DEPARTMENT OF ENVIRONMENT,
TECHNOLOGY AND TECHNOLOGY MANAGEMENT

I-optimal versus D-optimal split-plot response surface designs

Bradley Jones & Peter Goos

UNIVERSITY OF ANTWERP
Faculty of Applied Economics

Stadscampus
Prinsstraat 13, B.226
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

I-optimal versus D-optimal split-plot response surface designs

Bradley Jones & Peter Goos

RESEARCH PAPER 2012-002
JANUARY 2012

University of Antwerp, City Campus, Prinsstraat 13, B-2000 Antwerp, Belgium
Research Administration – room B.226
phone: (32) 3 265 40 32
fax: (32) 3 265 47 99
e-mail: joeri.nys@ua.ac.be

The papers can be also found at our website:
www.ua.ac.be/tew (research > working papers) &
www.repec.org/ (Research papers in economics - REPEC)

D/2012/1169/002

I-optimal versus D-optimal split-plot response surface designs

Bradley Jones
SAS Institute Inc.
Universiteit Antwerpen

Peter Goos
Universiteit Antwerpen
Erasmus Universiteit Rotterdam

Abstract

Response surface experiments often involve only quantitative factors, and the response is fit using a full quadratic model in these factors. The term response surface implies that interest in these studies is more on prediction than parameter estimation since the points on the fitted surface are predicted responses. When computing optimal designs for response surface experiments, it therefore makes sense to focus attention on the predictive capability of the designs. However, the most popular criterion for creating optimal experimental designs is the D-optimality criterion, which aims to minimize the variance of the factor-effect estimates in an omnibus sense. Because I-optimal designs minimize the average variance of prediction over the region of experimentation, their focus is clearly on prediction. Therefore, the I-optimality criterion seems to be a more appropriate one than the D-optimality criterion for generating response surface designs. Here, we introduce I-optimal design of split-plot response surface experiments. We show through several examples that I-optimal split-plot designs provide substantial benefits in terms of prediction compared to D-optimal split-plot designs, while also performing very well in terms of the precision of the factor-effect estimates.

Keywords: coordinate-exchange algorithm, D-optimality, hard-to-change factors, I-optimality, IV-optimality, multi-stratum design, split-plot design, V-optimality.

Introduction

Complete randomization of designs for industrial experiments may be difficult due to cost constraints and logistic problems. The split-plot design, which involves a restricted randomization, often provides a reasonable alternative. The main reason many experimenters avoid complete randomization in industrial experimentation is the presence of factors whose levels are hard to set or hard to change, in addition to factors whose levels are easy to set or change.

Split-plot designs were originally used for agricultural experimentation where plots of land were subdivided in relatively large portions known as *whole plots*. Each of the possible levels of the whole-plot factors were then randomly assigned to these plots. Whole plots were further divided into smaller portions known as *sub-plots*, to which sub-plot factors were applied. Thus, the levels of the whole-plot factors vary from whole plot to whole plot, while the levels of the sub-plot factors vary from sub-plot to sub-plot within each whole plot. In split-plot designs for industrial experiments, hard-to-change factors act as whole-plot factors, whereas easy-to-change factors act as sub-plot factors.

The design and analysis of split-plot industrial experiments has received considerable attention in the literature in recent years. Letsinger, Myers and Lentner (1996) discussed response surface methods for split-plot designs focusing on the data analysis. They recommended the use of generalized least squares and restricted maximum likelihood for estimating split-plot response surface models. Huang, Chen and Voelkel (1998), Bingham and Sitter (1999) and Bingham, Schoen and Sitter (2004) described the construction of two-level fractional factorial split-plot designs using the aberration criterion. Trinca and Gilmour (2001) presented a general methodology for constructing multi-stratum response surface designs, of which split-plot designs are special cases. Kulahci and Bisgaard (2005) illustrated how split-plot designs can be constructed from Plackett-Burman designs. Goos and Vandebroek (2001, 2003, 2004) and Jones and Goos (2007) propose point-exchange and coordinate-exchange algorithms for constructing D-optimal split-plot designs. Vining, Kowalski and Montgomery (2005) and Parker, Kowalski and Vining (2006, 2007a,b) presented various classes of equivalent-estimation split-plot designs for which ordinary least squares and generalized least squares estimation leads to the same point estimates. Anbari and Lucas (2008) discuss the use of full factorial two-level designs for split-plot experimentation. Follow-up split-plot designs are discussed by Almimi, Kulahci and Montgomery (2008) and McLeod and Brewster (2008). Macharia and Goos (2010) identify a range of equivalent-estimation split-plot designs that are either D-optimal or close to D-optimal. Schoen, Jones and Goos (2011) present a case study of a split-plot design for which a coordinate-exchange algorithm yields designs that are superior to those produced by a point-exchange algorithm. Reviews of the literature on the design of split-plot experiments can be found in Goos (2002, 2006) and Jones and Nachtsheim (2009).

Perhaps the most flexible approach to designing split-plot experiments is the optimal design approach of Goos and Vandebroek (2001, 2003, 2004) and Jones and Goos (2007), because that approach can handle any combination of number of runs and number of whole plots. Also, the optimal design approach can cope with various model types and different types of experimental factors (continuous factors, categorical factors and mixture factors). The emphasis in the published work on the optimal design of split-plot experiments is on the construction of D-optimal designs, which maximize the determinant of the information matrix. In this article, we demonstrate that D-optimal split-plot response surface designs perform poorly when it comes to making predictions and recommend the use of I-optimal split-plot designs, which minimize the average prediction variance. Re-

markably, the performance of the D-optimal split-plot designs in terms of the I-optimality criterion is substantially worse than the performance of the I-optimal designs in terms of the D-optimality criterion. Also, on average, the prediction-oriented I-optimal split-plot designs provide more precise factor-effect estimates than their estimation-oriented D-optimal counterparts.

In general, in the literature on the optimal design of experiments, there has been a strong focus on the generation of D-optimal designs. In recent years, however, the usefulness of prediction-oriented optimal design criteria such as the I-optimality criterion (also called V-optimality, IV-optimality or Q-optimality) and the G-optimality criterion have received increasing attention. This is because, in response surface experimentation, the goal is usually to make predictions. The generation of I-optimal completely randomized designs (which minimize the average prediction variance) is discussed in Haines (1987), Meyer and Nachtsheim (1988; 1995) and Hardin and Sloane (1993), while the generation and the performance of G-optimal completely randomized designs (which minimize the maximum prediction variance) is treated in Rodríguez, Jones, Borror and Montgomery (2010). A key finding of the latter authors is that, to minimize the maximum variance of prediction, it is often necessary to accept larger prediction variances over most of the region of interest. Hardin and Sloane (1993) demonstrated that D-optimal response surface designs perform poorly in terms of the I-optimality criterion while I-optimal designs perform reasonably well with respect to the D-optimality criterion, when the experimental region is cuboidal. When the design region is spherical, the differences between D- and I-optimal designs are less pronounced, but generally still in favor of the I-optimal designs. Goos and Jones (2011) report an example of a completely randomized response surface experiment involving a three-level categorical factor, where the performance of the I-optimal design in terms of the D-optimality criterion is much better than the performance of the D-optimal design in terms of the I-optimality criterion. These are the reasons why we prefer minimizing the average variance of prediction, and focus on I-optimal split-plot designs. This is unlike Anbari and Lucas (2008) and Goos and Lucas (2009), who use the G-optimality criterion for selecting two-level factorial and fractional factorial split-plot designs.

In this article, we first describe the model used for data from split-plot experiments and discuss the model estimation. Next, we define the D-optimality criterion as well as the I-optimality criterion, and we show how to quantify the relative performance of two designs using D- or I-efficiency. Finally, we demonstrate the benefits of using I-optimal designs using three small examples, and using a real-life protein extraction experiment. All the illustrations in this article involve a cuboidal experimental region. A discussion of our results concludes the paper.

Statistical model and analysis

The model we use for analyzing data from a split-plot experiment with b whole plots of k runs is

$$Y_{ij} = \mathbf{f}'(\mathbf{x}_{ij})\boldsymbol{\beta} + \gamma_i + \varepsilon_{ij}, \tag{1}$$

where Y_{ij} is the response measured at the jth run in the ith whole plot, $\mathbf{x}_{ij}$ is a vector that contains the levels of all the experimental factors at the jth run in the ith whole plot, $\mathbf{f}'(\mathbf{x}_{ij})$ is its model expansion, and $\boldsymbol{\beta}$ contains the intercept and all the factor effects that are in the model. The term γ_i represents the random effect of the ith whole plot and ε_{ij} is the random error associated with the jth run in whole plot i. Each whole plot corresponds to an independent setting of the hard-to-change factors, and each run or sub-plot corresponds to an independent setting of the easy-to-change factors. We denote the dimension of $\mathbf{f}'(\mathbf{x}_{ij})$ and $\boldsymbol{\beta}$ by p.

