Research Article

Quality and Reliability Engineering International

(www.interscience.wiley.com) DOI: 10.1002/qre.1047

Published online 8 July 2009 in Wiley InterScience

Design and Analysis of Industrial Strip-plot Experiments

Heidi Arnouts,[a] Peter Goos[a*†] and Bradley Jones[b]

The cost of experimentation can often be reduced by forgoing complete randomization. A well-known design with restricted randomization is a split-plot design, which is commonly used in industry when some experimental factors are harder to change than others or when a two-stage production process is studied. Split-plot designs are also often used in robust product design to develop products that are insensitive to environmental or noise factors. Another, lesser known, type of experimental design plan that can be used in such situations is the strip-plot experimental design. Strip-plot designs are economically attractive in situations where the factors are hard to change and the process under investigation consists of two distinct stages, and where it is possible to apply the second stage to groups of semi-finished products from the first stage. They have a correlation structure similar to row–column designs and can be seen as special cases of split-lot designs. In this paper, we show how optimal design of experiments allows for the creation of a broad range of strip-plot designs. Copyright © 2009 John Wiley & Sons, Ltd.

Keywords: D-optimal strip-plot design; experimental cost reduction; post-fractionated strip-plot design; robust product design; row–column design; split-lot design; split-plot design; two-stage processes

1. Introduction

To reduce the cost of industrial experimentation one can forgo complete randomization. An example of a design with restricted randomization is a split-plot design. This design, which has received much attention in the recent literature on industrial experimental design, is commonly used when some experimental factors are difficult to reset due to time and/or cost constraints. The levels of these hard-to-change factors are therefore held constant for several successive runs, leading to blocks of runs at one level of each of the hard-to-change factors. In the split-plot literature, these blocks are called whole plots and the hard-to-change factors are called whole-plot factors. The remaining factors, the easy-to-change factors, are reset independently for each run. They are called subplot factors. Split-plot designs also arise naturally in certain experiments involving two process steps. In such experiments, it is often the case that batches are produced in the first process step, and these batches are split in the second step to undergo further processing. Box and Jones[1] point out that split-plot designs are suitable for many robust product experiments, where some of the factors are control or design factors and others are noise or environmental factors. The main goal of these experiments is to identify control-by-noise interaction effects. A precise estimation of these interaction effects is obtained by treating either the control factors or the noise factors as whole-plot factors. The importance of the split-plot design as well as its construction is discussed in Ganju and Lucas[2], Goos and Vandebroek[3], Vining *et al.*[4], Jones and Goos[5], and Anbari and Lucas[6].

To reduce the experimental costs in two-stage experimentation or robust product design, however, the strip-plot design is an attractive alternative to the split-plot design. As pointed out by Federer and King[7], strip-plot experimental designs are known in the literature under various names, including split-block designs, strip-block designs, two-way whole-plot designs and criss-cross designs. Strip-plot designs have been known and applied in agricultural experiments since the late 1930s, but the number of published applications in industry is fairly limited. In the next section, we describe the features of an agricultural strip-plot design. In Section 3, we discuss industrial applications of the design. Next, we describe the linear mixed model used for data from strip-plot designs. In Section 5, we sketch a number of combinatorial construction methods for strip-plot designs and show that optimal design of experiments can be used to generate a wide variety of alternative design options. Finally, in Section 6, we re-analyze the data from the battery cell experiment in Vivacqua and Bisgaard[8] and provide some new insights.

[a] *Universiteit Antwerpen, Antwerpen, Belgium*
[b] *SAS Institute Inc., Cary, NC, U.S.A.*
* *Correspondence to: Peter Goos, Faculty of Applied Economics, Universiteit Antwerpen, Prinsstraat 13, 2000 Antwerpen, Belgium.*
† *E-mail: peter.goos@ua.ac.be*

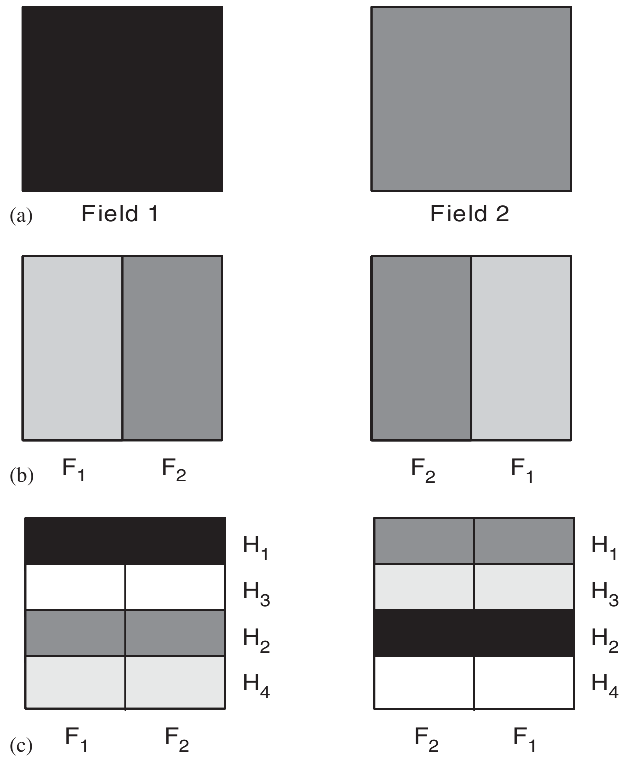

Figure 1. Strip-plot design with two replicates for studying two fertilizers (F_1 and F_2) and four herbicides (H_1–H_4): (a) two replicates; (b) random allocation of fertilizers to columns; and (c) random allocation of herbicides to rows

