1 Abstract

2 Introduction

Most experiments are conducted in order to gain understanding of the impact that different parameters of interest make on the outcome; the quantitative measure of this impact allows comparisons as well as making interpretable conclusions regarding the scale and the pattern of the relationships between process parameters and measured output. While the exact true nature of that relationship remains unknown, we need to have some form of approximation. Polynomial functions are well known to be able to provide an approximation of any required precision for functions from a certain class of differentiability (Rudin, 1987): obviously, higher desirable precision would require a polynomial of a higher order and, therefore, more experimental effort. Response Surface Methodology (RSM) that was first introduced by Box and Wilson (1951) is aimed at optimising the response by fitting a second order polynomial to the data obtained after an experiment on the points within the experimental region of interest has been performed. The response measured at each of n experimental runs would then be presented as a linear combination of experimental factors' values, their interactions and corresponding quadratic terms; of course, some of the terms might be excluded, but, unless stated otherwise, we will consider a full second-order polynomial model with $1 + 2k + \frac{k(k-1)}{2}$ parameters (including the intercept), where k is the number of factors. Together with the fixed polynomial part, the error term here would stand for all the variation in data that has not been captured by the model. At the stage of experimental planning, when there is little, if any, prior knowledge and no data available, there is still a lot that can and should be done to make sure the results of the experimentation and the following analysis are credible.

The design construction usually depends on the model in several ways. Firstly, the interest in the properties of the model parameters' estimation leads to the development of various optimality criteria (e.g. so-called 'alphabetic criteria'). Secondly, the fact of the model-dependence of the majority of the criteria implies the necessity of taking into account the possible lack-of-fit as well as the desirability of obtaining error variance estimates from replicated observations (model-independent, 'pure' error). The latter two are also in the list of the properties of what would be considered as a 'good design' (summarised by Box and Draper (1987), Chapter 14): it should "make it possible to detect lack of fit" and "provide an internal estimate of error from replication".

It is obvious that all the individual criteria are not interchangeable and can possibly be contradictory, however, they all are desired to be accounted for. There are a few ways of combining them, and here we will be working with the notion of a compound optimality criterion which is basically a weighted combination of the elementary criteria, where weights are generally arbitrary, and are expected to be chosen in accordance with the experimenter's beliefs and intentions.

We will focus in this work on factorial experiments with a relatively small number of runs and the fitted model being a polynomial regression. We work here with exact designs: for each run we are to provide a point as a combination of the treatment factors' levels; therefore, finding the corresponding optimal design is essentially a matrix optimisation problem, i.e. searching for a design maximising (minimising) the criterion function. When considering examples of different compound criteria with various weight allocations, we will examine the resulting optimal designs in terms of their performances with respect to individual components. This would allow evaluating how much is "lost" when achieving a compromise, what component criteria contradict each others' performances, and how they are affected by changes of weight allocations.

The first and main objective of the research presented is to develop a methodology for combining several desirable data inference properties in compound optimality criteria for factorial experiments. The properties are to correspond to the precision of inference based on the primary model and to the possibility of the specified model contamination presence; the components are aimed to comply with the 'pure error' strategy, where it is feasible.

The obvious following objective is to investigate the relationships between the individual components of the introduced criteria and, by considering some examples, provide some empirical recommendations for complex experimentation research that would benefit from applying these methods.

The criteria will also be adapted to be applied within experimental frameworks with restricted randomisation: blocked and multistratum experiments.

3 Background

Recall the polynomial regression model (1) for the unblocked experiment, containing initial factors, their powers and interactions is used as a good approximation to the relation between the variables:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \tag{1}$$

Here $\boldsymbol{X}$ is the $n \times p$ model matrix, $\boldsymbol{Y}$ is the $n \times 1$ vector of responses; $\boldsymbol{\beta}$ is the $p \times 1$ vector of parameters corresponding to the model terms and $\boldsymbol{\varepsilon}$ are independent normally distributed random error terms with constant variance, i.e. $\boldsymbol{\varepsilon} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n)$.

3.1 Pure error

The error variance estimation can be obtained in two ways. The first one is the mean square error: $\hat{\sigma}_{mse}^2 = \text{Residual SS}/(n-p)$, i.e. the residual sum of squares divided by the

corresponding number of degrees of freedom (obtained from the ANOVA decomposition, the details can be found in Draper and Smith (1998)). This estimator obviously depends on the model and on the number of its parameters. The other one is model-independent 'pure' error, derived from the further decomposition of the residual sum of squares into the 'pure' error and 'lack-of-fit' components: $\hat{\sigma}^2_{PE} = $ Pure error SS$/(n-t)$, where t is the number of unique treatments applied and $d = n - t$ is the pure error degrees of freedom, i.e. the number of replicated points. In other words, the error is estimated by fitting the full treatment model:

$$Y = X_t \mu_t + \varepsilon, \tag{2}$$

where X_t is the $n \times t$ full treatment model matrix, in which the $(i,j)^{th}$ element is equal to 1 if treatment j is applied to the ith unit, otherwise it is set to 0. Then the elements of the t-dimensional vector μ_t are the mean effects of each treatment. The vector of errors ε comprises the between-unit variation, such that $\mathrm{E}(\varepsilon) = 0$, $\mathrm{Var}(\varepsilon) = \sigma^2 I_n$.

Many authors advocate the use of the 'pure' error estimate of σ^2 instead of the one pooled with the lack-of-fit part from the model (1): Cox (1958) recommends using it for the estimation unless there are no replicate treatments, while Draper and Smith (1998) argue for the reliability of 'pure' error and recommend aiming for the presence of replicates at the stage of planning. Atkinson *et al.* (2007) also state that "If lack-of-fit of the model is of potential interest, σ^2 is better estimated from replicate observations" (page 22). For response surface experiments in blocks Gilmour and Trinca (2000) explicated the definition of pure error and its estimate which are compatible with the unblocked case: from replicates but also taking into account block effects additive to treatment effects. Finally, the work by Gilmour and Trinca (2012) that inspired this research, comprises a thorough analysis in favour of estimating the error from the full treatment model which is true regardless of what function is used to approximate the relationship of interest.

3.2 Combining criteria

The concept of the compound criteria is based on the notion of design efficiency, which can be defined for any design matrix and any criterion. For example, the D-efficiency of design X is

$$\mathrm{Eff}_D(X) = \left[\frac{|X'X|}{|X'_* X_*|} \right]^{1/p}, \tag{3}$$

where X_* is the D-optimum design. In this definition the power $1/p$ brings the efficiency to the scale of variances of model coefficients $\beta_i, i = 1 \ldots p$. It is obvious that the efficiency value may vary from 0 to 1 and is equal to 1 if and only if the design is optimal according to the criterion of interest.

The final criterion function to be maximised among all the possible designs is obtained

by combining the efficiencies for the component criteria $\text{Eff}_1, \ldots, \text{Eff}_m$ with the corresponding weights $\kappa_1, \ldots, \kappa_m$ such that each $\kappa_k > 0$ and $\sum_{k=1}^{m} \kappa_k = 1$:

$$\text{Eff}_1^{\kappa_1}(\boldsymbol{X}) \times \text{Eff}_2^{\kappa_2}(\boldsymbol{X}) \times \ldots \times \text{Eff}_m^{\kappa_m}(\boldsymbol{X}) \to \max_{\boldsymbol{X}}. \qquad (4)$$

The choice of weights is an arbitrary decision, made relying on both prior knowledge of the experimenter, the objectives of a specific experiment and the interpretation of the criterion components. The most general and intuitively sensible recommendation would be to obtain optimal designs with respect to several weight allocations, and then, after examining them, choose the one to be applied. Throughout the course of this research, when considering the examples of experiments layouts, we chose a set of weight allocations. Most of these sets were essentially classical design schemes for experiments with mixtures (Cornell, 2011). For example, if the criterion consists of three elements, we would first obtain designs with all the weight on each component, then all combinations of distributing it equally between two components, and finally put equal weights on all of the components. Some additional schemes are included when examining particular cases.

Some of the alternative approaches of considering several objectives would include the Pareto frontier approach, thoroughly described, for example, by Lu *et al.* (2011). When there are l individual criterion functions $\phi_1(\boldsymbol{X}), \ldots \phi_l(\boldsymbol{X})$ to be maximised, the Pareto optimal set is generated so that for each of the designs from this set there is no other possible design that provides better criterion values for all the criteria (or, as it was defined, the Pareto optimal set is formed of designs that are not "dominated" by any other design).

Another approach was presented by Stallings and Morgan (2015): the authors developed methodology for generalising eigenvalue-based criteria (e.g. A- and E-optimality) in a way that allows differing interest (expressed through the weights) among any set of estimable functions of the fitted model parameters. The introduced strategy reflects the aims of experimentation that are not traditionally accounted for but definitely are of interest.

4   Component criteria construction

Standard design optimality theory is developed under the assumption that the fitted model represents the true relationship of interest and that there is no misspecification. However, it is obviously quite a strong belief and in reality we need to take into account at least the possibility that the model might not provide a good fit to the data. In Section ?? a brief overview of various types of model misspecification is presented.

In this chapter we consider the case when the fitted polynomial model is nested within

a larger model that is assumed to provide a better fit for the data:

$$\boldsymbol{Y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}, \tag{5}$$

where $\boldsymbol{X}_2$ is an $n \times q$ is an extension of the initial model matrix representing those extra q potential terms that might represent the fitted model disturbance. The vector $\boldsymbol{\beta}_2$ denotes the corresponding parameters. This larger model has $p + q$ parameters in total and not all of them are estimable when the experiment is relatively small: $n < p + q$; this is the case we consider here.

