarios in Tables 3-5.

3. Find an optimum allocation of the blocks of the sub-plot design to the runs of the whole-plot design using the $D$-criterion which jointly considers the effects of the whole-plot and sub-plot factors.

Because of the limited number of whole plots, it is possible to enumerate all of the options for the whole-plot design in step 1; see Section 4. To construct a sub-plot design in six blocks (step 2), we used the algorithm of Cook and Nachtsheim (1989). This algorithm optimizes a $D$-criterion assuming fixed block sizes and a pre-specified model for the treatment factors. In our case, the block sizes were 14 or 7, as required for the various whole-plot options in Tables 3-5. For example, for design option 56 A, we generated sub-plot designs with two whole plots of size 14 and four whole plots of size 7. The model we used for the sub-plot factors included all linear and quadratic main effects and all linear-by-linear interaction effects. At this stage, we treated the sub-plot factor PP as a three-level factor, because the final assignment of the levels depend on the setting of the whole-plot factor process loop. For each of the run sizes 56, 63 and 70, step 2 results in a single sub-plot design.

In step 3, the blocks of each sub-plot design were linked in all possible ways to the runs of the whole-plot design. Thus, a complete enumeration of all possible combinations of the whole-plot and sub-plot designs from steps 1 and 2 was performed. Then, the settings of PP were relabeled such that a four-level factor satisfying the factor constraints in Table 2 resulted. The final designs were evaluated in terms of the $D$-criterion, and the design with the largest value was selected for each run size.

## 5.2 Integrated approach

Recently, Goos and Vandebroek (2003) and Jones and Goos (2007) proposed algorithms that jointly optimize the settings of the HTC and ETC factors. Both algorithms require modification to handle the factor-dependent whole-plot sizes in the coffee cream experiment.

### 5.2.1 Point-exchange algorithm

Roughly speaking, the original algorithm of Goos and Vandebroek (2003) involves the following steps:

1. Construct a candidate set containing all feasible combinations of levels for the nine factors in Table 1.

2. Specify the number of whole plots and the whole-plot sizes.

3. Create a random starting design obeying the split-plot structure.

4. For each run in the design, evaluate the effect on the $D$-criterion of exchanging its factor level combination by each of the other factor level combinations in the candidate set that have the same settings of the HTC factors. Carry out the best exchange and repeat this step until no further improvement in the $D$-criterion can be made.

5. For each run in the design, evaluate the effect on the $D$-criterion of interchanging its factor level combination with every other factor level combination in a different whole plot of the design that has the same settings of the HTC factors. Carry out the best interchange and repeat this step until no further improvement in the $D$-criterion can be made.

6. For each whole plot in the design, evaluate the effect on the $D$-criterion of exchanging its HTC factor level combination with any other feasible HTC factor level combination. Carry out the best exchange and repeat this step until no further improvement in the $D$-criterion can be made.

The algorithm is called a point-exchange algorithm because, in steps 4 and 5, it interchanges or exchanges combinations of levels of all the experimental factors and these combinations are usually called design points. The steps 4 up to 6 are repeated either until no further increases in the $D$-criterion occur, or until a specified number of cycles have been carried out. Then, the entire procedure is repeated for a prespecified number of starting designs. The best design according to the $D$-criterion is saved.

For the coffee cream experiment, the restriction on the levels of the process parameter PP did not require modification of the algorithm of Goos and Vandebroek (2003). It sufficed to exclude infeasible factor level combinations from the candidate set to satisfy the constraints. Therefore, the factor level combinations with 1 process loop and PP$= 2.5$ and with 2 process loops and PP$= 1$ were not included in the candidate set. This resulted in a candidate set with 3888 ($= 3^5 \times 2^4$) factor level combinations.

We dealt with the factor-dependent whole-plot sizes by skipping step 6 of the point-exchange algorithm of Goos and Vandebroek (2003). In addition, we fixed the whole-plot design in advance and modified the creation of a starting design so that only the sub-plot factor level combinations were generated at random. We repeated this procedure for each of the scenarios 56 A, 56 B, 63 A, 63 B, 70 A and 70 B, yielding six different designs for the coffee cream experiment.

### 5.2.2 Coordinate-exchange algorithm

Jones and Goos (2007) proposed a coordinate-exchange algorithm for the construction of $D$-optimum split-plot experiments. The algorithm proceeds as follows:

1. Specify number of whole plots and their sizes.

2. Create a random starting design that obeys the split-plot structure.

3. For each run in the design, perform the following steps, factor by factor:

   (a) For each HTC factor, evaluate the effect on the $D$-criterion of switching its level along a discrete range of other levels. Keep the best level change if it improves the $D$-criterion.