2. Strip-plot designs in agriculture

The first applications of strip-plot designs appeared in agricultural experimentation and involved a simple full factorial treatment design. Typically, these initial applications involved replicates of the design on λ different fields. Each field was then subdivided in ab cells arranged in a rows and b columns. The numbers a and b are the numbers of levels of the two experimental factors, A and B. The a levels of factor A were randomly applied to the rows and the b levels of factor B were randomly applied to the columns. A different random assignment was used for the different fields. The total number of observations in the resulting design is $n=\lambda ab$. A graphical illustration of the setup of a strip-plot design is given in Figure 1 for an experiment involving two fertilizers, four herbicides, and two fields. The two levels of the factor fertilizer are applied to the columns, whereas the four levels of the factor herbicide are applied to the rows. The rows and columns are sometimes referred to as strips, which explains the name of the design.

A key feature of the initial applications of strip-plot designs is that they are fully replicated on different fields. This ensures that the variance components associated with rows, columns and cells in the design are estimable, so that formal hypothesis tests can be conducted for the main effects of the two experimental factors and for the interaction effects.

3. Strip-plot designs in industrial settings

The strip-plot arrangement is not only applicable to agricultural experiments but also to industrial experiments. As the examples in this section show, it is particularly useful in situations where the process under investigation consists of two distinct stages, and where it is possible to apply the second stage to groups of semi-finished products from the first stage.

Miller[9] describes an industrial application of a strip-plot design conducted at a major manufacturer of household appliances to find methods for reducing the wrinkling of laundry. Some of the factors in the experiment were related to the configuration of the washing machines used, whereas other factors were related to the configuration of the dryers. The experiment was run in two replicates and each replicate involved four different washing machines and four different dryers. In each replicate, sets of cloth samples were run through the washing machines. Next, the samples were divided into groups such that each group contained one sample from each washer. In the following stage of the experiment, each group of samples was assigned to a dryer. Once dry, the extent of wrinkling on each sample was evaluated.

A key feature of the design is that each replicate requires only four washing machines and four dryers to be built. If the experiment had been run using a split-plot arrangement, with the washing factors as whole-plot factors and the drying factors as subplot factors, then each sample of cloth washed in the first stage of the experiment would be processed individually by a dryer. This would involve the configuration of 16 dryers, but only four washing machines. A completely randomized design would have been even more

Table I. Strip-plot design and average OCV, coded by (volts − 1.175) × 1000, for each sublot in the battery cell experiment

Assembly factors				Curing factors				
				E	−1	+1	−1	+1
A	*B*	*C*	*D*	*F*	−1	−1	+1	+1
−1	−1	−1	−1		39	40	4	4
+1	−1	−1	−1		36	33	7	7
−1	+1	−1	−1		49	48	6	10
+1	+1	−1	−1		30	28	9	4
−1	−1	+1	−1		46	41	1	10
+1	−1	+1	−1		46	50	6	10
−1	+1	+1	−1		43	45	12	11
+1	+1	+1	−1		40	43	6	2
−1	−1	−1	+1		48	55	13	14
+1	−1	−1	+1		38	38	9	12
−1	+1	−1	+1		28	35	13	29
+1	+1	−1	+1		40	36	6	5
−1	−1	+1	+1		47	53	5	17
+1	−1	+1	+1		53	52	23	9
−1	+1	+1	+1		51	52	17	23
+1	+1	+1	+1		38	37	7	14

cumbersome as it would require that each cloth sample is washed and dried individually, so that 16 washing machines and 16 dryers would have been necessary. Compared with the completely randomized design, the split-plot design thus saves costs in one stage of the experiment, whereas the strip-plot design saves costs in both stages.

Vivacqua and Bisgaard[8] describe an application of a strip-plot design at a battery manufacturer that faced problems keeping the open circuit voltage (OCV) within specification limits for one type of battery cells. The experiment focused on the battery assembly and the subsequent curing process. Four of the six experimental factors were associated with the assembly process. The two others were associated with the curing process. The design for the assembly factors was a 2^4 factorial design, and that for the curing factors was a 2^2 factorial design. These two designs were crossed, so that $2^4 \times 2^2 = 64$ responses were measured according to a complete 2^6 factorial design. For each of the 16 settings of the assembly factors, a lot of 2000 batteries was produced. Each of these lots was split into four sublots of 500 batteries. Then, all 16 sublots that were subjected to the same curing conditions were grouped and processed together. As a result, the assembly process had to be run for 16 combinations of the assembly factors and the curing process had to be run only four times. The design for the battery cell experiment is shown in Table I, along with the responses obtained. The tabular representation of the design involves rows and columns, just like the graphical representation of the agricultural strip-plot design in Figure 1. The four assembly factors are associated with the rows, whereas the two curing factors are associated with the columns.

Box and Jones[1] point out that, on top of the cost savings they generate, strip-plot arrangements also offer statistical advantages in experiments for robust product design. In robust product design, the goal is to discover how the design of a product can be modified to minimize the effect of variation from noise factors or environmental sources on the product quality. Box and Jones[1] address this issue using a cake recipe example with three control or design factors—flour, shortening and egg powder—and two noise factors—baking time and temperature. To develop a cake mix that is robust to the noise factors, it is crucial that the two-factor interaction effects between the control factors and the noise factors are estimated with maximum precision. Box and Jones[1] show that, compared to two split-plot designs (one with the control factors as whole-plot factors and one with the noise factors as whole-plot factors), a strip-plot design results in a smaller experimental cost and provides the most precise estimates of the control-by-noise interaction effects.