4.1 Lack-of-fit criterion

We consider the Generalised criterion presented by Goos et al.: [more details?]

$$\text{minimise} \ |\boldsymbol{X}_1'\boldsymbol{X}_1|^{-\frac{\kappa_D}{p}} \times \left| \boldsymbol{L} + \frac{\boldsymbol{I}_q}{\tau^2} \right|^{-\frac{\kappa_{LoF}}{q}} \times |\boldsymbol{A}'\boldsymbol{A} + \boldsymbol{I}_q|^{\frac{\kappa_{bias}}{q}},$$

where $\boldsymbol{L}$ is known in literature as dispersion matrix, and the matrix $\boldsymbol{L} + \boldsymbol{I}_q/\tau^2$ in the second component is essentially the inverse posterior variance-covariance matrix of the vector of potential terms (up to a factor of σ^2).

** Derivation first or the Generalised criterion

A diffuse prior shall be put on primary terms (as was done by DuMouchel and Jones (1994)) – an arbitrary mean and a variance going to infinity, and the prior on potential terms is the one specified before: $\boldsymbol{\beta}_2 \sim \mathcal{N}(0, \boldsymbol{\Sigma}_0)$, $\boldsymbol{\Sigma}_0 = \sigma^2\tau^2\boldsymbol{I}_q$, then the posterior distribution of the coefficients is (as in Koch (2007) and DuMouchel and Jones (1994)):

$$\boldsymbol{\beta}|\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{b}, \boldsymbol{\Sigma}), \ \text{where} \ \boldsymbol{b} = \boldsymbol{\Sigma}\boldsymbol{X}'\boldsymbol{Y}, \boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2],$$

$$\boldsymbol{\Sigma} = \left[\frac{\boldsymbol{K}}{\sigma^2\tau^2} + \sigma^{-2}(\boldsymbol{X}'\boldsymbol{X}) \right]^{-1} = \sigma^2[\boldsymbol{K}/\tau^2 + \boldsymbol{X}'\boldsymbol{X}]^{-1}$$

and

$$\boldsymbol{K} = \begin{pmatrix} \boldsymbol{0}_{p\times p} & \boldsymbol{0}_{p\times q} \\ \boldsymbol{0}_{q\times p} & \boldsymbol{I}_{q\times q} \end{pmatrix}.$$

The marginal posterior of $\boldsymbol{\beta}_2$ is $\mathcal{N}(\boldsymbol{b}_2, \boldsymbol{\Sigma}_{22})$, so using the general formula for the inverse of a block matrix

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix}^{-1} = \begin{bmatrix} \dots & \dots \\ \dots & (\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1} \end{bmatrix},$$

we can obtain the expression for $\boldsymbol{\Sigma}_{22}$:

$$[\boldsymbol{K}/\tau^2 + \boldsymbol{X}'\boldsymbol{X}]^{-1} = \begin{bmatrix} \boldsymbol{X}_1'\boldsymbol{X}_1 & \boldsymbol{X}_1'\boldsymbol{X}_2 \\ \boldsymbol{X}_2'\boldsymbol{X}_1 & \boldsymbol{X}_2'\boldsymbol{X}_2 + \boldsymbol{I}_q/\tau^2 \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} \ldots & \ldots \\ \ldots & (\boldsymbol{X}_2'\boldsymbol{X}_2 + \boldsymbol{I}_q/\tau^2 - \boldsymbol{X}_2'\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2)^{-1} \end{bmatrix},$$

$$\boldsymbol{\Sigma}_{22} = \sigma^2[(\boldsymbol{K}/\tau^2 + \boldsymbol{X}'\boldsymbol{X})^{-1}]_{22} = \sigma^2\left(\boldsymbol{L} + \frac{\boldsymbol{I}_q}{\tau^2}\right)^{-1},$$

and

$$(\boldsymbol{\Sigma}_{22})^{-1} = \frac{1}{\sigma^2}\left(\boldsymbol{L} + \frac{\boldsymbol{I}_q}{\tau^2}\right).$$

Therefore, the $(1-\alpha)\times100\%$ confidence region for the potential terms over the posterior distribution, when σ^2 is estimated by s^2 on ν degrees of freedom, is given by (Draper and Smith, 1998):

$$(\boldsymbol{\beta}_2 - \boldsymbol{b}_2)'(\boldsymbol{L} + \boldsymbol{I}_q/\tau^2)(\boldsymbol{\beta}_2 - \boldsymbol{b}_2) \leq qs^2F_{q,\nu;1-\alpha},$$

where $\mathrm{F}(q,\nu,\alpha)$ is the α-quantile of F-distribution with q and ν degrees of freedom.

4.2   MSE-based bias criterion

From the inference point of view, in the case of potential model contamination the bias in the parameters' estimates would be of substantial interest. A natural way of evaluating the quality of $\hat{\boldsymbol{\beta}}_1$, as noted by Montepiedra and Fedorov (1997) is the matrix of mean square error, which is the $\boldsymbol{L}_2$-distance between the true and estimated values with respect to the probability distribution measure of $\boldsymbol{Y}$ under the assumption of model (??):

$$\begin{aligned} \mathrm{MSE}(\hat{\boldsymbol{\beta}}_1|\boldsymbol{\beta}) &= \mathrm{E}_{\boldsymbol{Y}|\boldsymbol{\beta}}[(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1)(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1)'] \\ &= \sigma^2(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1} + \boldsymbol{A}\boldsymbol{\beta}_2\boldsymbol{\beta}_2'\boldsymbol{A}', \end{aligned} \quad (6)$$

where, as before, $\boldsymbol{A} = (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2$ is the alias matrix. The determinant-based criterion function is minimising

$$\mathrm{E}_\beta \log(\det[\mathrm{MSE}(\hat{\boldsymbol{\beta}}_1|\boldsymbol{\beta})]), \quad (7)$$

where by $\boldsymbol{\beta}$ we denote the joint vector of primary and potential model parameters: $\boldsymbol{\beta} = [\boldsymbol{\beta}_1, \boldsymbol{\beta}_2]$. Using the matrix determinant lemma (Harville, 2006)

$$\det[\boldsymbol{P} + \boldsymbol{u}\boldsymbol{v}'] = \det[\boldsymbol{P}]\det[1 + \boldsymbol{v}'\boldsymbol{P}^{-1}\boldsymbol{u}],$$

where $\boldsymbol{P}$ is an invertible square matrix and $\boldsymbol{u}$, $\boldsymbol{v}$ are column vectors, the determinant in (7) can be decomposed:

$$
\begin{aligned}
&\det[\mathrm{MSE}(\hat{\boldsymbol{\beta}}_1|\boldsymbol{\beta}_2)] \\
&= \det[\sigma^2 \boldsymbol{M}^{-1} + \boldsymbol{M}^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2\boldsymbol{\beta}_2\boldsymbol{\beta}_2'\boldsymbol{X}_2'\boldsymbol{X}_1\boldsymbol{M}^{-1}] \\
&= \sigma^{2p} \det[\boldsymbol{M}^{-1} + \boldsymbol{M}^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2\tilde{\boldsymbol{\beta}}_2\tilde{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{X}_1\boldsymbol{M}^{-1}] \\
&= \sigma^{2p} \det[\boldsymbol{M}^{-1}]\det[1 + \tilde{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{X}_1\boldsymbol{M}^{-1}\boldsymbol{M}\boldsymbol{M}^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2\tilde{\boldsymbol{\beta}}_2] \\
&= \sigma^{2p} \det[\boldsymbol{M}^{-1}](1 + \tilde{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{X}_1\boldsymbol{M}^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2\tilde{\boldsymbol{\beta}}_2), \qquad (8)
\end{aligned}
$$
where $\boldsymbol{M} = \boldsymbol{X}_1'\boldsymbol{X}_1$ and $\tilde{\boldsymbol{\beta}}_2 = \boldsymbol{\beta}_2/\sigma$.

Applying the logarithm function gives us the following sum:

$$
\begin{aligned}
&\log(\det[\mathrm{MSE}(\hat{\boldsymbol{\beta}}_1|\boldsymbol{\beta}_2)]) \\
&= p\log\sigma^2 + \log(\det[\boldsymbol{M}^{-1}]) + \log(1 + \tilde{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{X}_1\boldsymbol{M}^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2\tilde{\boldsymbol{\beta}}_2).
\end{aligned}
$$

The first summand does not depend on the design, so it will not be included in the criterion; the second one is the D-optimality component. Therefore, (7) is replaced by

$$
\log(\det[\boldsymbol{M}^{-1}]) + \mathrm{E}_{\tilde{\boldsymbol{\beta}}_2}\log(1 + \tilde{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{X}_1\boldsymbol{M}^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2\tilde{\boldsymbol{\beta}}_2).
$$

The second term needs to be evaluated numerically due to the obvious lack of information regarding the potential terms' coefficients. Note that as $\boldsymbol{\beta}_2 \sim \mathcal{N}(\boldsymbol{0}, \tau^2\sigma^2\boldsymbol{I}_q)$, then $\tilde{\boldsymbol{\beta}}_2 \sim \mathcal{N}(\boldsymbol{0}, \tau^2\boldsymbol{I}_q)$ and its prior distribution does not depend on the unknown σ^2.