   (b) For each ETC factor, evaluate the effect on the $D$-criterion of switching its level along a discrete range of other levels. Keep the best level change if it improves the $D$-criterion.

The algorithm is called a coordinate-exchange algorithm, because it involves an evaluation of changes to individual factor levels, i.e. changes to coordinates of the design points. Step 3 is repeated either until no further improvements in the $D$-criterion occur or until a specified number of cycles have been carried out. Then, the entire procedure is repeated for a prespecified number of starting designs. The best design according to the $D$-criterion is saved.

Table 6: Comparison of three design approaches for the coffee cream experiment in terms of the $D$-criterion

| $N$ | option | $D$ | | |
|---|---|---|---|---|
| | | 3-step | integrated, pe | integrated, ce |
| 56 | A | $3.47 \times 10^{10}$ | $4.94 \times 10^{13}$ | $1.42 \times 10^{14}$ |
| | B | $3.10 \times 10^{9}$ | $1.30 \times 10^{13}$ | $3.98 \times 10^{13}$ |
| 63 | A | $6.37 \times 10^{14}$ | $4.49 \times 10^{16}$ | $3.08 \times 10^{17}$ |
| | B | $2.43 \times 10^{14}$ | $4.00 \times 10^{16}$ | $2.47 \times 10^{17}$ |
| 70 | A | $1.03 \times 10^{17}$ | $8.68 \times 10^{18}$ | $6.78 \times 10^{19}$ |
| | B | $1.60 \times 10^{16}$ | $2.37 \times 10^{18}$ | $1.13 \times 10^{19}$ |

NOTES: pe: point exchange; ce: coordinate exchange

Table 7: Comparison of three design approaches for the coffee cream experiment in terms of the $A$-criterion

| $N$ | option | $A$ | | |
|---|---|---|---|---|
| | | 3-step | integrated, pe | integrated, ce |
| 56 | A | 143 | 88 | 87 |
| | B | 193 | 108 | 86 |
| 63 | A | 92 | 99 | 74 |
| | B | 107 | 79 | 64 |
| 70 | A | 72 | 79 | 61 |
| | B | 73 | 70 | 61 |

NOTES: pe: point exchange; ce: coordinate exchange

Accounting for the restriction on the levels of the process parameter PP was not as simple in the coordinate-exchange algorithm as it was in the point-exchange algorithm. The coordinate-exchange algorithm, as implemented in JMP, can easily handle linear constraints on the factor levels. However, the restriction on the levels of the process parameter PP could not be written in terms of a linear constraint. It therefore necessitated a modification of the algorithm's exchange procedure so that the setting PP $= 2.5$ was not considered for factor level combinations involving 1 process loop and the setting PP $= 1$ was not considered for combinations involving 2 process loops.

We dealt with the factor-dependent whole-plot sizes by skipping step 3a of the coordinate-exchange algorithm. Hence, as in the point-exchange algorithm, we fixed the whole-plot design. Additionally, we modified the creation of the random starting design so that only the levels of the ETC factors were generated at random. As for the point-exchange algorithm, we repeated this procedure for each of the scenarios 56 A, 56 B, 63 A, 63 B, 70 A and 70 B, yielding six new designs for the coffee cream experiment.

## 6 Results and Discussion

Tables 6 and 7 present the $D$- and $A$-criterion values for the designs we generated for the six scenarios 56 A, 56 B, 63 A, 63 B, 70 A and 70 B of the coffee cream experiment, using the three approaches outlined in the previous section and with all factor levels rescaled to the $[-1/2, +1/2]$ interval. When interpreting the $D$- and $A$-criterion values, it is important to realize that the $D$-criterion is a the-larger-the-better measure, while the $A$-criterion is a the-smaller-the-better measure.

In terms of the $D$-criterion, the three-step approach performs substantially worse than the other approaches. This comes as no surprise, because this approach does not use information on the HTC factor levels and ignores the factor level restrictions in Table 2 when creating the sub-plot design. For this reason, the levels of the ETC process parameter PP had to be relabeled after merging the whole-plot and sub-plot designs to obtain the required four levels satisfying the constraints in Table 2.