The industrial applications of strip-plot designs involve more factors than the traditional agricultural applications. In situations where the affordable number of runs is large enough to run a full factorial design (such as in the battery cell experiment in Table I), the construction of a strip-plot design is straightforward. However, in some scenarios, using all the factor level combinations of the full factorial design will not be possible due to cost constraints or other practical considerations. In such cases, it is important to select appropriate fractions of a factorial design. We propose to use a computerized-search approach to find D-optimal strip-plot designs, as an alternative to the combinatorial construction methods described in the literature. The use of the computerized-search approach requires the explicit specification of the model to be estimated.

4. Model

A consequence of using a strip-plot configuration is that the usual assumption of independence of observations is invalid. In the battery cell experiment, batteries that are assembled in the same lot are more alike than batteries from different lots. Batteries that are cured together are more similar than ones that are not cured together. As a result, responses within a given row of Table I are correlated, as are responses within a given column. The strip-plot model therefore contains random terms that represent variation due to the lots in which the batteries are assembled (the rows) and due to the runs of the curing process (the columns). In a strip-plot context, it is obvious that we can use the terms row factors and column factors to refer to factors that are applied to rows and columns, respectively.

If the model of interest is a main-effects-plus-two-factor-interactions model, the response in the ith row ($i=1,\ldots,r$) and the jth column ($j=1,\ldots,c$) can be written as

$$Y_{ij}=\beta_0+\sum_{k=1}^{f_1}\beta_k^R x_{ik}^R+\sum_{k=1}^{f_2}\beta_k^C x_{jk}^C+\sum_{k=1}^{f_1-1}\sum_{l=k+1}^{f_1}\beta_{kl}^R x_{ik}^R x_{il}^R + \sum_{k=1}^{f_2-1}\sum_{l=k+1}^{f_2}\beta_{kl}^C x_{jk}^C x_{jl}^C+\sum_{k=1}^{f_1}\sum_{l=1}^{f_2}\beta_{kl}^{RC} x_{ik}^R x_{jl}^C+\gamma_i+\delta_j+\varepsilon_{ij} \quad (1)$$

where f_1 and f_2 are the numbers of experimental factors in the first and second process stage (i.e. the numbers of row and column factors), x_{ik}^R is the level of the kth row factor in the ith row and x_{jk}^C is the level of the kth column factor in the jth column. The main effects of the row factors and the column factors are denoted by β_k^R and β_k^C, respectively. Finally, β_{kl}^R, β_{kl}^C and β_{kl}^{RC} are interaction effects between the kth and the lth row factor, between the kth and the lth column factor, and between the kth row factor and the lth column factor, respectively. The random effect of the ith row is denoted by γ_i, whereas the random effect of the jth column is denoted by δ_j. Finally, ε_{ij} is the random error for the response in the ith row and jth column.

In matrix notation, the strip-plot model in Equation (1) can be written as

$$\mathbf{Y}=\mathbf{X}\beta+\mathbf{Z}_\gamma\gamma+\mathbf{Z}_\delta\delta+\varepsilon \quad (2)$$

where $\mathbf{Y}$ is the $n\times 1$ vector containing the n responses of the experiment, β is a $p\times 1$ vector that contains the p model parameters and $\mathbf{X}$ is the corresponding $n\times p$ model matrix (containing the settings of all the factors and their cross-products). The matrix $\mathbf{Z}_\gamma$ is an $n\times r$ matrix with (i,j)th entry equal to 1 if the ith response was obtained in the jth row of the design, and equal to 0 otherwise. Likewise, $\mathbf{Z}_\delta$ is an $n\times c$ matrix with (i,j)th entry equal to 1 if the ith response was obtained in the jth column, and equal to 0 otherwise. Finally, γ and δ are the $r\times 1$ and $c\times 1$ vectors containing the random effects of the r rows and the c columns, respectively, and ε is the $n\times 1$ vector of the n random errors.

It is assumed that $\gamma\sim N(\mathbf{0}_r,\sigma_\gamma^2\mathbf{I}_r)$, $\delta\sim N(\mathbf{0}_c,\sigma_\delta^2\mathbf{I}_c)$, and $\varepsilon\sim N(\mathbf{0}_n,\sigma_\varepsilon^2\mathbf{I}_n)$, and that $\text{cov}(\gamma,\delta)=\mathbf{0}_{r\times c}$, $\text{cov}(\gamma,\varepsilon)=\mathbf{0}_{r\times n}$, and $\text{cov}(\delta,\varepsilon)=\mathbf{0}_{c\times n}$, where $\mathbf{0}_s$ and $\mathbf{I}_s$ represent an s-dimensional zero vector and identity matrix, respectively, and $\mathbf{0}_{s\times t}$ is a zero matrix of dimension $s\times t$. Under these assumptions, the variance–covariance matrix of the responses in $\mathbf{Y}$ is

$$\mathbf{V}=\sigma_\varepsilon^2\mathbf{I}_n+\sigma_\gamma^2\mathbf{Z}_\gamma\mathbf{Z}_\gamma'+\sigma_\delta^2\mathbf{Z}_\delta\mathbf{Z}_\delta' \quad (3)$$

