Improved Split-Plot and Multi-Stratum Designs

Luzia A. Trinca*
Department of Biostatistics, São Paulo State University, Botucatu, Brazil
and
Steven G. Gilmour
Southampton Statistical Sciences Research Institute, University of Southampton, UK

February 4, 2014

Abstract

Many industrial experiments involve some factors whose levels are harder to set than others. The best way to deal with these is to plan the experiment carefully as a split-plot, or more generally a multi-stratum, design. Several different approaches for constructing split-plot type response surface designs have been proposed in the literature since 2001, which has allowed experimenters to make better use of their resources by using more efficient designs than the classical balanced ones. One of these approaches, the stratum-by-stratum strategy, has been shown to produce designs that are less efficient than locally D-optimal designs. An improved stratum-by-stratum algorithm is given, which, though more computationally intensive than the old one, makes better use of the advantages of this approach, i.e. it can be used for any structure and does not depend on prior estimates of the variance components. This is shown to be almost as good as the locally optimal designs in terms of their own criteria and more robust across a range of criteria. Supplementary material is available online.

Keywords: A-optimality; D-optimality; hard-to-change factor; hard-to-set factor; mixed model; prediction variance; response surface.

*The authors gratefully acknowledge financial support from EPSRC grant EP/C541715/1: Unifying Approaches to Design of Experiments and FAPESP grant 2010/0250-08. This work was completed while the authors were participants in the Design and Analysis of Experiments programme at the Isaac Newton Institute in 2011. We thank two referees, an associate editor and the editor for helpful comments which led to improvements in the paper.

1 Introduction

Fractional factorial and response surface designs are widely used in industrial and laboratory based experiments. It has been increasingly recognized in recent years that many, perhaps most, industrial experiments and many laboratory experiments involve some factors whose levels are harder to set than others. It is clear that the best way to deal with such situations is to take account in a structured way, when designing the experiment, of the hard-to-set factors by ensuring that their levels do not have to be set for each run, but only less frequently. If there are only hard-to-set and easy-to-set factors, this leads to a (usually nonorthogonal) split-plot structure. If there are very-hard-to-set (VHS), fairly-hard-to-set (HS) and easy-to-set (ES) factors, we have a split-split-plot structure. Generally, each level of hardness-to-set in factors which is taken account of in the design defines a *stratum*, as does each level of blocking, and, following **?**, we refer to designs with factors in at least two strata as *multi-stratum* designs.

The restricted randomization in multi-stratum designs introduces additional random effects into the model. We will assume that there are s strata, with stratum i having n_i units within each unit at stratum $(i-1)$, stratum 0 being defined as the entire experiment ($n_0 = 1$). The model can then be written as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \sum_{i=1}^{s} \mathbf{Z}_i \boldsymbol{\epsilon}_i,$$

where $\mathbf{Y}$ is an $n \times 1$ random vector ($n = \prod_{j=1}^{s} n_j$), of which the observed responses $\mathbf{y}$ are assumed to be a realization, $\mathbf{X}$ is the $n \times p$ design matrix for the p-parameter treatment model, $\boldsymbol{\beta}$ is a $p \times 1$ vector of fixed treatment parameters, $\mathbf{Z}_i$ is an $n \times m_i$ indicator matrix for the units in stratum i, $m_i = \prod_{j=1}^{i} n_j$, $\boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \sigma_i^2 \mathbf{I}_{m_i})$ is an $m_i \times 1$ vector of random effects and all random effects are uncorrelated. The main aim is usually to estimate the treatment parameters $\boldsymbol{\beta}$ but, in order to estimate their standard errors, it is also necessary to estimate the variance components $\sigma_i^2,\ i = 1, \ldots, s$.

Following **?** and **?**, there is a large body of work on regular (mainly two-level) fractional factorial designs in multi-stratum structures - see **?** for recent comprehensive results. This work extends the concepts of resolution and aberration to orthogonal multi-stratum structures. The orthogonality means that all information on each effect appears in a single stratum and the parameters and their standard errors can be estimated by least squares using any standard analysis of variance program which deals with orthogonal multi-stratum structures.

Irregular fractional factorial and response surface designs require different procedures for the analysis of data, due to the nonorthogonality, which means that information on some parameters appears in more than one stratum. **?** recommended analyzing the data using residual maximum likelihood (REML) to estimate the variance components and generalized least squares (GLS) to estimate the fixed (treatment) effects. This has become accepted as the standard analysis method, although **?** showed that it can be unreliable when there are small numbers of units in the higher strata.

? and **?** studied the properties of standard response surface designs when they are run in split-plot structures, but the treatment designs were not specifically chosen to take account of the split-plot structure. The first paper to recommend choosing designs with a specific split-plot or other multi-stratum structure in mind was by **?**. They suggested a stratum-by-stratum strategy for building designs and then combining the designs from the different strata to optimize particular criteria for each step in the procedure.