The $D$-criterion values produced by the modified coordinate-exchange algorithm are 2.9-7.8 times larger than those obtained with the modified point-exchange algorithm. So, the modified coordinate-exchange algorithm beats the modified point-exchange algorithm. It is unclear why, for the coffee cream experiment, the modified point-exchange algorithm gets stuck in local optima so easily, at least compared to the coordinate-exchange algorithm. Our experience with various types of algorithms suggests that this is due to the fact that the interchange step (step 5 of the point-exchange algorithm outlined in the previous section) is not used very intensively when constructing a split-plot design. This is because only factor level combinations with the same settings for the HTC factors can be swapped between whole plots. This is different from block design problems where the interchange step is more heavily used (because any pair of factor level combinations can be swapped in block designs) to escape from local optima. To the best of our knowledge, the results for the coordinate-exchange and point-exchange algorithms in Table 6 are the first written account in the optimal design literature of such large differences in solution quality between the two kinds of algorithms. Most likely, this is because most scientific papers that compare algorithms for optimal design of experiments deal with relatively small design problems, with fewer runs and factors than in the coffee cream experiment.

The $D$-criterion values in Table 6 bear on run size and orthogonality of the whole-plot design. As expected, the larger the run size, the better the values of the $D$-criterion. We also observe that the design options of type A are generally better than those of type B. The explanation for this is that the whole-plot designs for the type A options are closer to being orthogonal than the type B options. This can be verified in Tables 3-5. Another way of putting this is to say that, in scenarios of type A, we extended an orthogonal and, hence, a $D$-optimum whole-plot design with sub-plots such that the overall value of the $D$ criterion is optimized.

The difference between the designs of types A and B is less pronounced in the 63-run case than in the 56-run and 70-run cases. To acquire some insight into the reason for this small difference, we show the proportions of whole plots with a + or a − sign for each of the three whole-plot effect contrasts in Table 8. For example, the 56-run design option A has four whole plots at the low processing speed and two whole plots at the high speed. This explains the 4:2 entry for the factor speed in the first row of Table 8. The design also has three whole plots for either setting of the factor process loop, which explains the 3:3 entry for the process loop. Finally, for the contrast corresponding to the interaction between the processing speed and the factor process loop, the design has three whole plots at the low level and three whole plots at the high level. As a matter of fact, the whole-plot design in Table 3a has three whole plots on each of its diagonals. This explains the 3:3 entry for the interaction in the row corresponding to design 56 A.

From basic design theory, it is well known that, all other things being equal, an equal distribution of runs over the different levels of an effect contrast minimizes the variance with which the effect is estimated. The message of Table 8 is that the difference between the A and B options for 56 runs and for 70 runs is a switch from two equal distributions (3:3) to two unequal ones (4:2). For 63 runs, moving from A to B requires a switch from one equal distribution to an unequal one as well as a switch from an unequal distribution to an equal one. This explains why the A and B options for the 63-run experiment differ less than those for the other run sizes.

In general, the results for the $A$-criterion show that the coordinate-exchange algorithm leads to smaller variances of the parameter estimates than the point-exchange algorithm and the three-step approach. The two 70-run designs produced by the coordinate-exchange algorithm perform equally well in terms of the $A$-criterion. For the 56- and 63-run design options, the designs of type B produced by the coordinate-exchange algorithm yield smaller $A$-criterion values than the designs of type A. This is surprising given the results we obtained for the $D$-criterion and the non-orthogonality of the whole-plot designs for the scenarios 56 B, 63 B, and 70 B.

The best designs in terms of the $D$-criterion for each of the run sizes considered for the coffee cream experiment are given in Table 9. All three designs were obtained using the modified coordinate-exchange algorithm. In the table, the levels of the two HTC factors processing speed and process loop are listed first for each design. Next, the levels of the ETC factors are given, starting with the levels of the process parameter PP.

The design eventually used for the coffee cream experiment was a 63-run design obtained with the point-exchange algorithm. The reason for this was simply that the design had to be delivered in a few weeks; the research for this paper took considerably longer. Fortunately, the design actually used was 97%

Table 8: Number of whole plots for the calculation of the three whole-plot effects

| run size | scenario | speed | loop | interaction |
|---|---|---|---|---|
| 56 | A | 4:2 | 3:3 | 3:3 |
| | B | 4:2 | 4:2 | 4:2 |
| 63 | A | 3:3 | 4:2 | 3:3 |
| | B | 3:3 | 3:3 | 4:2 |
| 70 | A | 4:2 | 3:3 | 3:3 |
| | B | 4:2 | 4:2 | 4:2 |

efficient in terms of the $D$-criterion when compared to the best design found using the coordinate-exchange algorithm.

