



Multi-level Block Designs for Comparative Experiments

Rodney N. EDMONDSON 

Complete replicate block designs are fully efficient for treatment effects and are the designs of choice for many agricultural field experiments. For experiments with a large number of treatments, however, they may not provide good control of variability over the whole experimental area. Nested incomplete block designs with a single level of nesting can then improve ‘within-block’ homogeneity for moderate sized experiments. For very large designs, however, a single level of nesting may not be adequate and this paper discusses multi-level nesting with hierarchies of nested blocks. Multi-level nested block designs provide a range of block sizes which can improve ‘within-block’ homogeneity over a range of scales of measurement. We discuss design and analysis of multi-level block designs for hierarchies of nested blocks including designs with crossed block factors. We describe an R language package for multi-level block design and we exemplify the design and analysis of multi-level block designs by a simulation study of block designs for cereal variety trials in the UK. Finally, we re-analyse a single large row-and-column field trial for 272 spring barley varieties in 16 rows and 34 columns assuming an additional set of multi-level nested column blocks superimposed on the existing design. For each example, a multi-level mixed blocks analysis is compared with a spatial analysis based on hierarchical generalized additive (HGAM) models. We discuss the combined analysis of random blocks and HGAM smoothers in the same model.

Key Words: Block designs; Mixed models; GAM models; HGAM models; Hierarchical nesting; Nested blocks; Row-and-column blocks; D-optimality; Trend analysis.

1. INTRODUCTION

Comparative experiments in agriculture often involve the estimation of treatment effects against a background of high plot variability. Effective control of plot variability is essential for good treatment estimation and the most common method of control is the randomized complete blocks design. Randomized complete blocks can be effective against a range of nuisance effects such as patchy fertility, row-and-column effects or even the residual effects of previous treatments. Randomized complete blocks, however, contain every treatment

R. N. Edmondson (✉) Rana House 18 Church Street, Wellesbourne, Warwick CV35 9LS, UK
(E-mail: rodney.edmondson@gmail.com).

© 2020 The Author(s)

Journal of Agricultural, Biological, and Environmental Statistics, Volume 25, Number 4, Pages 500–522
<https://doi.org/10.1007/s13253-020-00416-0>

in every block and may be too large to give good control of variability for experiments with many treatments. It is common practice, therefore, to subdivide large replicate blocks into incomplete blocks for improved intra-block homogeneity, Bailey (1999, 2008), Dean et al. (2015). Sometimes agricultural field trials have crossed blocks and then there are many additional options for nesting and crossing of blocks within the same design, see, for example, Piepho et al. (2015).

Complete replicate blocks can be chosen to account for known sources of non-treatment variability such as location or management effects but the sources of variability within large complete replicate blocks are likely to be unknown and may occur over a range of scales of measurement. In that situation, a single size of nested blocks may not capture all the important sources of variability and then multi-level nesting can be used to capture patchy variability over a range of scales of measurement. Incomplete blocks confound treatment information between blocks and an efficient analysis of treatment effects will require a suitable model for the block analysis. One commonly used method is the mixed model which assumes a mixture of fixed treatment effects and random block effects, see Pinheiro and Bates (2000) and Piepho et al. (2003). Mixed models are of proven value for field trials but other methods such as spatial trend models for spatially correlated plots are also feasible; see Wilkinson et al. (1983), Gilmour et al. (1997), Cullis et al. (2006) and Rodriguez-Alvarez et al. (2016, 2018). Spatial models are commonly fitted by GAM models, see Wood (2017), or HGAM models, see Pedersen et al. (2019).

This paper examines multi-level block designs for block and spatial effects in field experiments. We first make a simulation study of 3 replicates of 48 varieties in complete randomized blocks with four levels of nesting using simulation data based on a spatial correlation model from a study of 244 UK cereal variety trials by Patterson and Hunter (1983). We then generalize the model for a 3×48 row-and-column design with 5 levels of nesting for columns. Finally, we examine real data from a large 16×34 row-and-column spring barley variety trial (Durban et al. 2003). The original analysis of that trial assumed a row-and-column design for 16 rows and 34 columns and we examine the potential improvement that could have been achieved if a design with a hierarchy of nested column blocks had been used instead. All the examples are fitted by both a mixed model analysis and an HGAM smoothing model. HGAM's can be expected to perform well for spatial models hence the HGAM models provide useful standard reference models for the simulation examples. For the spring barley variety trial, the underlying data distribution is unknown but comparison by two independent methods of analysis provides a useful check on the consistency, or otherwise, of the two analyses.