? also outlined the possibility of finding a *globally* D- or A-optimum design using a modified exchange algorithm. By globally optimum we mean a design found by optimizing some property of the variance covariance matrix of the fixed effects considering all strata but with variance components fixed. They preferred the stratum-by-stratum construction because the globally optimum designs are only optimal for specific values of the ratios of variance components, whereas the stratum-by-stratum method is optimal in the situation in which obtaining informative data is most challenging, i.e. $\sigma_i^2/\sigma_j^2 \to \infty$, for all $1 \leq i < j \leq s$, and because it can be implemented using only standard designs and interchange algorithms, which are computationally less expensive than exchange algorithms.

Other authors followed up the suggestion of finding globally optimum designs for point prior estimates of the variance components in specific types of structure. In particular, **?**, **?**, **?** and Jones and Goos (2007, 2009) developed efficient exchange algorithms for split-plot response surface and mixtures designs and split-split-plot response surface designs. They found designs which, even though the search procedures depend on point prior estimates of the variance components, convincingly outperform the designs of **?** even in situations where the latter were claimed to be better. More recently **?** studied optimum split-plot designs for predicting the responses.

A different approach to split-plot response surface design, motivated by the equivalent-estimation

(EE) property, has been considered by Vining and co-authors. An equivalent-estimation design is one in which the GLS estimator of the fixed effects gives the same estimates as the ordinary least squares (OLS) estimator. **?** showed how to accommodate the treatments of central composite designs (CCDs) and Box-Benhken designs in the split-plot framework such that EE is satisfied. **?** proposed strategies for systematically constructing EE designs. However, in general, such construction methods result in very inefficient designs with respect to the usual design criteria as has been shown in **?**. In their search for globally D-optimum designs, Goos and co-authors noted many D-efficient designs also satisfy the EE property. **?** and **?** presented algorithms to select D-efficient EE split-plot designs.

The aim of the present paper is to re-examine the stratum-by-stratum strategy of **?** for design construction, to introduce a modification which is a considerable improvement and to compare the designs obtained with those resulting from existing approaches with respect to popular design criteria. For a range of design criteria see **?**.

The relative advantages of stratum-by-stratum and global construction methods are described in Section **??**. The new algorithm is described in Section **??** and examples of response surface designs are given in Section **??**. Some general recommendations are made in Section **??**.

2 Methods for Construction of Multi-Stratum Designs

In a GLS analysis, assuming that the ratios of variance components are known, the covariance matrix of the fixed effect estimators is given by $\mathbf{V}(\hat{\boldsymbol{\beta}}|\boldsymbol{\eta}) = \sigma^2(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$, where $\mathbf{V} = \sum_{i=1}^{s} \eta_i \mathbf{Z}_i \mathbf{Z}_i'$, $\boldsymbol{\eta}' = [\eta_1, \ldots, \eta_s]$, $\eta_i = \sigma_i^2/\sigma^2$ and $\sigma^2 = \sigma_s^2$. In practice, the variance components have to be estimated and the covariance matrix of the fixed effects is usually estimated by $\widehat{\mathbf{V}(\hat{\boldsymbol{\beta}})} = \hat{\sigma}^2(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}$, where $\hat{\mathbf{V}} = \sum_{i=1}^{s} \hat{\eta}_i \mathbf{Z}_i \mathbf{Z}_i'$, $\hat{\eta}_i = \hat{\sigma}_i^2/\hat{\sigma}^2$ and $\hat{\sigma}_i^2$ is usually the REML estimator of σ_i^2.

Two difficulties arise when experiments are being designed. First, neither η_i nor $\hat{\eta}_i$ are known, so we do not know $\mathbf{V}(\hat{\boldsymbol{\beta}})$ or $\widehat{\mathbf{V}(\hat{\boldsymbol{\beta}})}$ even up to the constant σ^2. Secondly, $\widehat{\mathbf{V}(\hat{\boldsymbol{\beta}})}$ is only an estimate of $\mathbf{V}(\hat{\boldsymbol{\beta}})$ which can be very poor, especially when there are few units in some strata, and can be better for some designs than for others. The global optimization and stratum-by-stratum algorithms deal with these difficulties in different ways.

The global optimization algorithms optimize some scalar function $\phi(\mathbf{X}|\boldsymbol{\eta})$ of $\mathbf{V}(\hat{\boldsymbol{\beta}}|\boldsymbol{\eta})$, such as the determinant (for D-optimality) or the trace (for A-optimality), for some point prior estimate of the ratios of variance components $\boldsymbol{\eta}$. The authors of these algorithms typically search for optimal designs for a few different values of $\boldsymbol{\eta}$ and then examine ϕ as a function of $\boldsymbol{\eta}$ for the different optimal designs found. They then choose one which is optimal across a wide range of values of $\boldsymbol{\eta}$ in the expectation that this design will perform well. Conceptually, it would be a small step to use a prior distribution for $\boldsymbol{\eta}$ and find a design which is optimal integrated across this prior. However, this is computationally expensive and not usually regarded as being worthwhile.

