Contents

Abstract

1 Introduction

Model-independent approach to building designs in case of model uncertainty

- *Planning a controlled intervention – reliance on a model (1) +*
- *Active assumption of a specific model misspecification (2) =*
- *Combining conflicting aims: model-based inference based on the model-independent variance estimate and distrust in the model sufficiency*
- *Briefly: extension on more complex experimental structures*

Designed experiments are commonly 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, some form of approximation is needed – and polynomial functions are able to provide any required precision for functions from a certain class of differentiability (Rudin, 1987). Response Surface Methodology (RSM, Box and Wilson (1951)) aims at optimising the approximated functions by fitting second-order polynomials, and higher desirable precision would require a polynomial of a higher order and, therefore, more experimental effort.

Whichever the chosen, 'primary', model is, planning a controlled intervention does rely on the approximating model assumptions in two main – and quite contradicting – directions(?). Firstly, it is desirable to ensure the goodness of what the model is essentially fitted for, that is, for example, the precision of model parameters and/or the prediction accuracy of the untested treatment combinations. The corresponding design aims are usually reflected in utilising well-known optimality criteria (D-, A-, ?prediction) and rely on the variance estimate, the way it is obtained. Secondly, treating the chosen model as the absolute truth, especially at the designing stage is at least too optimistic and, as in many cases fitting a more suitable, but complicated model might not be feasible due to various restrictions and limitations, could be dangerous for the results credibility. So having a particular model also means that at the stage of planning it is highly desirable to include some [sensitivity towards and robustness against] control over the model lack of fit and its effect on the inference.

We deal with such duality [of model-dependence and accounting for its misspecification] by developing compound optimality criteria, each constructed as a weighted combination of individual criterion functions, with the two main features:

1. Each corresponds to a specific desirable property: either accounting for an aim coming from trusting the model or mitigating the effects from its potential misfit. The relative importance of the components are reflected by the assigned weights – and we shall examine the performance of the resulting designs in terms of the individual criteria, and explore the role of weight allocation.

2. The concept of model-independent internal variance estimation, 'pure error' (Gilmour and Trinca (2012))) underlies each of the individual criteria – the most appropriate [sensible] strategy in the case of possible model insufficiency.

The general spirit of this work aims at aligning with the concept of a good design, which, as summarised by Box and Draper (1987), should "make it possible to detect lack of fit" and "provide an internal estimate of error from replication", among other properties.

We will focus on factorial experiments with a relatively small number of runs and the fitted model being a polynomial regression. Section 2 provides the background on the modelling, error estimation and fundamental individual criteria. Controlling the lack-of-fit and the bias arising from the model misspecification are introduced in Section 3, where they are combined with the primary model-driven ones in compound optimality criteria. Their adaptation to experimental frameworks with restricted randomisation is described in Section **??**, as well as the design construction procedure for the multistratum experiments. Finally, a series of examples are presented in Section 5 – examining dynamics across various optimal designs, shape of the constructed criteria and other properties, followed up by a Discussion in Section 6 with main conclusions and recommendations.

The R code used to search for the optimal designs are located at ¡...¿ [?github for now].

2 Model-dependent planning

- *Modelling and relevant error estimation*

- *Fundamental individual criteria*

- *Combining objectives: lit.review*

Assuming a smooth enough relationship between k experimental factors $X_1, \ldots, X_k \in \Theta \subset \mathrm{R}^k$ and the response variable $Y = \eta(X_1, \ldots, X_k) \in \mathrm{R}$, a suitable polynomial model is chosen to fit data obtained from n experimental runs:

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

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: $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{I}_n)$.

Any inference following the model fitting relies on the error variance estimate $\hat{\sigma}^2$. The most common one is the mean square error: $\hat{\sigma}^2_{mse} = \text{Residual SS}/(n-p)$ (e.g. Draper and Smith (1998)), which is model-dependent with the residual degrees of freedom $\nu = n - p$ containing the number of model parameters.

The other one is 'pure' error, independent from the parametric model, which is derived from the further decomposition of the residual sum of squares into the 'pure' error and 'lack-of-fit' components: $\hat{\sigma}^2_{PE} = \text{Pure error SS}/(n-t)$, where t is the number of unique treatments (combinations of factors' levels) applied and $d = n - t$ is the pure error degrees of freedom, that is the number of replicated points. In other words, the error is estimated as the mean square error from fitting the full treatment model:

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

where $\boldsymbol{X_t}$ 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 i^{th} unit, otherwise it is set to 0. Then the elements of the t-dimensional vector $\boldsymbol{\mu_t}$ are the mean effects of each treatment. The vector of errors $\boldsymbol{\varepsilon}$ comprises the between-unit variation, such that $\text{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$, $\text{Var}(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{I}_n$.

Many authors advocate the use of the 'pure' error estimate 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". 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.

Optimality criteria

Model-dependency at the stage of experimental planning is reflected in searching for a design that optimises a criterion function, that is a function of a design that translates a specific inference-driven objective. For example, among the most well-known, "alphabetic" optimality criteria, such as D-, C-, L-optimality and a series of others (Atkinson *et al.* (2007)) target the precision of parameters estimates in model (1); while others, like G- and V-optimality, deal with the prediction variance. Traditionally these criteria are formulated based on the mean square error estimate; citetGilmourTrinca2012 derived the alternative pure-error based criteria, which would guarantee the presence of replicates in the resulting designs.

Combining multiple desirable objectives in the design can be fulfilled through constructing a compound criterion. This concept is based on the notion of design efficiency, which can be defined for any design matrix X and any criterion $F(X)$ as the ratio with respect to the best (e.g. maximum, without the loss of generality) value achieved at the optimal design. For example, the D-efficiency of design $\boldsymbol{X}$ is

$$\mathrm{Eff}_D(X) = \frac{|\boldsymbol{X}'\boldsymbol{X}|^{1/p}}{|\boldsymbol{X}_*'\boldsymbol{X}_*|^{1/p}},$$

where $\boldsymbol{X}_*$ is the D-optimum design. In this definition the power $1/p$ brings the efficiency to the scale of variances of model coefficients $\boldsymbol{\beta}_i, i = 1 \dots p$. 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 compound criterion to be maximised among all the possible designs is obtained then as weighted product of the individual criteria efficiencies $\mathrm{Eff}_1, \dots, \mathrm{Eff}_m$ with the corresponding weights $\kappa_1, \dots, \kappa_m$ (s.t. $\kappa_k > 0$ and $\sum_{k=1}^m \kappa_k = 1$):

$$\mathrm{Eff}_1^{\kappa_1}(\boldsymbol{X}) \times \mathrm{Eff}_2^{\kappa_2}(\boldsymbol{X}) \times \ldots \times \mathrm{Eff}_m^{\kappa_m}(\boldsymbol{X}) \to \max_{\boldsymbol{X}}. \tag{3}$$

The choice of weights is arbitrary in general, although is driven by prior knowledge of the experimenter, the objectives of a specific experiment and the components' interpretation. In the examples considered further in this work, we chose a set of weight allocations, most of which are classical design schemes for experiments with mixtures (Cornell, 2011).

Some of the alternative approaches of combining several objectives would include generating the Pareto optimal set of designs – the approach thoroughly described by Lu *et al.* (2011). Stallings and Morgan (2015)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. [Add more references to compound optimality?]

3 Individual criteria construction

- *Model misspecification setup. What objectives to be reflected in criteria: PE*
- *Lack-of-fit criterion (and then the traditional LoF)*
- *MSE criterion*
- *Combining with DP/LP – the general compound criterion*

Standard design optimality theory is developed under the assumption that the primary model provides the most proper fit for the data: in many real applications this is quite a strong belief, and in reality we need to take into account at least the possibility that some misspecification is present at the planning stage.

In this work we consider the case when the fitted polynomial model with p parameters is nested within a larger model that is assumed to provide a better fit for the data:

$$\boldsymbol{Y} = \boldsymbol{X}_p\boldsymbol{\beta}_p + \boldsymbol{X}_q\boldsymbol{\beta}_q + \boldsymbol{\varepsilon}, \tag{4}$$

where $\boldsymbol{X}_q$ is an $n \times q$ extension of the primary model matrix containing extra q terms that we refer to as "potential terms" and that represent the fitted model disturbance, with vector $\boldsymbol{\beta}_q$ denoting the corresponding parameters. They are not of any inferential interest, and, moreover, not all of them are necessarily estimable when the experiment is relatively small, i.e. $n < p+q$ – this is the case we mainly consider here. As usually, we do assume independent and normally distributed error terms: $\boldsymbol{\varepsilon} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I}_n)$.

Lack-of-fit criterion

Aiming towards controlling the magnitude and scope of the potential terms, we adapt Bayesian approach regarding the full model parameters. Diffuse prior shall be put on primary terms – an arbitrary mean and a variance going to infinity, and the prior on potential terms is a normal distribution: $\boldsymbol{\beta}_q \sim \mathcal{N}(0, \boldsymbol{\Sigma}_0)$, with the prior variance scaled with respect to the error variation: $\boldsymbol{\Sigma}_0 = \sigma^2\tau^2\boldsymbol{I}_q$. Then, following the normality in the model (4), the posterior distribution of the joint vector of coefficients $\boldsymbol{\beta} = [\boldsymbol{\beta}_p^T, \boldsymbol{\beta}_q^T]^T$ is (Koch (2007) is multivariate normal DuMouchel and Jones (1994)):

$$\begin{gathered}
\boldsymbol{\beta}|\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{b}, \boldsymbol{\Sigma}), \\
\text{where } \boldsymbol{b} = \boldsymbol{\Sigma}\boldsymbol{X}^T\boldsymbol{Y} \text{ and } \boldsymbol{\Sigma} = \sigma^2[\boldsymbol{X}^T\boldsymbol{X} + \boldsymbol{K}/\tau^2],^{-1} \boldsymbol{X} = [\boldsymbol{X}_p, \boldsymbol{X}_q], \\
\boldsymbol{K} = \begin{pmatrix} \boldsymbol{0}_{p\times p} & \boldsymbol{0}_{p\times q} \\ \boldsymbol{0}_{q\times p} & \boldsymbol{I}_{q\times q} \end{pmatrix}.
\end{gathered}$$