Using the variance ratios $\eta_\gamma=\sigma_\gamma^2/\sigma_\varepsilon^2$ and $\eta_\delta=\sigma_\delta^2/\sigma_\varepsilon^2$ for the row factors and the column factors, respectively, this matrix can also be written as

$$\mathbf{V}=\sigma_\varepsilon^2(\mathbf{I}_n+\eta_\gamma\mathbf{Z}_\gamma\mathbf{Z}_\gamma'+\eta_\delta\mathbf{Z}_\delta\mathbf{Z}_\delta') \quad (4)$$

This variance–covariance structure of the responses is identical to that of row–column designs, which involve two crossed blocking factors. These kinds of designs are studied in Goos and Donev[10].

The parameters in β can most efficiently be estimated using the generalized least-squares estimator

$$\hat{\beta}=(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y} \quad (5)$$

the variance–covariance matrix of which can be expressed as

$$\text{cov}(\hat{\beta})=(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \quad (6)$$

The use of the generalized least-squares estimator requires the estimation of σ_γ^2, σ_δ^2 and σ_ε^2. For σ_γ^2 and σ_δ^2 to be estimable, it is required that $r>1+f_1$ and $c>1+f_2$ if the interest is in a main-effects model, and that $r>1+f_1+f_1(f_1-1)/2$ and $c>1+f_2+f_2(f_2-1)/2$ if the interest is in a main-effects-plus-interactions model. If these conditions are met, there is no need to replicate the design for the purpose of variance component estimation. The design in Table I does not satisfy the requirement that $c>1+f_2+f_2(f_2-1)/2$ so that it does not allow the estimation of σ_δ^2 if a model with interactions is fitted to the data.

The information matrix on the unknown parameter vector β is given by

$$\mathbf{M}=\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} \quad (7)$$

The information matrix forms the basis for the search for D-optimal strip-plot designs in the next section. D-optimal designs maximize the determinant of the information matrix, which is often referred to as the D-optimality criterion. It turns out that the D-optimal designs only depend on the relative magnitudes of the variance components σ_γ^2, σ_δ^2 and σ_ε^2. For the purpose of computing a D-optimal strip-plot design, it is therefore sufficient to specify the relative magnitudes of the variance components, η_γ and η_δ.

5. Construction of strip-plot designs

Several combinatorial construction methods for industrial strip-plot designs have already been described in the literature. We start this section by sketching these approaches and by explaining why a computerized-search approach for generating D-optimal strip-plot designs is useful. Then, we give several examples of D-optimal designs that cannot be constructed using the existing methodology.

5.1. Combinatorial construction methods

The literature on the combinatorial construction of strip-plot designs is less extensive than that on the construction of split-plot designs. Miller[9] suggested a general method for constructing strip-plot designs based on the use of latin-square designs. His method can be used to set up *m*-level designs as well as mixed-level fractional factorial designs, and for constructing strip-plot response surface designs. The method requires identifying a row design and a column design, both consisting of *b* blocks, and selecting a suitable latin-square fraction from the design obtained by crossing the row and column designs. The method is straightforward if an orthogonal row design and an orthogonal column design with the same number of blocks, *b*, can be found, and if these designs are orthogonally blocked. In such scenarios, a relatively simple stratum-by-stratum analysis can be done. However, it is no longer obvious what the best way is to set up a strip-plot design using Miller's method when, for example, the row factors and the column factors have unequal numbers of levels, or when the model of interest involves a second-order polynomial, because it is impossible then to find suitable row and column designs. This inspired Miller[9] to write that *'Finding a fraction of a $2^w 3^v$ design that could be used in a specified strip-plot arrangement would seem a challenging task'*.

Vivacqua and Bisgaard[11] introduce the idea of post-fractionation for two-level designs, which comes down to aliasing high-order interactions of row factors with high-order interactions of column factors to fractionate the full factorial design. This leads to more attractive confounding patterns between main effects and two-factor interactions than using a fractional factorial design for the row factors and another one for the column factors. In general, however, it will lead to larger numbers of rows and columns and, thus, to designs that are more expensive to run. Vivacqua and Bisgaard[11] also point out that there exist instances for which their approach does not allow them to find good designs and for which a computer search, similar to the one we present below, is needed.

Butler[12] focuses on two-level designs, and presents minimum aberration strip-plot designs for two-stage processes and minimum aberration split-lot designs (which are generalizations of strip-plot designs) for three- and four-stage processes. The minimum aberration property of the designs guarantees minimal confounding between main effects and two-factor interaction effects. Within the class of minimum aberration designs, Butler's designs have maximum precision for the main effects and the estimable interaction effects because they minimize the confounding of main effects and two-factor interaction effects with the sublots at each stage (i.e. with the row and columns if there are only two stages). Mee and Bates[13] discuss the construction of split-lot designs for processes involving up to nine stages.