We like to stress the fact that there are two types of factors in a split-plot experiment by using two different symbols. For the N_w hard-to-change factors, we use the symbols $w_1, \ldots, w_{N_w}$ or $\mathbf{w}$. For the N_s easy-to-change factors, we use the symbols $s_1, \ldots, s_{N_s}$ or $\mathbf{s}$. The split-plot model then is

$$Y_{ij} = \mathbf{f}'(\mathbf{w}_i, \mathbf{s}_{ij})\boldsymbol{\beta} + \gamma_i + \varepsilon_{ij}, \tag{2}$$

where $\mathbf{w}_i$ gives the settings of the hard-to-change factors in the ith whole plot and $\mathbf{s}_{ij}$ gives the settings of the easy-to-change factors at the jth run within the ith whole plot. This way of writing the model is instructive as it emphasizes the fact that the levels of the hard-to-change factors do not change within a whole plot. We call the hard-to-change factors whole-plot factors because their levels are applied to whole plots, and we use the term sub-plot factors for all the easy-to-change factors because their levels are applied to sub-plots.

For a split-plot experiment with sample size n and b whole plots, the model can be written in matrix notation as

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

where $\boldsymbol{Y}$ is the vector of responses, $\mathbf{X}$ represents the $n \times p$ model matrix containing the settings of both the whole-plot factors $\mathbf{w}$ and the sub-plot factors $\mathbf{s}$ and their model expansions, $\boldsymbol{\beta}$ is again the p-dimensional vector containing the p fixed effects in the model, $\mathbf{Z}$ is an $n \times b$ matrix of zeroes and ones assigning the n runs to the b whole plots, $\boldsymbol{\gamma}$ is the b-dimensional vector containing the random effects of the b whole plots, and $\boldsymbol{\varepsilon}$ is the n-dimensional vector containing the random errors. It is commonly assumed that

$$\mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}_n \text{ and } \mathrm{cov}(\boldsymbol{\varepsilon}) = \sigma_\varepsilon^2 \mathbf{I}_n, \tag{4}$$

$$\mathrm{E}(\boldsymbol{\gamma}) = \mathbf{0}_b \text{ and } \mathrm{cov}(\boldsymbol{\gamma}) = \sigma_\gamma^2 \mathbf{I}_b, \tag{5}$$

$$\mathrm{cov}(\boldsymbol{\gamma}, \boldsymbol{\varepsilon}) = \mathbf{0}_{b \times n}. \tag{6}$$

Under these assumptions, the covariance matrix of the responses, var($\boldsymbol{Y}$), is

$$\mathbf{V} = \sigma_\varepsilon^2 \mathbf{I}_n + \sigma_\gamma^2 \mathbf{Z}\mathbf{Z}'. \tag{7}$$

When the entries of $\boldsymbol{Y}$ are arranged per whole plot, then

$$\mathbf{V} = \text{diag}(\mathbf{V}^*, \ldots, \mathbf{V}^*), \tag{8}$$

where

$$\mathbf{V}^* = \sigma_\varepsilon^2 \mathbf{I}_k + \sigma_\gamma^2 \mathbf{1}_k \mathbf{1}_k', \; = \sigma_\varepsilon^2 (\mathbf{I}_k + \eta \mathbf{1}_k \mathbf{1}_k'), \tag{9}$$

k is the number of runs in each whole plot, and the variance ratio $\eta = \sigma_\gamma^2/\sigma_\varepsilon^2$ is a measure for the extent to which responses from runs within the same whole plot are correlated. The larger η, the more the responses within one whole plot are correlated.

When the random error terms as well as the whole-plot effects are normally distributed, the maximum likelihood estimator of the unknown model parameter vector $\boldsymbol{\beta}$ is the generalized least squares (GLS) estimator

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\boldsymbol{Y}, \tag{10}$$

with covariance matrix

$$\text{var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}. \tag{11}$$

The information matrix for the parameter vector $\boldsymbol{\beta}$ is given by

$$\mathbf{M} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}. \tag{12}$$

Criteria for selecting designs

D-optimality criterion

The most commonly used criterion to select experimental designs is the D-optimality criterion which seeks designs that maximize the determinant of the information matrix,

$$|\mathbf{M}| = |\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|.$$

We use the D-efficiency to compare the quality of two designs with information matrices $\mathbf{M}_1$ and $\mathbf{M}_2$. The D-efficiency of a design with information matrix $\mathbf{M}_1$ relative to a design with information matrix $\mathbf{M}_2$ is defined as

$$\text{D-efficiency} = \left(\frac{|\mathbf{M}_1|}{|\mathbf{M}_2|} \right)^{1/p}.$$

A D-efficiency larger than one indicates that Design 1 is better than Design 2 in terms of the D-optimality criterion. The D-optimality criterion has been used for constructing split-plot designs by Goos and Vandebroek (2001, 2003, 2004), Jones and Goos (2007), Macharia and Goos (2010) and Schoen, Jones and Goos (2011). In general, the D-optimal design depends on the variance ratio η through the covariance matrix $\mathbf{V}$ of the responses, as does the D-efficiency of one design relative to another.

I-optimality criterion

An I-optimal split-plot design minimizes the average prediction variance

$$\text{Average variance } = \frac{\int_{\chi} \mathbf{f}'(\mathbf{x})(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{f}(\mathbf{x})d\mathbf{x}}{\int_{\chi} d\mathbf{x}} \tag{13}$$

over the experimental region χ. It is not difficult to calculate this expression for an arbitrary model. If there are $N = N_w + N_s$ quantitative experimental variables and the experimental region is $[-1, +1]^N$, then the volume of the experimental region in the denominator is 2^N. To simplify the calculation of the numerator, we first observe that the variance of prediction, $\mathbf{f}'(\mathbf{x})(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{f}(\mathbf{x})$, is a scalar, so that

$$\mathbf{f}'(\mathbf{x})(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{f}(\mathbf{x}) = \text{tr } \big[\mathbf{f}'(\mathbf{x})(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{f}(\mathbf{x})\big].$$

We can now exploit the property that, when calculating the trace of a matrix product, we can cyclically permute the matrices. Therefore,

$$\text{tr } \big[\mathbf{f}'(\mathbf{x})(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{f}(\mathbf{x})\big] = \text{tr } \big[(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{f}(\mathbf{x})\mathbf{f}'(\mathbf{x})\big],$$

and

$$\begin{aligned}\int_{\chi} \mathbf{f}'(\mathbf{x})(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{f}(\mathbf{x})d\mathbf{x} &= \int_{\chi} \text{tr } \big[(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{f}(\mathbf{x})\mathbf{f}'(\mathbf{x})\big]d\mathbf{x}, \\ &= \text{tr } \big[\int_{\chi} (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{f}(\mathbf{x})\mathbf{f}'(\mathbf{x})d\mathbf{x}\big].\end{aligned}$$

Now, note that, since the factor level settings are fixed, the matrix $\mathbf{X}$, and hence $(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$, is constant as far as this integration is concerned. Therefore,

$$\int_{\chi} \mathbf{f}'(\mathbf{x})(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{f}(\mathbf{x})d\mathbf{x} = \text{tr } \big[(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \int_{\chi} \mathbf{f}(\mathbf{x})\mathbf{f}'(\mathbf{x})d\mathbf{x}\big],$$

so that we can rewrite the formula for the average prediction variance as

$$\text{Average variance } = 2^{-N} \text{ tr } \big[(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \int_{\chi} \mathbf{f}(\mathbf{x})\mathbf{f}'(\mathbf{x})d\mathbf{x}\big].$$

The integral in this expression is applied to a matrix of one-term polynomials (monomials). This notation is to be interpreted as the matrix of integrals of these monomials. If the experimental region is $\chi = [-1, +1]^N$, then these integrals are quite simple. Let

$$\mathbf{B} = \int_{\mathbf{x}\in[-1,+1]^N} \mathbf{f}(\mathbf{x})\mathbf{f}'(\mathbf{x})d\mathbf{x}, \tag{14}$$

then,

$$\text{Average variance } = 2^{-N} \text{ tr } \big[(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{B}\big]. \tag{15}$$