The example designs are built by a custom R (R Core Team 2020) design package, *blocksdesign* by Edmondson (2020). Some authors have developed block design algorithms based on an assumed known distribution for random block effects, Goos and Vandebroek (2001), Goos (2002), Goos and Donev (2006) and Shah and Sinha (2012) but such methods necessarily make strong assumptions about the distribution of block effects. As block variability will vary from field to field and year to year due to a range of factors such as weather, field effects and crop interaction effects, we prefer to assume a fixed effects blocks model at the design stage. The *blocksdesign* algorithm finds fixed block effects designs but we believe that fixed effects designs are robust against a range of alternative methods of

analysis so the assumption of fixed block effects at the design stage does not preclude the choice of an alternative model at the analysis stage. See Gilmour and Trinca (2006) Section 5.1.1. for further discussion of fixed versus random effects block models.

2. MULTI-LEVEL BLOCK DESIGNS

In general, analysis should be based only on the block structures that were included at the design stage of an experiment and, in this section, we discuss methods for the construction of such designs.

2.1. MODEL

Let \mathbf{t} be a vector of fixed treatment effects for a set of v treatments and let $\mathbf{b}_i, i = 1 \dots u$ be vectors of fixed block effects for u sets of blocks with $k_i, i = 1 \dots u$ block levels, respectively. Let \mathbf{T} be a $n \times v$ treatment indicator design matrix and let $\mathbf{B}_i, i = 1 \dots u$ be a set of $n \times k_i$ block indicator design matrices. Assuming \mathbf{y} is a vector of observations and \mathbf{e} is a vector of model residuals, a multi-stratum blocks model for an experiment with n observations and u sets of block factors can be written:

$$\mathbf{y} = \mathbf{T}\mathbf{t} + \mathbf{B}_1\mathbf{b}_1 + \dots + \mathbf{B}_u\mathbf{b}_u + \mathbf{e} \quad (1)$$

2.2. OPTIMIZATION

Let \mathbf{T}_f be a treatment design factor allocating treatments to plots, let $\mathbf{B}_{fi}, i = 1 \dots u$ be a set of block design factors allocating plots to blocks and let $\mathbf{D}_f = (\mathbf{B}_{f1}, \dots, \mathbf{B}_{fu})$ be a data frame for the block factors in order of fitting. Then \mathbf{T}_f can be optimised by the *blocksdesign* function:

$$\text{optimised}(\mathbf{T}_f) = \text{design}(\mathbf{T}_f, \mathbf{D}_f) \text{\$Design\$treatments} \quad (2)$$

The *blocksdesign* algorithm conditionally swaps pairs of treatments between blocks for each set of blocks $\mathbf{B}_i, i = 1 \dots u$, taken in order of fitting until no further improvement is possible. Conditional swapping means that the treatment swaps for any particular set of blocks are restricted within the levels of all previously optimized sets, which means that the blocks of each successive set must be nested within or crossed with the blocks of all previous sets. In particular, the algorithm cannot work if successive blocks are groupings of previous blocks in the sequence, which means that the algorithm cannot be used for agglomerating smaller blocks into larger blocks.

For crossed blocks factors, the relative importance of the various factorial block interaction effects should also be considered and the design algorithm can differentially weight the importance of 2-factor block interaction effects relative to block main effects. The relative importance of the 2-factor interactions in a crossed multi-factor block design decreases from full importance to zero importance as the weighting decreases from 1 to zero. Higher-order interactions, if any, are ignored. The most appropriate choice of weighting can be found

by comparing the relative efficiency of main effects and 2-factor interaction effects, see `help(design)` for details with further details in: `vignette('design_Vignette')`.