The marginal posterior distribution of $\boldsymbol{\beta}_q$ is also multivariate normal with mean $\boldsymbol{b}_q$ – the last q elements of $\boldsymbol{b}$ and the covariance matrix $\boldsymbol{\Sigma}_{qq}$– the bottom right $q \times q$ submatrix of $\boldsymbol{\Sigma}$:

$$\begin{aligned}
\boldsymbol{\Sigma}_{qq} = \sigma^2[(\boldsymbol{X}^T\boldsymbol{X} + \boldsymbol{K}/\tau^2)^{-1}]_{[q,q]} &= \sigma^2 \begin{bmatrix} \boldsymbol{X}_p^T\boldsymbol{X}_p & \boldsymbol{X}_p^T\boldsymbol{X}_q \\ \boldsymbol{X}_q^T\boldsymbol{X}_p & \boldsymbol{X}_q^T\boldsymbol{X}_q + \boldsymbol{I}_q/\tau^2 \end{bmatrix}^{-1}_{[q,q]} \\
&= \sigma^2[\boldsymbol{X}_q^T\boldsymbol{X}_q + \boldsymbol{I}_q/\tau^2 - \boldsymbol{X}_q^T\boldsymbol{X}_p(\boldsymbol{X}_p^T\boldsymbol{X}_p)^{-1}\boldsymbol{X}_p^T\boldsymbol{X}_q]^{-1} \\
&= \sigma^2\left[\boldsymbol{L} + \boldsymbol{I}_q/\tau^2\right],^{-1}
\end{aligned}$$

where $\boldsymbol{L} = \boldsymbol{X}_q^T\boldsymbol{X}_q - \boldsymbol{X}_q^T\boldsymbol{X}_p(\boldsymbol{X}_p^T\boldsymbol{X}_p)^{-1}\boldsymbol{X}_p^T\boldsymbol{X}_q$ is known in model-sensitivity design literature as the "dispersion matrix" [references], which provides a point-wise measure

of the distance between column vectors of $\boldsymbol{X}_q$ and linear subspace defined by the column vectors of the primary model matrix $\boldsymbol{X}_p$, that is how well each of the potential terms could be approximated by a linear span of the primary ones [?].

Reducing the primary model's lack-of-fit in the direction of the potential terms can be translated into a criterion function of the design by utilising the posterior distribution for $\boldsymbol{\beta}_q$ derived above and constructing a $(1-\alpha)\times 100\%$ confidence region for the parameters depending on the model matrices and the variance estimate s^2 on ν degrees of freedom (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. Minimising the volume of the confidence region is equivalent to

$$\left|\boldsymbol{L}+\boldsymbol{I}_q/\tau^2\right|^{-1/q}F_{q,d;1-\alpha_{LoF}}\longrightarrow\ \min, \tag{5}$$

which we refer to as "Lack-of-fit DP-criterion", that is directly related to [1] the lack-of-fit component in the Generalised D-optimality developed by Goos *et al.* (2005) – where the residual number of degrees of freedom ν does not depend on the design, and [2] DP-optimality (Gilmour and Trinca, 2012), with the F-quantile preserved from $\nu=d$ being the number of replicates in the design.

Another way of formulating a criterion that we consider in this work – minimising the average of posterior variances of linear functions of $\boldsymbol{\beta}_q$ defined by matrix $\boldsymbol{J}$, and we define the Lack-of-fit LP-criterion as mean of the squared lengths of the $(1-\alpha)\times 100\%$ confidence intervals for these linear functions:

$$\frac{1}{q}\mathrm{trace}\left[\boldsymbol{J}\boldsymbol{J}^T\left(\boldsymbol{L}+\frac{\boldsymbol{I}_q}{\tau^2}\right)^{-1}\right]F_{1,d;1-\alpha}\longrightarrow\ \min. \tag{6}$$

This trace-based criterion is linked to the lack-of-fit part of the Generalised L-optimality (Goos *et al.*, 2005), and the pure error estimation approach retains the corresponding F-quantile. [Henceforth we mainly consider the case when $\boldsymbol{J}$ is the identity matrix, that is we work with the analogue of AP-optimality. In other words, the lack-of-fit component in the generalised AP-criterion stands for the minimisation of the $\boldsymbol{L}_2$-norm of the q-dimensional vector of the posterior confidence intervals' lengths for the potential parameters.]

MSE-based bias criterion

Together with controlling the magnitude of model contamination, it is also desirable to "protect" the quality of inference that is to be drawn through fitting the primary model, from the potential presence of extra terms. From this point of view, the bias

of the parameters' estimates $\hat{\boldsymbol{beta}}_p$ would be of substantial interest; a natural way of evaluating their quality is the matrix of mean square error (Montepiedra and Fedorov, 1997) , 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 (4):

$$\begin{aligned}\mathrm{MSE}(\hat{\boldsymbol{\beta}}_p|\boldsymbol{\beta}) =& \mathtt{E}_{\boldsymbol{Y}|\boldsymbol{\beta}}[(\hat{\boldsymbol{\beta}}_p - \boldsymbol{\beta}_p)(\hat{\boldsymbol{\beta}}_p - \boldsymbol{\beta}_p)^T] \\ =& \sigma^2(\boldsymbol{X}_p^T\boldsymbol{X}_p)^{-1} + \boldsymbol{A}\boldsymbol{\beta}_q\boldsymbol{\beta}_q^T\boldsymbol{A}^T, \end{aligned} \tag{7}$$

where $\boldsymbol{A} = (\boldsymbol{X}_p^T\boldsymbol{X}_p)^{-1}\boldsymbol{X}_p^T\boldsymbol{X}_q$ denotes the $p \times q$ alias matrix, which elements reflect the linearity scale of the relationship between the primary (rows) and potential (columns) terms.

The determinant-based criterion that would correspond to the overall simultaneous minimisation of the bias above is constructed as log-determinant of the MSE matrix averaged across [?] the values of the full model parameters $\boldsymbol{\beta}$:

$$\mathtt{E}_{\boldsymbol{\beta}} \log(\det[\mathrm{MSE}(\hat{\boldsymbol{\beta}}_p|\boldsymbol{\beta})]) \longrightarrow \text{ min.} \tag{8}$$

Using the matrix determinant lemma (Harville, 2006) and setting $\boldsymbol{M} = \boldsymbol{X}_1'\boldsymbol{X}_1$ and $\tilde{\boldsymbol{\beta}}_2 = \boldsymbol{\beta}_2/\sigma$, the determinant and the applied logarithm in (8)) can be decomposed:

$$\begin{aligned}&\det[\mathrm{MSE}(\hat{\boldsymbol{\beta}}_p|\boldsymbol{\beta}_q)] \\ &= \det[\sigma^2\boldsymbol{M}^{-1} + \boldsymbol{M}^{-1}\boldsymbol{X}_p^T\boldsymbol{X}_q\boldsymbol{\beta}_q\boldsymbol{\beta}_q^T\boldsymbol{X}_q^T\boldsymbol{X}_p\boldsymbol{M}^{-1}] \\ &= \sigma^{2p}\det[\boldsymbol{M}^{-1} + \boldsymbol{M}^{-1}\boldsymbol{X}_p'\boldsymbol{X}_q\tilde{\boldsymbol{\beta}}_q\tilde{\boldsymbol{\beta}}_q^T\boldsymbol{X}_q^T\boldsymbol{X}_p\boldsymbol{M}^{-1}] \\ &= \sigma^{2p}\det[\boldsymbol{M}^{-1}]\det[1 + \tilde{\boldsymbol{\beta}}_q^T\boldsymbol{X}_q^T\boldsymbol{X}_p\boldsymbol{M}^{-1}\boldsymbol{M}\boldsymbol{M}^{-1}\boldsymbol{X}_p^T\boldsymbol{X}_q\tilde{\boldsymbol{\beta}}_q] \\ &= \sigma^{2p}\det[\boldsymbol{M}^{-1}](1 + \tilde{\boldsymbol{\beta}}_q^T\boldsymbol{X}_q^T\boldsymbol{X}_p\boldsymbol{M}^{-1}\boldsymbol{X}_p^T\boldsymbol{X}_q\tilde{\boldsymbol{\beta}}_q); \end{aligned} \tag{9}$$

$$\begin{aligned}&\log(\det[\mathrm{MSE}(\hat{\boldsymbol{\beta}}_p|\boldsymbol{\beta}_q)]) \\ &= p\log\sigma^2 + \log(\det[\boldsymbol{M}^{-1}]) + \log(1 + \tilde{\boldsymbol{\beta}}_q^T\boldsymbol{X}_q^T\boldsymbol{X}_p\boldsymbol{M}^{-1}\boldsymbol{X}_p^T\boldsymbol{X}_q\tilde{\boldsymbol{\beta}}_q). \end{aligned} \tag{10}$$

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 criterion function. Therefore, the criterion in (8) which we refer to becomes

$$\log(\det[\boldsymbol{M}^{-1}]) + \mathtt{E}_{\tilde{\boldsymbol{\beta}}_q}\log(1 + \tilde{\boldsymbol{\beta}}_q^T\boldsymbol{X}_q^T\boldsymbol{X}_p\boldsymbol{M}^{-1}\boldsymbol{X}_p^T\boldsymbol{X}_q\tilde{\boldsymbol{\beta}}_q) \longrightarrow min.$$

Due to the obvious lack of information regarding $\tilde{\boldsymbol{\beta}}_q$, the second term needs to be evaluated numerically. Expressing the prior variance of $\boldsymbol{\beta}_q$ as a scaled error variance $\boldsymbol{\beta}_q \sim \mathcal{N}(\boldsymbol{0}, \tau^2\sigma^2\boldsymbol{I}_q)$ means that $\tilde{\boldsymbol{\beta}}_q \sim \mathcal{N}(\boldsymbol{0}, \tau^2\boldsymbol{I}_q)$, and that, quite conveniently, its prior distribution does not depend on the unknown σ.[2] Then a primitive Monte-Carlo can be used to evaluate that term: drawing a sample of large size N from the prior, and approximating the expectation above by the average across the sampled values of $\tilde{\boldsymbol{\beta}}_{q_i}$,