Finally, for split-plot designs whose whole-plot sizes depend on the setting of a whole-plot factor, we recommend the modified coordinate exchange algorithm in conjunction with a $D$-optimum whole-plot design.


## Acknowledgements

We are grateful to the management of FrieslandCampina, Leeuwarden, the Netherlands for their consent to disclose some details of the experimental factors. We thank Harald de Jonge of FrieslandCampina for his enthusiasm in formulating and conducting the experiment, and for his help in preparing this article.


Table 9: Best 56-, 63- and 70-run designs in terms of the $D$-criterion

| 56 runs | | 63 runs | | 70 runs | |
|---|---|---|---|---|---|
| HTC | ETC | HTC | ETC | HTC | ETC |
| -1 -1 | -1 1 1 -1 1 1 -1 | -1 -1 | 1/3 -1 -1 0 -1 -1 -1 | -1 -1 | 1/3 -1 -1 1 -1 -1 1 |
| | -1 1 -1 1 -1 -1 1 | | -1 -1 1 1 1 -1 1 | | -1 1 -1 1 1 1 1 |
| | -1 -1 -1 0 0 -1 -1 | | 1/3 1 -1 1 1 1 1 | | -1 1 -1 -1 -1 -1 -1 |
| | 1/3 1 -1 -1 -1 -1 -1 | | 1/3 -1 1 -1 1 1 -1 | | 1/3 1 1 1 1 -1 -1 |
| | -1 1 1 0 -1 -1 -1 | | 1/3 1 1 -1 -1 -1 1 | | -1 -1 1 1 -1 1 -1 |
| | -1/3 0 0 -1 1 -1 1 | | -1 1 -1 1 -1 1 -1 | | 1/3 1 1 -1 -1 1 1 |
| | 1/3 -1 1 1 0 1 1 | | -1 -1 -1 -1 -1 1 1 | | -1 -1 1 -1 1 -1 1 |
| 1 -1 | -1/3 1 1 1 -1 1 1 | 1 -1 | -1 1 1 -1 1 -1 1 | 1 -1 | 1/3 1 -1 1 1 1 1 |
| | -1 1 -1 0 1 -1 1 | | -1/3 0 0 1 -1 -1 -1 | | -1/3 -1 -1 -1 1 1 -1 |
| | -1 -1 0 1 -1 1 -1 | | -1 1 1 1 1 1 -1 | | -1 1 1 -1 -1 1 -1 |
| | 1/3 -1 1 0 1 1 -1 | | -1 1 -1 1 1 -1 -1 | | 1/3 0 1 1 -1 -1 -1 |
| | -1 1 -1 -1 -1 1 -1 | | -1 -1 -1 -1 1 1 1 | | 1/3 1 0 -1 -1 -1 1 |
| | 1/3 -1 -1 -1 1 -1 -1 | | -1 1 -1 -1 -1 -1 -1 | | -1 1 -1 1 -1 1 1 |
| | 1/3 -1 -1 1 -1 1 -1 | | 1/3 -1 -1 1 0 1 -1 | | -1 1 0 1 1 -1 -1 |
| | -1/3 -1 -1 1 0 -1 1 | | -1 -1 1 -1 1 -1 -1 | | 1/3 1 1 0 1 -1 1 |
| | -1/3 1 1 -1 1 -1 -1 | | 1/3 0 1 1 0 -1 1 | | 1/3 -1 0 0 1 -1 -1 |
| | 1/3 0 0 1 1 1 1 | | 1/3 1 -1 -1 1 1 -1 | | -1 -1 -1 1 -1 -1 -1 |
| | -1 -1 -1 -1 1 1 1 | | 1/3 -1 0 -1 -1 1 -1 | | -1 -1 1 -1 -1 1 1 |
| | 1/3 0 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 | | -1/3 1 -1 0 -1 1 1 | | 1/3 0 -1 -1 0 -1 1 |
| | 1/3 -1 0 -1 -1 1 1 | | 1/3 -1 -1 1 1 -1 1 | | 1/3 -1 1 1 0 1 -1 |
| -1 1 | -1/3 -1 1 1 1 1 -1 | -1 1 | -1/3 -1 1 -1 -1 -1 -1 | -1 1 | 1 1 -1 -1 1 -1 1 |
| | 1 -1 -1 -1 -1 -1 1 | | -1/3 1 -1 0 1 -1 1 | | -1/3 1 1 -1 1 1 -1 |
| | 1 1 0 1 -1 1 -1 | | 1 -1 -1 1 -1 -1 1 | | -1/3 1 1 1 -1 -1 1 |
| | -1/3 -1 1 -1 0 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 |
| | -1/3 -1 -1 1 1 -1 1 | | -1/3 1 1 1 -1 1 1 | | -1/3 -1 -1 -1 -1 1 1 |
| | 1 1 1 -1 1 -1 -1 | | -1/3 -1 -1 1 1 1 -1 | | -1/3 -1 -1 1 1 -1 -1 |

