ARTICLE IN PRESS
JID: ECOSTA [m3Gsc;May 1, 2022;12:42]

Econometrics and Statistics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Econometrics and Statistics

journal homepage: www.elsevier.com/locate/ecosta

Multi-objective optimisation of split-plot designs

Matteo Borrotti[a], Francesco Sambo[b], Kalliopi Mylona[c,*]

[a] Department of Economics, Management and Statistics, University of Milan-Bicocca, Italy
[b] Verizon Connect Research, Italy
[c] Department of Mathematics, King's College London, United Kingdom

ARTICLE INFO

Article history:
Received 20 July 2021
Revised 1 April 2022
Accepted 3 April 2022
Available online xxx

Keywords:
Pareto front
pure-error estimation
restricted randomised experiments
variance components

ABSTRACT

Modern experiments allow scientists to tackle scientific problems of increasing complexity. Often experiments are characterised by factors that have levels which are harder to set than others. A possible solution is the use of a split-plot design. Many solutions are available in the literature to find optimal designs that focus solely on optimising a single criterion. Multi-criteria approaches have been developed to overcome the limitations of the one-objective optimisation, however they mainly focus on estimating the precision of the fixed factor effects, ignoring the variance component estimation. The Multi-Stratum Two-Phase Local Search (MS-TPLS) algorithm for multi-objective optimisation of designs of experiments is extended, in order to ensure pure-error estimation of the variance components. The proposed solution is applied to two motivating problems and the final optimal Pareto front and related designs are compared with other designs from the relevant literature. Experimental results show that the designs from the obtained Pareto front represent good candidate solutions based on the different objectives.

1. Introduction

In many areas, improvements in the way experiments are designed have the potential to produce significant increases in knowledge as well as time and money savings. Better experiments, for example, result in fewer humans being involved in clinical studies (see Yi and Wang (2021)), fewer expensive components being required for others, fewer prototypes being required in engineering experiments, and faster product and process development timeframes. Many industrial experiments involve one or more restrictions on the randomisation (e.g., Trinca and Gilmour (2001), Goos and Vandebroek (2001), Jones and Goos (2009), Vining et al. (2005), Großmann and Gilmour (2021)). In such instances, a split-plot design layout, in which the experimental runs are conducted in groups, is a standard cost-effective method for reducing the number of independent settings of the hard-to-change factors.

Split-plot designs were originated in agricultural experiments, where large plots of land were subdivided into relatively large sections, known as whole plots. These plots were then randomly allocated to each of the possible levels of the whole-plot factors. Whole plots were further divided into smaller portions known as subplots, to which subplot factors were ap-

* Corresponding author.
E-mail address: kalliopi.mylona@kcl.ac.uk (K. Mylona).

https://doi.org/10.1016/j.ecosta.2022.04.001
2452-3062/

Please cite this article as: M. Borrotti, F. Sambo and K. Mylona, Multi-objective optimisation of split-plot designs, Econometrics and Statistics, https://doi.org/10.1016/j.ecosta.2022.04.001

plied. As a result, whole-plot factor levels differ from plot to plot, while subplot factor levels differ from subplot to subplot. Hard-to-change factors serve as whole-plot factors in industrial experimentation, while easy-to-change factors act as subplot factors.

Computerised design search algorithms have gained popularity for constructing split-plot designs for response surface models. A sequential method for constructing multi-stratum designs, from stratum to stratum and beginning from the highest stratum, is presented in Trinca and Gilmour (2001), and an improved version is presented in Trinca and Gilmour (2015). Goos and Vandebroek (2001) and Goos and Vandebroek (2003) proposed point-exchange algorithms for constructing D-optimal designs for split-plot response surface experiments, and Jones and Goos (2007) and Jones and Goos (2009) defined a coordinate-exchange algorithm for split-plot and split-split-plot experiments.

The aforementioned algorithms have the weakness of focusing solely on optimising a single criterion. The Pareto front method is used to construct split-plot designs in Sambo et al. (2014): the newly developed coordinate exchange–two step local search (CE-TPLS) algorithm extends the Jones and Goos (2007) and Jones and Goos (2012) coordinate-exchange (CE) algorithm with a two-phase local search approach (Paquete and Stützle (2007)). In Borrotti et al. (2017), the Multi-Stratum Two-Phase Local Search (MS-TPLS) algorithm was proposed, an extention of the CE-TPLS algorithm for multi-stratum designs and with more optimality criteria. In Cao and Robinson (2017), a multi-criteria approach is being considered as well, where the degrees of freedom for the whole plots and subplots are combined with the overall D-criterion. A weakness of the majority of the existing approaches is that they focus entirely on estimating the precision of the fixed factor effects, and ignore the variance component estimation required to compute these effects' generalised least squares (GLS) estimates and to make statistical inference. In Mylona et al. (2014) and Mylona et al. (2020) a family of alternative *D*-optimality criteria that focuses on the estimation of fixed effects, as well as the estimation of variance components are proposed. In this paper, we construct the Pareto front of optimal split-plot designs that ensure pure-error estimation of the variance components, by incorporating in our multi-objective optimisation, the determinant of the information matrix of the variance components, as presented in Mylona et al. (2020). Moreover, as it is being stated in Cao and Robinson (2017), there are advantages of having designs with degrees of freedom for both pure error and for lack-of-fit, hence our multi-objective optimisation will consider these degrees of freedom for the whole plots and subplots combined with the determinants of the information matrices for the fixed effects and the variance components.

This article is organised as follows. In the next section, we introduce the model assumed when analysing data from split-plot response surface experiments and the relative statistical inference. In Section 3, we discuss our multiple criterion and algorithmic approach. In Section 4, we compare designs constructed with our approach to designs from the state-of-the-art literature. We end the article with a discussion.