The stratum-by-stratum algorithm, on the other hand, takes a minimax approach and aims to optimize the information in stratum i when $\sigma_{i-1}^2/\sigma_i^2 \to \infty$. The justification for this approach is that we are ensuring that the design is optimal for each stratum in the situation in which the higher order variance components are large, which makes it most difficult to obtain useful information from the experiment. This is easiest to see in a two-stratum, i.e. split-plot, structure. If η_1 is large, the variances of parameters estimated in the whole plots stratum will be very large compared with the variances of the parameters estimated in the subplots stratum; the variances of the parameters estimated in the subplots stratum will be essentially identical to those obtained by treating the whole plots as fixed block effects. By concentrating first on the whole plots stratum, we get as good precision as possible for the parameter estimates which will inevitably have highest variance. If η_1 is so large that, despite this good design, we get no useful information from the whole plots stratum, then we have chosen a design in the subplots stratum which is optimal for the parameters estimated in this stratum, with respect to the fixed block effects. If η_1 is small, on the other hand, the variances of the parameter estimates in the whole plots stratum will be typically only slightly bigger than the variances of the parameter estimates in the subplots stratum. The philosophy of stratum-by-stratum construction is that it is better to ensure that we get variances as small as possible in the case that they are very large and accept that, when they are small, it might have been possible to make them smaller.

3 An Improved Stratum-By-Stratum Method

The algorithm of **?** did not implement the stratum-by-stratum construction in the simplest or most effective way. Motivated by computational efficiency, they chose the treatment combinations in each stratum separately, usually based on central composite, subset (**?**) or other classical designs, and then arranged them in blocks using interchange algorithms. A second interchange algorithm was then used to match the designs from neighboring strata and then a third to adjust the design for even higher strata. By using only interchange, rather than exchange, algorithms, the method was very fast and could deal with very large problems where other methods struggled. However, Goos and his co-workers have shown that the designs obtained are often quite inefficient. In this section, we describe an improved procedure, which makes use of exchange algorithms, either point exchange or coordinate exchange. Given the increased computing power in the last fifteen years, it is now possible to easily realize the full benefits of the stratum-by-stratum approach. The choice between a point exchange and a coordinate exchange algorithm is not clear cut. Given enough time, experimenters should probably try both. Our examples were found using a point exchange algorithm. We later tried Examples 1-4 with coordinate exchange algorithms, but did not find any improved designs. The coordinate exchange is most likely to be better in experiments with more factors than our examples.

Consider the general multi-stratum unit structure with s strata, where there may or may not be treatment factors to be applied in any particular stratum. Let f_i be number of factors to be applied to stratum i and p_i the number of parameters to be estimated in that stratum. We construct designs from the highest stratum to the lowest. For the highest stratum i ($i \in \{1,\ 2,\ \ldots,\ s\}$) for which there are factors to be applied, proceed as follows:

1. If $i = 1$ choose the treatment design for the factors to be applied to the units in stratum i considering the efficiency for estimating the model parameters involving the factors in this stratum only. Otherwise treat the units in stratum $i-1$ as blocks with fixed effects. Choose the treatments for the factors to be applied in this stratum and their blocking arrangement considering the efficiency for estimating the model parameters involving the factors in this stratum only.

2. Set $i = i + 1$. Maintaining the design chosen in the last step, treat the units in $i - 1$ as blocks with fixed effects. Choose the treatment combinations and their arrangement in the units in stratum i considering the efficiency for estimating the model parameters involving the factors in this stratum and the interactions between the factors in this stratum and the factors in higher strata.

3. If $i > 2$ rearrange the blocks just created within the units of stratum $i-2$, exchanging only

between stratum $i-2$ units with the same treatment, such that the efficiency of parameters estimated in stratum i is maximized when we treat these stratum $i-2$ units as blocks. Repeat this step for the units in strata $i-3, \ldots, 1$.

4. If $i = s$ stop; otherwise repeat 2-4, always considering efficiency for estimating parameters in stratum i and interactions between the factors in the current stratum and all higher strata.

The main modification of the new method from that of **?** is that the treatment set at each stage is not chosen independently of the structures formed in the previous stage. Thus we use a candidate treatment set for each stratum. The treatments that are actually chosen in each stratum are optimized by an exchange algorithm rather than an interchange algorithm. Simultaneous optimization of treatments and their blocking arrangement is performed. The method can be used for any design criteria based on the variance matrix for blocked designs with fixed number and sizes of blocks. Any algorithm for blocked designs can be used with slight modification of construction of the model matrix in each step and, in particular, a candidate set free coordinate exchange might sometimes find better designs. We will refer to this method as the MSS (modified stratum-by-stratum) approach.