The matrix $\mathbf{B}$ is called the moments matrix. As pointed out by Hardin and Sloane (1991b), $\mathbf{B}$ has a very specific structure for a full quadratic model:

$$\mathbf{B} = 2^N \begin{bmatrix} 1 & \mathbf{0}'_N & \mathbf{0}'_{N^*} & \frac{1}{3}\mathbf{1}'_N \\ \mathbf{0}_N & \frac{1}{3}\mathbf{I}_N & \mathbf{0}_{N\times N^*} & \mathbf{0}_{N\times N} \\ \mathbf{0}_{N^*} & \mathbf{0}_{N^*\times N} & \frac{1}{9}\mathbf{I}_{N^*} & \mathbf{0}_{N^*\times N} \\ \frac{1}{3}\mathbf{1}_N & \mathbf{0}_{N\times N} & \mathbf{0}_{N\times N^*} & \frac{1}{45}(4\mathbf{I}_N + 5\mathbf{J}_N) \end{bmatrix}, \tag{16}$$

where $N = N_w + N_s$ is the number of factors and $N^* = N(N-1)/2$ is the number of two-factor interaction effects.

If P_1 is the average variance of prediction of one design and P_2 is the average variance of prediction of a second design, then the I-efficiency of the former design compared to the latter is computed as

$$\text{I-efficiency} = P_2/P_1.$$

An I-efficiency larger than one indicates that Design 1 is better than Design 2 in terms of the average prediction variance. The use of the I-optimality criterion for constructing split-plot designs is new to the literature. Note that the I-optimal design and the I-efficiency of one design relative to another depend on the variance ratio η through the covariance matrix $\mathbf{V}$.

Dependence on the variance components

A technical problem with finding a D- or I-optimal split-plot design is that the matrix $\mathbf{V}$, and therefore also the D- and I-optimality criteria, depends on the unknown variances σ^2_γ and σ^2_ε. Fortunately, the optimal split-plot designs do not depend on the absolute magnitude of these two variances, but only on their relative magnitude. Therefore, generating optimal split-plot designs requires input only on the relative magnitude of σ^2_γ and σ^2_ε. For the purpose of generating an optimal design, an educated guess of the variance ratio is good enough because a design that is optimal for one variance ratio is also optimal for a broad range of variance ratios around the specified one. Moreover, whenever different variance ratios lead to different designs, the quality of these designs is almost identical. Goos (2002) recommends using a variance ratio of one for finding optimal split-plot designs in the absence of detailed *a priori* information about it.

In this article, we used three different η values to generate D- and I-optimal designs to study the sensitivity of the results. We report detailed results for the first three examples in the paper. These results show that the relative D- and I-efficiencies of competing designs are not very sensitive to the value of η. Therefore, for practical purposes, knowledge of the value of η is unimportant.

We computed I-optimal designs using the coordinate-exchange algorithm described in Jones and Goos (2007), where we modified the objective function to minimize the average

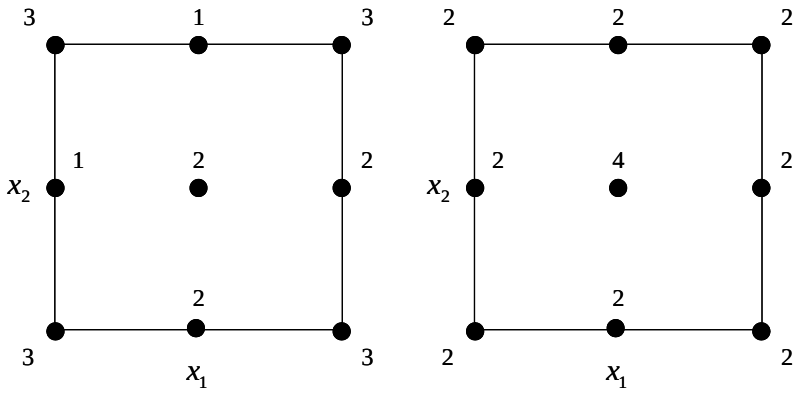
prediction variance in Equation (15), incorporating the update formulas in Arnouts and Goos (2010).

Theoretical examples

Two factors

Completely randomized designs

We first compare the D-optimal and I-optimal completely randomized designs for a full quadratic model in two factors x_1 and x_2, and restrict each design to place points on the 3×3 factorial grid. The D-optimal design places three runs at each vertex and two runs in the center. The remaining six points are allocated to the four edge midpoints. Two adjacent midpoints are duplicated, while the other two are not. The I-optimal design places two runs at each vertex, two runs at each edge midpoint, and four runs in the center of the design region. These D- and I-optimal designs are shown in Figure 1.



(a) D-optimal design (b) I-optimal design

Figure 1: D- and I-optimal 20-run completely randomized designs for estimating a full quadratic model in two factors x_1 and x_2.

The D-optimal design has an average relative variance of prediction of 0.233, while the corresponding diagnostic for the I-optimal design is 0.183. Here, our use of the word relative is with respect to σ_ε^2, which is generally unknown prior to experimentation. So, the D-optimal design has an I-efficiency of $0.183/0.233 = 78.5\%$. In most of the design region, the I-optimal design gives lower relative variances of prediction than the D-optimal design. This is clearly shown in Figure 2, where the light-shaded area represents the factor-level combinations for which the I-optimal design gives the better predictions and the dark-shaded area represents the factor-level combinations for which the D-optimal design results in more precise predictions.

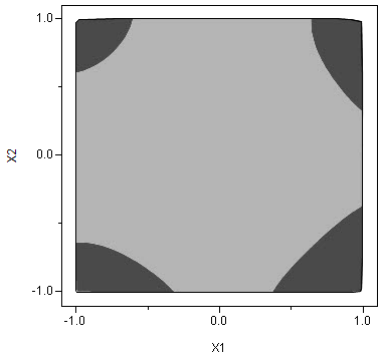


Figure 2: Plot showing the factor-level combinations for which the I-optimal (light-shaded) and D-optimal (dark-shaded) completely randomized designs from Figure 1 result in the most precise predictions.

Table 1: Relative variances of factor-effect estimates obtained from the D- and I-optimal completely randomized designs in Figure 1.

Effect	D-optimal	I-optimal
Intercept	0.302	0.179
x_1	0.068	0.083
x_2	0.068	0.083
x_1x_2	0.083	0.125
x_1^2	0.282	0.214
x_2^2	0.282	0.214

While the I-efficiency of the D-optimal design is only 78.5%, the D-efficiency of the I-optimal design is 94.9%.

Table 1 shows the relative variances of the factor-effect estimates for the full quadratic model for each design. Note that the I-optimal design estimates the quadratic effects and the intercept more precisely than the D-optimal design. The D-optimal design provides less variable estimates of the main effects and the two-factor interaction effect. This is the usual pattern whether the experiment is conducted using a completely randomized design or a split-plot design.

Split-plot designs

Suppose that logistic considerations make changing the level of the first factor very difficult, so that it is desirable to divide the 20 runs into four whole plots of five runs where, within each group of runs, the first factor's level remains constant. This restriction results in a split-plot structure for the design. Table 2 shows the four whole plots of five runs of the D-optimal split-plot design in the middle two columns, and the I-optimal split-plot

Table 2: D- and I-optimal 20-run split-plot designs in four whole plots of size five for estimating a full quadratic model in one whole-plot factor w and one sub-plot factor s.

	D-Optimal		I-Optimal	
Whole Plot	w	s	w	s
1	−1	−1	−1	−1
1	−1	−1	−1	−1
1	−1	0	−1	0
1	−1	1	−1	1
1	−1	1	−1	1
2	0	−1	0	−1
2	0	−1	0	0
2	0	0	0	0
2	0	0	0	0
2	0	1	0	1
3	1	−1	0	−1
3	1	−1	0	0
3	1	0	0	0
3	1	1	0	0
3	1	1	0	1
4	1	−1	1	−1
4	1	−1	1	−1
4	1	0	1	0
4	1	1	1	1
4	1	1	1	1

design in the final two columns. Note that, in the table, we label the first factor w to emphasize it is a whole-plot factor and that we label the sub-plot factor s to stress it is a sub-plot factor. The D-optimal design has two whole plots at the high level of the first factor, while the I-optimal design repeats the middle level of the first factor. The I-optimal design thus has a symmetric whole-plot design. In addition, within each whole plot, the sub-plot factors' design is also symmetric. The D- and the I-optimal design are optimal for η values of 0.1, 1 and 10. Note that the I-optimal design involves six center runs, while the D-optimal design has only two runs at the center of the design region.