*continued on next page*

| 56 runs | | 63 runs | | 70 runs | |
|---|---|---|---|---|---|
| HTC | ETC | HTC | ETC | HTC | ETC |
| 1 1 | 1 -1 1 -1 1 -1 1 | 1 1 | -1/3 -1 0 1 0 -1 1 | 1 1 | -1/3 0 -1 0 -1 -1 1 |
| | -1/3 -1 -1 -1 -1 -1 1 | | 1 1 -1 1 1 1 1 | | 1 1 -1 1 -1 -1 1 |
| | -1/3 1 1 1 1 -1 1 | | 1 -1 -1 -1 0 1 1 | | 1/3 0 0 -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 |
| | -1/3 1 -1 1 -1 1 1 | | 1 1 -1 1 -1 -1 -1 | | -1/3 1 1 1 -1 1 -1 |
| | 1 -1 1 -1 -1 -1 -1 | | -1/3 0 1 1 1 -1 -1 | | -1/3 1 1 1 1 -1 1 |
| | 1 1 0 0 -1 -1 1 | | 1 -1 1 1 -1 1 -1 | | 1/3 -1 -1 1 -1 1 -1 |
| | 1 1 1 1 1 1 -1 | | -1/3 1 -1 -1 1 1 -1 | | 1 1 1 -1 -1 -1 -1 |
| | -1/3 -1 1 -1 1 -1 -1 | | -1/3 -1 1 -1 1 1 1 | | 1 1 -1 -1 1 -1 -1 |
| | -1/3 1 -1 -1 1 1 -1 | | -1/3 -1 -1 0 -1 1 -1 | | 1 -1 1 1 1 -1 -1 |
| | 1 -1 -1 1 1 -1 -1 | | 1 -1 -1 -1 1 -1 -1 | | -1/3 -1 0 -1 0 -1 -1 |
| | -1/3 1 1 -1 -1 1 -1 | | 1 0 0 0 1 1 -1 | | 1 -1 -1 0 1 1 1 |
| | -1/3 1 -1 1 -1 -1 -1 | | -1/3 0 -1 -1 -1 -1 1 | | -1/3 1 1 -1 0 1 1 |
| | 1 0 -1 -1 0 1 -1 | | 1/3 1 1 -1 0 1 -1 | | 1 -1 1 0 -1 -1 1 |
| -1 -1 | -1/3 -1 1 -1 -1 1 -1 | -1 -1 | 1/3 1 0 1 1 -1 -1 | 1 -1 | 1/3 -1 0 1 -1 1 1 |
| | -1/3 0 -1 1 1 1 -1 | | -1 1 -1 1 -1 -1 1 | | 1/3 1 -1 -1 -1 1 -1 |
| | 1/3 1 0 1 1 -1 -1 | | -1 -1 -1 -1 1 -1 -1 | | 1/3 0 -1 1 1 1 -1 |
| | -1 0 -1 0 -1 1 1 | | 1/3 -1 1 0 0 1 1 | | -1/3 -1 -1 1 1 -1 1 |
| | 1/3 1 -1 -1 1 1 1 | | -1 1 1 -1 1 1 -1 | | -1 -1 1 1 1 1 1 |
| | 1/3 1 1 -1 0 -1 1 | | -1/3 -1 1 1 -1 1 -1 | | -1/3 -1 1 -1 -1 -1 -1 |
| | -1 -1 1 1 1 -1 1 | | 1/3 -1 0 -1 1 -1 1 | | -1 1 1 1 -1 -1 1 |
| -1 1 | 1 -1 0 -1 1 1 -1 | 1 -1 | -1 0 1 -1 -1 1 -1 | | 1/3 1 -1 1 -1 -1 -1 |
| | 1/3 1 1 0 1 1 1 | | 1/3 1 1 1 -1 1 -1 | | -1 1 -1 -1 1 -1 1 |
| | -1/3 -1 1 1 -1 -1 1 | | 1/3 1 -1 -1 0 -1 1 | | 1/3 -1 1 -1 1 1 1 |
| | 1 1 -1 1 1 -1 1 | | -1/3 -1 1 -1 -1 -1 1 | | -1 1 0 -1 1 1 1 |
| | -1/3 1 -1 -1 0 -1 1 | | -1 1 1 1 -1 -1 1 | | -1 0 1 -1 1 -1 -1 |
| | 1 1 1 -1 -1 1 1 | | 1/3 -1 -1 1 -1 1 1 | | -1 -1 -1 0 0 1 -1 |
| | -1/3 -1 -1 -1 -1 1 -1 | | 1/3 1 1 -1 1 1 1 | | -1/3 1 1 1 1 1 -1 |
| | | | -1 -1 0 0 1 1 -1 | 1 1 | 1/3 1 1 0 0 -1 -1 |
| | | | -1 0 -1 1 1 1 1 | | -1/3 1 1 -1 -1 -1 1 |
| | | | 1/3 1 1 -1 1 -1 -1 | | 1 0 0 1 1 -1 1 |
| | | | -1/3 -1 1 1 1 1 1 | | -1/3 -1 -1 -1 1 -1 1 |
| | | | -1 1 0 -1 -1 1 1 | | 1 1 -1 -1 -1 1 1 |
| | | | -1 -1 -1 1 -1 -1 -1 | | -1/3 1 -1 -1 -1 1 -1 |
| | | | 1/3 -1 1 1 1 -1 -1 | | 1 1 1 -1 1 1 -1 |
| | | | | | -1/3 -1 -1 1 0 1 1 |
| | | | | | -1/3 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 |
| | | | | | -1/3 -1 1 0 1 1 -1 |
| | | | | | -1/3 -1 1 1 -1 -1 1 |