In the illustrations in the next section we used D_S- and A_S-optimality criteria, the intercept and block effects being considered as nuisance parameters. Note that the D_S and D criteria give an identical ordering of designs in this setup, so that our designs are comparable with D-optimal designs in the literature. However, D-efficiencies are not the same as D_S-efficiencies and we use the latter. Following **?**, for second order models we used A_S on a scale such that the relative weights are 1/4 for each quadratic effect and 1 for other effects, whenever the design region is a hypercube. With this weight pattern we bring the different effects to the same scale. An unblocked design will be needed only when there are factors to be applied to the units in stratum 1. Let $\boldsymbol{\beta}_i$ be the model parameter vector ($p_i - 1$ parameters, excluding the intercept) to be estimated in stratum i. Let $\mathbf{X}_i$ be the $m_i \times (p_i - 1)$ associated model matrix where m_i is the number of units in this stratum. The partition of interest of the variance covariance matrix of $\hat{\boldsymbol{\beta}}_i$ is $(\mathbf{M}_i^{-1})_{22} = (\mathbf{X}_i'\mathbf{Q}_i\mathbf{X}_i)^{-1}$. For unblocked structures, $\mathbf{Q}_i = \mathbf{I} - \frac{1}{m_i}\mathbf{1}\mathbf{1}'$ while for blocked structures $\mathbf{Q}_i = \mathbf{I} - \mathbf{B}_i(\mathbf{B}_i'\mathbf{B}_i)^{-1}\mathbf{B}_i'$ with $\mathbf{B}_i$ being a $m_i \times m_{i-1}$ indicator matrix for blocks in stratum i. Thus for D_S we minimise $|(\mathbf{M}_i^{-1})_{22}|$ and for A_S-optimality we minimise $\mathsf{trace}\{\mathbf{W}_i(\mathbf{M}_i^{-1})_{22}\}$ where $\mathbf{W}_i$ is a diagonal matrix with the weights scaled so that $\mathsf{trace}(\mathbf{W}_i) = 1$.

4 Examples

In this section we present several illustrations comparing designs constructed by the MSS approach and other existing methods. For constructing the designs, for each stratum, in general, the candidate treatment set was the full 3-level factorial or 2-level factorial, depending on the underlying model. In some cases, the designs are compared with respect to properties of the variance-covariance matrix of the GLS estimator, $\hat{\boldsymbol{\beta}}$, $(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$ for a range of $\boldsymbol{\eta}$ values. A_S values are calculated as $\mathsf{trace}[\mathbf{W}\{(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\}_{22}]$ in which $\mathbf{W}$ is a diagonal matrix of weights as in Section **??** re-scaled such that $\mathsf{trace}(\mathbf{W}) = 1$. D_S values are calculated as $|\{(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\}_{22}|^{\frac{1}{p-1}}$ (eliminating the row and column relating to the intercept) such that for both properties the smaller the criterion function value the better the designs. We find it useful to show differences between designs on a variance scale, rather than a relative efficiency scale, since it is variances which are important in practice. One design might be only 50% efficient with respect to another, but if they both give very small variances, this is unimportant; conversely, one design might have only slightly less than 100% efficiency relative to another, but if they both give very high variances, the difference could still be important in practice. However, for the sake of quick comparisons we also show the efficiencies of alternative designs calculated with respect to the globally D-optimal design, as best known from the literature, which is used as a baseline. The efficiency of one particular design is defined as the ratio between the criterion value (as defined above) of the baseline design and the particular design, the larger the ratio the more efficient the design is compared with the baseline.

Since we find it useful to study several properties of a given design before recommendation to the experimenter we also compare all the designs with respect to their prediction performances. The prediction performance is evaluated by the average or integrated variance of the estimated mean response, the I-efficiency, sometimes called IV- or V-efficiency, and by the integrated variances of the estimated differences of responses, across the design region. For a multi-stratum design involving a total of q factors, the average variance is proportional to

$$\text{Average Variance} \propto \frac{\int_{\mathbf{x}\in\mathcal{X}} \mathbf{f}'(\mathbf{x})(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{f}(\mathbf{x})d\mathbf{x}}{\int_{\mathbf{x}\in\mathcal{X}} d\mathbf{x}}, \tag{1}$$

where $\mathcal{X} \subset \mathbb{R}^q$ is the experimental region of interest and $\mathbf{f}(\mathbf{x})$ is the model expansion of $\mathbf{x}$, the combination of the levels of the q factors. The numerator of (**??**) can be simplified to

$\mathsf{trace}\{\mathcal{M}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\}$ where $\mathcal{M} = \int_{\mathbf{x}\in\mathcal{X}} \mathbf{f}(\mathbf{x})\mathbf{f}'(\mathbf{x})d\mathbf{x}$ is the region moment matrix of the region of interest. For spherical and cuboidal regions the calculations of the integrals are exact (**?**).