Here, the I-optimal design is 93.4% D-efficient, independent of the η value. The average relative variance of prediction for the D-optimal design is 0.973 when η takes the value 1, compared to 0.717 for the I-optimal design. This makes the D-optimal design 73.8% I-efficient for $\eta = 1$. The I-efficiency of the D-optimal design is 75.9% when η equals 0.1 and 72.9% when η is 10. Hence, as with the completely randomized design, the I-optimal split-plot design does better with respect to the D-optimality criterion than the D-optimal design fares with the I-optimality criterion.

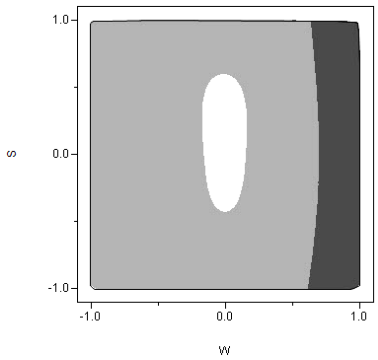


Figure 3: Plot showing the factor-level combinations for which the I-optimal (light-shaded and white areas) and D-optimal (dark-shaded area) split-plot designs from Table 2 result in the most precise predictions.

Figure 3 compares the performance of the D- and I-optimal designs in terms of prediction variance over the entire design region. The dark-shaded area, which covers about 17% of the design region, indicates the factor-level combinations for which the D-optimal design results in more precise predictions than the I-optimal design. In the remaining 83% of the design region, the I-optimal design outperforms the D-optimal one in terms of the relative variance of prediction. In 7% of the design region, indicated in white in the plot, the I-optimal design has a prediction variance that is less than half the prediction variance given by the D-optimal design. The reason why the I-optimal design predicts so well in the center of the design region is that it involves six center runs, versus two only for the D-optimal design.

Table 3 shows the relative variances of the factor-effect estimates for the full quadratic model for each design for η values of 0.1, 1 and 10. Again, note that the D-optimal design provides more precise estimates of the main effects and the two-factor interaction effects, while the I-optimal design gives more precise estimates of the quadratic effects.

Assuming an η value of 1, the average relative variance of the estimates of the main effects, the interaction effects and the quadratic effects for the D-optimal design is 0.51, compared to 0.46 for the I-optimal design. So, in that case, the I-optimal design also scores 10.2% better than the D-optimal design in terms of the average precision of the factor-effect estimates. For η values of 0.1 and 10, the I-optimal design is, respectively, 3.7% and 13.7% better than the D-optimal design in terms of the average precision of the factor-effect estimates. This is despite the fact that the D-optimality criterion is an estimation-oriented optimality criterion, while the I-optimality criterion is a prediction-oriented criterion. Note that, if we take into account the relative variance of the intercept estimate, the difference in favor of the I-optimal design is even larger.

Table 3: Relative variances of factor-effect estimates obtained from the D- and I-optimal split-plot designs in Table 2.

	$\eta = 0.1$		$\eta = 1$		$\eta = 10$	
Effect	D-Opt	I-Opt	D-Opt	I-Opt	D-Opt	I-Opt
Intercept	0.401	0.190	1.301	0.640	10.301	5.140
w	0.113	0.150	0.450	0.600	3.825	5.100
s	0.075	0.083	0.075	0.083	0.075	0.083
ws	0.092	0.125	0.092	0.125	0.092	0.125
w^2	0.427	0.340	1.665	1.240	14.040	10.240
s^2	0.279	0.250	0.279	0.250	0.279	0.250
Average (incl. intercept)	0.231	0.190	0.643	0.490	4.768	3.490
Average (excl. intercept)	0.197	0.190	0.512	0.460	3.662	3.160

Three factors

Seven whole plots of four runs

In this example, we consider a three-factor design with one whole-plot factor w and two sub-plot factors s_1 and s_2. Suppose that the investigator wishes to group the runs into seven whole plots of four runs each, so that the total number of runs is 28.

Unlike in the two-factor scenario we discussed in the previous section, the D-optimal design as well as the I-optimal design now depends on the value of the variance ratio η. In its left panel, Table 4 shows the factor-level combinations of the two D-optimal designs that we obtained. The first design, Design I, is D-optimal for η values smaller than 3.10, while the second design, Design II, is D-optimal for η values larger than 3.10. In its right panel, Table 4 shows the factor-level combinations of the two I-optimal designs that we obtained. The first of these designs, Design III, is I-optimal for η values smaller than 2.05, while the second design, Design IV, is I-optimal for η values larger than 2.05.

Each of the D-optimal designs has three whole plots each at the low and high levels of the whole-plot factor and one whole plot at the middle level. The I-optimal designs have three whole plots at the middle level and two whole plots each at the low and high levels of the whole-plot factor. The I-optimal design for small η values has six center runs in total, while the I-optimal design for large η values has four center runs. In contrast, the D-optimal designs have only one such run. This greater emphasis at the center of the design region is typical of I-optimal designs, whether they are completely randomized or split-plot designs.

The two D-optimal designs exhibit very similar behavior in terms of I-efficiency. Likewise,

Table 4: D- and I-optimal 28-run split-plot designs in seven whole plots of size four for estimating a full quadratic model in one whole-plot factor w and two sub-plot factors s_1 and s_2.

	D-Optimal Designs						I-Optimal Designs					
	DESIGN I $\eta < 3.10$			DESIGN II $\eta > 3.10$			DESIGN III $\eta < 2.05$			DESIGN IV $\eta > 2.05$		
Whole Plot	w	s_1	s_2	w	s_1	s_2	w	s_1	s_2	w	s_1	s_2
1	−1	1	−1	−1	−1	0	−1	−1	1	−1	0	1
1	−1	1	1	−1	0	1	−1	0	−1	−1	1	−1
1	−1	−1	−1	−1	1	1	−1	1	0	−1	1	0
1	−1	−1	1	−1	1	−1	−1	−1	−1	−1	−1	−1
2	−1	−1	−1	−1	−1	1	−1	1	−1	−1	1	1
2	−1	−1	1	−1	1	1	−1	1	1	−1	0	−1
2	−1	1	0	−1	0	0	−1	0	1	−1	−1	1
2	−1	0	−1	−1	−1	−1	−1	−1	0	−1	−1	0
3	−1	1	1	−1	0	−1	0	−1	1	0	0	−1
3	−1	0	1	−1	−1	−1	0	0	0	0	0	0
3	−1	1	−1	−1	−1	1	0	0	0	0	1	1
3	−1	−1	0	−1	1	0	0	0	0	0	−1	0
4	0	1	−1	0	1	−1	0	1	1	0	1	−1
4	0	0	0	0	1	1	0	0	0	0	−1	1
4	0	−1	−1	0	0	0	0	0	0	0	0	0
4	0	1	1	0	−1	−1	0	−1	−1	0	0	0
5	1	1	1	1	−1	−1	0	0	1	0	0	1
5	1	1	−1	1	−1	1	0	−1	0	0	1	0
5	1	−1	1	1	1	−1	0	0	0	0	−1	−1
5	1	−1	−1	1	1	1	0	1	−1	0	0	0
6	1	−1	0	1	0	−1	1	−1	0	1	−1	−1
6	1	1	1	1	1	0	1	0	1	1	0	0
6	1	0	1	1	1	1	1	1	1	1	1	−1
6	1	1	−1	1	−1	1	1	1	−1	1	0	1
7	1	1	0	1	0	1	1	−1	1	1	0	−1
7	1	−1	−1	1	−1	0	1	−1	−1	1	0	0
7	1	−1	1	1	1	−1	1	0	−1	1	1	1
7	1	0	−1	1	−1	−1	1	1	0	1	−1	1

Table 5: D- and I-efficiencies of the D- and I-optimal designs in Table 4 for three values of the variance ratio η.