5.2. Constructing D-optimal strip-plot designs

The combinatorial construction methods are extremely valuable and provide important insights into the problem of setting up strip-plot designs. However, each of the methods lacks flexibility in the sense that they cannot be used to design strip-plot experiments in every practical situation. As a matter of fact, many practical problems involve continuous and categorical factors with more than two levels, categorical factors acting at different numbers of levels and/or constraints on the factor levels. Moreover, the budget for experimental studies is often limited, leaving the experimenter with very few options for determining the total number of runs, the number of rows and the number of columns. For example, the feasible numbers of runs, rows and columns for two-level designs are usually not powers of two, so that the attractive combinatorial construction methods outlined above cannot be used. Also, as Miller[9] pointed out, it is often impossible to find row and columns designs that are suitable building blocks for his approach. Vivacqua and Bisgaard[11] also signal limitations to their combinatorial construction method, even in certain scenarios involving two-level factors only and numbers of runs that are powers of two.

For these reasons, we believe that there is a need for a generic approach to designing strip-plot designs. The computerized-search algorithm we have developed generates a D-optimal strip-plot design for a given experimental scenario. The input to our algorithm requires the total number of runs, rows and columns to be specified, as well as a prior point estimate of the variance ratios η_γ and η_δ.

Our algorithm for generating D-optimal designs is a modification of the coordinate-exchange algorithm which Meyer and Nachtsheim[14] proposed for completely randomized designs. The original algorithm was modified so that it can handle strip-plot designs. The resulting algorithm is similar to that by Jones and Goos[5] for D-optimal split-plot designs in that it also involves two groups of experimental factors (one group for each processing stage studied in the experiment), and to that by Jones and Goos[15] for D-optimal split-split-plot designs in that it also involves two variance ratios, η_γ and η_δ. Our algorithm differs from these two because, in a strip-plot design, the row and column factors are crossed whereas, in split-plot and split-split-plot designs, the sub-subplot and subplot factor levels are nested within the factors applied to the higher strata.

The main strength of the computerized-search approach to the design of strip-plot experiments is its generic character. However, there are also two weaknesses. First, design construction algorithms such as the coordinate-exchange algorithm and its modification for strip-plot designs cannot guarantee that a true D-optimal design will be found. In other words, it is possible that design construction algorithms produce suboptimal designs. However, the coordinate-exchange algorithm and its modifications have been shown to produce theoretically known optimal designs in many instances. In instances where suboptimal designs are produced and where the theoretically optimal designs are known, the suboptimal designs are only marginally worse than the optimal ones. The difference in performance is usually so small that it is negligible in practical applications. Because of the possibility that the

Table II. 24-run strip-plot design with four rows and eight columns to estimate a main-effects model in seven factors. The design is D-optimal for all $(\eta_\gamma, \eta_\delta) \in [0.1, 10]^2$

First stage factors		Second stage factors								
x_1^R	x_2^R									
		x_1^C	−1	+1	+1	−1	+1	−1	−1	+1
		x_2^C	−1	+1	−1	+1	+1	−1	+1	−1
		x_3^C	+1	+1	+1	+1	−1	−1	−1	−1
		x_4^C	−1	+1	+1	−1	−1	+1	+1	−1
		x_5^C	+1	−1	+1	−1	+1	−1	+1	−1
−1	+1		√	√	√	√	√	√		
+1	−1		√	√	√	√			√	√
−1	−1		√	√			√	√	√	√
+1	+1				√	√	√	√	√	√

A check mark represents a factor level combination for which a response is measured, whereas an empty cell indicates a factor level combination that is not executed.

computerized-search produces suboptimal designs, we recommend to run coordinate-exchange algorithms a large number of times to increase the probability of finding the true optimal design. Following the literature on optimal experimental design, we name the designs produced by a computerized search optimal, even though there is no guarantee that the true optimal design is found.

A second weakness of our computerized-search approach is that, in theory, the designs it produces are optimal only for one specific set of values for η_γ and η_δ. Our experience, however, suggests that, in many instances, the designs are optimal for broad ranges of values for η_γ and η_δ. This is in line with the results in Goos and Donev[10], who studied D-optimal designs for blocked experiments with the same sort of variance–covariance structure, $\mathbf{V}$, as strip-plot designs and concluded that the D-optimal designs were robust to miss-specifications of η_γ and η_δ. The robustness is twofold: (i) D-optimal designs that are optimal for one set of values of η_γ and η_δ are often also optimal for other sets of values and (ii) designs that are optimal for one set of values of η_γ and η_δ but not for another set of values are usually only marginally worse than the optimal design for that other set. Obviously, it represents good practice to verify the robustness of the optimal designs in every new strip-plot design problem by comparing optimal designs constructed assuming different values for the variance ratios. We performed such comparisons for the two D-optimal designs presented below, and it turned out that they are optimal for every practical set of values for η_γ and η_δ. This demonstrates that precise knowledge of the two variance ratios is usually not required to set up a D-optimal strip-plot design.

5.3. A D-optimal main-effects design

We start by discussing the construction of a D-optimal strip-plot design for a main-effects model in two row factors, corresponding to the first stage of the process under investigation, and five column factors, corresponding to the second stage of the process,

$$Y_{ij} = \beta_0 + \sum_{k=1}^{2} \beta_k^R x_{ik}^R + \sum_{k=1}^{5} \beta_k^C x_{jk}^C + \gamma_i + \delta_j + \varepsilon_{ij}$$

In the model, x_{ik}^R represents the level of the kth row factor in the ith row and x_{jk}^C is the level of the kth column factor in the jth column.