Difference variance dispersion graphs were suggested by **?** based on the argument that often differences in response from some particular point, such as the expected position of the optimum or standard operating conditions, are more important than the response itself. Here we apply the concept of integrated variance for the difference between the estimated mean response at each point in the design region and the estimated mean response at the center of the region, since this is often the best guess of the optimum conditions at the design stage. We define the I_D, the integrated variance for differences, criterion function by

$$\text{Average Difference Variance} = \frac{\int_{\mathbf{x}\in\mathcal{X}} \mathsf{var}(\hat{y}(\mathbf{x}) - \hat{y}(\mathbf{0}))d\mathbf{x}}{\int_{\mathbf{x}\in\mathcal{X}} d\mathbf{x}},$$

which is proportional to

$$\frac{\int_{\mathbf{x}\in\mathcal{X}}[\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{0})]'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}[\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{0})]d\mathbf{x}}{\int_{\mathbf{x}\in\mathcal{X}} d\mathbf{x}} = \frac{\mathsf{trace}\{\mathcal{M}_0(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\}}{\int_{\mathbf{x}\in\mathcal{X}} d\mathbf{x}}, \tag{2}$$

where $\mathcal{M}_0 = \int_{\mathbf{x}\in\mathcal{X}}[\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{0})][\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{0})]'d\mathbf{x}$ and $\mathbf{f}(\mathbf{0})$ is the vector whose first element is one and all others are zero. The designs are compared with respect to these properties by using the relative efficiency, the baseline being the best design with respect to the property in question.

4.1 Example 1 (1 HS and 4 ES factors, 21 whole plots with 2 subplots each)

This example was described in **?** and served as motivation for some other publications. Five factors were to be investigated in an experiment on protein extraction from a mixture of two types of proteins and other components. The runs were to be executed sequentially and one of the factors, the feed position for the inflow of the mixture, was hard to set (HS). Fixing it for a day, two runs could be done per day and 21 days experimentation were considered reasonable. The primary model proposed was the second order polynomial. We compare five designs for this problem, three of them previously proposed, one by **?**, referred to as SS, one by **?**, referred to as D, a D-optimum design for $1 \leq \eta \leq 10$, and the third by **?**, referred to as I, an I-optimum design for $\eta = 1$. The two other designs were constructed by the approach proposed in this paper, MSS_A (using A_S) and MSS_D (using D_S). Our design construction follows two steps:

Table 1: Prediction performances for alternative designs for Example 1

η	Efficiency	Design				
		SS	D	I	MSS$_A$	MSS$_D$
1	I	77.18	60.14	100.00	70.58	67.35
1	I_D	75.57	80.75	100.00	84.61	84.50
10	I	82.50	53.33	100.00	83.80	81.66
10	I_D	83.51	74.23	100.00	94.84	94.68
20	I	85.37	51.92	100.00	86.71	85.28
20	I_D	88.60	72.27	100.00	97.02	96.86

Figure 1: A_S and D_S values, as functions of η, for alternative designs for Example 1. Inset: Efficiencies, as described at the start of Section **??**.

1. Find an unblocked design with 21 units and the quadratic model for 1 HS factor. The best design according to either criterion is obviously level-balanced, the same as in **?**;
2. Each unit with its assigned treatment in step 1 is duplicated and treated as a block. The $\mathbf{B}_2$ matrix of indicators for blocks is formed with dimension 42×21. The design in stratum 2 is chosen as a D_S- or A_S-optimum blocked design considering the model terms for the 4 ES factors and their interactions with the HS factor whose levels were fixed at step 1.

The new designs are shown in Supplementary (Supp.) Table A.

Properties of the designs, such as A_S and D_S values and efficiencies, for several values of η, are plotted in Figure **??**. For very small η values design D has the best and design SS the worst performance, in terms of A_S efficiency. As η increases design D becomes less efficient. The newer designs (MSS$_A$ and MSS$_D$) become more efficient for η larger than about 1.7. In terms of the determinant, designs D, MSS$_A$ and MSS$_D$ have similar performances, design D being better for the range of η studied with the efficiencies of the new designs ranging from about 94.0% to 98.7%. It can be seen in Table **??** that in terms of predicting the responses, the I-optimal design gives the best performance for the range of η studied followed by the new designs that are clearly better than the D-optimal design. For estimating differences in response, the new designs show high efficiencies that increase with η. We also note that the MSS algorithm which uses A_S is better than that which uses D_S. Square root of the mean of the variances for the model parameter estimators averaged according to the type of effects are displayed in Supp. Table B. We note that using an exchange algorithm in stratum 2 improves considerably the design compared with the SS approach that fixed the treatment set to be a CCD. The D-optimal design penalizes the quadratic effects of HS factors, but gives very good estimation of the corresponding linear effects, as usual, while the I-optimal design goes in the opposite direction.