2. Models and estimation

In general, the model suggested analysing data from split-plot response surface experiments with n runs and b whole plots (Letsinger et al. (1996), Gilmour and Trinca (2000)), is

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon}, \tag{1}$$

where $\mathbf{Y}$ is the n-dimensional vector of responses, $\mathbf{X}$ is the $n \times p$ model matrix corresponding to the assumed model, $\boldsymbol{\beta}$ is the p-dimensional vector of fixed model parameters (often including an intercept, main effects, interaction effects and quadratic effects), $\boldsymbol{\gamma}$ is a b-dimensional vector of whole-plot effects, $\mathbf{Z}$ is the $n \times b$ design matrix for these random effects and $\boldsymbol{\epsilon}$ is the n-dimensional vector of random errors. It is further assumed that $\boldsymbol{\gamma} \sim N\left(\mathbf{0}_b, \sigma_\gamma^2 \mathbf{I}_b\right)$, $\boldsymbol{\epsilon} \sim N\left(\mathbf{0}_n, \sigma_\epsilon^2 \mathbf{I}_n\right)$, and that $\boldsymbol{\gamma}$ and $\boldsymbol{\epsilon}$ are independent.

In the assumed response surface model, the generalised least square (GLS) estimator of the parameter vector $\boldsymbol{\beta}$ is given by

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

where

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

From the GLS estimator, it is obvious that for obtaining good estimates for the fixed effects ensuring good estimates of the variance components is essential. Following Letsinger et al. (1996), the two variance components σ_ϵ^2 and σ_γ^2 are generally estimated using residual maximum likelihood (REML) estimation. However, in Gilmour et al. (2017), starting from the complete treatment model is suggested, to obtain estimates for the variance components that are robust against misspecification of the response surface model.

$$\mathbf{Y} = \mathbf{X}_t\boldsymbol{\tau} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon}, \tag{2}$$

where the full treatment design matrix is being denoted as $\mathbf{X}_t$, and $\boldsymbol{\tau}$ is the corresponding vector of treatment means. The (i,t)th element of the full treatment design matrix is equal to 1 if treatment t is used for run i and 0 otherwise. There are as many fixed model parameters for the full treatment model as there are distinct treatments in the design. The model therefore does not indicate any lack of fit, and pure-error estimates are the variance component estimates obtained from it.

Therefore, using REML estimation in conjunction with the complete treatment model results in pure-error estimates. This includes maximising the log likelihood function

$$l_R^* = -\frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}\log|\mathbf{X}_t'\mathbf{V}^{-1}\mathbf{X}_t| - \frac{1}{2}\mathbf{r}_t'\mathbf{V}^{-1}\mathbf{r}_t - \frac{n-n_t}{2}\log(2\pi),$$

where $\mathbf{r}_t = \mathbf{y} - \mathbf{X}_t(\mathbf{X}_t'\mathbf{V}^{-1}\mathbf{X}_t)^{-1}\mathbf{X}_t'\mathbf{V}^{-1}\mathbf{y}$ and n_t is the number of distinct treatments in the design (or, equivalently, the number of parameters in $\boldsymbol{\tau}$).

Moreover, in Cao and Robinson (2017), with the use of the ANOVA table, formulae to calculate the pure error (PE) degrees of freedom for whole-plot (WP) and for subplot effects (SP), in a split-plot design are provided. Specifically, the partition demonstrated in Cao and Robinson (2017) using (Harville, 1997, p. 395) is

$$df_{\mathrm{PE}}(\mathrm{WP}) = \mathrm{rank}(\mathbf{Z}, \mathbf{X_t}) - \mathrm{rank}(\mathbf{X_t}) \quad \text{and} \quad df_{\mathrm{PE}}(\mathrm{SP}) = n - \mathrm{rank}(\mathbf{Z}, \mathbf{X_t}),$$

where the sum of these equations is the number of replicated observations (r). Hence, $df_{\mathrm{PE}}(\mathrm{WP})$ is the number of whole plots containing design runs that have been replicated in another whole plot and $df_{\mathrm{PE}}(\mathrm{SP})$ is the total number of replicated design runs minus the number of whole plots containing design runs that have been replicated in another whole plot. Additionally, the pure error plus lack-of-fit (PE+LoF) degrees of freedom are provided for whole plots and subplots and they are respectively:

$$df_{\mathrm{PE+LoF}}(\mathrm{WP}) = \mathrm{rank}(\mathbf{Z}, \mathbf{X}) - \mathrm{rank}(\mathbf{X}) \quad \text{and} \quad df_{\mathrm{PE+LoF}}(\mathrm{SP}) = n - \mathrm{rank}(\mathbf{Z}, \mathbf{X}).$$

3. Optimal design and algorithm

3.1. Optimality Criterion

The D-optimality criterion, which seeks designs that minimise the generalised variance of the parameter estimators, is the most widely used optimality criterion for constructing experimental designs. This is achieved by minimising the variance-covariance matrix determinant of the estimates of the factor impact or, equivalently, by maximising the determinant of the information matrix about β. For a blocked experiment and a split-plot experiment, the information matrix is given by

$$\mathbf{M} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}, \tag{3}$$

when the GLS estimator is used. A D-optimal design therefore maximises

$$|\mathbf{M}| = |\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|. \tag{4}$$

D-optimality criterion emphasises on the estimation of the fixed model parameters.

In Mylona et al. (2014), a family of alternative D-optimality criteria is proposed that focuses on the estimation of fixed effects, as well as the estimation of variance components. It assumes that the variance components are estimated using REML, in combination with the response surface model.

An alternative composite D-optimality criterion was proposed in Mylona et al. (2020), provided that model-robust pure-error estimates of the variance components can be obtained by applying REML to the complete treatment model in (2), in which the information on the variance components is quantified in a way that is insensitive to the specification of the response surface model for the treatment effects and thus requires pure-error estimates of the variance components. In Cao and Robinson (2017), a multi-criteria approach is being considered, where the pure error degrees of freedom for the whole plots and subplots are combined with the overall D-criterion: $(D, df_{\mathrm{PE}}(\mathrm{WP}), df_{\mathrm{PE}}(\mathrm{SP}))$.