	D-efficiency			I-efficiency		
	$\eta = 0.1$	$\eta = 1$	$\eta = 10$	$\eta = 0.1$	$\eta = 1$	$\eta = 10$
Design I (D-optimal for $\eta < 3.10$)	1.0000	1.0000	0.9997	0.5426	0.5156	0.5022
Design II (D-optimal for $\eta > 3.10$)	0.9963	0.9992	1.0000	0.5895	0.5363	0.5053
Design III (I-optimal for $\eta < 2.05$)	0.8966	0.8897	0.8869	1.0000	1.0000	0.9999
Design IV (I-optimal for $\eta > 2.05$)	0.8900	0.8906	0.8905	0.9906	0.9993	1.0000

the two I-optimal designs perform very similarly in terms of the D-optimality criterion. This can be seen from Table 5, where we display the D- and I-efficiencies of the D- and I-optimal designs for η values of 0.1, 1 and 10. The table shows that the I-efficiencies of the D-optimal designs I and II lie between 50.22% and 58.95% for the η values considered, whereas the D-efficiencies of the I-optimal designs III and IV lie between 88.69% and 89.66%. As a result, independent of the exact η value, both I-optimal designs do much better with respect to the D-optimality criterion than the two D-optimal designs perform in terms of the I-optimality criterion.

Note also that the D-efficiency of design I, which is D-optimal for small values of η, is 99.97% when η is large, and that the D-efficiency of design II, which is D-optimal for large values of η, is 99.63% when η is small. The I-efficiency of design III, which is I-optimal for small values of η, is 99.99% when η is large, and the I-efficiency of design IV, which is I-optimal for large values of η, is 99.06% when η is small. This shows that the dependence of the D- and I-optimal designs on the exact value of η is of little practical relevance: a design that is optimal for one value of η also performs well for another value of η.

Table 6 provides a comparison of the relative variances of the factor-effect estimates of the four designs for η values of 0.1, 1 and 10. For any given η value, the D-optimal designs I and II estimate the main effects and the two-factor interaction effects more precisely than the I-optimal designs III and IV. The I-optimal designs estimate the quadratic effects more precisely, as well as the intercept. This is due to the I-optimal designs' greater emphasis on the center of the design region. Finally, whether or not ignoring the variance of the intercept, the average relative variance of the parameter estimates for the I-optimal designs is smaller than that for the D-optimal designs. In the sense of having lower average relative variance for the factor-effect estimates, therefore, the I-optimal designs are outperforming the D-optimal designs in terms of precision of the parameter estimation.

Table 6: Relative variances of factor-effect estimates obtained from the D- and I-optimal designs in Table 4.

	$\eta = 0.1$				$\eta = 1$				$\eta = 10$			
	D-Optimal		I-Optimal		D-Optimal		I-Optimal		D-Optimal		I-Optimal	
Effect	I	II	III	IV	I	II	III	IV	I	II	III	IV
Intercept	0.637	0.573	0.169	0.181	1.563	1.480	0.470	0.481	10.573	10.483	3.470	3.481
w	0.058	0.060	0.088	0.090	0.208	0.210	0.313	0.315	1.708	1.710	2.563	2.565
s_1	0.045	0.048	0.061	0.064	0.046	0.049	0.063	0.065	0.046	0.049	0.064	0.066
s_2	0.045	0.048	0.061	0.056	0.046	0.049	0.063	0.058	0.046	0.049	0.064	0.059
ws_1	0.051	0.055	0.085	0.104	0.052	0.057	0.090	0.106	0.053	0.057	0.091	0.107
ws_2	0.051	0.055	0.085	0.085	0.052	0.057	0.090	0.089	0.053	0.057	0.091	0.091
s_1s_2	0.054	0.058	0.086	0.086	0.054	0.058	0.093	0.093	0.054	0.058	0.095	0.095
w^2	0.423	0.421	0.238	0.215	1.473	1.471	0.764	0.740	11.973	11.971	6.015	5.990
s_1^2	0.251	0.221	0.175	0.163	0.263	0.225	0.178	0.163	0.268	0.226	0.178	0.163
s_2^2	0.251	0.221	0.175	0.177	0.263	0.225	0.178	0.177	0.268	0.226	0.178	0.177
Avg (incl int)	0.187	0.176	0.122	0.122	0.402	0.388	0.230	0.229	2.504	2.489	1.281	1.279
Avg (excl int)	0.137	0.132	0.117	0.116	0.273	0.267	0.203	0.201	1.608	1.600	1.038	1.035

We generated 10,000 random points within the $[-1, 1]^3$ cube comprising the design region and calculated the relative variance of prediction for the D-optimal and I-optimal designs at each point. For the scenario where η is one, for example, the prediction variance produced by the first I-optimal design, labeled III in Table 4, was lower than that obtained from the first D-optimal design, labeled I in Table 4, for greater than 98% of our sample. The median ratio of the prediction variance of the D-optimal design to the prediction variance of the I-optimal design was 2.47. This means that, for half of the design region, the relative variance of prediction of the D-optimal design was at least 2.47 times as large as the relative variance of prediction for the I-optimal design. Another way of putting this is to say that the confidence intervals for means obtained from the D-optimal design are at least 1.57 times as wide as those obtained from the I-optimal design in 50% of the cases.

Other whole-plot sizes and numbers of runs

In the three-factor scenario involving seven whole plots of four runs, the D- and I-efficiencies of the different designs in Table 4 are clearly in favor of the I-optimal design criterion. In this section, we investigate to what extent the evidence in favor of the I-optimal designs depends on the number of whole plots and the number of runs within a whole plot. To this end, we computed D- and I-optimal designs for the nine scenarios defined by a 3^2 factorial design in two factors, the number of whole plots, b, and the number of runs within a whole plot, k. We studied designs with six, seven and eight whole plots, as well as designs with three, four and five runs per whole plot. In Table 7, we report the computational results of this study for an η value of one.

The table clearly shows that the D-efficiency of the I-optimal design is substantially higher than the I-efficiency of the D-optimal design in each of the scenarios. This is especially

Table 7: D-efficiency of the I-optimal design and I-efficiency of the D-optimal design for nine experimental scenarios, assuming η is one. The parameters b, k and n are the number of whole plots, the number of runs within a whole plot and the total number of runs, respectively.

b	k	n	D-efficiency	I-efficiency
6	3	18	0.8348	0.6417
6	4	24	0.8008	0.6201
6	5	30	0.8228	0.6047
7	3	21	0.8484	0.5535
7	4	28	0.8898	0.5359
7	5	35	0.9125	0.5233
8	3	24	0.8306	0.7772
8	4	32	0.8608	0.7274
8	5	40	0.8621	0.7679

so for the scenarios involving six or seven whole plots.

Four factors

In this section, we consider a four-factor split-plot design with two whole-plot factors and two sub-plot factors. We group the runs into ten whole plots of three runs each, so that the total number of runs is 30. In this scenario, we obtained a different D-optimal design for each of the η values 0.1, 1 and 10, an I-optimal design that is optimal for $\eta = 0.1$ and another design that is I-optimal for η values of 1 and 10.

We limit ourselves, however, to discussing the designs that are optimal for an η value of one for three reasons. Firstly, the differences in D-efficiency between the different D-optimal designs were smaller than 0.53%, while the differences in I-efficiency between the different I-optimal designs were smaller than 0.13%. These are differences in efficiency that have no practical significance. Secondly, the different D-optimal designs have an identical whole-plot design. Also the two I-optimal design possess an identical whole-plot design. Finally, like in the three-factor example, our conclusions regarding the D- and I-efficiencies are not sensitive to the exact D- and I-optimal designs used in the comparisons.

Table 8 shows the factor-level combinations for the D-optimal design on the left and for the I-optimal design on the right, for the case where $\eta = 1$. The ten settings for the whole-plot factors w_1 and w_2 in the D-optimal design involve the 2×2 factorial design with replicates at three of the four corners. The three remaining whole plots are at the center point of the whole-plot factors' design region and at two adjacent edge midpoints. By contrast, the I-optimal design in the whole plots is a 3×3 full factorial design, or, equivalently, a face-centered central composite design, with a replicated center point.

The I-optimal design is 88.6% D-efficient when η is one. The D-optimal design, however, is only 66.9% I-efficient. Table 9 provides a comparison of the relative variances of the factor-effect estimates for the two designs. Again, the D-optimal design estimates the main effects and the two-factor interaction effects more precisely than the I-optimal design. The I-optimal design estimates the quadratic effects more precisely, as well as the intercept. Ignoring the variance of the intercept, the average relative variance of the factor-effect estimates obtained from the I-optimal design is 0.202, compared to 0.261 for the D-optimal design. In the sense of having lower average relative variance for the factor-effect estimates, therefore, the I-optimal design is outperforming the D-optimal design in terms of parameter estimation.