A simple strip-plot arrangement for this model would be obtained by crossing a 2^2 factorial design for the row factors with a 2^5 factorial design for the column factors. This would result in 128 observations, and require the first-stage process factors to be set four times and the second-stage process factors to be set 32 times. As an important reason for choosing a strip-plot design is cost, this $2^2 \times 2^5$ full factorial strip-plot design is unattractive. An economical 16-run design can be constructed using the post-fractionation method proposed by Vivacqua and Bisgaard[11]. This requires the selection of a 2^2 factorial design for the row factors and a 2^{5-2} factorial design for the column factors, and taking a half fraction of the $2^2 \times 2^{5-2}$ design obtained by crossing these two designs. It is, however, not obvious how economical strip-plot designs can be constructed using existing methods when the number of runs is not a power of two.

Using our computerized-search algorithm, it is possible to construct the D-optimal 24-run strip-plot design displayed in Table II. This strip-plot design can be implemented with four settings of the factors in the first process stage and only eight settings of the factors in the second process stage. It therefore offers a large cost reduction in comparison with the 128-run full factorial strip-plot design. A key aspect of the 24-run design is that each of the four lots produced in the first stage of the experiment is split into six sublots that are assigned to six of the eight factor settings in the second stage. An attractive feature of the D-optimal design is that it is optimal for any value of η_γ and η_δ between 0.1 and 10. This means that it is optimal for any practical value of the two variance ratios and, hence, that it is robust to miss-specification of η_γ and η_δ. Another attractive feature of the D-optimal design is that it yields a

Table III. Information matrix of the 24-run design in Table II assuming $\sigma_\gamma^2 = \sigma_\delta^2 = \sigma_\varepsilon^2 = 1$

	β_0	β_1^R	β_2^R	β_1^C	β_2^C	β_3^C	β_4^C	β_5^C
β_0	2.400	0	0	0	0	0	0	0
β_1^R	0	3.385	0	0	0	0	0	0
β_2^R	0	0	3.385	0	0	0	0	0
β_1^C	0	0	0	6.000	0	0	0	0
β_2^C	0	0	0	0	6.000	0	0	0
β_3^C	0	0	0	0	0	5.846	0	0
β_4^C	0	0	0	0	0	0	6.000	0
β_5^C	0	0	0	0	0	0	0	6.000

Table IV. A $2^4 \times 2^2$ post-fractionated strip-plot design suggested by Vivacqua and Bisgaard[11]

Assembly factors				Curing factors				
				E	−1	+1	−1	+1
A	*B*	*C*	*D*	*F*	−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		√			√

diagonal information matrix for any value of σ_γ^2, σ_δ^2 and σ_ε^2, so that all the main effects in the model can be estimated independently. The diagonal information matrix assuming $\sigma_\gamma^2 = \sigma_\delta^2 = \sigma_\varepsilon^2 = 1$ is displayed in Table III.

The D-optimal strip-plot design in Table II demonstrates the capability of our algorithm to generate attractive designs in situations involving numbers of runs that are not a power of two. Note that the row design in Table II is a simple 2^2 design and the column design is a regular quarter fraction of a 2^5 design. The quarter fraction has $x_1^C x_3^C x_4^C = +1$ and $x_2^C x_3^C x_5^C = +1$. Crossing the 2^2 and the 2^{5-2} design and dropping eight points for which $x_1^R x_2^R \prod_{i=1}^{5} x_i^C = -1$ yields the D-optimal design. The value of the computerized-search algorithm is that it assists the experimenter in selecting the row and column designs, and in choosing the best fraction of the design obtained by crossing the row and column designs.

5.4. A D-optimal main-effects-plus-two-factor-interaction-effects design

To illustrate the usefulness and flexibility of a computerized-search algorithm for a model involving main effects and two-factor interactions, we revisit the battery cell experiment described in Vivacqua and Bisgaard[8]. In that experiment, the factors *A* to *D* were associated with the assembly and the factors *E* and *F* were associated with the curing. The goal of the experiment was to estimate the main effects as well as the two-factor interaction effects of these six factors. The original design for the experiment is shown in Table I and involves 64 runs. Vivacqua and Bisgaard[11] show how a more economical design, with only 32 runs, can be obtained by selecting the half fraction with defining relation $ABCDEF = +1$ of the original design. Table IV shows the resulting design, which is especially useful if only eight sublots can be cured together.

A problem with both the original design in Table I and the fractionated design in Table IV is that there are only four columns. These four columns allow the main effects and the two-factor interaction effect of the two curing factors, *E* and *F*, to be estimated, but not the variance σ_δ^2. As a result, no formal significance test can be performed for these two main effects and the interaction effect. It is

Table V. D-optimal 48-run strip-plot design with 16 rows and six columns. The design is optimal for all $(\eta_\gamma, \eta_\delta) \in [0.1, 10]^2$

Assembly factors					Curing factors					
				E	−1	+1	−1	+1	+1	+1
A	B	C	D	F	−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		√		√	√		

Table VI. Nonzero correlations between the parameter estimates for the strip-plot design in Table V assuming $\sigma_\gamma^2 = \sigma_\delta^2 = \sigma_\varepsilon^2 = 1$

Effects	Correlation
β_1^R and β_{12}^{RC}	−0.174
β_2^R and β_{22}^{RC}	−0.174
β_3^R and β_{32}^{RC}	−0.174
β_4^R and β_{42}^{RC}	−0.174
β_1^C and β_{12}^C	−0.371
β_2^C and β_0	−0.293

therefore advisable to design a strip-plot experiment that involves more than four columns and, thus, more than four independent settings of the curing factors. One option might be to set up a 48-run design involving 16 rows, as before, and six instead of four columns. Because the number of runs and the number of columns of that design are not powers of two, it is not obvious how to construct a good strip-plot design in a combinatorial fashion. However, with our algorithmic approach, it is not difficult to construct such a design. The D-optimal design we obtained is displayed in Table V. The 16 rows contain a 2^4 factorial design, while the six columns are formed by the settings of a 2^2 factorial design two of which are duplicated. In the D-optimal design, each of the 16 batches produced using the 16 settings of the assembly factors is split into three sublots. The resulting 48 sublots are then partitioned into six groups of eight. Each of the six groups is cured together. An added value of the D-optimal design, which is optimal for any values of η_γ and η_δ between 0.1 and 10, is that it allows the estimation of σ_δ^2 so that proper significance tests can be done for the effects of the curing factors, E and F.