In this work, we construct Pareto optimal designs using as well a multi-criteria approach and taking into consideration the estimation of fixed effects and variance components at split-plot designs, together with the pure error and lack-of-fit degrees of freedom for whole plots and subplots. In accordance with the MS-TPLS algorithm, we minimise

$$\left(\frac{1}{|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|}, \frac{1}{|\mathbf{N}_t|}, \frac{1}{df_{\mathrm{PE}}(\mathrm{WP})}, \frac{1}{df_{\mathrm{PE}}(\mathrm{SP})}, \frac{1}{df_{\mathrm{LoF}}(\mathrm{WP})}, \frac{1}{df_{\mathrm{LoF}}(\mathrm{SP})}\right),$$

where $\mathbf{N}_t$ is the information matrix about the variance components assuming that the full treatment model is used for estimating them, given in (5)

$$\mathbf{N}_t = \frac{1}{2}\begin{bmatrix} \mathrm{tr}\{(\mathbf{P}_t\mathbf{Z}\mathbf{Z}')^2\} & \mathrm{tr}(\mathbf{P}_t^2\mathbf{Z}\mathbf{Z}') \\ \mathrm{tr}(\mathbf{P}_t^2\mathbf{Z}\mathbf{Z}') & \mathrm{tr}(\mathbf{P}_t^2) \end{bmatrix}, \tag{5}$$

with

$$\mathbf{P}_t = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}_t(\mathbf{X}_t'\mathbf{V}^{-1}\mathbf{X}_t)^{-1}\mathbf{X}_t'\mathbf{V}^{-1}.$$

```
MS-Opt(criteria, ᾱ, iterations, initDesign)
bestScore = +INF
for it = 1 to iterations
    if initDesign == NULL
        curDesign = SampleDesign()
    else curDesign = initDesign
    curScore = ᾱ · Scores(curDesign, criteria)
    improvement = TRUE
    while improvement
        improvement = FALSE
        for each stratum s
            for each unit u of s
                for each factor f in s
                    for each of the other available values of f
                        nextDesign = curDesign
                        Assign the value to all elements of matrix tmpDesign
                        at column f and at the rows corresponding to u
                        nextScore = ᾱ · Scores(nextDesign, criteria)
                        if nextScore < curScore
                            curDesign = nextDesign
                            curScore = nextScore
                            improvement = TRUE
    if curScore < bestScore
        bestDesign = curDesign
        bestScore = curScore
return bestDesign
```

Fig. 1. Pseudo-code of MS-Opt as presented in Borrotti et al. (2017).

3.2. Algorithm

Multi-objective optimisation problems involve the optimisation of several, typically conflicting objectives. When considering more than one criterion for identifying optimal design of experiments, the final result is no longer a unique solution but, more often, a set of optimal designs that simultaneously compete in the optimisation of the concurrent objectives (Paquete and Stützle (2007)).

Considering a multi-criteria design of experiments where the aim is to minimise n optimality criteria, candidate designs are evaluated according to a criterion vector of *objective functions* $\overline{f} = (f_{c_1}, f_{c_2}, \ldots, f_{c_n})$, where $c_1 \ldots c_n$ are the different criteria. Now, given two designs $\mathbf{d}$ and $\mathbf{d}'$, we say that $\mathbf{d}$ *dominates* $\mathbf{d}'$ ($\mathbf{d} \prec \mathbf{d}'$) iff $\overline{f}(\mathbf{d}) \neq \overline{f}(\mathbf{d}')$ and $f_c(\mathbf{d}) \leq f_c(\mathbf{d}')$, $\forall c \in \{c_1 \ldots c_n\}$, *i.e.* if $\mathbf{d}$ is equal or better than $\mathbf{d}'$ according to all criteria.

If no $\mathbf{d}'$ exists such that $\mathbf{d}' \prec \mathbf{d}$, the design $\mathbf{d}$ is called *Pareto-optimal*. The ultimate goal of multi-criteria optimal design is to determine (or approximate) the set of all Pareto-optimal designs, whose image in the multi-objective space is called the *Pareto front*.

In this work we exploit the multi-stratum two-phase local search algorithm (MS-TPLS) proposed in Borrotti et al. (2017). MS-TPLS extends the coordinate-exchange two-phase local search (CE-TPLS) algorithm (Sambo et al. (2014)) to any type of multi-stratum designs and it is basically composed by two main components: (i) the coordinate-exchange (CE) algorithm (Jones and Goos (2007, 2012)) and (ii) the two-phase local search (TPLS) approach (Dubois-Lacoste et al. (2011)). The CE algorithm is codified in the function MS-Opt, whose pseudo-code is reported in Fig. 1 for completeness. It searches for an optimal design according to a specific *scalarisation* of the multi-objective problem, i.e. a weighted sum of all the objectives according to a weight vector α, by sampling an initial solution at random and iteratively adjusting it with minimal changes to its factors. The TPLS algorithm uses the CE algorithm as subroutine and, by iteratively generating new scalarisations and solving them with CE, explores the multi-objective space in search for the optimal set of non-dominating solutions, *i.e.* the Pareto front.

More precisely, the single-objective local search component (*i.e.* MS-Opt) of the MS-TPLS algorithm minimizes a scalarisation of the objective functions in the form

$$f_W = \sum_{c \in C} \alpha_c f_c = \overline{\alpha} \cdot \overline{f}, \qquad \sum_{c \in C} \alpha_c = 1, \tag{6}$$

Table 1
Set of free parameters.