To sum up, the resulting criterion function comprising the DP-criterion and the $LoF(DP)$-component from the generalised DP-criterion as well as the MSE-based part, all brought to the scale of efficiencies, is

$$
\begin{aligned}
\text{minimise } &\left[\left|\boldsymbol{X}_1'\boldsymbol{X}_1\right|^{-1/p} F_{p,d;1-\alpha_{DP}}\right]^{\kappa_{DP}} \times \\
&\left[\left|\boldsymbol{L} + \frac{\boldsymbol{I}_q}{\tau^2}\right|^{-1/q} F_{q,d;1-\alpha_{LoF}}\right]^{\kappa_{LoF}} \times \\
&\left[\left|\boldsymbol{X}_1'\boldsymbol{X}_1\right|^{-1} \exp\left(\frac{1}{N}\sum_{i=1}^{N}\log(1 + \tilde{\boldsymbol{\beta}}_{2i}'\boldsymbol{X}_2'\boldsymbol{X}_1\boldsymbol{M}^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2\tilde{\boldsymbol{\beta}}_{2i})\right)\right]^{\kappa_{MSE}/p} \qquad (9)
\end{aligned}
$$

Here $\tilde{\boldsymbol{\beta}}_{2i}$, $i = 1..N$ – independent random vectors sampled from $\mathcal{N}(\boldsymbol{0}, \tau^2\sigma^2\boldsymbol{I}_q)$.

4.2.1 Point-based estimate

One of the alternatives to the MC approach we would be willing to consider is to use the point prior for $\boldsymbol{\beta}_2$, that is $\boldsymbol{\beta}_2 = \sigma\tau\boldsymbol{1}_q$ (where $\boldsymbol{1}_q$ is a q-dimensional vector with all components being equal to 1), which is the standard deviation of the initial (normal)

prior. Therefore, $\tilde{\boldsymbol{\beta}}_2 = \tau\mathbf{1}_q$, and the MSE(D)-part is

$$|\boldsymbol{X}_1'\boldsymbol{X}_1|^{-1}(1 + \tilde{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{X}_1\boldsymbol{M}^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2\tilde{\boldsymbol{\beta}}_2)$$
$$=|\boldsymbol{X}_1'\boldsymbol{X}_1|^{-1}(1 + \tau^2\mathbf{1}_q'\boldsymbol{X}_2'\boldsymbol{X}_1\boldsymbol{M}^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2\mathbf{1}_q)$$
$$=|\boldsymbol{X}_1'\boldsymbol{X}_1|^{-1}\left(1 + \tau^2\sum_{i,j=1}^{q}[\boldsymbol{X}_2'\boldsymbol{X}_1\boldsymbol{M}^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2]_{(i,j)}\right),$$

so essentially the last part of the equation is the summation of matrix elements, which obviously takes less computational time even for large dimensions q.

The normal prior for $\boldsymbol{\beta}_2$ shall be preserved in the lack-of-fit component in (9), and so the criterion with the point prior estimate of the MSE(D)-component is the following function:

$$\text{minimise } \left[|\boldsymbol{X}_1'\boldsymbol{X}_1|^{-1/p} F_{p,d;1-\alpha_{DP}}\right]^{\kappa_{DP}} \times$$
$$\left[\left|\boldsymbol{L} + \frac{\boldsymbol{I}_q}{\tau^2}\right|^{-1/q} F_{q,d;1-\alpha_{LoF}}\right]^{\kappa_{LoF}} \times$$
$$\left[|\boldsymbol{X}_1'\boldsymbol{X}_1|^{-1}\left(1 + \tau^2\sum_{i,j=1}^{q}[\boldsymbol{X}_2'\boldsymbol{X}_1\boldsymbol{M}^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2]_{(i,j)}\right)\right]^{\kappa_{MSE}/p} \quad (10)$$

Later we will refer to this as the "MSE(D)P"-criterion, and in Section ?? we will explore how much efficiency is lost when it is used instead of the computationally expensive original MSE(D)-based criterion (9).

To infer the trace-based form of the criterion, we use the pseudo-Bayesian approach to calculating the trace function of the MSE matrix (6):

$$\text{E}_{\beta_2}\text{trace}[\text{MSE}(\hat{\boldsymbol{\beta}}_1|\boldsymbol{\beta}_2)] = \text{trace}[\text{E}_{\beta_2}\text{MSE}(\hat{\boldsymbol{\beta}}_1|\boldsymbol{\beta}_2)] =$$
$$= \text{trace}[\sigma^2(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1} + \text{E}_{\beta_2}(\boldsymbol{A}\boldsymbol{\beta}_2\boldsymbol{\beta}_2'\boldsymbol{A}')] =$$
$$= \text{trace}[\sigma^2(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1} + \sigma^2\tau^2\boldsymbol{A}\boldsymbol{A}'] =$$
$$= \sigma^2\text{trace}[(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1} + \tau^2\boldsymbol{A}\boldsymbol{A}']$$
$$= \sigma^2[\text{trace}\{(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\} + \tau^2\text{trace}\boldsymbol{A}\boldsymbol{A}'].$$

The transitions above are possible due to commutativity of the operations of trace and expectation. The main advantage is the absence of the necessity of any additional numerical evaluations, and in this case of the trace-based criterion using the point prior for $\boldsymbol{\beta}_2$ defined earlier would lead to the same function. By minimising the whole function the sum of primary terms' variance and the expected squared norm of the bias vector in the direction of the potential terms are minimised simultaneously. The second part here contains the scaling parameter τ^2 which regulates the magnitude of the potential terms' variation relatively to the error variance.

Joining up the derived criterion function, the LP-criterion and the $LoF(LP)$-component from the generalised LP-criterion (**??**), we obtain the compound criterion:

$$\text{minimise } \left[\frac{1}{p}\text{trace}(\boldsymbol{W}\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}F_{1,d;1-\alpha_{LP}}\right]^{\kappa_{LP}} \times$$

$$\left[\frac{1}{q}\text{trace}\left(\boldsymbol{L}+\frac{\boldsymbol{I}_q}{\tau^2}\right)^{-1}F_{1,d;1-\alpha_{LoF}}\right]^{\kappa_{LoF}} \times$$

$$\left[\frac{1}{p}\text{trace}\{(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}+\tau^2\boldsymbol{A}\boldsymbol{A}'\}\right]^{\kappa_{MSE}}. \tag{11}$$

5   Examples

Here we will study the optimal designs in terms of the criteria (9), (10) and (11) (referred to as $MSE(D)$, $MSE(D)P$ and $MSE(L)$ criteria respectively) in the framework a factorial experiment, with 5 factors, each is at three levels. The small number of runs (40) allows estimation of the full-second order polynomial model ($p = 21$), but we assume that the 'true' model contains also all third-order terms (linear-by-linear-by-linear and quadratic-by-linear interactions), $q = 30$ of them in total.. [Some comparison with other criteria is to be added]

The intercept is a nuisance parameter, and so the criteria are adapted in such a way that the DP- and LP-components are replaced by DPs and LPs. After considering the mean square error matrix (6) for $p-1$ parameters, in the MSE-based components the full information matrix is replaced by the one excluding the intercept – $\boldsymbol{X}_1'\boldsymbol{Q}_0\boldsymbol{X}_1$ (see (**??**)), where $\boldsymbol{X}_1$ is the model matrix without the intercept, $\boldsymbol{Q}_0 = \boldsymbol{I}_n - \frac{1}{n}\boldsymbol{1}\boldsymbol{1}'$, and $\tilde{\boldsymbol{M}}$ is the full information matrix for the model with the intercept. Otherwise, the procedure of obtaining the $MSE(D)$- and $MSE(L)$-based components remains the same.

Below are expressions of the MSE-based criteria functions that have been amended according to the intercept exclusion from the set of parameters of interest:

MSE(D): $\left[\left|\boldsymbol{X}_1'\boldsymbol{Q}_0\boldsymbol{X}_1\right|^{-1/(p-1)}F_{p-1,d;1-\alpha_{DP}}\right]^{\kappa_{DP}} \times$

$\left[\left|\boldsymbol{L}+\frac{\boldsymbol{I}_q}{\tau^2}\right|^{-1/q}F_{q,d;1-\alpha_{LoF}}\right]^{\kappa_{LoF}} \times$