One drawback of the D-optimal 48-run design in Table V, compared with the original 64-run design and the fractionated 32-run design, is that it does not have a diagonal information matrix, so that not all factor effects can be estimated independently from each other. It turns out, however, that only six of the 231 pairs of parameter estimates in the main-effects-plus-two-factor-interactions model are correlated. The six non-zero correlations occur between the main-effect estimate of the first five factors and the estimate of their two-factor interaction effect with the sixth factor, and between the main effect estimate of the sixth factor and the estimate of the intercept. As a result, all main effects can be estimated independently of each other, and most interaction effects can also be estimated independently. As the largest absolute correlation is only 0.371 (see Table VI for a list of the non-zero correlations), the D-optimal design does not lead to serious inferential problems for the estimates that are correlated so that the D-optimal design is suitable for practical use.

Table VII. Parameter estimates, standard errors and *p* values for the finalmodel in the battery cell experiment

Term	Estimate	Std. Error	df	t Ratio	*p* Value
Intercept	26.2969	0.7490	2.575	36.17	0.0001
A	−2.1094	0.7490	10	−2.82	0.0183
B	−0.7656	0.7490	10	−1.02	0.3308
C	2.1406	0.7490	10	2.86	0.0170
D	2.3594	0.7490	10	3.15	0.0103
F	−16.1406	0.6428	45	−25.11	<0.0010
BF	1.4844	0.6428	45	2.31	0.0256
CF	−1.4844	0.6428	45	−2.31	0.0256
BA	−1.8594	0.7490	10	−2.48	0.0324

6. Data analysis

In this section, we revisit the battery cell experiment the data of which are shown in Table I. Vivacqua and Bisgaard[8] analyzed the data from the experiment and considered using three separate normal probability plots to decide which of the effects were significant. One plot was for the main effects of the four assembly factors and their two-, three- and four-factor interactions. Another plot was for the effects of the two curing factors and their two-factor interaction. The third plot was for the interaction effects between assembly factors and curing factors. The justification for this is that the effects are estimated in three different strata and, thus, with three different precisions. The use of normal probability plots is common in industrial experimentation where saturated or nearly saturated designs are often used. This implies that there are no degrees of freedom for estimating the error variance(s), so that no formal significance tests can be done. Based on one of the plots, Vivacqua and Bisgaard[8] designated the main effects of *A*, *C* and *D*, and the two-factor interaction effect between *A* and *B* as significant. A second plot led them to conclude that no interaction effects between assembly factors and curing factors were significant. The third plot, for the curing factors' effects, could not be used because it contained too few effects. The main effect of *F*, one of the curing factors, was declared significant because it had the largest estimate.

There are two main problems with this graphical approach. First, the interpretation of the plots is highly subjective. Second, plots that contain only a small number of factor-effect estimates are not informative at all. Schoen[16] suggests using normal probability plots only when they show a minimum of seven contrasts. Because the plot for the effects of the curing factors in the battery cell experiment contained only three effects, it cannot be used to decide which of the curing factors' effects was significant.

As three- and four-factor interactions are uncommon and as we prefer formal significance tests, we fitted a strip-plot main-effects-plus-two-factor-interactions model to the data. The final model we obtained, using stepwise backward elimination, includes the main effects of the factors *A*, *C*, *D* and *F*, and the interaction effect between *A* and *B*, which are the effects designated as significant by Vivacqua and Bisgaard[8]. Additionally, we also found that the two-factor interaction effects between *B* and *F* and between *C* and *F* are significant. Thus, we found evidence that there are potentially important interactions between two of the assembly factors and one of the curing factors. The estimates of the model coefficients, their standard errors and *p*-values are displayed in Table VII.

Before concluding, a few aspects of our analysis deserve further mention. When the full model is estimated, there are not enough degrees of freedom in the column stratum to estimate σ_δ^2. This is due to the facts that the design only has four columns and that the resulting three degrees of freedom are used up by the estimation of the main effects of the column factors *E* and *F*, and their interaction effect. Because of this, the strip-plot analysis for the full model reduces to a split-plot analysis where the assembly factors act as whole-plot factors and the curing factors act as subplot factors. In the final model, the three column degrees of freedom are used for estimating only the main effect of *F* so that there are two degrees of freedom for estimating σ_δ^2. The problem with that model is that σ_δ^2 is now estimated to be negative. The interpretation of that negative estimate is that the batteries that are cured together are more different from each other than batteries that are not cured together. This is counterintuitive. Therefore, we reran the analysis and bounded the estimate of σ_δ^2 to zero. This approach is default in many software packages. In the battery cell experiment, this also leads to a split-plot analysis with the assembly factors as whole-plot factors and the curing factors as subplot factors. The problem of negative variance component estimates is relatively common if there are only a few degrees of freedom and the true variance component is small. The problem was discussed in detail in the context of split-plot designs by Goos *et al.*[17] and Gilmour and Goos[18], but the data from the battery cell experiment show that the same kind of problem occurs in strip-plot designs. An undesirable aspect of the default settings of statistical packages in this scenario is that the degrees of freedom used for hypothesis tests involving the variance component that is estimated to be zero are much too optimistic. This is the case for the significance test for the main effect of factor *F* in Table VII, where 45 degrees of freedom are used. It would be more appropriate to utilize two degrees of freedom, because the design's four columns yield three column degrees of freedom, one of which is used for estimating the main effect of *F*. Using two degrees of freedom instead of 45, however, would still yield a *p* value smaller than 0.001.