Parameter	Name	Short description
1	ChooseInitialDesign	Function responsible of selecting the initial design for the n objective functions
2	ComputeWeight	Function responsible of computing the scalarisation weight α used in Eq. 6
3	Number of iterations	Number of single-objective iterations for generating initial solutions for each criterion involved in the optimisation
4	Number of scalarisations	Number of times that the n-objective optimisation problem is transformed in a single objective problem based on a weighted sum of the n criteria (Eq. 6)
5	Number of restarts	Number of times that the algorithm is started from the beginning

where C is the set of criteria to be minimized, f_c is the objective function for criterion c and the parameters $\overline{\alpha}$ control the relative weight of each objective function. For the scalarisation to be unbiased, all objective functions should lie in the range [0,1]: before computing each scalarisation, thus, MS-TPLS dynamically normalizes each objective function value, as in Borrotti et al. (2017):

$$f_c^{norm} = \frac{f_c - f_c^{min}}{f_c^{max} - f_c^{min}},$$

where f_c^{min} and f_c^{max} are the minimum and maximum values of the objective function, among all the designs encountered by the algorithm from the beginning of the run. Given an initial design, the solution is then improved by iteratively replacing values for each factor in each unit of each stratum, until no further local change can increase the weighted sum f_W of the objective functions. The procedure is repeated for a given number of iterations and the best design is returned. At each iteration, all the designs generated during the search are used to identify the set of non-dominated designs forming the Pareto front. For more details about the algorithm please see Borrotti et al. (2017).

From an algorithmic point of view, in this work we developed and integrated two new functions for computing the $\mathbf{Z}$ and $\mathbf{X_t}$ matrices. These new functions lead to the extension that allows the integration of the new criteria at the function Scores that contains the implementation of all the possible criteria. Scores is the main function used by function MS-Opt, in line 6 of Fig 1 (Borrotti et al. (2017)). More precisely, we added the N_t criterion, the pure error degrees of freedom for whole-plot ($df_{PE}(WP)$) and for subplot effects ($df_{PE}(SP)$) and the lack-of-fit degrees of freedom for whole-plot ($df_{LoF}(WP)$) and for subplot effects ($df_{LoF}(SP)$). Given that, the MS-TPLS is now able to optimise any combination of seven design optimality criteria (D, A, I, I_d, A_s, D_s, N_t) plus any of the following degrees of freedom ($df_{PE}(WP)$, $df_{PE}(SP)$, $df_{LoF}(WP)$, $df_{LoF}(SP)$).

3.3. Parameter settings

The MS-TPLS algorithm has several free parameters that have been set in accordance with the results obtained in Sambo et al. (2014). Table 1 summarises the free parameters. For all possible options of the free parameters, please check Sambo et al. (2014).

The ChooseInitialDesign function is set in order to select the initial designs from the Pareto front. The ComputeWeight function is set to sample the α weight at random from $[0, 1]$. This combination represents the best balance between diversification of the solutions and intensification of the search as empirically demonstrated in Sambo et al. (2014). Furthermore, the results were obtained with $n \times 10$ restarts, where n is the number of criteria simultaneously optimized by the MS-TPLS algorithm, each composed of $n \times 2$ initial single-objective iterations (*i.e.* number of iterations) followed by $100 - (n \times 2)$ scalarisations (*i.e.* number of scalarisations), for a total of $n \times 1000$ calls to the single-objective optimisation function. As an example, if we want to optimise 4 criteria (*i.e.*D, N_t, $df_{\mathrm{PE}}(\mathrm{WP})$, $df_{\mathrm{PE}}(\mathrm{SP})$) then we have 40 restarts each composed of 2 + 2 + 2 + 2 initial single-objective iterations followed by 92 scalarisations. The value of η (*i.e.* ratio between variance components) is 1 for all the reported results.

4. Result and comparisons

In this section, the proposed criteria are used for the construction of designs based on two motivating examples presented in Vining et al. (2005) and Mylona et al. (2020).

4.1. A 48-run split-plot response surface example

In this example, in Vining et al. (2005) a central composite design was modified to obtain an equivalent estimation design (EED). We will call this design, VKM design. The number of whole plots is 12 and the number of sub-plot runs for each whole plot is 4, for a total of 48 runs ($n = 48$). There are two whole-plot factors (w_1, w_2) and two subplot factors (s_1, s_2). The considered model is the usual response surface model that includes an intercept, main effects, two-factor interaction effects and quadratic effects.

In Cao and Robinson (2017) a multi-objective split-plot design problem was considered where they simultaneously optimised the criterion vector (D, $df_{\mathrm{PE}}(\mathrm{WP})$, $df_{\mathrm{PE}}(\mathrm{SP})$) for a design with the same features. From the Pareto front that they have

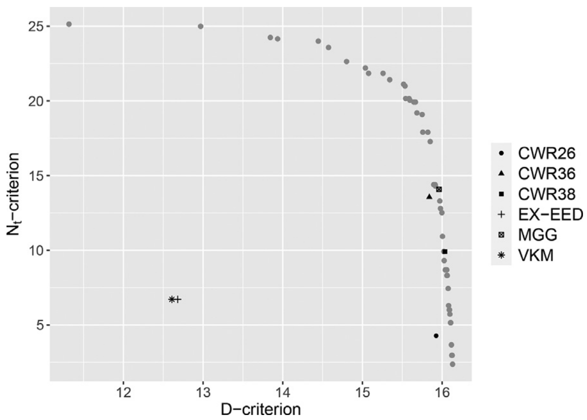

Fig. 2. Pareto front for example in Vining et al. (2005).

constructed (Table 10 in Cao and Robinson (2017)), we will use for comparisons the Pareto optimal designs based upon just the *D*-criterion and pure error degrees of freedom for whole plots, denote them as CWR26, CWR36 and CWR38. Furthermore, we considered the extra EED design, EX, proposed for comparison purposes in Table 10 in Cao and Robinson (2017), denoted here by EX-EED.

The same motivating example was also used in Mylona et al. (2020). The optimality criterion used in that work, as mentioned in Section 3.1, guarantees model-robust pure-error estimates of the variance components. For this reason our results are also compared with the design proposed in Table 1 in Mylona et al. (2020), denoted here as MGG.