$\left[|\boldsymbol{X}_1'\boldsymbol{Q}_0\boldsymbol{X}_1|^{-1}\exp\left(\frac{1}{N}\sum_{i=1}^{N}\log(1+\tilde{\boldsymbol{\beta}}_{2i}'\boldsymbol{X}_2'\boldsymbol{Q}_0\boldsymbol{X}_1\tilde{\boldsymbol{M}}^{-1}\boldsymbol{X}_1'\boldsymbol{Q}_0\boldsymbol{X}_2\tilde{\boldsymbol{\beta}}_{2i})\right)\right]^{\frac{\kappa_{MSE}}{(p-1)}},$

$$\text{MSE(D)P: } \left[\left| \boldsymbol{X}_1' \boldsymbol{Q}_0 \boldsymbol{X}_1 \right|^{-1/(p-1)} F_{p-1,d;1-\alpha_{DP}} \right]^{\kappa_{DP}} \times$$

$$\left[\left| \boldsymbol{L} + \frac{\boldsymbol{I}_q}{\tau^2} \right|^{-1/q} F_{q,d;1-\alpha_{LoF}} \right]^{\kappa_{LoF}} \times$$

$$\left[|\boldsymbol{X}_1' \boldsymbol{Q}_0 \boldsymbol{X}_1|^{-1} \left(1 + \tau^2 \sum_{i,j=1}^{q} [\boldsymbol{X}_2' \boldsymbol{Q}_0 \boldsymbol{X}_1 \tilde{\boldsymbol{M}}^{-1} \boldsymbol{X}_1' \boldsymbol{Q}_0 \boldsymbol{X}_2]_{(i,j)} \right) \right]^{\frac{\kappa_{MSE}}{(p-1)}},$$

$$\text{MSE(L): } \left[\frac{1}{p} \text{trace}(\boldsymbol{W} \boldsymbol{X}_1' \boldsymbol{Q}_0 \boldsymbol{X}_1)^{-1} F_{1,d;1-\alpha_{LP}} \right]^{\kappa_{LP}} \times$$

$$\left[\frac{1}{q} \text{trace} \left(\boldsymbol{L} + \frac{\boldsymbol{I}_q}{\tau^2} \right)^{-1} F_{1,d;1-\alpha_{LoF}} \right]^{\kappa_{LoF}} \times$$

$$\left[\frac{1}{p-1} \text{trace}\{ \tilde{\boldsymbol{M}}^{-1} + \tau^2 \boldsymbol{A} \boldsymbol{A}' \}_{22} \right]^{\kappa_{MSE}}.$$

In the third component of the MSE(L)-criterion the $\{ \tilde{\boldsymbol{M}}^{-1} + \tau^2 \boldsymbol{A} \boldsymbol{A}' \}_{22}$ stands for the submatrix corresponding to the parameters of interest, i.e. the first column and the first row are to be excluded. It is worth noting that these criteria are a special case of a blocked experiment framework (with one block), which will be explicitly described in the next chapter.

Considering two values of the variance scaling parameter $\tau^2 = 1$ and $\tau^2 = 1/q$, for each compound criterion we obtain two sets of optimal designs, given in Tables 1, 3 and 4.

The first three columns of the tables contain information about the weight distribution among the criterion components, so that every row corresponds to the design that has been obtained as optimal according to the criterion with the given weights. The next two columns show how the available residual degrees of freedom were allocated between the pure error and lack-of-fit components. Finally, the efficiencies of the resulting designs with respect to the individual criterion components are given in the last columns. In the case of the $MSE(D)P$-optimal designs, in Table 4, the difference between their $MSE(D)$-efficiency values and the $MSE(D)$-efficiencies of the corresponding designs in Table 1 indicates how much we lose in terms of the performance when the point prior is used for the $MSE(D)$-component estimation.

It is worth noting that when this criterion part is being estimated using a stochastic method (MC in this case), despite the error being small, the efficiency values are still approximate. Another disadvantage is the increased time consumption of the algorithm in comparison to the previously considered criteria: with 150 random starts, it can take up to several days to obtain an $MSE(D)$-optimal design; whilst $MSE(D)P$- and $MSE(L)$-based optimal designs are usually obtained within 4–8 hours, even with 500 random starts.

The resulting designs tend to have a lot of the available degrees of freedom allocated to the pure error, except, for example, for the $MSE(D)$-optimal designs (#3). However, for the smaller value of $\tau^2 = 1/q$ the imbalance is more extreme for the determinant-

Table 1: Properties of MSE(D)-optimal designs

	Criteria, $\tau^2 = 1$			DoF		Efficiency,%					
	DP	LoF(DP)	MSE(D)	PE	LoF	DP	LoF(DP)	MSE(D)	LP	LoF(LP)	MSE(L)
1	1	0	0	18	1	100.00	47.77	91.05	96.38	94.92	10.70
2	0	1	0	8	11	43.70	100.00	54.74	0.75	89.99	2.08
3	0	0	1	0	19	0.00	0.00	100.00	0.00	0.00	22.32
4	0.5	0.5	0	11	8	78.50	87.61	88.56	73.88	98.85	17.66
5	0.5	0	0.5	15	4	97.26	56.51	93.77	97.55	50.04	12.74
6	0	0.5	0.5	8	11	64.72	96.84	87.53	57.04	36.17	29.33
7	1/3	1/3	1/3	10	9	79.45	84.14	93.23	81.06	43.42	16.71
8	0.5	0.25	0.25	13	6	93.38	64.35	95.55	95.76	48.58	14.77
9	0.25	0.5	0.25	9	10	69.52	95.76	87.36	63.13	40.41	25.46
	Criteria, $\tau^2 = 1/q$			DoF		Efficiency,%					
	DP	LoF(DP)	MSE(D)	PE	LoF	DP	LoF(DP)	MSE(D)	LP	LoF(LP)	MSE(L)
1	1	0	0	18	1	100.00	94.41	90.60	96.42	98.36	44.94
2	0	1	0	16	3	39.66	100.00	37.95	0.13	100.00	0.12
3	0	0	1	0	19	0.00	0.00	100.00	0.00	0.00	77.97
4	0.5	0.5	0	18	1	100.00	94.41	90.60	96.42	98.36	44.94
5	0.5	0	0.5	17	2	97.31	91.28	93.98	96.21	94.32	50.53
6	0	0.5	0.5	15	4	96.10	92.31	93.29	99.48	95.09	57.05
7	1/3	1/3	1/3	18	1	100.00	94.41	90.60	96.42	98.36	44.94
8	0.5	0.25	0.25	18	1	100.00	94.41	90.60	96.42	98.36	44.94
9	0.25	0.5	0.25	18	1	99.96	94.33	90.65	96.24	98.31	44.84

based criteria, where almost no degrees of freedom are left for lack of fit.

There are no repeated designs for $\tau^2 = 1$, but for $\tau^2 = 1/q$, in Table 1, designs #4, #7 and #8 are the same as the DP-optimal design, and have quite large values of other efficiencies, and design #9, although it is different, has quite similar efficiency values.

The DP-optimal design is given in Table 2; it performs well in terms of the $MSE(D)$-component for both values of τ^2, however, its $LoF(DP)$-efficiency drops by roughly half when the scaling parameter goes from $1/q$ to 1; also, $LoF(DP)$-optimal designs provide the lowest DP-efficiency values (around 40%) for any value of τ^2.

Table 2: MSE(D)-criterion, DP-optimal design

1	-1	-1	-1	-1	-1		21	0	-1	-1	1	0
2	-1	-1	-1	-1	-1		22	0	0	0	-1	1
3	-1	-1	0	1	1		23	0	0	0	-1	1
4	-1	-1	1	-1	1		24	0	1	-1	0	-1
5	-1	-1	1	-1	1		25	1	-1	-1	-1	1
6	-1	-1	1	1	-1		26	1	-1	-1	-1	1
7	-1	-1	1	1	-1		27	1	-1	-1	1	-1
8	-1	0	-1	0	1		28	1	-1	-1	1	-1
9	-1	0	-1	0	1		29	1	-1	1	-1	-1
10	-1	1	-1	-1	1		30	1	-1	1	-1	-1
11	-1	1	-1	-1	1		31	1	-1	1	1	1
12	-1	1	-1	1	-1		32	1	-1	1	1	1
13	-1	1	-1	1	-1		33	1	0	-1	-1	0
14	-1	1	0	0	0		34	1	1	-1	-1	-1
15	-1	1	0	0	0		35	1	1	-1	1	1
16	-1	1	1	-1	-1		36	1	1	-1	1	1
17	-1	1	1	-1	-1		37	1	1	1	-1	1
18	-1	1	1	1	1		38	1	1	1	-1	1
19	-1	1	1	1	1		39	1	1	1	1	-1
20	0	-1	-1	1	0		40	1	1	1	1	-1

As for the trace-based criterion and the optimal designs studied in Table 3, in gen-

eral, all of them tend to have larger LP- and $MSE(L)$-efficiencies in case of smaller τ^2, i.e. the decreased scale of the potentially missed contamination results in a more easily achievable compromise between the contradicting parts of the criteria (the same happens with the trace-based efficiencies of the MSE determinant-based optimal designs). It is also notable that, as it was observed in the case of generalised criteria, the $LoF(LP)$-optimal design is also $LoF(DP)$-optimal for $\tau^2 = 1/q$ (design #2 in the corresponding tables).

The $MSE(L)$-component seems to be much more sensitive to the weight allocations than the $MSE(D)$ component. For example, in the case of $\tau^2 = 1$ some decent efficiency values are gained only when the whole weight is on the 'potential terms' criterion components, i.e. designs #3 and #6.

It does not seem to work agreeably with the $LoF(LP)$-component either: the $LoF(LP)$-optimal design is 0.00% LP- and $MSE(L)$-efficient in the case of $\tau^2 = 1$, and the efficiencies are close to 0% in the case of smaller τ^2.