$i = 1, \ldots, N$:

$$\mathrm{E}_{\tilde{\boldsymbol{\beta}}_q} \log(1 + \tilde{\boldsymbol{\beta}}_q^T \boldsymbol{X}_q^T \boldsymbol{X}_p \boldsymbol{M}^{-1} \boldsymbol{X}_p^T \boldsymbol{X}_q \tilde{\boldsymbol{\beta}}_q) \approx \frac{1}{N} \sum_{i=1}^{N} \log(1 + \tilde{\boldsymbol{\beta}}_{q_i}^T \boldsymbol{X}_q^T \boldsymbol{X}_p \boldsymbol{M}^{-1} \boldsymbol{X}_p^T \boldsymbol{X}_q \tilde{\boldsymbol{\beta}}_{q_i}).$$

One of the alternatives to this more computationally demanding approach is to use the point prior for $\boldsymbol{\beta}_q$, that is setting $\boldsymbol{\beta}_q = \sigma\tau\mathbf{1}_q$ (where $\mathbf{1}_q$ is a q-dimensional vector of 1s), which is the standard deviation of the initial normal prior. Then $\tilde{\boldsymbol{\beta}}_q = \tau\mathbf{1}_q$ with probability 1 and the expectation above becomes:

$$\begin{aligned}\mathrm{E}_{\tilde{\boldsymbol{\beta}}_q} \log(1 + \tilde{\boldsymbol{\beta}}_q^T \boldsymbol{X}_q^T \boldsymbol{X}_p \boldsymbol{M}^{-1} \boldsymbol{X}_p^T \boldsymbol{X}_q \tilde{\boldsymbol{\beta}}_q) &\approx \log(1 + \tau^2 \mathbf{1}_q^T \boldsymbol{X}_q^T \boldsymbol{X}_p \boldsymbol{M}^{-1} \boldsymbol{X}_p^T \boldsymbol{X}_q \mathbf{1}_q) \\ &= \log(1 + \tau^2 \sum_{i,j=1}^{q} [\boldsymbol{X}_q^T \boldsymbol{X}_p \boldsymbol{M}^{-1} \boldsymbol{X}_p^T \boldsymbol{X}_q][i,j]),\end{aligned}$$

with summation of matrix elements taking less computational time even for large q.

To infer the trace-based form of the criterion that would minimise the mean squared bias, we use the pseudo-Bayesian approach of setting priors on the model's parameters to calculating the trace function of the MSE matrix (7):

$$\begin{aligned}\mathrm{E}_{\boldsymbol{\beta}_q} \mathrm{trace}[\mathrm{MSE}(\hat{\boldsymbol{\beta}}_p \boldsymbol{\beta}_q)] &= \mathrm{trace}[\mathrm{E}_{\boldsymbol{\beta}_q} \mathrm{MSE}(\hat{\boldsymbol{\beta}}_p | \boldsymbol{\beta}_q)] = \\ &= \mathrm{trace}[\sigma^2 (\boldsymbol{X}_p^T \boldsymbol{X}_p)^{-1} + \mathrm{E}_{\boldsymbol{\beta}_q} (\boldsymbol{A} \boldsymbol{\beta}_q \boldsymbol{\beta}_q^T \boldsymbol{A}^T)] = \\ &= \mathrm{trace}[\sigma^2 (\boldsymbol{X}_p^T \boldsymbol{X}_p)^{-1} + \sigma^2 \tau^2 \boldsymbol{A} \boldsymbol{A}^T] = \\ &= \sigma^2 \mathrm{trace}[(\boldsymbol{X}_p^T \boldsymbol{X}_p)^{-1} + \tau^2 \boldsymbol{A} \boldsymbol{A}^T] \\ &= \sigma^2 [\mathrm{trace}\{(\boldsymbol{X}_p^T \boldsymbol{X}_p)^{-1}\} + \tau^2 \mathrm{trace} \boldsymbol{A} \boldsymbol{A}^T].\end{aligned}$$

The operations of calculating trace and expectation are commutative, hence there is no necessity of any additional numerical evaluations, and in this case of the trace-based criterion using the point prior for $\boldsymbol{\beta}_q$ defined earlier would lead to the same resulting function. By minimising the whole function above, we simultaneously minimise both the sum of primary terms' variance and the expected squared norm of the bias vector in the direction of the potential terms. The second part here contains the scaling parameter τ^2 which regulates the magnitude of the potential terms' variation relatively to the error variance.

Compound criteria

The compound criterion is constructed to account for the three main objectives – precision (or accuracy?) of the primary model parameters, control for the lack-of-fit and minimising the inferential bias from the potential model contamination – as the weighted product of efficiencies (3) with respect to DP-criterion, and Lack-of-Fit DP functions(5, 6) and the MSE-based ones. All of them are brought to the scale of

efficiencies [?and exp]:

$$\begin{aligned}\text{minimise } & \left[\left|\boldsymbol{X}_p^T\boldsymbol{X}_p\right|^{-1/p} F_{p,d;1-\alpha_{DP}}\right]^{\kappa_{DP}} \times \\ & \left[\left|\boldsymbol{L}+\frac{\boldsymbol{I}_q}{\tau^2}\right|^{-1/q} F_{q,d;1-\alpha_{LoF}}\right]^{\kappa_{LoF}} \times \\ & \left[\left|\boldsymbol{X}_p^T\boldsymbol{X}_p\right|^{-1} \exp\left(\mathrm{E}_{\tilde{\boldsymbol{\beta}}_q} \log(1+\tilde{\boldsymbol{\beta}}_q^T\boldsymbol{X}_q^T\boldsymbol{X}_p\boldsymbol{M}^{-1}\boldsymbol{X}_p^T\boldsymbol{X}_q\tilde{\boldsymbol{\beta}}_q)\right)\right]^{\kappa_{MSE}/p}. \end{aligned} \tag{11}$$

This criterion is later referred to as compound "MSE-DP-criterion", where α_{DP} and α_{LoF} denote the confidence intervals' probability levels for the primary $\boldsymbol{\beta}_p$ and potential coefficients $\boldsymbol{\beta}_q$ – in the DP and LoF-DP elementary criteria; usually they are set to 0.05 or 0.01. As mentioned before, non-negative weights κ_X that sum up to 1 define the compound criterion and are chosen to reflect the experimenter's priorities.

Similarly, we obtain the compound "MSE-LP-criterion" by joining the LP criterion with trace-based Lack-of-fit (6)) and MSE components :

$$\begin{aligned}\text{minimise } & \left[\frac{1}{p}\text{trace}(\boldsymbol{W}\boldsymbol{X}_p^T\boldsymbol{X}_p)^{-1}F_{1,d;1-\alpha_{LP}}\right]^{\kappa_{LP}} \times \\ & \left[\frac{1}{q}\text{trace}\left(\boldsymbol{L}+\frac{\boldsymbol{I}_q}{\tau^2}\right)^{-1} F_{1,d;1-\alpha_{LoF}}\right]^{\kappa_{LoF}} \times \\ & \left[\frac{1}{p}\text{trace}\{(\boldsymbol{X}_p^T\boldsymbol{X}_p)^{-1}+\tau^2\boldsymbol{A}\boldsymbol{A}^T\}\right]^{\kappa_{MSE}}. \end{aligned} \tag{12}$$

4 Restricted randomisation

- *General framework of multistratum experiments. [add more on model uncertainty]*
- *Error estimation: REML and pure error REML. Yates procedure*
- *Adapting optimality criteria and design construction*

In a large number of industrial, engineering and laboratory-based experiments, either the nature of the study or certain restrictions result in the necessity of considering a multi-level structure of experimental units. 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).

In this work we go on to consider experimental framework comprising such a hierarchical

structure of experimental units and treatments, which is referred to as multistratum experiment, 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 – [Figure .with Hasse diagram?] . In the general case of two strata, we deal with what is called a "split-plot" experiment, in case of three strata a "split-split-plot" experiment. Our aim is to adapt the MSE-based criteria derived before to the factorial experiments with units distributed in several strata. The most important aspects are (1) estimating the error components at all of the levels and (2) formulating the compliant criteria and provide the suitable implementation strategy for the optimal design search.

4.1 Response-Surface Methodology for multistratum framework

[OE: introduce the notion of "block"?] Here we are 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. As randomisation occurs at each stratum, the model accounting for the hierarchical error structure and correlated observations can be written as

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \sum_{i=1}^{s} \boldsymbol{Z}_i \boldsymbol{\varepsilon}_i, \tag{13}$$

where $\boldsymbol{Y}$ represents the n-dimensional vector of responses, $\boldsymbol{X}$ stands for the $n \times p$ model matrix and $\boldsymbol{\beta}$ for the p-dimensional vector of corresponding model coefficients. Each row of the $n \times m_i$ matrix $\boldsymbol{Z}_i$ corresponds to a single experimental run and indicates the unit in stratum i containing this run. $\boldsymbol{\varepsilon}_i$ is a vector of random effects occurring due to the randomisation at level i, and these effects are assumed to be independent and identically distributed around zero mean and variance σ_i^2. Generalised Least Squared (GLS) estimators of $\boldsymbol{\beta}$ are calculated as

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{Y}. \tag{14}$$

Error Estimate

An extensive amount of research has been conducted on the design of and analysis of data from split-plot and split-split-plot experiments, with one of the first works by Letsinger *et al.* (1996), emphasising the necessity of adapting the design and analysis strategy with respect to the more complicated error structure. In the most general case of non-orthogonal unit structures Residual Maximum Likelihood (REML) methodology has been acknowledged to be the most sensible approach to estimating the variance components by maximising the part of the likelihood function corresponding to the

"random" part of the model; the details are provided, for example, in McCulloch and Searle (2001).

In some case, though, REML tends to underestimate the variance components in the higher stratum, for example Goos *et al.* (2006) considered some split-plot experiments with true whole-plot variances being non-negligibly far from zero, for which REML, however, provided zero estimates. A possible alternative of following Bayesian strategy (Gilmour and Goos, 2009), however, this requires a careful choice of the prior which is often not possible, especially in the case of more than two strata.