We generated 10,000 random points within the four-dimensional hypercube comprising the design region and calculated the relative variance of prediction for both the D-optimal and I-optimal designs at each point. For 92% of our sample, the I-optimal design's variance was lower than the D-optimal design's variance when η equals one. The median ratio of the two variances was 1.57. This means that, for half of the design region, the relative variance of prediction of the D-optimal design was at least 57% larger than the relative variance of prediction for the I-optimal design.

Protein extraction experiment

Trinca and Gilmour (2001) describe an experiment covering 21 days to investigate the effect of five factors on protein extraction. The factors were the feed position for the inflow of a mixture, the feed flow rate, the gas flow rate, the concentration of protein A and the concentration of protein B. Three levels were used for each factor. Setting the feed position involved taking the equipment apart and then reassembling it. This was time-consuming. Therefore, the feed position was changed only between days of experimentation. This meant that two runs instead of one could be performed each day, allowing for a total of 42 experimental runs.

Trinca and Gilmour (2001) apply a general design construction method for multi-stratum designs, of which split-plot designs are a special case, to the protein extraction experiment, and obtain the design in the left panel of Table 10. The design combines an equireplicated three-level design for the whole-plot factor w and a modified face-centered central composite design for the sub-plot factors s_1, s_2, s_3 and s_4. The way Trinca and Gilmour modify the central composite design is by duplicating the center point and the axial points, and by adding a regular half fraction to the complete 2^4 factorial design. The whole-plot design and the sub-plot design are combined so that the sub-plot factors' effects and the whole-plot-by-sub-plot interaction effects are as orthogonal as possible to the whole plots. The criterion they used for combining the designs is a weighted A-optimality criterion.

Table 8: D- and I-optimal 30-run designs in ten whole plots of size three for estimating a full quadratic model in two whole-plot factors w_1 and w_2 and two sub-plot factors s_1 and s_2 when $\eta = 1$.

	D-Optimal				I-Optimal			
Whole Plot	w_1	w_2	s_1	s_2	w_1	w_2	s_1	s_2
1	−1	−1	1	−1	−1	−1	0	1
1	−1	−1	1	1	−1	−1	−1	0
1	−1	−1	−1	1	−1	−1	1	−1
2	−1	0	0	−1	−1	0	1	0
2	−1	0	1	1	−1	0	0	1
2	−1	0	−1	0	−1	0	−1	−1
3	−1	1	−1	−1	−1	1	0	−1
3	−1	1	−1	1	−1	1	−1	1
3	−1	1	1	−1	−1	1	1	1
4	−1	1	−1	−1	0	−1	−1	−1
4	−1	1	1	0	0	−1	0	0
4	−1	1	0	1	0	−1	1	1
5	0	−1	1	1	0	0	0	0
5	0	−1	0	0	0	0	0	0
5	0	−1	−1	−1	0	0	−1	1
6	0	0	0	0	0	0	0	0
6	0	0	−1	1	0	0	0	0
6	0	0	1	−1	0	0	1	−1
7	1	−1	1	−1	0	1	1	0
7	1	−1	−1	−1	0	1	0	1
7	1	−1	0	1	0	1	−1	−1
8	1	−1	1	0	1	−1	1	0
8	1	−1	−1	−1	1	−1	−1	1
8	1	−1	−1	1	1	−1	0	−1
9	1	1	0	−1	1	0	−1	−1
9	1	1	1	1	1	0	0	0
9	1	1	−1	0	1	0	1	1
10	1	1	1	−1	1	1	0	1
10	1	1	1	1	1	1	−1	0
10	1	1	−1	1	1	1	1	−1

Table 9: Relative variances of factor-effect estimates obtained from the D- and I-optimal split-plot designs in Table 8, assuming $\eta = 1$.

Effect	D-Optimal	I-Optimal
Intercept	1.129	0.526
w_1	0.196	0.223
w_2	0.196	0.223
s_1	0.044	0.058
s_2	0.044	0.057
w_1w_2	0.210	0.337
w_1s_1	0.054	0.084
w_1s_2	0.054	0.080
w_2s_1	0.055	0.084
w_2s_2	0.055	0.080
s_1s_2	0.062	0.110
w_1^2	1.091	0.581
w_2^2	1.094	0.581
s_1^2	0.251	0.158
s_2^2	0.251	0.172

Goos and Vandebroek (2003) revisit the protein extraction experiment and suggest the D-optimal split-plot design in the middle panel of Table 10 for it. It turns out that the D-efficiency of Trinca and Gilmour's design is 76.8% when η equals one. In terms of the I-optimality criterion, however, Trinca and Gilmour's design performs better. The average relative variance of prediction is 0.510 for the design of Trinca and Gilmour, whereas it is 0.655 for the D-optimal design. Hence, the I-efficiency of the D-optimal design relative to Trinca and Gilmour's design is 77.9%.

We compare the predictive performance of the two designs to that of the I-optimal design, obtained for an η value of one and displayed in the right panel of Table 10, in Figure 4, using a Fraction of Design Space (FDS) plot. In the plot, the dashed line corresponds to the D-optimal design. The dash-dotted line shows the predictive performance of the Trinca and Gilmour design. The solid line shows the predictive performance of the I-optimal design. Each point on the horizontal axis of the FDS plot, which is scaled from 0 to 1, corresponds to a fraction of the design space or the design region. The vertical axis covers the range from the minimum relative prediction variance to the maximum relative prediction variance. Suppose, for example, that the point $(0.65, 2.7)$ is on the FDS curve. Then, the relative variance of prediction is less than or equal to 2.7 over 65% of the experimental region. For a design to be good, its curve in the FDS plot should be as low as possible. This means that the design results in small prediction variances in large fractions of the design region.

Table 10: Three alternative split-plot designs for the protein extraction experiment involving 21 whole plots of size two for estimating a full quadratic model in one whole-plot factor w and four sub-plot factors s_1, s_2, s_3 and s_4.

	Trinca & Gilmour					D-Optimal					I-Optimal				
WP	w	s_1	s_2	s_3	s_4	w	s_1	s_2	s_3	s_4	w	s_1	s_2	s_3	s_4
1	−1	−1	−1	−1	−1	−1	1	−1	−1	1	−1	1	0	0	0
1	−1	1	−1	−1	1	−1	1	1	1	−1	−1	−1	1	−1	−1
2	−1	−1	1	−1	1	−1	−1	0	−1	−1	−1	0	1	1	0
2	−1	0	0	0	0	−1	1	1	−1	1	−1	−1	−1	−1	1
3	−1	−1	0	0	0	−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
4	−1	−1	1	1	−1	−1	−1	−1	1	1	−1	1	−1	−1	−1
4	−1	1	1	−1	1	−1	1	−1	−1	−1	−1	−1	0	1	1
5	−1	0	0	0	−1	−1	−1	−1	1	−1	−1	1	−1	1	1
5	−1	0	1	0	0	−1	0	1	0	0	−1	0	1	−1	1
6	−1	0	0	0	0	−1	−1	1	−1	1	−1	1	1	1	−1
6	−1	1	−1	1	−1	−1	1	−1	1	0	−1	0	−1	0	−1
7	−1	0	−1	0	0	−1	−1	−1	−1	1	0	0	0	0	0
7	−1	1	1	1	1	−1	−1	1	1	−1	0	−1	1	1	1
8	0	−1	−1	−1	1	−1	−1	1	−1	−1	0	0	0	−1	0
8	0	1	−1	1	−1	−1	1	−1	1	−1	0	1	1	1	1
9	0	−1	−1	1	−1	−1	−1	−1	0	0	0	0	0	0	0
9	0	1	−1	1	1	−1	1	0	1	1	0	1	−1	1	−1
10	0	−1	−1	1	1	0	0	0	−1	0	0	1	0	−1	1
10	0	1	−1	−1	−1	0	1	−1	0	1	0	0	1	0	0
11	0	−1	1	−1	−1	0	0	−1	0	−1	0	−1	0	0	−1
11	0	1	1	1	1	0	1	0	−1	0	0	0	−1	−1	0
12	0	−1	1	1	1	0	0	−1	1	1	0	0	1	0	−1
12	0	1	1	1	−1	0	1	1	1	−1	0	−1	0	−1	0
13	0	−1	0	0	0	1	−1	−1	1	−1	0	1	0	0	0
13	0	0	0	0	1	1	1	1	−1	−1	0	0	−1	1	1
14	0	0	0	1	0	1	−1	0	0	1	0	0	0	1	−1
14	0	1	0	0	0	1	1	1	1	1	0	−1	1	0	0
15	1	−1	−1	−1	−1	1	−1	−1	−1	−1	0	1	1	−1	−1
15	1	1	0	0	0	1	0	1	1	−1	0	0	0	0	0
16	1	−1	−1	1	1	1	−1	−1	1	1	1	−1	1	−1	1
16	1	0	1	0	0	1	1	1	−1	1	1	1	−1	0	−1
17	1	−1	1	1	−1	1	0	−1	−1	1	1	−1	0	0	1
17	1	0	0	−1	0	1	1	−1	1	0	1	1	1	−1	0
18	1	−1	1	−1	1	1	−1	1	−1	−1	1	0	−1	0	1
18	1	1	1	−1	−1	1	1	0	0	1	1	−1	1	1	−1
19	1	0	0	−1	0	1	−1	0	0	−1	1	1	0	1	1
19	1	0	−1	0	0	1	1	−1	−1	1	1	−1	−1	−1	−1
20	1	1	−1	−1	1	1	−1	1	1	1	1	−1	−1	1	0
20	1	0	0	0	−1	1	1	−1	1	−1	1	0	0	−1	−1
21	1	0	0	0	1	1	−1	1	−1	1	1	0	0	1	0
21	1	0	0	1	0	1	1	−1	−1	−1	1	1	−1	−1	1