Table 3: Properties of MSE(L)-optimal designs

	Criteria, $\tau^2 = 1$			DoF		Efficiency,%					
	LP	LoF(LP)	MSE(L)	PE	LoF	DP	LoF(DP)	MSE(D)	LP	LoF(LP)	MSE(L)
1	1	0	0	16	3	97.54	53.86	92.54	100.00	96.87	12.08
2	0	1	0	13	6	35.43	81.99	36.72	0.00	100.00	0.00
3	0	0	1	4	15	18.67	38.43	51.84	11.73	34.79	100.00
4	0.5	0.5	0	15	4	95.14	60.37	92.78	99.80	98.12	13.99
5	0.5	0	0.5	12	7	77.77	72.05	84.91	81.10	98.72	25.19
6	0	0.5	0.5	9	10	36.80	70.91	51.12	28.13	91.60	83.52
7	1/3	1/3	1/3	11	8	69.53	73.16	79.71	70.59	97.47	27.98
8	0.5	0.25	0.25	12	7	77.20	72.83	84.44	81.47	98.80	23.88
9	0.25	0.5	0.25	12	7	70.90	69.80	78.15	72.16	98.49	26.19
	Criteria, $\tau^2 = 1/q$			DoF		Efficiency,%					
	LP	LoF(LP)	MSE(L)	PE	LoF	DP	LoF(DP)	MSE(D)	LP	LoF(LP)	MSE(L)
1	1	0	0	16	3	97.54	92.13	92.18	100.00	95.61	52.55
2	0	1	0	16	3	39.66	100.00	37.95	0.13	100.00	0.12
3	0	0	1	3	16	0.77	0.93	83.48	0.02	0.01	100.00
4	0.5	0.5	0	17	2	96.87	94.29	89.84	97.97	97.80	51.17
5	0.5	0	0.5	12	7	79.66	87.18	86.32	84.81	88.22	79.60
6	0	0.5	0.5	13	6	76.46	89.87	80.78	79.23	91.48	79.11
7	1/3	1/3	1/3	13	6	81.53	90.09	85.52	85.96	91.56	76.65
8	0.5	0.25	0.25	15	4	90.60	91.94	88.43	95.09	94.75	63.82
9	0.25	0.5	0.25	15	4	84.12	92.88	83.39	87.86	95.51	72.66

Regarding the designs' performances with respect to the DP- and LP-components, it can be observed that the designs tend to be quite DP-efficient, overall more efficient in the case of $MSE(D)$-efficient designs with smaller τ^2. DP-efficient designs are not bad in terms of LP-efficiency and vice versa; again, the same cannot be said for the lack-of-fit components and seems not to be true at all for the MSE components, especially, for the $MSE(L)$-optimal design when $\tau^2 = 1$.

LP- and $MSE(L)$-components seem to be in a conflict, though for $\tau^2 = 1/q$ $MSE(L)$-efficiency values are stable across the designs, therefore, making it possible to find compromise designs which at least perform not too badly with regard to both of these criterion parts; for example, designs #5 – #7 are more than 75%-efficient (Table 3).

Table 4 provides the performance summary of the designs constructed in the same way as the designs given in Table 1, but with the point prior of $\boldsymbol{\beta}_2 = \sigma\tau\mathbf{1}_q$ used for the estimation of the $MSE(D)$-component of the criterion. This approach allowed reducing the computational times considerably, from as long as 2 days to just a few hours, with other parameters being equal.

Table 4: Properties of MSE(D)-optimal designs, with point prior

| | Criteria, $\tau^2 = 1$ | | | DoF | | Efficiency,% | | | | | | |
	DP	LoF(DP)	MSE(D)P	PE	LoF	DP	LoF(DP)	MSE(D)	MSE(D)P	LP	LoF(LP)	MSE(L)
1	1	0	0	18	1	100.00	47.77	91.05	90.06	96.38	94.92	10.70
2	0	1	0	8	11	43.70	100.00	54.74	53.51	0.75	89.99	2.08
3	0	0	1	0	19	0.00	0.00	98.18	100.00	0.00	0.00	23.72
4	0.5	0.5	0	11	8	78.50	87.61	88.56	87.04	73.88	98.85	17.66
5	0.5	0	0.5	18	1	100.00	47.77	90.70	93.28	96.38	94.92	10.70
6	0	0.5	0.5	8	11	65.84	94.78	88.51	88.72	58.15	89.54	23.63
7	1/3	1/3	1/3	11	8	83.32	80.22	92.71	94.22	84.52	98.34	17.21
8	0.5	0.25	0.25	13	6	92.07	66.67	94.27	94.92	94.02	98.91	13.88
9	0.25	0.5	0.25	9	10	69.69	95.64	87.48	86.46	61.06	94.29	25.30

| | Criteria, $\tau^2 = 1/q$ | | | DoF | | Efficiency,% | | | | | | |
	DP	LoF(DP)	MSE(D)P	PE	LoF	DP	LoF(DP)	MSE(D)	MSE(D)P	LP	LoF(LP)	MSE(L)
1	1	0	0	18	1	100.00	94.41	90.60	90.23	96.42	98.36	44.94
2	0	1	0	16	3	39.66	100.00	37.95	37.87	0.13	100.00	0.12
3	0	0	1	0	19	0.00	0.00	98.30	100.00	0.00	0.00	77.96
4	0.5	0.5	0	18	1	100.00	94.41	90.60	90.23	96.42	98.36	44.94
5	0.5	0	0.5	18	1	100.00	94.41	90.60	90.23	96.42	98.36	44.94
6	0	0.5	0.5	15	4	96.16	92.23	92.85	95.16	96.19	95.04	51.06
7	1/3	1/3	1/3	18	1	99.98	94.34	90.64	93.04	95.95	98.31	44.71
8	0.5	0.25	0.25	18	1	99.98	94.34	90.64	93.04	95.95	98.31	44.71
9	0.25	0.5	0.25	18	1	98.84	94.72	89.71	92.34	94.92	98.59	46.43

The $MSE(D)$ column contains the estimation of the $MSE(D)$-efficiency based on the MC calculation with the sample size of 1000. The main observation to be made here is that the resulting designs perform very similarly to the designs seen above, and the maximum loss in the MSE(D)-efficiency does not exceed 2.5% across both values of τ^2; losses in the efficiencies with respect to the $MSE(D)$-optimal designs presented in Table 1 are also very small and do not exceed 3.5%.

The observed relationships between the components of the criteria provide some ideas on how the designs obtained as optimal with respect to the various allocations of weights differ, what they have in common, and how the components interact when the weights are reallocated. A compromise can be found between the criterion components corresponding to the properties of the primary terms and the components responsible for the reducing the negative impact from the assumed potential model misspecification. Thereby, the careful choice of the prior parameters (τ^2 in this case) turned out to be of a certain importance as well.

6   Criteria for blocked experiments

We will start by reviewing the 'full' model for blocked experiments, similar to (??):

$$\boldsymbol{Y} = \boldsymbol{Z}\boldsymbol{\beta}_B + \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}.$$

Denote the $n \times (b+p)$ model matrix of the block and primary terms by $\tilde{X}_1 = [Z, X_1]$, where the columns of Z contain block indicators, and the columns of X_1 correspond to primary terms. Let also $\tilde{\beta}_1 = [\beta_B, \beta_1]'$ be the joint vector of fixed block effects and primary model terms, and by $\hat{\tilde{\beta}}_1$ we denote the vector of the corresponding estimates. It is worth noting that the number of primary terms p does not include the intercept, as it is aliased with the block effects.

Then the overall mean square error matrix is

$$
\begin{aligned}
\text{MSE}(\hat{\tilde{\beta}}_1 | \tilde{\beta}) =& \text{E}_{Y|\beta}[(\hat{\tilde{\beta}}_1 - \tilde{\beta}_1)(\hat{\tilde{\beta}}_1 - \tilde{\beta}_1)'] \\
=& \sigma^2 (\tilde{X}_1' \tilde{X}_1)^{-1} + \tilde{A} \beta_2 \beta_2' \tilde{A}',
\end{aligned}
\tag{12}
$$

where $\tilde{A} = (\tilde{X}_1' \tilde{X}_1)^{-1} \tilde{X}_1' X_2$ is the alias matrix in the case of a blocked experiment.

Now consider the partition of the MSE matrix with respect to block and primary effects:

$$
\begin{aligned}
&\text{E}_{Y|\beta}[(\hat{\tilde{\beta}}_1 - \tilde{\beta}_1)(\hat{\tilde{\beta}}_1 - \tilde{\beta}_1)'] = \\
&\text{E}_{Y|\beta}\{[\hat{\tilde{\beta}}_{11} - \tilde{\beta}_{11}, \ldots, \hat{\tilde{\beta}}_{1b} - \tilde{\beta}_{1b}, \hat{\tilde{\beta}}_{1b+1} - \tilde{\beta}_{1b+1}, \ldots, \hat{\tilde{\beta}}_{1b+p} - \tilde{\beta}_{1b+p}] \times \\
&[\hat{\tilde{\beta}}_{11} - \tilde{\beta}_{11}, \ldots, \hat{\tilde{\beta}}_{1b} - \tilde{\beta}_{1b}, \hat{\tilde{\beta}}_{1b+1} - \tilde{\beta}_{1b+1}, \ldots, \hat{\tilde{\beta}}_{1b+p} - \tilde{\beta}_{1b+p}]'\} = \\
&\text{E}_{Y|\beta}\{[\hat{\beta}_B - \beta_B, \hat{\beta}_1 - \beta_1][\hat{\beta}_B - \beta_B, \hat{\beta}_1 - \beta_1]'\} = \\
&\begin{bmatrix}
\text{E}_{Y|\beta}(\hat{\beta}_B - \beta_B)(\hat{\beta}_B - \beta_B)' & \text{E}_{Y|\beta}(\hat{\beta}_B - \beta_B)(\hat{\beta}_1 - \beta_1)' \\
\text{E}_{Y|\beta}(\hat{\beta}_1 - \beta_1)(\hat{\beta}_B - \beta_B)' & \text{E}_{Y|\beta}(\hat{\beta}_1 - \beta_1)(\hat{\beta}_1 - \beta_1)'
\end{bmatrix}.
\end{aligned}
\tag{13}
$$