In the context of model uncertainty, it would be appropriate to consider instead the direct relationship between the response and the set of treatments, similar to (2):

$$\boldsymbol{Y} = \boldsymbol{X_t}\boldsymbol{\mu_t} + \sum_{i=1}^{s} \boldsymbol{Z}_i\boldsymbol{\varepsilon}_i. \tag{15}$$

Each column of $\boldsymbol{X_t}$ corresponds to a treatment defined as a combination of the experimental factors; $\boldsymbol{Z}_i$ is the indicator matrix of random effects at stratum i.

Gilmour *et al.* (2017b) introduced the notion of "Pure Error REML", where random effects are estimated from the treatment model (15), and then for the sake of making inference regarding parameters of the response model (13), they are substituted in the GLS estimators of the fixed effects (14). It is also argued that some appropriate corrections are to be adapted and applied to the obtained estimates, i.e. the ones suggested by Kenward and Roger (1997).

A careful definition of "pure error" is necessary when considering the treatment and response surface models, especially in the presence of nested random block effects. Gilmour and Trinca (2000) discuss the pure error estimation issues in the context of blocked experiments.

[OE: The following is to be paraphrased – made much shorter or removed overall?] Pure error is expected to measure the variability between experimental units regardless of the treatments applied, and therefore, the assumption of treatment-unit additivity is essential in unblocked experiments. In the presence of one or several blocking factors, the existence of block-treatment interaction would imply that pure error in the lower strata can only be estimated from the replicates within blocks, so that inter-block information is not taken into consideration. Hence, in the blocked cases the assumption of treatment-block additivity (Draper and Smith, 1998) is desirable. Gilmour and Trinca (2000) argue for the preservation of treatment-unit additivity when either fixed or random block effects are in the model in order to preserve consistency with the unblocked experiments; that would also conveniently imply treatment-block additivity. Adoption of these assumptions would then lead to the following representation of the relationship between responses and treatments in multi-stratum experiments.

The outcome of applying treatment i $(i = 1 \dots T)$ to the experimental run located in

the units j_1 of the first stratum, j_2 of the second stratum, ... , and unit j_s of the $s-$th stratum can be expressed as follows:

$$y_{ij_1\ldots j_s} = \mu + t_i + b_{j_1} + \ldots + b_{j_s}, \tag{16}$$

where b_{j_s} is usually denoted as between-run variation.

Under the stronger assumption of the response surface model (13), the treatment effects t_i are represented as a set of polynomial model terms.

Based on the interrelation of the two models, and the desirability of being able to provide the possibility for testing for the lack-of-fit together with obtaining robust estimates of the variance components, the approach of using pure error is adopted, Gilmour *et al.* (2017a) explored several approaches to estimating the variance components.

[OE: the end of the part to be re-written/excluded]

The first approach is stratum-by-stratum data analysis, where each randomisation level is considered separately, and at each stratum i units are treated as runs aggregated in m_{i-1} blocks with fixed effects. The lower stratum variance is estimated from within blocks (using intra-block residual mean square), and the higher stratum (inter-block) variance from the difference between the intra- and inter-block residual mean squares scaled by the number of runs per block (Hinkelmann and Kempthorne, 2005). This strategy requires only the treatment-unit additivity assumption and provides unbiased treatment effects' estimators. However, it would be desirable to make use of combining information not just from within-strata (intra-block) treatment replicates, but also from between-strata (inter-block) ones, especially in the common cases of relatively small experiments when the amount of runs available does not allow for obtaining higher stratum variance components from only considering blocks as whole units. Yates' procedure, first suggested by Yates (1939) and described in detail by Hinkelmann and Kempthorne (2005), provides more degrees of freedom for the variance estimate in the higher stratum – and this is the approach we are adopting in this work.

In the case of two levels of randomisation, variance components are estimated in two variance decomposition steps:

1. Total SS = SS(Blocks) + SS(Treatments|Blocks) + SS(Residual)
   From fitting this full treatment-after-blocks model the residual mean square S^2 is taken as the estimate of the intra-block variance. SS(Residual) is then substituted in the following partition.

2. Total SS = SS(Treatments) + SS(Blocks|Treatments) + SS(Residual)
   Therefore, the sums of squares for the blocks after the treatment effects have been accounted for is obtained and the corresponding mean square

$$S_b^2 = \text{SS(Blocks|Treatments)}/\nu_b$$

is then used to get an estimate of the inter-block variance: $\hat{\sigma}_b^2 = \frac{\nu_b(S_b^2 - S^2)}{d}$, where pure error degrees of freedom for inter-block variance is $d = n - \text{trace}[\boldsymbol{Z}'\boldsymbol{X_t}(\boldsymbol{X_t}'\boldsymbol{X_t})^{-1}\boldsymbol{X_t}'\boldsymbol{Z}]$, $\nu_b = \text{rank}([\boldsymbol{X_t}\boldsymbol{Z}]) - \text{rank}(\boldsymbol{X_t})$, and $\boldsymbol{X_t}$ and $\boldsymbol{Z}$ are as in (15).

Allocation of the degrees of freedom in practice, following Yates' procedure, is provided in the Appendix.

[OE: to shorten]
Gilmour *et al.* (2017a) show that in addition to the treatment-unit additivity and randomisation the assumption of the experimental units being a random sample from an infinite population is necessary for the inter-block variance estimate to be unbiased. In this particular context, when the normality and independence of responses in (13) is a standard assumption, although a potential presence of contamination effects in the fixed part of the model is also accounted for, Yates' approach seems to be the most appropriate technique. However, relying on the distribution assumption to the extent that makes REML the most sensible approach means that the fixed part of the fitted model is assumed to be absolutely true, that is the parameters of the population distribution are known, which does not comply with the model uncertainty framework.

Together with the specifications regarding estimation of the variance components, it is necessary to carefully establish the way the pure error degrees of freedom should be determined and how the available remaining number of treatment degrees of freedom would be distributed between the polynomial, lack-of-fit and inter-block information components.

4.2 MSE-based compound criteria for multistratum experiments

In the presence of potential model disturbance which is expressed as additional polynomial terms, the full model for a multistratum experiment is then:

$$\boldsymbol{Y} = \boldsymbol{X}_p\boldsymbol{\beta}_p + \boldsymbol{X}_q\boldsymbol{\beta}_q + \sum_{i=1}^{s} \boldsymbol{Z}_i\boldsymbol{\varepsilon}_i, \tag{17}$$

where $\boldsymbol{X}_p$ is the primary terms matrix, and bmX_q contains potential terms.

The design construction will follow the approach developed by Trinca and Gilmour (2015): an iterative, stratum-by-stratum algorithm, that treats the higher stratum units as fixed blocks; such an approach does not require any prior assumptions regarding the values of the variance components.

At each step a candidate set of treatments applied at the current stratum is to be generated, and the point exchange algorithm is applied. The model matrix to be used in the optimisation comprises model terms from all higher strata (with the factor values for individual runs obtained at previous steps), terms from the current stratum and interactions between these two groups of terms, if there are any to be considered.

The allocation of the degrees of freedom shall be implemented according to the Yates procedure, and some illustrative examples will be given later on.

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 $\tilde{\boldsymbol{X}}_1 = [\boldsymbol{Z}, \boldsymbol{X}_1]$, where the columns of $\boldsymbol{Z}$ contain block indicators, and the columns of $\boldsymbol{X}_1$ correspond to primary terms. Let also $\tilde{\boldsymbol{\beta}}_1 = [\boldsymbol{\beta}_B, \boldsymbol{\beta}_1]'$ be the joint vector of fixed block effects and primary model terms, and by $\hat{\tilde{\boldsymbol{\beta}}}_1$ 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

$$\begin{aligned}\mathrm{MSE}(\hat{\tilde{\boldsymbol{\beta}}}_1|\tilde{\boldsymbol{\beta}}) =& \mathtt{E}_{\boldsymbol{Y}|\boldsymbol{\beta}}[(\hat{\tilde{\boldsymbol{\beta}}}_1 - \tilde{\boldsymbol{\beta}}_1)(\hat{\tilde{\boldsymbol{\beta}}}_1 - \tilde{\boldsymbol{\beta}}_1)'] \\ =& \sigma^2(\tilde{\boldsymbol{X}}_1'\tilde{\boldsymbol{X}}_1)^{-1} + \tilde{\boldsymbol{A}}\boldsymbol{\beta}_2\boldsymbol{\beta}_2'\tilde{\boldsymbol{A}}', \end{aligned} \tag{18}$$

where $\tilde{\boldsymbol{A}} = (\tilde{\boldsymbol{X}}_1'\tilde{\boldsymbol{X}}_1)^{-1}\tilde{\boldsymbol{X}}_1'\boldsymbol{X}_2$ 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:

$$\begin{aligned}&\mathtt{E}_{\boldsymbol{Y}|\boldsymbol{\beta}}[(\hat{\tilde{\boldsymbol{\beta}}}_1 - \tilde{\boldsymbol{\beta}}_1)(\hat{\tilde{\boldsymbol{\beta}}}_1 - \tilde{\boldsymbol{\beta}}_1)'] = \\ &\mathtt{E}_{\boldsymbol{Y}|\boldsymbol{\beta}}\{[\hat{\tilde{\beta}}_{11} - \tilde{\beta}_{11}, \ldots, \hat{\tilde{\beta}}_{1b} - \tilde{\beta}_{1b}, \hat{\tilde{\beta}}_{1b+1} - \tilde{\beta}_{1b+1}, \ldots, \hat{\tilde{\beta}}_{1b+p} - \tilde{\beta}_{1b+p}]\times \\ &[\hat{\tilde{\beta}}_{11} - \tilde{\beta}_{11}, \ldots, \hat{\tilde{\beta}}_{1b} - \tilde{\beta}_{1b}, \hat{\tilde{\beta}}_{1b+1} - \tilde{\beta}_{1b+1}, \ldots, \hat{\tilde{\beta}}_{1b+p} - \tilde{\beta}_{1b+p}]'\} = \\ &\mathtt{E}_{\boldsymbol{Y}|\boldsymbol{\beta}}\{[\hat{\boldsymbol{\beta}}_B - \boldsymbol{\beta}_B, \hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1][\hat{\boldsymbol{\beta}}_B - \boldsymbol{\beta}_B, \hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1]'\} = \\ &\begin{bmatrix} \mathtt{E}_{\boldsymbol{Y}|\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}_B - \boldsymbol{\beta}_B)(\hat{\boldsymbol{\beta}}_B - \boldsymbol{\beta}_B)' & \mathtt{E}_{\boldsymbol{Y}|\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}_B - \boldsymbol{\beta}_B)(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1)' \\ \mathtt{E}_{\boldsymbol{Y}|\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1)(\hat{\boldsymbol{\beta}}_B - \boldsymbol{\beta}_B)' & \mathtt{E}_{\boldsymbol{Y}|\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1)(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1)' \end{bmatrix}. \end{aligned} \tag{19}$$