The lesson to learn from the analysis of the data from the battery cell experiment is that, ideally, a strip-plot design has a sufficiently large number of rows and columns. Otherwise, the estimation of the model of interest and the inference becomes problematic. As larger numbers of rows and columns result in more expensive designs, requiring the estimability of all the variance components in the model for inference purposes increases the experimental cost in many situations.

7. Summary

In this paper, we have reviewed the agricultural roots of the strip-plot design and the early work on the design of industrial strip-plot experiments. To overcome the limitations of the combinatorial methods for constructing strip-plot designs, we presented a computerized-search approach that allows researchers to set up experiments where the numbers of rows, columns and observations is not a power of two. The approach is also flexible in that it is able to handle experimental design problems involving factors acting at different numbers of levels, unlike the existing methodology.

We have also described the model for strip-plot data and we have re-analyzed a data set from a battery cell experiment. Compared with an earlier published analysis, our analysis revealed extra significant effects and suggests that there is an interaction between the assembly and the curing in the production of batteries. Our data analysis also highlights that the strip-plot design that was used for the experiment had too few columns to allow formal significance tests concerning the curing factors. We therefore strongly advise experimenters to make sure that their strip-plot designs have sufficiently large numbers of rows and columns, so that a proper statistical inference is possible.

References

1. Box GEP, Jones S. Split-plot designs for robust product experimentation. *Journal of Applied Statistics* 1992; **19**:3–26.
2. Ganju J, Lucas JM. Detecting randomization restrictions caused by factors. *Journal of Statistical Planning and Inference* 1999; **81**:129–140.
3. Goos P, Vandebroek M. D-optimal split-plot designs with given numbers and sizes of whole plots. *Technometrics* 2003; **45**:235–245.
4. Vining GG, Kowalski SM, Montgomery DC. Response surface designs within a split-plot structure. *Journal of Quality Technology* 2005; **37**:115–129.
5. Jones B, Goos P. A candidate-set-free algorithm for generating D-optimal split-plot designs. *Applied Statistics* 2007; **56**:347–364.
6. Anbari FT, Lucas JM. Designing and running super-efficient experiments: Optimum blocking with one hard-to-change factor. *Journal of Quality Technology* 2008; **40**:31–45.
7. Federer WT, King F. *Variations on Split Plot and Split Block Experiment Designs*. Wiley: New York, 2007.
8. Vivacqua CA, Bisgaard S. Strip-block experiments for process improvement and robustness. *Quality Engineering* 2004; **16**:495–500.
9. Miller A. Strip-plot configurations of fractional factorials. *Technometrics* 1997; **39**:153–161.
10. Goos P, Donev AN. Blocking response surface designs. *Computational Statistics and Data Analysis* 2006; **51**:1075–1088.
11. Vivacqua CA, Bisgaard S. Post-fractionated strip-block designs. *Technometrics* 2009; **51**:47–55.
12. Butler NA. Construction of two-level split-lot fractional factorial designs for multistage processes. *Technometrics* 2004; **46**:445–451.
13. Mee RW, Bates RL. Split-lot designs: Experiments for multistage batch processes. *Technometrics* 1998; **40**:127–140.
14. Meyer RK, Nachtsheim CJ. The coordinate-exchange algorithm for constructing exact optimal experimental designs. *Technometrics* 1995; **37**:60–69.
15. Jones B, Goos P. D-optimal design of split–split-plot experiments. *Biometrika* 2009; **96**:67–82.
16. Schoen ED. Designing fractional two-level experiments with nested error structures. *Journal of Applied Statistics* 1999; **26**:495–508.
17. Goos P, Langhans I, Vandebroek M. Practical inference from industrial split-plot designs. *Journal of Quality Technology* 2006; **38**:162–179.
18. Gilmour SG, Goos P. Bayesian analysis of data from multi-stratum response surface designs. *Applied Statistics* 2009; **58**:467–484.

Authors' biographies

Heidi Arnouts is a researcher and teaching assistant in the Department of Mathematics, Statistics and Actuarial Science of the Universiteit Antwerpen. Her research deals with the design and analysis of cost-efficient experiments for industrial applications.

Dr Peter Goos is an associate professor in the Department of Mathematics, Statistics and Actuarial Science of the Universiteit Antwerpen. His work deals with the optimal design of experiments in industry and marketing. His book, 'The Optimal Design of Blocked and Split-Plot Experiments', is recommended reading for experimentalists in industry.

Dr Bradley Jones is the Director of Research & Development of the JMP Division of SAS Institute where he guides the implementation of new statistical design and analysis features of the JMP software.