For the sake of clarity, in what follows all the final numbers are appropriately transformed in order to show the results as in a maximisation problem. Furthermore in both examples, the values for *D*-criterion are reported as $\log_{10}(|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|)$. From now on, we refer to $\log_{10}(|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|)$ simply as *D*.

For demonstrative reasons, we initially construct designs that optimise the two-objective criterion (D, N_t). Nevertheless, as in multi-objective optimisation, it is expected that there would be trade-offs among these criteria. The criterion vector allows one to evaluate and compare these trade-offs. The criterion vector also depends on the variance ratio η. In this work η is set to 1, a commonly used search value for a D-optimal designs (Macharia and Goos (2010), Goos (2002)). We have constructed designs for other values of η, with similar conclusions.

Figure 2 shows the Pareto front obtained by taking the set of non-dominated designs from the results of 20 random restarts of the MS-TPLS algorithm for the instance considered (Vining et al. (2005)). The CWR26, CWR36, CWR38, EX-EED, MGG and VKM designs are also reported. As in Sambo et al. (2014), the best compromise between the D, N_t-criteria is identified as the closest design to the *utopia* point, which is the ideal point in the two-objective space with the maximum value of the D, N_t-criteria (Lu et al. (2011)). This design is named the D, N_t-*symmetrical* design.

In what follows, the *symmetrical* design together with the optimal designs according to each separate criterion are selected from the analysed Pareto fronts for further investigation and comparisons. More precisely, we found the Pareto optimal designs for the following vectors of criteria:

1. $(D, N_t, df_{\mathrm{PE}}(\mathrm{WP}), df_{\mathrm{PE}}(\mathrm{SP}))$;
2. $(D, N_t, df_{\mathrm{PE}}(\mathrm{WP}), df_{\mathrm{PE}}(\mathrm{SP}), df_{\mathrm{LoF}}(\mathrm{WP}), df_{\mathrm{LoF}}(\mathrm{SP}))$.

For each vector, we compare the CWR26, CWR36, CWR38, EX-EED, MGG and VKM designs with the *symmetrical*-design and the optimal designs according to D and N_t respectively, obtained by the modified MS-TPLS algorithm and from the Coordinate-Exchange (CE) algorithm in Jones and Goos (2007) (run for 1000 restarts for each criterion). In Table 2, the 6-symmetrical (6-sym) design and the optimal D- and N_t- designs from Pareto Front are reported. By looking at the values in Table 3, one can see that our symmetrical designs outperform all other designs with respect to estimation of the variance components while they maintain a high value for the D-optimality criterion.

Similarly to Borrotti et al. (2017), we compute the percentage efficiency gains by the 6-symmetrical, with respect to CWR26, CWR36, CWR38, EX-EED, MGG and VKM designs. Furthermore, we also compute the efficiency with respect to the D- and N_t- optimal designs obtained by the CE algorithm in Jones and Goos (2007). The efficiency gain is calculated as

$$\%EFF_{gain}^{c,\mathrm{design}} = 100 \times \left(1 - \frac{f_c^{\mathrm{design}}}{f_c^{sym}}\right), \tag{7}$$

where c identifies the criterion (D or N_t) for which efficiency is computed, f_c^{design} is the value of this criterion for the compared design and f_c^{sym} is the value of the same criterion for the symmetrical design. Table 4 reports the percentage of gained efficiency in terms of D and N_t.

Table 2
6-symmetrical (6-sym), *D*- and N_t- optimal designs from Pareto Front identified by the modified version of MS-TPL algorithm.

6-sym				*D*-optimal				N_t-optimal			
w_1	w_2	s_1	s_2	w_1	w_2	s_1	s_2	w_1	w_2	s_1	s_2
-1	1	0	0	1	-1	1	-1	-1	1	1	-1
-1	1	1	1	1	-1	1	1	-1	1	1	-1
-1	1	-1	1	1	-1	-1	-1	-1	1	-1	0
-1	1	-1	-1	1	-1	-1	1	-1	1	-1	0
-1	-1	-1	1	-1	-1	-1	-1	0	1	-1	-1
-1	-1	1	-1	-1	-1	1	1	0	1	1	1
-1	-1	1	1	-1	-1	-1	1	0	1	-1	-1
-1	-1	-1	-1	-1	-1	1	-1	0	1	-1	-1
1	1	1	1	-1	1	1	1	-1	-1	-1	-1
1	1	-1	-1	-1	1	-1	0	-1	-1	-1	-1
1	1	-1	1	-1	1	1	-1	-1	-1	-1	-1
1	1	1	-1	-1	1	0	1	-1	-1	1	1
1	1	-1	1	-1	1	-1	-1	-1	-1	0	1
1	1	1	1	-1	1	1	-1	-1	-1	0	1
1	1	-1	-1	-1	1	1	1	-1	-1	1	0
1	1	1	-1	-1	1	-1	1	-1	-1	0	1
-1	1	1	-1	-1	0	0	-1	1	-1	1	-1
-1	1	-1	1	-1	0	-1	1	1	-1	-1	-1
-1	1	1	1	-1	0	1	0	1	-1	-1	1
-1	1	0	0	-1	0	-1	-1	1	-1	1	1
0	0	-1	-1	1	1	-1	1	1	-1	1	1
0	0	-1	-1	1	1	0	-1	1	-1	1	-1
0	0	0	0	1	1	1	1	1	-1	-1	1
0	0	-1	-1	1	1	1	0	1	-1	-1	-1
1	0	-1	0	1	-1	-1	0	1	-1	-1	1
1	0	1	1	1	-1	1	-1	1	-1	1	1
1	0	-1	0	1	-1	1	1	1	-1	-1	-1
1	0	0	-1	1	-1	0	1	1	-1	1	-1
-1	-1	1	-1	1	1	1	1	1	-1	-1	-1
-1	-1	-1	1	1	1	1	-1	1	-1	1	-1
-1	-1	1	1	1	1	-1	1	1	-1	-1	1
-1	-1	-1	-1	1	1	-1	-1	1	-1	1	1
1	-1	1	-1	1	0	0	-1	-1	0	0	0
1	-1	1	1	1	0	-1	1	-1	0	1	1
1	-1	-1	0	1	0	1	0	-1	0	1	1
1	-1	0	1	1	0	-1	-1	-1	0	1	1
1	-1	1	-1	-1	-1	-1	-1	1	-1	1	-1
1	-1	-1	0	-1	-1	-1	1	1	-1	-1	1
1	-1	1	1	-1	-1	1	1	1	-1	1	1
1	-1	0	1	-1	-1	0	0	1	-1	-1	-1
0	-1	1	0	0	1	1	-1	1	-1	1	1
0	-1	-1	1	0	1	-1	-1	1	-1	-1	-1
0	-1	1	0	0	1	-1	0	1	-1	1	-1
0	-1	0	-1	0	1	0	1	1	-1	-1	1
-1	1	1	-1	0	-1	-1	-1	1	1	0	1
-1	1	1	1	0	-1	1	-1	1	1	1	0
-1	1	-1	-1	0	-1	-1	1	1	1	1	0
-1	1	-1	1	0	-1	0	0	1	1	0	1