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 $\mathrm{MSE}(\hat{\tilde{\boldsymbol{\beta}}}_1|\tilde{\boldsymbol{\beta}})_{22}$. Then the corresponding submatrix of the first summand in (18) is

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

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

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

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

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

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 (18):

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

Therefore, the submatrix of (18) corresponding to the primary terms is

$$\begin{aligned} \mathrm{MSE}(\hat{\tilde{\boldsymbol{\beta}}}_1|\tilde{\boldsymbol{\beta}})_{22} =& \sigma^2(\boldsymbol{X}_1'\boldsymbol{Q}\boldsymbol{X}_1)^{-1} + (\boldsymbol{X}_1'\boldsymbol{Q}\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{Q}\boldsymbol{X}_2\boldsymbol{\beta}_2\boldsymbol{\beta}_2'\boldsymbol{X}_2'\boldsymbol{Q}\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{Q}\boldsymbol{X}_1)^{-1} = \\ & \sigma^2\tilde{\boldsymbol{M}}^{-1} + \tilde{\boldsymbol{M}}^{-1}\boldsymbol{X}_1'\boldsymbol{Q}\boldsymbol{X}_2\boldsymbol{\beta}_2\boldsymbol{\beta}_2'\boldsymbol{X}_2'\boldsymbol{Q}\boldsymbol{X}_1\tilde{\boldsymbol{M}},^{-1} \end{aligned} \tag{20}$$

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

As in the unblocked case (8), 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:

$$\begin{aligned} \det[\mathrm{MSE}(\hat{\tilde{\boldsymbol{\beta}}}_1|\tilde{\boldsymbol{\beta}})_{22}] =& \det[\sigma^2\tilde{\boldsymbol{M}}^{-1} + \tilde{\boldsymbol{M}}^{-1}\boldsymbol{X}_1'\boldsymbol{Q}\boldsymbol{X}_2\boldsymbol{\beta}_2\boldsymbol{\beta}_2'\boldsymbol{X}_2'\boldsymbol{Q}\boldsymbol{X}_1\tilde{\boldsymbol{M}}^{-1}] = \\ & \sigma^{2p}\det[\tilde{\boldsymbol{M}}^{-1} + \tilde{\boldsymbol{M}}^{-1}\boldsymbol{X}_1'\boldsymbol{Q}\boldsymbol{X}_2\tilde{\boldsymbol{\beta}}_2\tilde{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{Q}\boldsymbol{X}_1\tilde{\boldsymbol{M}}^{-1}] = \\ & \sigma^{2p}\det[\tilde{\boldsymbol{M}}^{-1}](1 + \tilde{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{Q}\boldsymbol{X}_1'\tilde{\boldsymbol{M}}^{-1}\boldsymbol{X}_1'\boldsymbol{Q}\boldsymbol{X}_2\tilde{\boldsymbol{\beta}}_2). \end{aligned} \tag{21}$$

The expression in (21) is similar to the one we have in the unblocked case (9):

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

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 $\mathcal{N}(\mathbf{0}, \tau^2 \boldsymbol{I}_q)$, so that this prior does not depend on the error variance $\sigma^2_.$.

Taking the expectation of the logarithm of (21) over the set prior distribution is completely identical to the derivations shown on page 8; and the MSE-component in the determinant criterion is

$$\log(\det[\tilde{\boldsymbol{M}}^{-1}]) + \mathtt{E}_{\tilde{\boldsymbol{\beta}}_2} \log(1 + \tilde{\boldsymbol{\beta}}_2' \boldsymbol{X}_2' \boldsymbol{Q} \boldsymbol{X}_1 \tilde{\boldsymbol{M}}^{-1} \boldsymbol{X}_1' \boldsymbol{Q} \boldsymbol{X}_2 \tilde{\boldsymbol{\beta}}_2). \tag{22}$$

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

$$\begin{aligned} \text{minimise} \quad & \left[\left| (\boldsymbol{X}_1' \boldsymbol{Q} \boldsymbol{X}_1)^{-1} \right|^{1/p} F_{p,d_B;1-\alpha_{DP}} \right]^{\kappa_{DP}} \times \\ & \left[\left| \tilde{\boldsymbol{L}} + \frac{\boldsymbol{I}_q}{\tau^2} \right|^{-1/q} F_{q,d_B;1-\alpha_{LoF}} \right]^{\kappa_{LoF}} \times \\ & \left[|\boldsymbol{X}_1' \boldsymbol{Q} \boldsymbol{X}_1|^{-1} \exp\left(\frac{1}{N} \sum_{i=1}^{N} \log(1 + \tilde{\boldsymbol{\beta}}_{2i}' \boldsymbol{X}_2' \boldsymbol{Q} \boldsymbol{X}_1 \tilde{\boldsymbol{M}}^{-1} \boldsymbol{X}_1' \boldsymbol{Q} \boldsymbol{X}_2 \tilde{\boldsymbol{\beta}}_{2i}) \right) \right]^{\kappa_{MSE}/p}_{.} \end{aligned} \tag{23}$$

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 $\sigma\tau\mathbf{1}_q$, so that the scaled version of it incorporated in (23) $\tilde{\boldsymbol{\beta}}_2$ is equal to $\tau\mathbf{1}_q$. Then the determinant-based criterion function with the point prior estimate, which is referred to as "MSE(D)P", is

$$\begin{aligned} \text{minimise} \quad & \left[\left| (\boldsymbol{X}_1' \boldsymbol{Q} \boldsymbol{X}_1)^{-1} \right|^{1/p} F_{p,d_B;1-\alpha_{DP}} \right]^{\kappa_{DP}} \times \\ & \left[\left| \tilde{\boldsymbol{L}} + \frac{\boldsymbol{I}_q}{\tau^2} \right|^{-1/q} F_{q,d_B;1-\alpha_{LoF}} \right]^{\kappa_{LoF}} \times \\ & \left[|\boldsymbol{X}_1' \boldsymbol{Q} \boldsymbol{X}_1|^{-1} \left(1 + \tau^2 \sum_{i,j=1}^{q} [\boldsymbol{X}_2' \boldsymbol{Q} \boldsymbol{X}_1 \tilde{\boldsymbol{M}}^{-1} \boldsymbol{X}_1' \boldsymbol{Q} \boldsymbol{X}_2]_{(i,j)} \right) \right]^{\kappa_{MSE}/p}_{.} \end{aligned} \tag{24}$$

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 (20):

$$\begin{aligned} \mathtt{E}_{\beta_2} \text{trace}[\text{MSE}(\hat{\tilde{\boldsymbol{\beta}}}_1 | \tilde{\boldsymbol{\beta}})_{22}] = & \text{trace}[\mathtt{E}_{\beta_2} \text{MSE}(\hat{\tilde{\boldsymbol{\beta}}}_1 | \tilde{\boldsymbol{\beta}})_{22}] = \\ & \text{trace}[\sigma^2 \tilde{\boldsymbol{M}}_{22}^{-1} + \mathtt{E}_{\beta_2}(\tilde{\boldsymbol{A}} \boldsymbol{\beta}_2 \boldsymbol{\beta}_2' \tilde{\boldsymbol{A}})_{22}] = \\ & \sigma^2 \text{trace}[\tilde{\boldsymbol{M}}_{22}^{-1} + \tau^2 \{\tilde{\boldsymbol{A}} \tilde{\boldsymbol{A}}'\}_{22}] = \\ & \sigma^2 [\text{trace}(\boldsymbol{X}_1' \boldsymbol{Q} \boldsymbol{X}_1)^{-1} + \tau^2 \text{trace}\{\tilde{\boldsymbol{A}} \tilde{\boldsymbol{A}}'\}_{22}], \end{aligned} \tag{25}$$

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

$$\begin{aligned}\text{minimise } & \left[\frac{1}{p}\text{trace}(\boldsymbol{W}\boldsymbol{X}_1'\boldsymbol{Q}\boldsymbol{X}_1)^{-1}F_{1,d_B;1-\alpha_{LP}}\right]^{\kappa_{LP}} \times \\ & \left[\frac{1}{q}\text{trace}\left(\tilde{\boldsymbol{L}} + \frac{\boldsymbol{I}_q}{\tau^2}\right)^{-1}F_{1,d_B;1-\alpha_{LoF}}\right]^{\kappa_{LoF}} \times \\ & \left[\frac{1}{p}\text{trace}\{(\boldsymbol{X}_1'\boldsymbol{Q}\boldsymbol{X}_1)^{-1} + \tau^2[\tilde{\boldsymbol{A}}\tilde{\boldsymbol{A}}']_{22}\}\right]^{\kappa_{MSE}}. \end{aligned} \tag{26}$$

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.

Then the optimal multistratum design search is implemented following the steps below:

1. Starting from the first stratum, if there are any factors applied at this level, a candidate set of treatments is formed together with the fitted model matrix comprising the primary terms and the matrix of potential terms. The optimal unblocked design X_1 is then obtained using the usual point-exchange algorithm, by minimising (**??**) or (**??**). The labels of the treatments applied at the current stratum are saved at this stage as well in order to calculate the number of pure error degrees of freedom at the lower strata.

   If there are no factors applied at the first stratum, we move to the second one, and conduct the optimal design search for a blocked experiment, where the number of blocks is equal to the number of units in the first stratum n_1, with n_2 runs per block; in this case criterion function (**??**) or (**??**) is used, and number of pure error degrees of freedom d_B is calculated according to the usual blocked experiment framework. Treatment labels are saved as well.