4.2 Example 2 (7 HS and 4 ES factors, 20 whole plots with 5 subplots each)

The second example compares alternative designs for the polypropylene experiment described in **?**. There are 7 two-level HS and 4 ES factors, 3 continuous and 1 a three-level qualitative factor. The model includes linear main effects for all factors, quadratic effects for the 3 ES continuous factors and 50 two-factor interactions (only one of the HS factors, W_1, is expected to interact with the others). There was a constraint among two of the HS factors, W_3 and W_4, which were not allowed to both appear at the highest level, and this was taken into account when specifying the candidate set for the exchange algorithm. **?** compared two designs for these factors in 20 whole plots of 5 subplots each, one constructed by the SS approach (D criterion in each phase, the separate treatment sets also being chosen by the D criterion) and the other by the global D-optimum approach ($\eta = 1$). We found two other designs for this experiment, MSS_A and MSS_D designs, which are shown in Supp. Tables C and D. We compare the designs in terms of D_S and A_S values in Figure **??** and Supp. Table E. The new designs are better than SS and D designs with respect to the A_S criterion and even the old SS design is better than D for $\eta > 10$. We note that even for $\eta = 1$ design MSS_D outperforms design D with respect to D_S values. As can happen, especially for such a large experiment, the optimization procedure failed to find the globally D-optimum design. As one factor is qualitative, for evaluating the prediction capabilities we calculated variances for each level and averaged them. The new designs are not so impressive in terms of I-efficiency, although their advantage is clearer with respect to the difference-based prediction criterion (Table **??**).

Figure 2: A_S and D_S values, as functions of η, for alternative designs for Example 2. Inset: Efficiencies, as described at the start of Section **??**.

Table 2: Prediction performances for alternative designs for Example 2

		Design			
η	Efficiency	SS	D	MSS_A	MSS_D
1	I	96.52	86.21	100.00	87.83
1	I_D	87.50	88.31	100.00	90.28
10	I	100.00	94.49	97.52	94.04
10	I_D	95.25	92.59	100.00	96.91
20	I	100.00	95.46	96.47	94.54
20	I_D	96.90	93.47	100.00	98.25

4.3 Example 3 (2 HS and 2 ES factors, 12 whole plots with 4 subplots each)

? gave a second-order equivalent-estimation design (EE) for 2 HS and 2 ES factors in 12 whole plots of size 4, based on the Box-Behnken treatment set. It is not clear what was the original region of experimentation, but we assume it was a hypercube, as did **?**, who constructed a D-optimum design (D) for the same problem. We constructed the MSS_A and MSS_D designs shown in Supp. Table F.

The performances of the four designs are shown in Figure **??** and in Table **??**. Designs MSS_A and MSS_D show very similar performances and are barely distinguishable in the graphs. The graphs highlight the inefficiency of the equivalent-estimation design. The other three designs have similar performances with a loss of efficiency of about 4% from design D, with respect to the A_S criterion, for $\eta > 0.7$. The newer designs are slightly more efficient than the D-optimum design in terms of variances and almost as efficient in terms of the determinant. Again we find that the new designs outperform all others in terms of predicting differences in response and are competitive

Table 3: Prediction performances for alternative designs for Example 3

η	Efficiency	Design			
		EE	D	MSS_A	MSS_D
1	I	100.00	59.86	98.21	97.84
1	I_D	53.53	77.59	100.00	99.33
10	I	100.00	48.54	82.21	82.17
10	I_D	57.75	75.12	100.00	99.91
20	I	100.00	47.69	80.97	80.95
20	I_D	58.13	74.92	100.00	99.95

Figure 3: A_S and D_S values, as functions of η, for alternative designs for Example 3. Inset: Efficiencies, as described at the start of Section **??**.

in terms of predicting the response. The EE designs are very efficient under the I criterion, but not under I_D, showing that these criteria are not always the same. Supp. Table G shows the low precision for estimating all effects of the EE design, except quadratic effects of the HS factors.

4.4 Example 4 (3 HS and 3 ES factors, 12 whole plots with 4 subplots each)

? found that some D-optimum designs also satisfy the equivalent-estimation property and that for a given structure there can be many equivalent-estimation designs, some of them with high efficiency in terms of the D criterion. They compared D-optimal designs (considering $\eta = 1$) and D-efficient equivalent-estimation designs (EE_D) for several structures including the situation with 3 HS and 3 ES factors in 12 whole plots with 4 subplots each. Here we compare their designs and A_S and D_S optimal designs obtained by the MSS approach.

The new designs are given in Supp. Table H. The graphs in Figure **??** compare the performances of the designs. In terms of A_S values designs D, MSS_A and MSS_D have almost the same

Figure 4: A_S and D_S values, as functions of η, for alternative designs for Example 4. Inset: Efficiencies, as described at the start of Section ??.