The $[2,2]$-submatrix (bottom-right) represents the entity we are interested in – the bias of the estimators of the primary terms. Denote it by $\text{MSE}(\hat{\tilde{\beta}}_1 | \tilde{\beta})_{22}$. Then the corresponding submatrix of the first summand in (12) is

$$
[\sigma^2 (\tilde{X}_1' \tilde{X}_1)^{-1}]_{22} = \sigma^2 (X_1' Q X_1)^{-1}, \text{ where } Q = I - Z(Z'Z)^{-1} Z'.
$$

Using the matrix inversion rule for block matrices (Harville, 2006):

$$
\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} =
\begin{bmatrix}
(A - BD^{-1}C)^{-1} & -(A - BD^{-1}C)^{-1} BD^{-1} \\
-D^{-1}C(A - BD^{-1}C)^{-1} & D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1} BD^{-1}
\end{bmatrix},
$$

under the conditions of invertability of A, D and $A - BD^{-1}C$, we now consider $\tilde{A}$:

$$
\begin{aligned}
\tilde{A} =& \left(\begin{bmatrix} Z' \\ X_1' \end{bmatrix} \begin{bmatrix} Z & X_1 \end{bmatrix} \right)^{-1} \begin{bmatrix} Z' \\ X_1' \end{bmatrix} X_2 = \begin{pmatrix} Z'Z & Z'X_1 \\ X_1'Z & X_1'X_1 \end{pmatrix}^{-1} \begin{bmatrix} Z' \\ X_1' \end{bmatrix} X_2 = \\
&\begin{bmatrix}
(Z'PZ)^{-1} & -(Z'PZ)^{-1} Z'X_1 (X_1'X_1)^{-1} \\
-(X_1'QX_1)^{-1} X_1'Z(Z'Z)^{-1} & (X_1'QX_1)^{-1}
\end{bmatrix} \begin{bmatrix} Z' \\ X_1' \end{bmatrix} X_2 = \\
&\begin{bmatrix}
(Z'PZ)^{-1} Z'PX_2 \\
(X_1'QX_1)^{-1} X_1'QX_2
\end{bmatrix},
\end{aligned}
$$

where $\boldsymbol{P} = \boldsymbol{I} - \boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'$ is symmetric.

Here $\boldsymbol{Z}\boldsymbol{Z}'$, $\boldsymbol{X}_1'\boldsymbol{X}_1$ and $\boldsymbol{Z}'\boldsymbol{P}\boldsymbol{Z}$ are all invertible and, therefore, the operations are legitimate.

Now consider the second summand in (12):

$$\tilde{\boldsymbol{A}}\boldsymbol{\beta}_2\boldsymbol{\beta}_2'\tilde{\boldsymbol{A}}' = \begin{bmatrix} (\boldsymbol{Z}'\boldsymbol{P}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{P}\boldsymbol{X}_2\boldsymbol{\beta}_2 \\ (\boldsymbol{X}_1'\boldsymbol{Q}\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{Q}\boldsymbol{X}_2\boldsymbol{\beta}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_2'\boldsymbol{X}_2'\boldsymbol{P}\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{P}\boldsymbol{Z})^{-1} & \boldsymbol{\beta}_2'\boldsymbol{X}_2'\boldsymbol{Q}\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{Q}\boldsymbol{X}_1)^{-1} \end{bmatrix} =$$

$$\begin{bmatrix} (\boldsymbol{Z}'\boldsymbol{P}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{P}\boldsymbol{X}_2\boldsymbol{\beta}_2\boldsymbol{\beta}_2'\boldsymbol{X}_2'\boldsymbol{P}\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{P}\boldsymbol{Z})^{-1} & (\boldsymbol{Z}'\boldsymbol{P}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{P}\boldsymbol{X}_2\boldsymbol{\beta}_2\boldsymbol{\beta}_2'\boldsymbol{X}_2'\boldsymbol{Q}\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{Q}\boldsymbol{X}_1)^{-1} \\ (\boldsymbol{X}_1'\boldsymbol{Q}\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{Q}\boldsymbol{X}_2\boldsymbol{\beta}_2\boldsymbol{\beta}_2'\boldsymbol{X}_2'\boldsymbol{P}\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{P}\boldsymbol{Z})^{-1} & (\boldsymbol{X}_1'\boldsymbol{Q}\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{Q}\boldsymbol{X}_2\boldsymbol{\beta}_2\boldsymbol{\beta}_2'\boldsymbol{X}_2'\boldsymbol{Q}\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{Q}\boldsymbol{X}_1)^{-1} \end{bmatrix}.$$

Therefore, the submatrix of (12) corresponding to the primary terms is

$$\begin{aligned} \mathrm{MSE}(\hat{\tilde{\boldsymbol{\beta}}}_1|\tilde{\boldsymbol{\beta}})_{22} =& \sigma^2(\boldsymbol{X}_1'\boldsymbol{Q}\boldsymbol{X}_1)^{-1} + (\boldsymbol{X}_1'\boldsymbol{Q}\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{Q}\boldsymbol{X}_2\boldsymbol{\beta}_2\boldsymbol{\beta}_2'\boldsymbol{X}_2'\boldsymbol{Q}\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{Q}\boldsymbol{X}_1)^{-1} = \\ & \sigma^2\tilde{\boldsymbol{M}}^{-1} + \tilde{\boldsymbol{M}}^{-1}\boldsymbol{X}_1'\boldsymbol{Q}\boldsymbol{X}_2\boldsymbol{\beta}_2\boldsymbol{\beta}_2'\boldsymbol{X}_2'\boldsymbol{Q}\boldsymbol{X}_1\tilde{\boldsymbol{M}},^{-1} \end{aligned} \tag{14}$$

where $\tilde{\boldsymbol{M}} = \boldsymbol{X}_1'\boldsymbol{Q}\boldsymbol{X}_1$.

6.1 Determinant-based criterion

As in the unblocked case (7), we construct the determinant-based criterion function to be minimised as the expectation of the logarithm of the determinant of the submatrix of the MSE matrix which was outlined above. First we see how its determinant looks:

$$\begin{aligned} \det[\mathrm{MSE}(\hat{\tilde{\boldsymbol{\beta}}}_1|\tilde{\boldsymbol{\beta}})_{22}] =& \det[\sigma^2\tilde{\boldsymbol{M}}^{-1} + \tilde{\boldsymbol{M}}^{-1}\boldsymbol{X}_1'\boldsymbol{Q}\boldsymbol{X}_2\boldsymbol{\beta}_2\boldsymbol{\beta}_2'\boldsymbol{X}_2'\boldsymbol{Q}\boldsymbol{X}_1\tilde{\boldsymbol{M}}^{-1}] = \\ & \sigma^{2p}\det[\tilde{\boldsymbol{M}}^{-1} + \tilde{\boldsymbol{M}}^{-1}\boldsymbol{X}_1'\boldsymbol{Q}\boldsymbol{X}_2\tilde{\boldsymbol{\beta}}_2\tilde{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{Q}\boldsymbol{X}_1\tilde{\boldsymbol{M}}^{-1}] = \\ & \sigma^{2p}\det[\tilde{\boldsymbol{M}}^{-1}](1 + \tilde{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{Q}\boldsymbol{X}_1'\tilde{\boldsymbol{M}}^{-1}\boldsymbol{X}_1'\boldsymbol{Q}\boldsymbol{X}_2\tilde{\boldsymbol{\beta}}_2). \end{aligned} \tag{15}$$

The expression in (15) is similar to the one we have in the unblocked case (8):

$$\sigma^{2p}\det[\boldsymbol{M}^{-1}](1 + \tilde{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{X}_1\boldsymbol{M}^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2\tilde{\boldsymbol{\beta}}_2).$$

The only two differences are the amended, "blocked" information matrix $\tilde{\boldsymbol{M}}$ for the primary terms, and the presence of matrix $\boldsymbol{Q}$ which indicates the inclusion of the blocked effects in the fitted model. The q-dimensional random vector $\tilde{\boldsymbol{\beta}}_2$, as before, follows $\mathcal{N}(\boldsymbol{0}, \tau^2\boldsymbol{I}_q)$, so that this prior does not depend on the error variance $\sigma_.^2$.