2. When moving from stratum $i-1$ to stratum i, all factors applied in the higher strata are now treated as "whole-plot" factors. The corresponding model matrix $\boldsymbol{Xw.m}$ containing terms inherited from the higher strata is expanded accordingly, as treatments applied at each unit in stratum $i-1$ are now applied to n_i units of the current stratum nested within it. A similar expansion procedure is carried out for the vector (or matrix, if there are two or more higher strata with factors applied) of the treatment labels, so that for each current unit we are able to see what treatment has been applied to it at each stratum.

Blocking with no factors applied may occur at any stratum, not only at the first one. In such cases the procedure remains the same: skipping to the next stratum with some treatment applied, expanding the design matrices and vectors with treatment labels corresponding to the higher strata.

3. Once there is a model matrix with the "whole-plot" terms is formed, the search procedure might be started for the current stratum. The candidate set of treatments is set for the factors applied in the current stratum; parameters of interest include not only the ones formed by these factors but also their interactions with the higher strata terms. As nesting within the previous strata is treated as fixed block effects, the criteria used are the ones given in (**??**) and (**??**). However, there are a few features worth noting:

   - For each design under consideration during the extensive search procedure, its model matrix $\boldsymbol{X}_1$ is now constructed by binding the "whole-plot" model matrix $\boldsymbol{Xw.m}$, the model comprising the terms formed from the factors applied at the current stratum $\boldsymbol{Xi.m}$, and the matrix formed of the interaction terms (if any are to be included) between the two. The same relates to the construction of the potential terms matrix $\boldsymbol{X}_2$: it needs to be recalculated every time a design point is swapped between the candidate set of the current stratum terms and the current design if it contains any interactions involving terms inherited from the previous strata. If not, it then only comprises terms from the current stratum factors.

   - Presence of the potential terms matrix in the criteria also implies that each stratum i will "have" its own number of potential terms q_i and, therefore, in the cases when the value of the variance scaling parameter τ^2 depends on it, at each stratum the criterion function will be evaluated with the respective values of τ_i^2 instead of some common one for all levels. In this work we consider common values of τ^2, however, it is a case-sensitive parameter, and it is to be discussed in each particular case.

   - As the numbers of primary and potential terms vary from stratum to stratum, so do the significance levels α_{LP} and α_{LoF} in the case of trace-based criterion (**??**):

   $$\alpha_{LP} = 1 - (1 - \alpha_1)^{\frac{1}{p}},$$
   $$\alpha_{LoF} = 1 - (1 - \alpha_2)^{\frac{1}{q}},$$

   as the corrected confidence levels depend on the dimension of the confidence regions (as in (**??**)).

   - We use the same values of weights in the criteria for all strata; however, the flexibility of the algorithm allows changing weights (and even criteria) between the strata.

4. If there are at least 3 strata with some factors applied, and when the current stratum number is 3 or more, an additional swapping procedure is performed (the same as that described by Trinca and Gilmour (2001)). By looking at the $i-2$ stratum units that have the same treatments applied to them, and interchange the $i-1$ stratum units within those, the performance of the design evaluated with respect to the performance at the current stratum i. The same swapping is performed for all the higher strata up to the first one.

5. It is all then repeated from step number 2, until current stratum i reaches the lowest stratum s.

5 Examples

[OE: To be edited]

Here we will study the optimal designs in terms of the criteria (11) and (12) (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 (7) 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, $\boldsymbol{Q}_0 = \boldsymbol{I}_n - \frac{1}{n}\mathbf{1}\mathbf{1}'$, 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:

$$\text{MSE(D):}\quad \left[\left|\boldsymbol{X}_1'\boldsymbol{Q}_0\boldsymbol{X}_1\right|^{-1/(p-1)} F_{p-1,d;1-\alpha_{DP}}\right]^{\kappa_{DP}} \times$$
$$\left[\left|\boldsymbol{L} + \frac{\boldsymbol{I}_q}{\tau^2}\right|^{-1/q} F_{q,d;1-\alpha_{LoF}}\right]^{\kappa_{LoF}} \times$$
$$\left[|\boldsymbol{X}_1'\boldsymbol{Q}_0\boldsymbol{X}_1|^{-1} \exp\left(\frac{1}{N}\sum_{i=1}^{N}\log(1+\tilde{\boldsymbol{\beta}}_{2i}'\boldsymbol{X}_2'\boldsymbol{Q}_0\boldsymbol{X}_1\tilde{\boldsymbol{M}}^{-1}\boldsymbol{X}_1'\boldsymbol{Q}_0\boldsymbol{X}_2\tilde{\boldsymbol{\beta}}_{2i})\right)\right]^{\frac{\kappa_{MSE}}{(p-1)}},$$

$$
\begin{aligned}
\text{MSE(D)P: } & \left[\left|\boldsymbol{X}_1'\boldsymbol{Q}_0\boldsymbol{X}_1\right|^{-1/(p-1)} F_{p-1,d;1-\alpha_{DP}}\right]^{\kappa_{DP}} \times \\
& \left[\left|\boldsymbol{L}+\frac{\boldsymbol{I}_q}{\tau^2}\right|^{-1/q} F_{q,d;1-\alpha_{LoF}}\right]^{\kappa_{LoF}} \times \\
& \left[|\boldsymbol{X}_1'\boldsymbol{Q}_0\boldsymbol{X}_1|^{-1}\left(1+\tau^2\sum_{i,j=1}^{q}[\boldsymbol{X}_2'\boldsymbol{Q}_0\boldsymbol{X}_1\tilde{\boldsymbol{M}}^{-1}\boldsymbol{X}_1'\boldsymbol{Q}_0\boldsymbol{X}_2]_{(i,j)}\right)\right]^{\frac{\kappa_{MSE}}{(p-1)}}, \\
\text{MSE(L): } & \left[\frac{1}{p}\text{trace}(\boldsymbol{W}\boldsymbol{X}_1'\boldsymbol{Q}_0\boldsymbol{X}_1)^{-1}F_{1,d;1-\alpha_{LP}}\right]^{\kappa_{LP}} \times \\
& \left[\frac{1}{q}\text{trace}\left(\boldsymbol{L}+\frac{\boldsymbol{I}_q}{\tau^2}\right)^{-1} F_{1,d;1-\alpha_{LoF}}\right]^{\kappa_{LoF}} \times \\
& \left[\frac{1}{p-1}\text{trace}\{\tilde{\boldsymbol{M}}^{-1}+\tau^2\boldsymbol{A}\boldsymbol{A}'\}_{22}\right]^{\kappa_{MSE}}.
\end{aligned}
$$

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.

5.1 Example. Case study

A company specialising in food supplements production for some animals wanted to conduct an experiment with the main aim of figuring out whether a slight decrease in the recommended dosages of three particular products taken together would make a significant impact on the resulting "performance", which is expressed in terms of some

continuous response. The dosage range of interest for each supplement is from 90% to 100% of the standard recommendation; it is desired that there would be 3 levels (i.e. taking the values of 90%, 95% and 100%). Carrying out the experiment with more than 3 levels was possible but more complicated, as measuring, for example, 92.5% of the recommended dosage might be quite tricky. However, the search was performed among a larger 5-level candidate set of points (which, obviously, contains all 3-level design points), but due to the form of the fitted model and the criteria used, the resulting designs' points consist of only 3-level points.

In total such treatments are to be applied to 36 cages of animals, which would be allocated in 2 equal sized blocks. The response surface model, therefore, would contain all linear, quadratic and interaction terms, so, since there are 3 factors, the number of terms in the fitted model is $p = 9$. However, there were some doubts regarding the expected curvature of the fitted function, in other words, the presence of third-order terms was anticipated: a linear-by-linear-by-linear term, quadratic-by-linear and cubic terms ($q = 10$ overall).

Another restriction was that the experimenters wanted to have at least two centre points in each block to ensure representation of the conditions thought a priori most likely to be best (with dosages of 95% for each supplement), i.e. 4 runs in total were fixed beforehand; this constraint was built directly into the search procedure.

5.2 Design search

Taking into account the possibility of the fitted second-order polynomial providing not the best fit, and the true relationship between the response and the dosages being influenced by the third-order terms, we can express this in terms of the full model

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

and the fitted model:

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

where $\boldsymbol{\varepsilon} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I}_n)$.

The notation here and further on is the same as before:

- $\boldsymbol{X}_1$ – model matrix of primary terms
- $\tilde{\boldsymbol{X}}_1 = [\boldsymbol{Z}, \boldsymbol{X}_1]$, where columns of $\boldsymbol{Z}$ contain block indicators
- $\boldsymbol{X}_2$ – model matrix of potential terms
- $\tilde{\boldsymbol{M}} = \boldsymbol{X}_1'\boldsymbol{Q}\boldsymbol{X}_1$, $\boldsymbol{Q} = \boldsymbol{I}_n - \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'$
- $\tilde{\boldsymbol{L}} = \boldsymbol{X}_2'\boldsymbol{X}_2 - \boldsymbol{X}_2'\tilde{\boldsymbol{X}}_1(\tilde{\boldsymbol{X}}_1'\tilde{\boldsymbol{X}}_1)^{-1}\tilde{\boldsymbol{X}}_1'\boldsymbol{X}_2$ – dispersion matrix, and
- $\tilde{\boldsymbol{A}} = (\tilde{\boldsymbol{X}}_1'\tilde{\boldsymbol{X}}_1)^{-1}\tilde{\boldsymbol{X}}_1'\boldsymbol{X}_2$ – alias matrix for the blocked model.

The experimenters preferred using the determinant-based criterion as the primary inferential interest is on the overall input of the model terms. Together with good precision of the parameters to be estimated ($\boldsymbol{\beta}_1$), other objectives are to minimise the impact of the potential terms' possible presence and the bias and variance of the primary terms' estimators.