Table 4: Prediction performances for alternative designs for Example 4

		Design			
η	Efficiency	EE_D	D	MSS_A	MSS_D
1	I	75.21	91.40	98.82	100.00
1	I_D	69.75	96.00	99.57	100.00
10	I	70.79	93.63	94.77	100.00
10	I_D	65.06	96.82	97.37	100.00
20	I	70.42	93.84	94.43	100.00
20	I_D	64.67	96.90	97.18	100.00

performance, though for very small η the D-optimum design is slightly more efficient. In terms of D_S values, both MSS_A and MSS_D designs perform very similarly to design D and do somewhat better in terms of prediction criteria - see Table ??. The D-efficient equivalent-estimation design (EE_D) is clearly poor in terms of both A_S and D_S values, as well as the prediction criteria. Supp. Table I give more detailed comparisons of these designs.

4.5 Example 5 (2 VHS, 1 HS and 3 ES factors, 8 whole plots, each with 2 subplots, each with 2 sub-subplots)

In this example we consider the design problem with three strata presented in ?. The model has linear main effects and two-factor interactions with 2 VHS, 1 HS and 3 ES factors. The unit structure is 8 whole plots, each with 2 subplots, each with 2 sub-subplots. ? constructed a D-optimum design fixing $\eta_1 = \eta_2 = 1$. As the number of units in each stratum is a power of 2 and the model is supported by a two-level factorial, they also presented an alternative design constructed by fractionating and aliasing high order terms. Our design construction follows four steps:

1. Find an unblocked design with 8 units for the model with 4 parameters for the two VHS factors. The best design is the duplicated 2^2 factorial;

2. Each unit with its assigned treatment in step 1 is duplicated and treated as a block. The $\mathbf{B}_2$ matrix of indicators for blocks is formed with dimension 16×8. A D_S- or A_S-optimal blocked design is obtained considering the model terms for the HS factor and its interactions with the VHS factors whose levels were fixed at step 1. The number of parameters of interest is 3.

3. Each unit with its assigned treatment in steps 1 and 2 is duplicated and treated as a block. The $\mathbf{B}_3$ matrix indicators for blocks is formed with dimension 32×16. A D_S- or A_S-optimal blocked design is obtained considering the model terms for the three ES factors and their interactions with the factors whose levels were fixed at steps 1 and 2. The number of parameters of interest is 15.

4. Treat each unit in stratum 1 as a *superblock* and form the $\mathbf{B}$ now with column indicators for 8 blocks of size 4 each. Re-arrange the 16 blocks found in step 3 between the superblocks with the same levels of w_1, w_2 and s_1 to be D_S- or A_S-optimally blocked.

Our approach resulted in the same design for both A_S and D_S criteria (Supp. Table J).

The efficiencies of the MSS design relative to the globally D optimum design of **?**, design D, are shown in Figure **??**. In the plots, σ^2 (the third stratum variance) is fixed to be 1 and σ_2^2 (the second stratum variance component) and σ_1^2 (the first stratum variance component) are varied. For very small values of the variance ratios the new design is less efficient than design D but it becomes more efficient as the ratios increase. We note that, with respect to D efficiencies, both designs (the globally D-optimum and the MSS design) are robust to changes in σ_1^2, especially when σ_2^2 is large. **?** also noted the robustness of D-optimum designs for changes in σ_1^2. Although the two designs show similar performances in terms of efficiencies, design D has one interaction term between the ES factors that is fully estimated in stratum 2. Our designs distributed the loss of information among all terms and thus none of the terms is sacrificed as shown in Supp. Table K. It should be noted that the alternative design of **?**, constructed by fractionating and aliasing terms, has two interactions of ES factors fully estimated in stratum 2 and one in stratum 1. Our

Table 5: Prediction performances for alternative designs for Example 5

			Design	
η_1	η_2	Efficiency	D	MSS
1	1	I	100.00	94.33
1	1	I_D	100.00	92.68
1	10	I	98.64	100.00
1	10	I_D	97.05	100.00
1	20	I	97.26	100.00
1	20	I_D	95.00	100.00
20	1	I	100.00	99.33
20	1	I_D	100.00	98.87
20	10	I	99.57	100.00
20	10	I_D	98.86	100.00
20	20	I	98.76	100.00
20	20	I_D	97.37	100.00

Figure 5: A_S and D_S efficiencies, relative to design D, of design MSS for Example 5, as functions of σ_1^2 and σ_2^2 (fixing $\sigma^2 = 1$). Far left: the same for $0 < \sigma_1^2,\ \sigma_2^2 < 1$.

new design also improves on the old one in terms of prediction variances, except when η_2 is small (Table **??**).

4.6 Example 6 (2 HS and 2 ES factors, 5 blocks, each with 3 whole plots, each with 3 subplots)

In this last example we re-design the experiment for the blocked split-plot structure presented in **?**. This is aimed at a response surface model for 2 HS and 2 ES factors. In the first stratum there are 5 units (blocks) to which no factors are applied. In the second stratum each block has 3 whole plots and 2 HS factors are to be applied. In the third stratum each whole plot is divided into three subplots and the 2 ES factors are to be applied. In this case the design to start with is a blocked design for the whole plots and this example is aimed at showing the flexibility of our methodology. The designs we constructed are shown in Supp. Table L. Note that, although in the third stratum the number of units would allow 5 replicates of the 3^2, it is not that treatment set that comes out of the search, for either criterion.