Taking the expectation of the logarithm of (15) over the set prior distribution is completely identical to the derivations shown on page 7; and the MSE-component in the determinant criterion is

$$\log(\det[\tilde{\boldsymbol{M}}^{-1}]) + \mathrm{E}_{\tilde{\boldsymbol{\beta}}_2}\log(1 + \tilde{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{Q}\boldsymbol{X}_1\tilde{\boldsymbol{M}}^{-1}\boldsymbol{X}_1'\boldsymbol{Q}\boldsymbol{X}_2\tilde{\boldsymbol{\beta}}_2). \tag{16}$$

The estimation of the second part of the expression above will be carried out using Monte Carlo sampling, as in the unblocked case. Therefore, the resulting determinant-based compound criterion for a blocked experiments (with the DP- and $LoF(DP)$-optimality as the first two components) is

$$\text{minimise} \quad \left[\left| (\boldsymbol{X}_1'\boldsymbol{Q}\boldsymbol{X}_1)^{-1} \right|^{1/p} F_{p,d_B;1-\alpha_{DP}} \right]^{\kappa_{DP}} \times$$
$$\left[\left| \tilde{\boldsymbol{L}} + \frac{\boldsymbol{I}_q}{\tau^2} \right|^{-1/q} F_{q,d_B;1-\alpha_{LoF}} \right]^{\kappa_{LoF}} \times$$
$$\left[|\boldsymbol{X}_1'\boldsymbol{Q}\boldsymbol{X}_1|^{-1} \exp\left(\frac{1}{N} \sum_{i=1}^{N} \log(1 + \tilde{\boldsymbol{\beta}}_{2i}' \boldsymbol{X}_2'\boldsymbol{Q}\boldsymbol{X}_1 \tilde{\boldsymbol{M}}^{-1} \boldsymbol{X}_1'\boldsymbol{Q}\boldsymbol{X}_2 \tilde{\boldsymbol{\beta}}_{2i}) \right) \right]^{\kappa_{MSE}/p}. \tag{17}$$

We also consider a point prior estimate of the MSE-component as was done in Section **??**: the vector of potential parameters $\boldsymbol{\beta}_2$ will be set to $\sigma\tau\boldsymbol{1}_q$, so that the scaled version of it incorporated in (17) $\tilde{\boldsymbol{\beta}}_2$ is equal to $\tau\boldsymbol{1}_q$. Then the determinant-based criterion function with the point prior estimate, which is referred to as "MSE(D)P", is

$$\text{minimise} \quad \left[\left| (\boldsymbol{X}_1'\boldsymbol{Q}\boldsymbol{X}_1)^{-1} \right|^{1/p} F_{p,d_B;1-\alpha_{DP}} \right]^{\kappa_{DP}} \times$$
$$\left[\left| \tilde{\boldsymbol{L}} + \frac{\boldsymbol{I}_q}{\tau^2} \right|^{-1/q} F_{q,d_B;1-\alpha_{LoF}} \right]^{\kappa_{LoF}} \times$$
$$\left[|\boldsymbol{X}_1'\boldsymbol{Q}\boldsymbol{X}_1|^{-1} \left(1 + \tau^2 \sum_{i,j=1}^{q} [\boldsymbol{X}_2'\boldsymbol{Q}\boldsymbol{X}_1 \tilde{\boldsymbol{M}}^{-1} \boldsymbol{X}_1'\boldsymbol{Q}\boldsymbol{X}_2]_{(i,j)} \right) \right]^{\kappa_{MSE}/p}. \tag{18}$$

When constructing the trace-based criterion, we take the expectation of the trace of the submatrix of the mean square error matrix, corresponding to the primary terms (14):

$$\begin{aligned} \mathrm{E}_{\beta_2}\text{trace}[\text{MSE}(\hat{\tilde{\boldsymbol{\beta}}}_1|\tilde{\boldsymbol{\beta}})_{22}] =& \text{trace}[\mathrm{E}_{\beta_2}\text{MSE}(\hat{\tilde{\boldsymbol{\beta}}}_1|\tilde{\boldsymbol{\beta}})_{22}] = \\ & \text{trace}[\sigma^2\tilde{\boldsymbol{M}}_{22}^{-1} + \mathrm{E}_{\beta_2}(\tilde{\boldsymbol{A}}\boldsymbol{\beta}_2\boldsymbol{\beta}_2'\tilde{\boldsymbol{A}})_{22}] = \\ & \sigma^2\text{trace}[\tilde{\boldsymbol{M}}_{22}^{-1} + \tau^2\{\tilde{\boldsymbol{A}}\tilde{\boldsymbol{A}}'\}_{22}] = \\ & \sigma^2[\text{trace}(\boldsymbol{X}_1'\boldsymbol{Q}\boldsymbol{X}_1)^{-1} + \tau^2\text{trace}\{\tilde{\boldsymbol{A}}\tilde{\boldsymbol{A}}'\}_{22}], \tag{19} \end{aligned}$$

where, as in the previous section, $\tilde{\boldsymbol{M}} = \boldsymbol{X}_1'\boldsymbol{Q}\boldsymbol{X}_1$ and $\tilde{\boldsymbol{A}} = (\tilde{\boldsymbol{X}}_1'\tilde{\boldsymbol{X}}_1)^{-1}\tilde{\boldsymbol{X}}_1'\boldsymbol{X}_2$.

Only the amended forms of the information and alias matrices make the expression above different from the similar one in the "unblocked" case:

$$\sigma^2[\text{trace}\{(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\} + \tau^2\text{trace}\boldsymbol{A}\boldsymbol{A}'].$$

So the final trace-based compound criterion function for a blocked experiment is

$$\text{minimise } \left[\frac{1}{p}\text{trace}(\boldsymbol{W}\boldsymbol{X}_1'\boldsymbol{Q}\boldsymbol{X}_1)^{-1}F_{1,d_B;1-\alpha_{LP}}\right]^{\kappa_{LP}} \times$$
$$\left[\frac{1}{q}\text{trace}\left(\tilde{\boldsymbol{L}} + \frac{\boldsymbol{I}_q}{\tau^2}\right)^{-1}F_{1,d_B;1-\alpha_{LoF}}\right]^{\kappa_{LoF}} \times$$
$$\left[\frac{1}{p}\text{trace}\{(\boldsymbol{X}_1'\boldsymbol{Q}\boldsymbol{X}_1)^{-1} + \tau^2[\tilde{\boldsymbol{A}}\tilde{\boldsymbol{A}}']_{22}\}\right]^{\kappa_{MSE}}. \quad (20)$$

In both criteria d_B stands for the number of pure error degrees of freedom in the case of units gathered into equal-sized blocks. It can be calculated as $d_B = n - \text{rank}(\boldsymbol{Z}:\boldsymbol{T})$ (as in Section **??**), where $\boldsymbol{T}$ is the matrix containing treatment indicators for each of the experimental runs. In other words, the total number of replicates needs to be adjusted for the number of comparisons between blocks that are to be taken into account.

6.2 Example

[A usual example or the case-study?]

7 Future work

7.1 Multistratum experiments

In a large number of industrial, engineering and laboratory-based experiments, either the nature of the process under study or certain technical restrictions result in the necessity of considering a multi-level unit structure. For example, a chemical process consisting of applying treatments to the material batches of different sizes at each stage; or one or several experimental factors' values can be changed only once per a certain amount of runs whereas values of other factors are varied between runs. Therefore, different factors are applied at different levels, and randomisation is performed at each level, thus the whole process of allocating treatments to experimental units should be amended accordingly (Mead *et al.*, 2012).

Experiments comprising a hierarchical structure of experimental units and treatments are referred to as multistratum experiments, and each stratum is defined as a level in the unit structure. Units are grouped into whole-plots, each of them divided into sub-plots, which contain a certain number of sub-sub-plots, and so on up to the smallest units — runs of the experiment. The corresponding factors are often called "very-hard-to-change", "hard-to-change", "easy-to-change", "very-easy-to-change" factors; these labels are quite arbitrary and certainly determined by a particular experimental framework. In the case of two strata, we deal with what is called a "split-plot" experiment,

in case of three strata a "split-split-plot" experiment. The names come from the terminology of agricultural experiments where such setups were originally implemented. Nelder (1965) introduced the notion of simple orthogonal block structures, comprising 'chains' of nested and/or crossed experimental units together with the corresponding factors. Sometimes the "split" notation is used exceptionally for orthogonal block design structures, however, we employ it for general nested unit structures where the orthogonality property does not necessarily hold.

Some visualisation tools allowing a comprehensive presentation of the outline of such experiments can be found in Goos and Gilmour (2012): Hasse diagrams, that were described extensively and in great detail by Bailey (2008). Hasse diagrams are graphs containing the information about the experimental units, factors applied at every stratum, together with a scheme of the available degrees of freedom allocation, so that they appear to be a much useful tool providing a full and clear picture of a complicated structured experiment.

A generalised example of a multistratum experiment with a simple nested unit structure is presented in Figure 1. Each node of the graph corresponds to a blocking factor and edges are drawn from one factor to another that is nested within it. The "universe" node U stands for the whole experiment.