Hence it was decided to optimise the mean square error based criterion function, in the same form as it was derived in (23):

$$\begin{aligned}
&\left[\left|\tilde{\boldsymbol{M}}\right|^{-1/p} F_{p,d_B;1-\alpha_{DP}}\right]^{\kappa_{DP}} \times \\
&\left[\left|\tilde{\boldsymbol{L}} + \frac{\boldsymbol{I}_q}{\tau^2}\right|^{-1/q} F_{q,d_B;1-\alpha_{LoF}}\right]^{\kappa_{LoF}} \times \\
&\left[|\tilde{\boldsymbol{M}}|^{-1} \exp\left(\frac{1}{N}\sum_{i=1}^{N} \log(1 + \tilde{\boldsymbol{\beta}}_{2i}' \boldsymbol{X}_2' \boldsymbol{Q} \boldsymbol{X}_1 \tilde{\boldsymbol{M}}^{-1} \boldsymbol{X}_1' \boldsymbol{Q} \boldsymbol{X}_2 \tilde{\boldsymbol{\beta}}_{2i})\right)\right]^{\kappa_{MSE}/p}.
\end{aligned} \tag{27}$$

Recall that here d_B is the number of pure error degrees of freedom for the blocked experiment, as was described in Section **??**.

Although the primary choice of the number of levels of each factor is 3, the number of experimental runs would allow for 5 levels so that all of the potential terms can be estimated. So we considered one more compound criterion which has a different lack-of-fit component: DP_S– optimality for potential terms in the full (third-order polynomial) model, arising from the D_S-optimality that was suggested by Atkinson *et al.* (2007) (page 360) in the context of model discrimination:

$$\begin{aligned}
&\left[\left|\tilde{\boldsymbol{M}}\right|^{-1/p} \mathrm{F}(p, d, \alpha_{DP})\right]^{\kappa_{DP}} \times \left[|\tilde{\boldsymbol{L}}|^{-1/q} F_{q,d_B;1-\alpha_{LoF}}\right]^{\kappa_{LoF}} \times \\
&\left[|\tilde{\boldsymbol{M}}|^{-1} \exp\left(\frac{1}{N}\sum_{i=1}^{N} \log(1 + \tilde{\boldsymbol{\beta}}_{2i}' \boldsymbol{X}_2' \boldsymbol{Q} \boldsymbol{X}_1 \tilde{\boldsymbol{M}}^{-1} \boldsymbol{X}_1' \boldsymbol{Q} \boldsymbol{X}_2 \tilde{\boldsymbol{\beta}}_{2i})\right)\right]^{\kappa_{MSE}/p}.
\end{aligned} \tag{28}$$

This criterion further on will be referred to as the "full" criterion, and we shall explore how this would affect the resulting designs.

Regarding the parameters of the criteria function, they need to be determined for the design search:

- τ^2 is the scaling parameter of the variance of the potential terms such that $\boldsymbol{\beta}_2 \sim \mathcal{N}(\mathbf{0}, \tau^2\sigma^2\boldsymbol{I}_q)$ and $\tilde{\boldsymbol{\beta}}_2 \sim \mathcal{N}(\mathbf{0}, \tau^2\boldsymbol{I}_q)$. For each criterion we will consider two cases: $\tau^2 = 1$ and $\tau^2 = 1/q$.

- N is the number of Monte Carlo samples used to estimate the third criterion component. For $\tau^2 = 1$ $N = 500$, and for $\tau^2 = 1/q$ we set $N = 1000$ in order to have a sufficiently small relative estimation error.

- The number of random starts in the point exchange algorithm is set to 50.

One of the pre-specified requirements was the presence of at least two centre points in each block. In order to implement that, the point exchange algorithm was slightly amended: at each random initial non-singular designs the first two points in each block are set to be centre points, and every loop with the exchange steps would then start from the third point onwards, so that whatever changes are to be accepted by the algorithm, it would always keep these two points; such an amendment actually reduced a bit the number of exchanges to be performed.

For the sake of assessing the quality of the designs obtained with this restriction, we will obtain also designs without any and will evaluate the efficiency losses.

5.3 Results

In this section we will give a summary of the designs we looked at while searching for the best one for this particular experimental setup. We considered three schemes of weight allocations: the first one is with the weight being equally distributed between the components, the second one – with a bit more weight (0.4) put on minimising the variation of the primary terms' estimators and their bias, with the rest allocated to the lack-of-fit component. The third one has half of the weight on the MSE-component with the rest of it distributed evenly among the other two components. For the sake of reference we also provided the performances of the DP-, $LOF(DP)$- and $MSE(D)$-optimal designs (#4 – #6 in the tables below).

Table 5 provides the performances of the optimal designs with respect to the criteria given in (27), without forcing the presence of two centre points per block. The main feature observed is that in general individual efficiency values are quite large (especially in comparison to the previously seen example); this might be explained by the fact that the number of available residual degrees of freedom ($n - b - p = 25$) is quite large compared to the number of primary terms ($p = 9$) that are to be estimated, and this flexibility results in better compromises achievable between the three criterion components.

Now we introduce the compulsory centre points, and see how the resulting designs perform. Table 6 contains two types of efficiencies: "Global" efficiencies are calculated with respect to the optimal designs obtained without any fixed points, whilst "Local" efficiencies compare the obtained designs with the best designs among the blocked designs with at least two centre points per block. The latter ones will be, obviously, larger, and the differences between the two for the same components represent the measure of the loss by restricting the set of designs to be considered. The last column of this table, "Relative Efficiency," is essentially the efficiency of a given design with respect to the optimal one in terms of the same compound criterion (i.e. with the same allocation of weights), obtained without any restrictions; this value gives a "single-

Table 5: Case-study. Properties of MSE(D)-optimal blocked designs

	Criteria, $\tau^2 = 1$			DoF		Efficiency, %		
	DP	LoF(DP)	MSE(D)	PE	LoF	DP	LoF(DP)	MSE(D)
1	1/3	1/3	1/3	16	9	90.58	95.60	97.34
2	0.4	0.2	0.4	16	9	93.09	89.23	98.68
3	0.25	0.25	0.5	14	11	89.10	94.04	99.52
4	1	0	0	21	4	100.00	71.70	98.84
5	0	1	0	14	11	64.62	100.00	76.95
6	0	0	1	8	17	69.48	75.00	100.00

	Criteria, $\tau^2 = 1/q$			DoF		Efficiency, %		
	DP	LoF(DP)	MSE(D)	PE	LoF	DP	LoF(DP)	MSE(D)
1	1/3	1/3	1/3	21	4	100.00	99.90	98.70
2	0.4	0.2	0.4	21	4	99.64	99.71	98.88
3	0.25	0.25	0.5	21	4	100.00	99.90	100.53
4	1	0	0	21	4	100.00	98.57	99.83
5	0	1	0	21	4	86.31	100.00	87.26
6	0	0	1	8	17	69.38	72.21	100.00

value" perspective on the efficiency loss.

Table 6: Case-study. Properties of MSE(D)-optimal blocked designs, with two centre points per block

	Criteria, $\tau^2 = 1$			DoF		'Global' Efficiency,%			'Local' Efficiency,%			Relative
	DP	LoF(DP)	MSE(D)	PE	LoF	DP	LoF(DP)	MSE(D)	DP	LoF(DP)	MSE(D)	Efficiency,%
1	1/3	1/3	1/3	14	11	88.63	90.03	99.97	92.89	97.59	100.75	98.18
2	0.4	0.2	0.4	14	11	88.35	90.14	99.15	92.60	97.70	99.92	98.31
3	0.25	0.25	0.5	14	11	88.63	90.03	99.75	92.89	97.59	100.53	98.90
4	1	0	0	20	5	95.41	62.13	95.24	100.00	67.34	95.98	95.58
5	0	1	0	14	11	66.49	92.26	79.91	69.69	100.00	80.53	92.26
6	0	0	1	14	11	88.31	88.35	99.23	92.55	95.77	100.00	99.23

	Criteria, $\tau^2 = 1/q$			DoF		'Global' Efficiency,%			'Local' Efficiency,%			Relative
	DP	LoF(DP)	MSE(D)	PE	LoF	DP	LoF(DP)	MSE(D)	DP	LoF(DP)	MSE(D)	Efficiency,%
1	1/3	1/3	1/3	18	7	92.52	95.13	95.86	96.97	98.00	96.55	94.95
2	0.4	0.2	0.4	18	7	94.19	92.19	96.54	98.72	94.98	97.24	95.34
3	0.25	0.25	0.5	17	8	92.44	93.70	97.05	96.88	96.53	97.75	95.72
4	1	0	0	20	5	95.41	90.63	95.66	100.00	93.37	96.35	95.41
5	0	1	0	22	3	76.11	97.07	77.10	79.77	100.00	77.65	97.07
6	0	0	1	14	11	88.37	90.37	99.29	92.62	93.10	100.00	99.29

The first observation to be made here is that the imbalance in the distribution of the residual degrees of freedom is not as strong as in the unrestricted case. However, the prevalence towards the pure error degrees of freedom is still evident. Designs' efficiencies in terms of individual components are still quite large, both "local" and "global" ones. In general, the performances are better for smaller τ^2 (except for the $MSE(D)$ part), as in this case the model disturbance effect is assumed to be quite small, and the compromise might be more feasible. Overall, relative efficiencies are quite good, losses among the first three designs (optimal with respect to the compound criteria) do not exceed 1.82% for τ^2 and 5.05% for $\tau^2 = 1/q$.

It is notable that designs #1 and #3 (for the larger τ^2) are the same. Its $MSE(D)$-value turned out to be better than of the design #6, which was constructed as being optimal with respect to this component. This design was chosen to be used when carrying out the experiment; the experiment was run successfully and useful conclusions were drawn from the obtained data. The design can be found in Table 7 below, the points in each block have been ordered for the sake of easier perception, and, of course, they have to

be and were randomised in each block before running the experiment.