Using a script supplied by a referee we also obtained the globally D optimum design constructed by JMP® (**?**). Our MSS designs have very similar properties as shown in Supp. Table M. For the MSS designs very little information comes from the highest stratum (inter-block information) no matter what are the sizes of the variance components. That was also true for most effects for SS and D designs except, respectively, the interaction and quadratic effects involving the HS factors. Plots of efficiencies are shown only for MSS_A with respect to the D optimum design in Figure **??**, the graphs for MSS_D being similar. For the range of variance component values studied the gain in A_S-efficiency of our designs varies from 8 to 25% while the loss in D_S-efficiency is less than 2%. MSS designs are also more efficient than design D in terms of prediction with the gain being up to around 10%. The SS design is better for prediction for all values studied (Table **??**).

Figure 6: A_S and D_S efficiencies, relative to design D, of design MSS for Example 6, as functions of σ_1^2 and σ_2^2 (fixing $\sigma^2 = 1$). Far left: the same for $0 < \sigma_1^2,\ \sigma_2^2 < 1$.

Table 6: Prediction performances for alternative designs for Example 6

			Design			
η_1	η_2	Efficiency	SS	D	MSS_A	MSS_D
1	1	I	100.00	73.27	81.25	81.87
1	1	I_D	100.00	84.33	90.27	91.11
1	10	I	100.00	66.97	76.23	76.31
1	10	I_D	100.00	83.97	88.57	88.70
1	20	I	100.00	66.48	75.82	75.86
1	20	I_D	100.00	84.35	88.61	88.68
20	1	I	100.00	94.36	96.47	96.61
20	1	I_D	100.00	83.65	91.26	92.09
20	10	I	100.00	79.51	86.50	86.51
20	10	I_D	100.00	80.71	88.62	88.75
20	20	I	100.00	74.16	82.42	82.40
20	20	I_D	100.00	80.74	88.03	88.09

5 Discussion

We have modified the stratum-by-stratum method of construction of multi-stratum response surface designs and compared it with several other approaches from the literature. The procedure produces efficient designs that are competitive with other popular designs. The step-by-step design construction makes the method quite attractive due to its direct application to designing experiments for any number of strata. The same program code can be run sequentially, once for each stratum, as long as the entries are correctly specified. The step-by-step approach does not experience the problems with storage of large candidate treatment sets and thus the usual point exchange algorithm is used. However the approach can also be used with the coordinate exchange algorithm of **?**. As the construction basis is a blocked design in each stratum, the updating formulae of **?** can be used to speed the search. Another important practical advantage is that it does not require prior estimates of variance component ratios.

Although the examples show that globally optimum designs for fixed η are quite robust to the variance component ratios, this method does not share the generality of the stratum-by-stratum approach. Algorithms are only widely available for some specific multi-stratum structures (split-plot, split-split plot). The advantages of designs constructed using the modified stratum-by-stratum approach for prediction properties was not anticipated and is not immediately easy to explain. The I criterion concentrates on estimating the intercept and therefore, in completely randomized structures, concentrates points near the center of the design. Consequently, designs with many points near the center, such as those obtained from classical designs using the original SS approach, tend to do well in terms of this criterion. However, it is not obvious that the MSS approach gives any more points near the center than the single stage D-optimal designs have. A possible explanation is that prediction at every point, whether of the response or differences in response, depends on all parameters and therefore those estimated in higher strata have more impact, especially when the variance components are large. In contrast, in the D and A criteria, the poor estimation of a few effects in high strata are swamped by good estimation of many effects in low strata. For prediction, therefore, by far the most important thing is precise estimation in the higher strata. By optimizing this first, the MSS algorithm achieves exactly what is needed.

The designs produced by the MSS method are highly D-efficient and much better across a range of criteria than many D-optimal designs. They are robust to high stratum variances and avoid completely confounding some parameter estimates with blocks. In addition, the method is simple to use and understand. We believe that, along with other algorithms, the stratum-by-stratum approach deserves a place in the experimental designer's toolbox and should be seriously considered for producing designs for any experiment that involves factors that are hard to set.

Supplementary Materials

Designs and results Supplementary Tables A-M containing designs identified for the six examples, and corresponding comparisons with other designs. (pdf file)

Code and designs General R code for the design algorithm, plus specific settings for the six examples in the paper. The identified designs are also included as an excel spreadsheet. (zip file)

Acknowledgments

The authors gratefully acknowledge financial support from EPSRC grant EP/C541715/1: Unifying Approaches to Design of Experiments and FAPESP grant 2010/0250-08. This work was completed while the authors were participants in the Design and Analysis of Experiments programme at the Isaac Newton Institute in 2011. We thank two referees, an associate editor and the editor for helpful comments which led to improvements in the paper.