Table 3
Comparison with designs in the literature. Entries in bold correspond to the best values for individual criteria.

		D	N_t	$df_{PE}(WP)$	$df_{PE}(SP)$	$df_{LoF}(WP)$	$df_{LoF}(SP)$
MS-TPLS 6-criteria	6-sym	15.85	14.10	5	18	1	9
	D-opt	**16.13**	2.38	4	5	2	**22**
	N_t-opt	13.21	21.60	5	**27**	1	0
MS-TPLS 4-criteria	4-sym	13.55	24.22	**6**	26	0	1
	D-opt	**16.13**	2.38	4	5	2	**22**
	N_t-opt	5.97	25.46	**6**	**27**	0	0
CE	*D*-opt	**16.13**	2.37	4	5	2	**22**
	N_t-opt	9.91	**25.77**	**6**	**27**	0	0
	CWR26	15.92	4.28	4	8	2	19
	CWR36	15.84	13.57	**6**	15	0	12
	CWR38	16.04	9.92	5	13	1	14
	EX-EED	12.68	6.72	2	21	**3**	7
	MGG	15.96	14.08	5	18	2	9
	VKM	12.61	6.72	2	21	**3**	7

Table 4
Percentage of gained efficiency by the 6-symmetrical (6-sym) design, with respect to designs in the literature and optimal designs constucted with CE algorithm.

		6-sym	
		D	N_t
CE	D-opt	-1.76	83.14
	N_t	37.46	-82.98
	CWR26	-0.46	69.58
	CWR36	0.08	3.67
	CWR38	-1.15	29.57
	EX-EED	20.00	52.29
	MGG	-0.68	0.004
	VKM	20.46	52.29

Table 5
Three designs from the Pareto front of the 6-objectives optimality criterion.

	D	N_t	$df_{PE}(WP)$	$df_{PE}(SP)$	$df_{LoF}(WP)$	$df_{LoF}(SP)$
6-sym	15.85	14.10	5	18	1	9
Design (a)	15.38	10.08	3	21	3	6
Design (b)	14.88	11.04	3	23	3	4

For example, considering CWR36 design, we compute the efficiency gain of the 6-symmetrical design, according to the D- and N_t-criteria. From Table 3, we have that $f_D^{sym} = 15.85$ and $f_{N_t}^{sym} = 14.10$ and the corresponding values for the CWR36 design are $f_D^{\text{CWR26}} = 15.92$ and $f_{N_t}^{\text{CWR26}} = 4.28$. From Table 4, we can see that even though the 6-symmetrical design has a loss of 0.46% in terms of D-efficiency, this is compensated by a gain of 69.58% in N_t-efficiency. By looking at the table, the percentage of efficiency lost by the 6-symmetrical design on the D-criterion is always compensated by an higher percentage gain on N_t-criterion, and vice versa. The MGG and the 6-sym designs are very competitive. The small amount of efficiency lost on the D-criterion is compensated by a slight increase of percentage gain on N_t-criterion. The 6-sym design has a positive percentage of gained efficiency for both criteria for the CWR36, EX-EED and VKM designs.

Analysing the Pareto front, the N_t-criterion seems to be responsible for the low values of $df_{LoF}(WP)$. In fact, if we consider a design with higher values for $df_{LoF}(WP)$ then the N_t-criterion value decreases. The Pareto front for criterion vector (ii) is composed by 3610 designs. Given that, practitioners can select the most suitable design for their applications and/or needs. For example, Table 5 shows two other designs from the Pareto front where the $df_{\text{LoF}}(\text{WP})$ is higher than 1, however some efficiency in terms of D- and N_t-criteria is being sacrificed.

4.2. A 24-run split-plot screening example

In this example, the new version of the MS-TPLS algorithm will be used for the construction of optimal designs for screening experiments. As in Mylona et al. (2020), we consider a 24-run split-plot designs with eight whole plots of three runs for a five-factor main-effects-plus-two-factor-interactions model. For comparison purposes, we consider three different designs. The first design is constructed with the traditional D-optimality criterion for fixed effects only (Mylona et al. (2020)). As pointed out in Mylona et al. (2020), this design is equivalent with the design constructed by using a Bayesian optimal design criterion which focuses on both the variance components and the fixed effects proposed by Mylona et al. (2014). From now on, this design is called OriCC-TradD. The second design is constructed with the alternative Bayesian optimality design criterion proposed in (Mylona et al. (2020)), denoted here impCC. To note here that the Bayesian designs at the original paper were constructed using as prior the log-normal distribution, where its hyperparameters were set under the assumption that the variance ratio η is around 1, and that it highly likely that η is in the interval [0.1, 10]. Also, we consider the more classical design based on a 2_{IV}^{5-1} fractional factorial design enriched with replicated runs as in (Mylona et al. (2020)), named Classical.