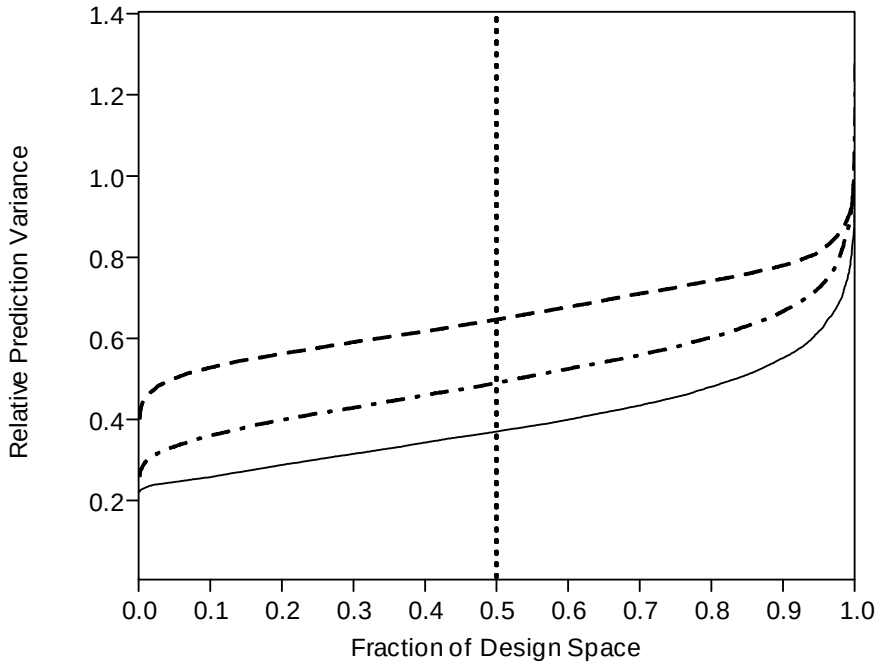


Figure 4: Fraction of Design Space (FDS) plot for the three designs in Table 10, constructed assuming $\eta = 1$. The dashed, dash-dotted and solid lines represent the D-optimal design, the Trinca and Gilmour design and the I-optimal design, respectively.

It is clear from Figure 4 that the Trinca and Gilmour design performs better than the D-optimal design in terms of prediction variance, since the dash-dotted line is substantially lower than the dashed line. However, the I-optimal design exhibits even better predictive performance. The average relative variance of prediction of the I-optimal design is 0.394. Compared to the I-optimal design, the I-efficiencies of the D-optimal design and the Trinca and Gilmour design are $0.394/0.655 = 60.2\%$ and $0.394/0.510 = 77.3\%$, respectively. The D-efficiency of the I-optimal design, relative to the D-optimal design, is 85.3%. So, once more, we observe that the I-optimal design performs better in terms of the D-optimality criterion than the D-optimal design performs in terms of the I-optimality criterion.

Various quantiles of the relative variances of prediction obtained from the three designs for the protein extraction experiment are given in Table 11. Perhaps the most striking observation from that table is that the I-optimal design gives more precise predictions over 50% of the design region than the D-optimal design does for the best possible factor-level combination. Numerically, the median relative variance of prediction is 0.373 for the I-optimal design, while the minimum relative variance of prediction is 0.404 for the D-optimal design.

In Figure 5, we show a Prediction Variance Ratio plot comparing the D-optimal to the I-optimal design. The line in the plot shows the ratio of the D-optimal prediction variance

Table 11: Minima, maxima and percentiles of the relative variance of prediction obtained from the three design in Table 10 for $\eta = 1$.

	Trinca & Gilmour	D-Optimal	I-Optimal
Minimum	0.262	0.404	0.224
1st percentile	0.301	0.461	0.236
5th percentile	0.336	0.504	0.249
1st quartile	0.418	0.579	0.304
Median	0.493	0.650	0.373
3rd quartile	0.582	0.728	0.458
95th percentile	0.729	0.817	0.622
99th percentile	0.876	0.903	0.759
Maximum	1.280	1.075	1.234

to the I-optimal prediction variance for a random sample of 10,000 factor-level combinations in the design region, $[-1, 1]^5$. When this ratio exceeds one, then the I-optimal design has lower variance. This is true for 93.2% of the design region. For half of the design region, the D-optimal design has a prediction variance that is at least 75% larger than the prediction variance of the I-optimal design. For some of the factor-level combinations in our sample, the D-optimal design results in prediction variances that are more than three times larger than those produced by the I-optimal design. The smallest ratios we obtained were around 0.7, indicating that, in the worst cases, the I-optimal design results in prediction variances that are $1/0.7 = 1.43$ times as large as those for the D-optimal design.

In Figure 6, we show a Prediction Variance Ratio plot comparing the Trinca and Gilmour design to the I-optimal design. The line in the plot shows the ratio of the prediction variance of Trinca and Gilmour's design to that of the I-optimal design. When this ratio exceeds one, then the I-optimal design has lower variance. This is true for 90.3% of the design region. For half of the design region, the Trinca and Gilmour design results in a prediction variance that is at least 32% larger than the prediction variance of the I-optimal design.

Discussion

D-optimal designs have been the most popular optimal designs because the first commercial software tools for creating optimal completely randomized designs were limited to D-optimal designs. D-optimal designs are also quite useful for screening experiments because of their focus on precise parameter estimation. Not surprisingly, the first computer generated split-plot designs were also D-optimal designs.

The theoretical properties of prediction-oriented design criteria such as G-optimality and

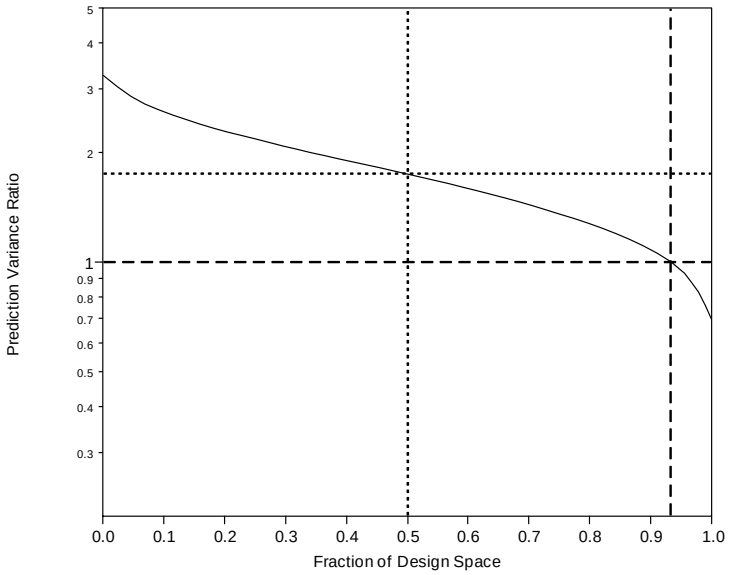


Figure 5: Prediction Variance Ratio plot comparing the D-optimal to the I-optimal design for the protein extraction experiment, constructed assuming $\eta = 1$.