The HTC factors are processing speed and process loop; the first ETC factor is the process parameter PP; second up to fourth ETC factors have three-levels; remaining ETC factors have two-levels.

# References


Bingham, D. R., Schoen, E. D., and Sitter, R. R. (2004). Designing fractional factorial split-plot experiments with few whole-plot factors. *Journal of the Royal Statistical Society Series C* 53, 325-339. Corrigendum, 54, 955-958.

Bisgaard, S. (2000). The Design and Analysis of $2^{k-p} \times 2^{q-r}$ Split Plot Experiments. *Journal of Quality Technology* 32, 3956.

Cook, R.D., and Nachtsheim, C.J. (1989) Computer-aided blocking of factorial and response-surface designs. *Technometrics* 31, 339-346.

Goos, P. (2002). *The optimal design of blocked and split-plot experiments* (New York: Springer).

Goos, P., and Vandebroek, M. (2003) D-optimal split-plot designs with given numbers and sizes of whole plots. *Technometrics* 45, 235-245.

Goos, P., Langhans, I., and Vandebroek, M. (2006) Practical inference from industrial split-plot experiments. *Journal of Quality Technology* 38, 162-179.

Jones, B., and Goos, P. (2007) A candidate-set-free algorithm for generating D-optimal split-plot designs. *Journal of the Royal Statistical Society Series C* 56, 347-364.

Jones, B., and Nachtsheim, C. (2009) Split-plot designs: what, why and how. *Journal of Quality Technology* 41, 340-361.

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

Letsinger, J.D., Myers, R.H., and Lentner, M. (1996) Response surface methods for bi-randomization structures. *Journal of Quality Technology* 28, 381-397.

Littell, R. C., Milliken, G. A., Stroup, W. W., and Wolfinger, R. D. (1996) *SAS System for Mixed Models* (Cary: SAS Institute).

Parker, P., Kowalski, S. M., and Vining, G. G. (2007) Construction of balanced equivalent estimation second-order split-plot designs. *Technometrics* 49, 56-65.

Schoen, E.D. (1999) Designing fractional two-level experiments with nested error structures. *Journal of Applied Statistics* 26, 495-508.

Trinca, L., and Gilmour, S.G. (2001) Multistratum response surface designs. *Technometrics* 43, 25-33.

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

Webb, D. F., Lucas, J. M., and Borkowski, J. J. (2004) Factorial experiments when factor levels are not necessarily reset. *Journal of Quality Technology* 36, 1-11.