From here onwards we will be considering experiments with s strata in total, stratum i being nested within units of the stratum $i-1$, and stratum 0 will be seen as the whole experiment (following the notation set by Trinca and Gilmour (2015)). The number of units in stratum i within every unit in stratum $i-1$ is denoted by n_i, such that $m_j = \prod_{i=1}^{j} n_i$ is the number of units in stratum j and, therefore, $n = m_s = \prod_{i=1}^{s} n_i$ is the total number of runs.
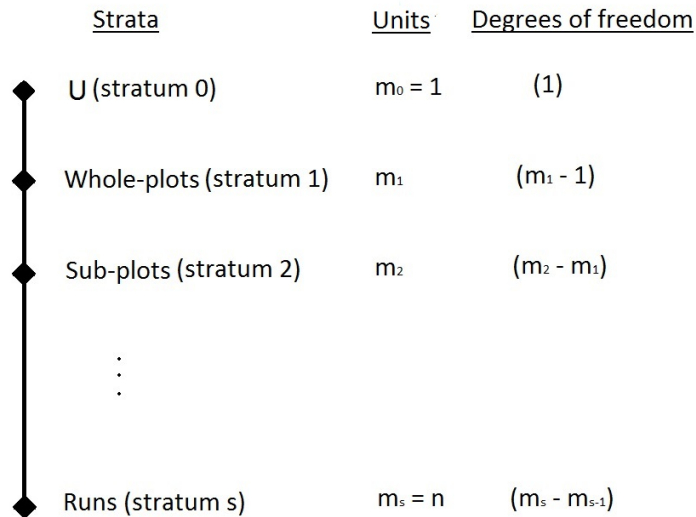
Strata	Units	Degrees of freedom
U (stratum 0)	$m_0 = 1$	(1)
Whole-plots (stratum 1)	m_1	$(m_1 - 1)$
Sub-plots (stratum 2)	m_2	$(m_2 - m_1)$
$\vdots$		
Runs (stratum s)	$m_s = n$	$(m_s - m_{s-1})$

Figure 1: Hasse diagram for factors

Recall the expression for the linear polynomial hierarchical model from (**??**):

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \sum_{i=1}^{s} \boldsymbol{Z}_i\boldsymbol{\varepsilon}_i, \tag{21}$$

where the first part $\boldsymbol{X}\boldsymbol{\beta}$ contains the fixed effects, and $\sum_{i=1}^{s} \boldsymbol{Z}_i\boldsymbol{\varepsilon}_i$ stands for the random part, i.e. it comprises information on variation occurring at each level of randomisation. All of the errors are assumed to be independent, having zero means and constant variances σ_i^2 and, as seen from the model formulation, additive.

In this case the fixed effects coefficients, $\boldsymbol{\beta}$ are usually estimated using the Generalised Least Square (GLS) formula:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}Y, \tag{22}$$

so that the variance of the estimators is

$$Var(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X})^{-1}, \tag{23}$$

where $\boldsymbol{V}$ is the variance-covariance matrix of the (normally distributed) responses, has a block-diagonal structure, and can be presented as:

$$\boldsymbol{V} = \sum_{i=1}^{s} \sigma_i^2 \boldsymbol{Z}_i\boldsymbol{Z}_i' = \sigma_s^2 \left(\boldsymbol{I}_n + \sum_{i=1}^{s-1} \eta_i \boldsymbol{Z}_i\boldsymbol{Z}_i' \right). \tag{24}$$

Here $\eta_i = \sigma_i^2/\sigma_s^2$ is defined as a variance ratio, denoting the magnitude of variability occurring at higher strata scaled with respect to the between-run variance.

At the analysis stage, as the variance components are unknown, their estimates are obtained and substituted in (23) to obtain an estimated variance-covariance matrix of the parameters' estimators, $\hat{\boldsymbol{V}}(\hat{\boldsymbol{\beta}})$; the estimation methods are discussed in more detail in Chapter **??**.

However, when an experiment is being planned, and none of the η_i or σ_s^2 are known, there are two main approaches to deal with it. The first one is to search for designs assuming some point prior values of the η parameter and, for these, evaluate the variance-based criteria, considering all strata simultaneously.

The case of split-plot experiments, with two strata, $\sigma_s^2 = \sigma^2$ and $\eta = \sigma_1^2/\sigma^2$, has been extensively considered in the literature. Goos and Vandebroek (2001) considered the three cases when D-optimal designs for split-plot experiments do not depend on the value of η; in other practical cases an estimate of the variance ratio is to be provided. Goos and Vandebroek (2003) and Jones and Goos (2007) developed algorithms for finding D-optimal split-plot designs and for some examples demonstrated the robustness to the different values of η. An algorithm for constructing D-optimal split-split-plot designs can be found in the paper by Jones and Goos (2009).

The Bayesian alternative to the 'exhaustive search' approach across the unknown parameter space allows specifying a prior distribution on the unknown parameters, and then evaluating the criteria by integrating over that prior. Arnouts and Goos (2012) presented a coordinate-exchange algorithm for constructing D-optimal designs for a staggered experimental structure (i.e. values of hard-to-change factors' are changed at different time points), with log-normal prior distributions being put on the variance ratios. Later Arnouts and Goos (2015) studied the staggered structured designs in the context of response surface modelling, and considered a few examples of D- and I-optimal designs.

Mylona *et al.* (2014) introduced a composite criterion, combining D-optimality for the fixed and variance components; in such an approach both the criterion formula and the form of the prior distributions define the way of evaluation the criterion function, which often leads to the necessity of choosing a computationally efficient numerical methodology. Gilmour and Goos (2009) proposed an appropriate Bayesian analysis strategy of data from multistratum experiments, that is robust to designs non-orthogonality and is shown to be more reliable than REML approach.

However, in the general case a lot of computational effort might be required in order to be confident in the goodness of the design obtained by choosing from a set of assumed possible values of η_i, either using point priors or adapting general Bayesian strategy, especially when there are more than two strata and, therefore, the range of unknown parameters becomes multi-dimensional. The stratum-by-stratum approach (as first suggested by Trinca and Gilmour (2001) and more recently improved by Trinca and Gilmour (2015)) implies going from the highest stratum to the lowest, at each level choosing the set of treatments, and the units of higher level being treated as fixed block effects. Such methodology allows 'protecting' against the case of large higher level variances, and eliminates the need for any prior information on unknown parameters. The authors later adapted the stratum-by-stratum strategy for the inference criteria (Trinca and Gilmour, 2017), providing tools for constructing efficient designs for relatively small experiments in the presence of restricted randomisation.

8   Acknowledgements

References

Arnouts, H. and Goos, P. (2012) Staggered-level designs for experiments with more than one hard-to-change factor. *Technometrics*, **54**, 355–366.

Arnouts, H. and Goos, P. (2015) Staggered-level designs for response surface modeling. *Journal of Quality Technology*, **47**, 156.

Atkinson, A. C., Donev, A. N. and Tobias, R. D. (2007) *Optimum Experimental Designs, with SAS*. Oxford University Press.

Bailey, R. A. (2008) *Design of Comparative Experiments*, vol. 25. Cambridge University Press.

Box, G. E. P. and Draper, N. R. (1987) *Empirical Model-Building and Response Surfaces*. John Wiley & Sons.

Box, G. E. P. and Wilson, K. B. (1951) On the experimental attainment of optimum conditions (with discussion). *Journal of the Royal Statistical Society, Series B*, **13**, 1–45.

Cornell, J. A. (2011) *Experiments with Mixtures: Designs, Models, and the Analysis of Mixture Data*. John Wiley & Sons.

Cox, D. R. (1958) *Planning of Experiments*. John Wiley & Sons.

Draper, N. R. and Smith, H. (1998) *Applied Regression Analysis, 3rd edn*. John Wiley & Sons.

DuMouchel, W. and Jones, B. (1994) A simple Bayesian modification of D-optimal designs to reduce dependence on an assumed model. *Technometrics*, **36**, 37–47.

Gilmour, S. G. and Goos, P. (2009) Analysis of data from non-orthogonal multistratum designs in industrial experiments. *Journal of the Royal Statistical Society: Series C*, **58**, 467–484.

Gilmour, S. G. and Trinca, L. A. (2000) Some practical advice on polynomial regression analysis from blocked response surface designs. *Communications in Statistics-Theory and Methods*, **29**, 2157–2180.

Gilmour, S. G. and Trinca, L. A. (2012) Optimum design of experiments for statistical inference (with discussion). *Journal of the Royal Statistical Society: Series C*, **61**, 345–401.

Goos, P. and Gilmour, S. G. (2012) A general strategy for analyzing data from split-plot and multistratum experimental designs. *Technometrics*, **54**, 340–354.

Goos, P. and Vandebroek, M. (2001) D-optimal response surface designs in the presence of random block effects. *Computational Statistics & Data Analysis*, **37**, 433–453.

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

Harville, D. A. (2006) *Matrix Algebra From a Statistician's Perspective*. Springer New York.

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

Jones, B. and Goos, P. (2009) D-optimal design of split-split-plot experiments. *Biometrika*, **96**, 67–82.

Koch, K. R. (2007) *Introduction to Bayesian Statistics*. Springer, 2nd edn.

Lu, L., Anderson-Cook, C. M. and Robinson, T. J. (2011) Optimization of designed experiments based on multiple criteria utilizing a Pareto frontier. *Technometrics*, **53**, 353–365.

Mead, R., Gilmour, S. G. and Mead, A. (2012) *Statistical Principles for the Design of Experiments: Applications to Real Experiments*. Cambridge University Press.

Montepiedra, G. and Fedorov, V. V. (1997) Minimum bias designs with constraints. *Journal of Statistical Planning and Inference*, **63**, 97–111.

Mylona, K., Goos, P. and Jones, B. (2014) Optimal design of blocked and split-plot experiments for fixed effects and variance component estimation. *Technometrics*, **56**, 132–144.

Nelder, J. A. (1965) The analysis of randomized experiments with orthogonal block structure. i. block structure and the null analysis of variance. *Proceedings of the Royal Society of London, Series A*, **283**, 147–162.

Rudin, W. (1987) *Real and Complex Analysis*. Tata McGraw-Hill Education.

Stallings, J. W. and Morgan, J. P. (2015) General weighted optimality of designed experiments. *Biometrika*, 1–11.

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

Trinca, L. A. and Gilmour, S. G. (2015) Improved split-plot and multistratum designs. *Technometrics*, **57**, 145–154.

Trinca, L. A. and Gilmour, S. G. (2017) Split-plot and multi-stratum designs for statistical inference. *Technometrics*. In press.