Table 7: Case-study. MSE(D)-optimal design #1 with two centre points, $\tau^2 = 1$

-1	-1	-1	-1	-1	-1
-1	-1	0	-1	-1	0
-1	-1	1	-1	-1	1
-1	0	-1	-1	0	1
-1	1	-1	-1	1	-1
-1	1	1	-1	1	0
-1	1	1	0	-1	-1
0	-1	1	0	-1	1
0	0	0	0	0	0
0	0	0	0	0	0
0	1	-1	0	1	1
1	-1	-1	1	-1	-1
1	-1	0	1	-1	0
1	-1	1	1	-1	1
1	0	1	1	0	-1
1	1	-1	1	0	1
1	1	0	1	1	-1
1	1	1	1	1	1

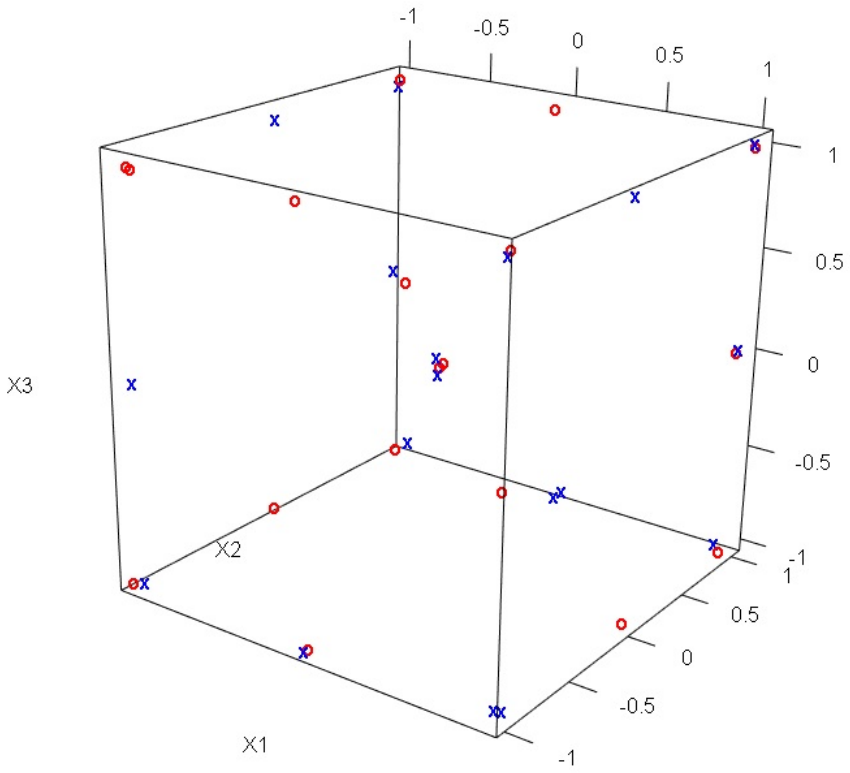

Figure 1: MSE(D)-optimal design #1 with two centre points, $\tau^2 = 1$

Figure 1 provides a graphical representation of the design: the three axes correspond to the experimental factors, each is at three levels (i.e. scaled to $[-1, 1]$ dosages): -1, 0 and 1. Each dot is a design point, and two colours (blue and red) and two symbols ('x' and 'o') serve as block indicators. We can see that there are only two centre points in each block, as required, and replicates of other points are generally split between blocks (except for the $(-1, 1, 1)$ point which is duplicated in the first block only).

All other designs may be found in the Supplementary material. Time-wise, for the

search algorithm's parameters outlined in the beginning of this section, on average an optimal design was found in $13-15$ hours, which was acceptable in this particular case. Sometimes, however, it took up to $20-24$ hours, so some extra time allowance should be accounted for when using these criteria and this search algorithm.

Further we decided to check what designs we would get if there were 5 levels of each factor rather than 3, so that all of the third-order terms are estimable (which is also feasible due to a large enough number of runs), i.e. the lack-of-fit component minimising the posterior confidence region for the potential terms would be replaced by the DP_S-optimality for them, as in (28).

The summary of the corresponding optimal designs is given in Table 8: the 'new' lack-of-fit component is denoted by DP*s. The notions of 'global' and 'local' efficiencies are the same as previously. For each design we calculated its $LoF(DP)$-value, so that we could assess how they would perform in terms of the criterion (27). The performances of the optimal designs constructed without the restriction of two centre points per block are summarised in Table **??** in Appendix **??**; all of the designs are also provided in the Supplementary material. In order to illustrate general tendencies in the designs' appearances, designs optimal with respect to the criterion with equal weights, for two values of τ^2 are provided in Table 9.

Table 8: Case-study. Properties of "Full" MSE(D)-optimal blocked designs, with two centre points per block

	Criteria, $\tau^2=1$			**DoF**		**'Global' Efficiency,%**				**'Local' Efficiency,%**				**Relative**
	DP	**DP*s**	**MSE(D)**	**PE**	**LoF**	**DP**	**DP*s**	**LoF(DP)**	**MSE(D)**	**DP**	**DP*s**	**LoF(DP)**	**MSE(D)**	**Efficiency,%**
1	1/3	1/3	1/3	14	11	80.58	69.38	87.05	91.74	84.46	79.90	94.36	92.46	94.42
2	0.4	0.2	0.4	14	11	83.90	61.18	85.36	57.47	87.93	70.46	92.52	57.92	95.83
3	0.25	0.25	0.5	13	12	79.93	68.24	87.06	94.05	83.77	78.59	94.37	94.79	96.32
4	1	0	0	20	5	95.41	0.00	62.13	95.24	100.00	0.00	67.34	95.98	95.41
5	0	1	0	14	11	62.01	86.83	92.98	74.95	64.99	100.00	100.78	75.53	86.83
6	0	0	1	14	11	88.31	0.00	88.35	99.23	92.55	0.00	95.77	100.00	99.23

	Criteria, $\tau^2=1/q$			**DoF**		**'Global' Efficiency,%**				**'Local' Efficiency,%**				**Relative**
	DP	**DP*s**	**MSE(D)**	**PE**	**LoF**	**DP**	**DP*s**	**LoF(DP)**	**MSE(D)**	**DP**	**DP*s**	**LoF(DP)**	**MSE(D)**	**Efficiency,%**
1	1/3	1/3	1/3	14	11	80.14	69.01	87.70	91.41	83.99	81.43	90.34	92.07	93.97
2	0.4	0.2	0.4	14	11	83.76	61.29	88.28	94.41	87.79	72.31	90.95	95.09	96.04
3	0.25	0.25	0.5	12	13	79.66	64.43	84.44	94.84	83.49	76.02	86.99	95.52	95.20
4	1	0	0	20	5	95.41	0.00	90.63	95.66	100.00	0.00	93.37	96.35	95.41
5	0	1	0	14	11	59.81	84.75	86.15	72.05	62.68	100.00	88.75	72.57	84.75
6	0	0	1	14	11	88.73	0.00	90.57	99.29	92.99	0.00	93.30	100.00	99.29

In case of $\tau^2 = 1$ all pure error degrees of freedom (except for the 2 coming from the replicated centre points) occur from 12 points duplicated in different blocks; in case of $\tau^2 = 1/q$ (i.e. $\tau^2 = 0.1$) two 'corner' points are replicated within the same block (they are highlighted in Table 9). Quite a few experimental units would receive an 'intermediate' ± 0.5 dosage of at least one product, and as this did not comply with the demands of the experimenters, the choice was still made in favour of the three-level design in Table 7.

Table 9: Case-study. Designs #1 from Table 8, $\tau^2 = 1$ (left) and $\tau^2 = 1/q$ (right)

-1	-1	0	-1	-1	-1		-1	-1	-1	-1	-1	-1
-1	-1	1	-1	-1	0		-1	-1	0.5	-1	-1	0.5
-1	0	-1	-1	-1	1		-1	-0.5	1	-1	-0.5	1
-1	0.5	1	-1	0.5	1		-1	0	-0.5	-1	1	-1
-1	1	-1	-1	1	-1		-1	1	-1	-1	1	1
-1	1	0.5	-1	1	0.5		-1	1	1	-0.5	-1	-0.5
-0.5	1	1	-0.5	-0.5	1		-0.5	1	0	-0.5	0.5	-1
0	-1	-1	-0.5	1	-0.5		0	-1	1	0	-1	1
0	0	0	0	-1	-1		0	0	0	0	0	0
0	0	0	0	0	0		0	0	0	0	0	0
0.5	-1	1	0	0	0		0	1	1	0	1	1
0.5	1	-1	0.5	-1	1		0.5	-0.5	-1	0.5	1	-0.5
1	-1	-1	0.5	1	-1		1	-1	-1	1	-1	-1
1	-1	0.5	1	-1	-1		1	-1	0	**1**	**-1**	**1**
1	-0.5	1	1	-1	0.5		1	0.5	1	**1**	**-1**	**1**
1	0.5	-1	1	-0.5	-0.5		**1**	**1**	**-1**	1	0	-0.5
1	1	-0.5	1	0.5	-1		**1**	**1**	**-1**	1	0.5	1
1	1	1	1	1	1		1	1	0.5	1	1	0.5

6 Discussion

7 Acknowledgements

References

Atkinson, A. C., Donev, A. N. and Tobias, R. D. (2007) *Optimum Experimental Designs, with SAS*. Oxford 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., Goos, P. and Großmann, H. (2017a) Distribution-robust methods for analyzing data from split-plot and multi-stratum designs. In preparation.

Gilmour, S. G., Goos, P. and Großmann, H. (2017b) Pure error REML for analyzing data from split-plot and multi-stratum designs. In preparation.

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., Kobilinsky, A., O'Brien, T. E. and Vandebroek, M. (2005) Model-robust and model-sensitive designs. *Computational Statistics & Data Analysis*, **49**, 201–216.

Goos, P., Langhans, I. and Vandebroek, M. (2006) Practical inference from industrial split-plot designs. *Journal of Quality Technology*, **38**, 162–179.

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

Hinkelmann, K. and Kempthorne, O. (2005) *Design and Analysis of Experiments. Volume 2: Advanced Experimental Design.* John Wiley & Sons.

Kenward, M. G. and Roger, J. H. (1997) Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, 983–997.

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

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

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.

McCulloch, C. E. and Searle, S. R. (2001) *Generalized, Linear and Mixed Models.* John Wiley & Sons.

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.

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.

Yates, F. (1939) The recovery of inter-block information in variety trials arranged in three-dimensional lattices. *Annals of Eugenics*, **9**, 136–156.