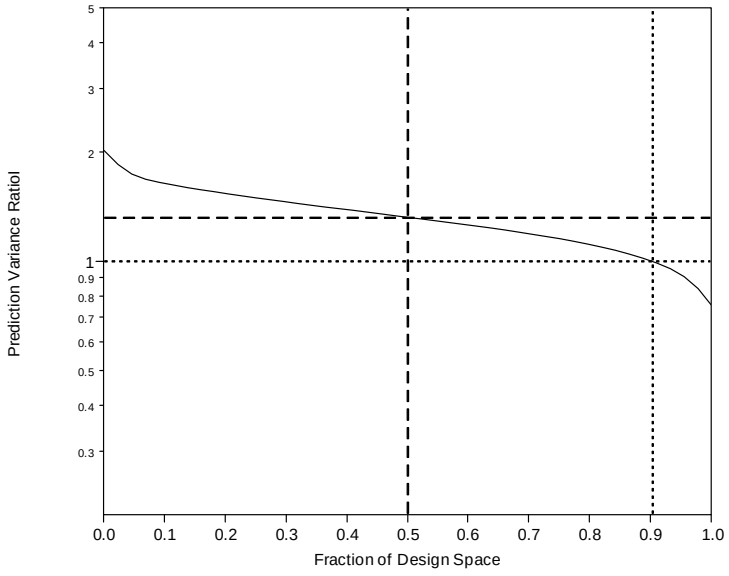


Figure 6: Prediction Variance Ratio plot comparing the Trinca and Gilmour design to the I-optimal design for the protein extraction experiment, constructed assuming $\eta = 1$.

I-optimality were recognized well before software was available to generate designs by any optimality criterion. One such theoretical property is the fact that G-optimality and D-optimality are equivalent for completely randomized designs under certain restrictive assumptions. This famous result, known as the General Equivalence Theorem, has perhaps caused some practitioners to believe that there is really not that much difference in designs created using different optimality criteria, since G-optimal designs focus on prediction and D-optimal designs focus on parameter estimation. Though G-optimal and D-optimal designs are indeed sometimes equivalent, this is far from true in general. Similarly, I-optimal designs are usually different from D-optimal designs.

Nowadays, there is more than one commercial software application capable of generating I-optimal completely randomized designs. In his popular text, Montgomery (2009) points out the potential utility of G- and I-optimal designs for response surface exploration by writing that "The G and I criteria are prediction-oriented criteria, so they would be most likely used for second-order models, as second-order models are often used for optimization, and good prediction properties are essential for optimization." So, for completely randomized designs, the use of prediction-oriented design criteria in response surface settings is recognized as appropriate.

The results we report here indicate that, for split-plot response surface designs, the choice of optimality criterion matters a great deal. Our examples show that the difference in predictive capability between D-optimal and I-optimal designs is as pronounced in the split-plot setting as in the completely randomized setting. While, in this paper, we focus entirely on cuboidal design regions, we obtained similar results for spherical design regions. The construction of I-optimal designs for spherical regions only differs from the construction for cuboidal regions in the form of the moments matrix $\mathbf{B}$. The moments matrix for spherical regions is given in Hardin and Sloane (1991a).

A recurring type of result in this paper is that the I-optimal designs outperform the D-optimal designs in terms of the A-optimality criterion, which seeks designs that minimize the trace of the covariance matrix of the parameter estimates in Equation (11), or, in other words, which seeks designs that minimize the average variance of the estimates of the model parameters. This inspired us to compute A-optimal split-plot designs for the scenarios discussed in this paper. It turns out that, in some cases, the A-optimal design is identical to the I-optimal design. In other cases, the A-optimal designs are very nearly I-optimal, and vice versa.

Finally, we also performed a small simulation study to investigate whether D- and I-optimal designs perform differently when it comes to detecting the point in the design region that provides the optimum response. It turns out that I-optimal designs do substantially better. The extent to which I-optimal designs outperform D-optimal designs strongly depends on the exact shape of the response surface, the number of runs and whole plots, and the error variances. A detailed study of this issue would be an interesting topic

for a future paper.

Because the prediction-oriented I-optimality criterion is more suitable for response surface designs, and because I-optimal designs perform well with respect to other design optimality criteria, we recommend the use of the I-optimality criterion for generating split-plot response surface designs instead of the D-optimality criterion.

Acknowledgement

The first author wishes to acknowledge the financial support of the Fonds voor Wetenschappelijk Onderzoek - Vlaanderen. Part of the work described in this article was carried out during the 2011 Design and Analysis of Experiments research programme at the Isaac Newton Institute for Mathematical Sciences at Cambridge, UK.

References

Almimi, A. A., Kulahci, M. and Montgomery, D. C. (2008). Follow-up designs to resolve confounding in split-plot experiments, *Journal of Quality Technology* **40**: 154–166.

Anbari, F. T. and Lucas, J. M. (2008). Designing and running super-efficient experiments: Optimum blocking with one hard-to-change factor, *Journal of Quality Technology* **40**: 31–45.

Arnouts, H. and Goos, P. (2010). Update formulas for split-plot and block designs, *Computational Statistics and Data Analysis* **54**: 3381–3391.

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, Ser. C (Applied Statistics)* **53**: 325–339. Corrigendum, **54**, 955-958.

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

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

Goos, P. (2006). Optimal versus orthogonal and equivalent-estimation design of blocked and split-plot experiments, *Statistica Neerlandica* **60**: 361–378.

Goos, P. and Jones, B. (2011). *Design of Experiments: A Case-Study Approach*, New York: Wiley.

Goos, P. and Lucas, J. M. (2009). Letter to the Editor, *Technometrics* **51**: 96–97.

Goos, P. and Vandebroek, M. (2001). Optimal split-plot designs, *Journal of Quality Technology* **33**: 436–450.

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. and Vandebroek, M. (2004). Outperforming completely randomized designs, *Journal of Quality Technology* **36**: 12–26.

Haines, L. M. (1987). The application of the annealing algorithm to the construction of exact optimal designs for linear-regression models, *Technometrics* **29**: 439–447.

Hardin, R. H. and Sloane, N. J. A. (1991a). Computer-generated minimal (and larger) response-surface designs: (I) The sphere. http://www2.research.att.com/∼njas/doc/doeh.pdf.

Hardin, R. H. and Sloane, N. J. A. (1991b). Computer-generated minimal (and larger) response-surface designs: (II) The cube. http://www2.research.att.com/∼njas/doc/meatball.pdf.

Hardin, R. H. and Sloane, N. J. A. (1993). A new approach to the construction of optimal designs, *Journal of Statistical Planning and Inference* **37**: 339–369.

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

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

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

Kulahci, M. and Bisgaard, S. (2005). The use of Plackett-Burman designs to construct split-plot designs, *Technometrics* **47**: 495–501.

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

Macharia, H. and Goos, P. (2010). D-optimal and D-efficient equivalent-estimation second-order split-plot designs, *Journal of Quality Technology* **42**: 358–372.

McLeod, R. G. and Brewster, J. F. (2008). Optimal foldover plans for two-level fractional factorial split-plot designs, *Journal of Quality Technology* **40**: 227–240.

Meyer, R. K. and Nachtsheim, C. J. (1988). Simulated annealing in the construction of exact optimal design of experiments, *American Journal of Mathematical and Management Science* **8**: 329–359.

Meyer, R. K. and Nachtsheim, C. J. (1995). The coordinate-exchange algorithm for constructing exact optimal experimental designs, *Technometrics* **37**: 60–69.

Montgomery, D. C. (2009). *Design and Analysis of Experiments*, New York: Wiley.

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

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

Parker, P. A., Kowalski, S. M. and Vining, G. G. (2007b). Unbalanced and minimal point equivalent estimation second-order split-plot designs, *Journal of Quality Technology* **39**: 376–388.

Rodríguez, M., Jones, B., Borror, C. M. and Montgomery, D. C. (2010). Generating and assessing exact G-optimal designs, *Journal of Quality Technology* **42**: 1–18.

Schoen, E., Jones, B. and Goos, P. (2011). A split-plot experiment with factor-dependent whole-plot sizes, *Journal of Quality Technology* **43**: 66–79.

Trinca, L. A. and Gilmour, S. G. (2001). Multi-stratum 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**: 115–128.