Table 6 reports the 6-symmetrical design for the 24-run split-plot screening example. The results in Table 7 demonstrate the main advantage of the 6-symmetrical design from the Pareto Front, which is the way it distributes the eight available degrees of freedom for the error, between subplot and whole-plot effects and between pure error and lack of fit, with a small sacrifice with respect to the D and N_t efficiencies. The results, in both Tables 3 and 7 are in accordance with the remark made in Gilmour and Trinca (2012), that the D-criterion tends to incorporate degrees of freedom for lack-of-fit as opposed to degrees of freedom for pure error. In Table 8, we show that alternative designs can be chosen from the Pareto Front according to the needs of the experimenter.

Table 6
6-symmetrical (6-sym) design from the Pareto Front identified by the modified version of MS-TPL algorithm for the 24-run split-plot screening example.

6-sym				
x_1	x_2	x_3	x_4	x_5
-1	-1	1	1	-1
-1	-1	-1	1	-1
-1	1	-1	-1	-1
1	-1	1	-1	1
1	1	-1	-1	1
1	1	1	-1	-1
1	-1	1	1	-1
1	1	-1	1	-1
1	1	1	-1	1
-1	1	-1	1	1
-1	-1	1	1	-1
-1	1	-1	-1	-1
-1	1	1	1	-1
-1	-1	1	1	1
-1	-1	-1	-1	-1
1	1	1	1	1
1	-1	-1	-1	-1
1	-1	-1	1	1
-1	-1	-1	1	-1
-1	-1	1	-1	-1
-1	1	1	-1	1
-1	-1	-1	-1	1
-1	-1	1	-1	-1
-1	1	1	-1	1

Table 7
Comparison with designs in the literature. Entries in bold correspond to the best values for individual criteria.

		D	N_t	$df_{PE}(WP)$	$df_{PE}(SP)$	$df_{LoF}(WP)$	$df_{LoF}(SP)$
MS-TPLS 6-criteria	6-sym	18.77	0.61	3	2	2	1
	D-opt	19.18	0.11	1	1	5	1
	N_t-opt	15.22	1.13	4	3	1	0
CE	*D*-opt	19.58	0	0	0	**6**	**2**
	N_t-opt	-0.45	**2.21**	4	**5**	0	0
	impCC	19.02	1	**6**	2	0	0
	OriCC-TradD	**19.64**	0	0	0	**6**	**2**
	Classical	18.77	1.78	4	4	0	0

Table 8
Alternative designs from the Pareto front of the 6-objectives optimality criterion

	D	N_t	$df_{PE}(WP)$	$df_{PE}(SP)$	$df_{LoF}(WP)$	$df_{LoF}(SP)$
6-sym	18.77	0.61	3	2	2	1
Design	16.84	0.35	4	1	1	2

5. Discussion

In this work, we extended the Multi-Stratum Two-Phase Local Search (MS-TPLS) algorithm in Borrotti et al. (2017). The extended version is now able to simultaneously optimise any of the following design criteria: D, A, I, I_d, A_s, D_s, N_t, $df_{PE}(WP)$, $df_{PE}(SP)$, $df_{LoF}(WP)$ and $df_{LoF}(SP)$. The new version of the algorithm includes the computation of the information matrix for the variance components assuming that the full treatment model is used for estimating them (N_t). Also it calculates the pure error degrees of freedom for the whole-plots ($df_{PE}(WP)$) and for the subplot effects ($df_{PE}(SP)$) and the lack-of-fit degrees of freedom for the whole-plot ($df_{LoF}(WP)$) and the subplot effects ($df_{LoF}(SP)$). We are now able to construct the Pareto front for split-plot designs that ensure pure-error estimation of the variance components.

Since the final solution is a Pareto front, *i.e.* a set of solutions representing different trade-offs between the optimisation criteria, practitioners can study the proposed solution and select the most suitable design for the problem under study. When one design can be implemented, we showed that the *symmetrical*-design is a good trade-off between all the competitive criteria. In fact, the symmetrical-design is the closest design to the ideal point that optimise all the considered objectives.

The performance of the proposed approach is evaluated considering the two motivating example presented in Vining et al. (2005) and Mylona et al. (2020), respectively. From the literature, we selected a set of designs for comparison purposes. Considering the following vector of criteria (D, N_t, $df_{\mathrm{PE}}(\mathrm{WP})$, $df_{\mathrm{PE}}(\mathrm{SP})$, $df_{\mathrm{LoF}}(\mathrm{WP})$, $df_{\mathrm{LoF}}(\mathrm{SP})$), we pointed out that our symmetrical-design represents the best compromised option with respect to all six criteria.

It is worth mentioning that the computation of $\mathbf{Z}$ and $\mathbf{X_t}$ matrices has lead to an increase of the computational time to find the final Pareto front. For instance considering the motivating example presented in Vining et al. (2005) and the simultaneously optimisation of six criteria, MS-TPLS had run for 1.66×10^4 seconds in order to do 60 restarts of the algorithm, each with 1000 calls to the MS-OPT function. The computational time is slightly lower while considering four criteria (1.25×10^4 seconds). Considering the second motivating example (Mylona et al. (2020)), MS-TPLS is faster to find the six criteria optimal Pareto front running for a total of 1.98×10^3 seconds. In order to understand the impact of the computation of $\mathbf{Z}$ and $\mathbf{X_t}$ matrices, we report the computational time required by MS-OPT function to find the best D- and N_t-optimal design (considering only 1000 restarts) for both the motivating examples. For the first motivating example (Vining et al. (2005)), 131.9 seconds were needed to find the D-optimal design and 1.61×10^3 seconds for finding the N_t-optimal design. In the second motivating example (Mylona et al. (2020)), the D-optimal design was found in only 39.8 seconds and the N_t-optimal design in 153.2 seconds. Since the computation of the $\mathbf{Z}$ and $\mathbf{X_t}$ is involved only in the N_t criterion, it is clear the impact of these two functions on the computational performance. All algorithms were implemented in MATLAB^{TM}, and all experiments were run on a 2.6 GHz Intel Core i5. Future efforts will be devoted to the improvement of the functions implemented for the computation of the $\mathbf{Z}$ and $\mathbf{X_t}$ in order to reduce the computational time required to find the Pareto front.

Furthermore, it is important to underline that the proposed algorithm depends on a set of hyper-parameters that should be tuned. In this work, we set the hyper-parameters following Sambo et al. (2014), however a sensitivity analysis about how the number of scalarisations affects the shape of the Pareto curve could be done as future research direction.

Acknowledgements

We greatly acknowledge the anonymous reviewers and the Editors, for their helpful comments and suggestions.

References

Borrotti, M., Sambo, F., Mylona, K., Gilmour, S., 2017. A multi-objective coordinate-exchange two-phase local search algorithm for multi-stratum experiments. Statistics and Computing 27, 469–481. doi:10.1007/s11222-016-9633-6.
Cao, Y., Robinson, S.S.W..T.J., 2017. Dp-optimality in terms of multiple criteria and its application to the split-plot design. Journal of Quality Technology 49, 2157–2180. doi:10.1080/00224065.2017.11918183.
Dubois-Lacoste, J., López-Ibáñez, M., Stützle, T., 2011. A hybrid TP+PLS algorithm for bi-objective flow-shop scheduling problems. Computers & Operations Research 38 (8), 1219–1236. doi:10.1016/j.cor.2010.10.008.
Gilmour, S., Goos, P., Grossmann, H., 2017. Pure error REML for analyzing data from split-plot and multi-stratum designs. Submitted for publication.
Gilmour, S.G., Trinca, L., 2000. Some practical advice on polynomial regression analysis from blocked response surface designs. Communications in Statistics: Theory and Methods 29, 2157–2180. doi:10.1080/03610920008832601.
Gilmour, S.G., Trinca, L., 2012. Optimum design of experiments for statistical inference. Journal of the Royal Statistical Society Series C-Applied Statistics 61, 345–401. doi:10.1111/j.1467-9876.2011.01000.x.
Goos, P., 2002. The Optimal Design of Blocked and Split- Plot Experiments. New York: Springer doi:10.1007/978-1-4613-0051-9.
Goos, P., Vandebroek, M., 2001. Optimal split-plot designs. Journal of Quality Technology 33, 436–450. doi:10.1080/00224065.2001.11980103.
Goos, P., Vandebroek, M., 2003. D-optimal split-plot designs with given numbers and sizes of whole plots. Technometrics 45, 235–245. doi:10.1198/004017003000000050.
Großmann, H., Gilmour, S.G., 2021. Partially orthogonal blocked three-level response surface designs. Econometrics and Statistics doi:10.1016/j.ecosta.2021.08.007.
Harville, D.A., 1997. Matrix Algebra from A Statistician 's Perspective. New York: Springer doi:10.1007/b98818.
Jones, B., Goos, P., 2007. A candidate-set-free algorithm for generating D-optimal split-plot designs. Journal of the Royal Statistical Society Series C-Applied Statistics 56, 347–364. doi:10.1111/j.1467-9876.2007.00581.x.
Jones, B., Goos, P., 2009. D-optimal design of split-split-plot experiments. Biometrika 96, 67–82. doi:10.1093/biomet/asn070.
Jones, B., Goos, P., 2012. I-optimal versus D-optimal split-plot response surface designs. Journal of Quality Technology 44, 85–101. doi:10.1080/00224065.2012.11917886.
Letsinger, J.D., Myers, R.H., Lentner, M., 1996. Response surface methods for bi-randomization structures. Journal of Quality Technology 28, 381–397. doi:10.1080/00224065.1996.11979697.
Lu, L., Anderson-Cook, C.M., Robinson, T.J., 2011. Optimization of design experiments based on multiple criteria utilizing a Pareto frontier. Technometrics 53 (4), 353–365. doi:10.1198/TECH.2011.10087.
Macharia, H., Goos, P., 2010. D-optimal and d-efficient equivalent-estimation second-order split-plot designs. Journal of Quality Technology 42 (4), 358–372. doi:10.1080/00224065.2010.11917833.
Mylona, K., Gilmour, S., Goos, P., 2020. Optimal blocked and split-plot designs ensuring precise pure-error estimation of the variance components. Technometrics 62, 57–70. doi:10.1080/00401706.2019.1595153.
Mylona, K., Goos, P., Jones, B., 2014. Optimal design of blocked and split-plot experiments for fixed effects and variance component estimation. Technometrics 56, 132–144. doi:10.1080/00401706.2013.818579.
Paquete, L., Stützle, T., 2007. Stochastic local search algorithms for multiobjective combinatorial optimization: A review. In: Gonzalez, T.F. (Ed.), Handbook of Approximation Algorithms and Metaheuristics. Chapman & Hall/CRC, Boca Raton, FL doi:10.5555/1199638.
Sambo, F., Borrotti, M., Mylona, K., 2014. A coordinate-exchange two-phase local search algorithm for the d- and i-optimal design of split-plot experiments. Computational Statistics and Data Analysis 71, 1193–1207. doi:10.1016/j.csda.2013.03.015.
Trinca, L.A., Gilmour, S.G., 2001. Multi-stratum response surface designs. Technometrics 43, 25–33. doi:10.1198/00401700152404291.
Trinca, L.A., Gilmour, S.G., 2015. Improved split-plot and multi-stratum designs. Technometrics 57 (2), 145–154. doi:10.1080/00401706.2014.915235.
Vining, G.G., Kowalski, S.M., Montgomery, D.C., 2005. Response surface designs within a split-plot structure. Journal of Quality Technology 37 (2), 115–129. doi:10.1080/00224065.2005.11980310.
Yi, Y., Wang, X., 2021. A Markov decision process for response adaptive designs. Econometrics and Statistics doi:10.1016/j.ecosta.2021.10.015.