2.3. OPTIMIZATION CRITERION

The optimization criterion used by `blocksdesign` is D-optimality, which maximizes the determinant of the treatment information matrix or, equivalently, minimizes the determinant of the treatment variance matrix. D-optimality is widely used and has the important property of scale-invariance, which means that it can be used for designs with a range of quantitative and qualitative treatments, see Mitchell (1974) and Atkinson et al. (2007). For unstructured treatments, an alternative criterion is A-optimality, see John and Williams (1995) Chapter 2. Assuming equal replication, A-optimality minimizes the average variance of the pairwise treatment differences and some authors consider that A-optimality is a better criterion for unstructured treatment sets than D-optimality, see Jones et al. (2020). Currently, A-optimality is not an option in `blocksdesign`.

3. MIXED MODELS AND HGAM MODELS

The standard package for mixed model analysis in R is `lme4`, Bates et al. (2015), which includes the `lmer()` function for fitting linear mixed-effects models via REML or maximum likelihood. There are various R packages available for fitting GAM models but the “recommended” package is `mgcv`, Wood (2017). The `mgcv` package has capabilities for fitting HGAM’s but, for our purpose, we prefer the `gamm4` package by Wood and Scheipl (2020), which is specifically intended for models with a combination of random and smoothed effects. The `gamm4` package is based on the `lme4` package and should, therefore, have good comparability with the `lme4` mixed-model analysis. Pedersen et al. (2019) give a full discussion of the analysis of HGAM models using R packages, including the `gamm4` package.

3.1. MODEL FORMULATION

The general formulation for a `gamm4` model has a response variable \mathbf{y} , a set of fixed effects \mathbf{X} , a smoothing model \mathbf{S} and a random blocks model \mathbf{U} :

$$Z = \text{gamm4}(\mathbf{y} \sim \mathbf{X} + \mathbf{S}, \text{random} = \sim \mathbf{U}) \quad (3)$$

Pedersen et al. (2019) outline five model specifications for hierarchically nested models but two of the specifications use factorial effects for the levels of the nesting factor. As we require a random effects model to allow for recovery of inter-block information, the factorial smoother models will not be discussed here. Assuming **blocks** is a nesting factor and **plots** is a nested factor, the remaining three possible smoothing models are summarized in Table 1, where k is the ‘basis dimension’ of the model, see Wood (2017); Section 5.9, and m is the order of the penalized derivatives of the model, see Pedersen et al. (2019).

Model G1 corresponds to Model G in Pedersen et al. (2019), Model G2 corresponds to Model GI and Model G3 corresponds to Model I. The random blocks model term \mathbf{U} in

Table 1. Specifications for three choices of smoother for HGAM models where **blocks** is the blocks nesting factor, **plots** is the plots-within-blocks nested factor, k is the basis dimension and m is the order of the penalized derivatives

	S	Description
G1	$s(\mathbf{plots}, k, m)$	Global smoother for plots-within-blocks.
G2	$s(\mathbf{plots}, k, m) + s(\mathbf{plots}, \mathbf{by} = \mathbf{blocks}, k, m)$	Global smoother plus individual plots-within-blocks smoothers.
G3	$s(\mathbf{plots}, \mathbf{by} = \mathbf{blocks}, k, m)$	Individual plots-within-blocks smoothers

Eq. (3) can be any random blocks structure and will be explained later for each example analysis. In this paper, all model fitting is done by REML, which is also the recommended method of Pedersen et al. (2019).

3.2. MODEL COMPARISON

Comparison of available block structures is sometimes necessary for model selection and one of the most widely used tools for model comparison is the AIC information criterion, Burnham and Anderson (2002). However, AIC calculated by REML can be used only for comparing models with the same fixed effects terms $\mathbf{X} + \mathbf{S}$ in Eq. (3), see Wolfinger (1993) and Zuur et al. (2009), therefore in this paper we cannot use AIC for comparing different classes of model, such as random blocks versus HGAM models. Some further discussion of model comparison methods is given by Muller et al. (2013). Throughout this paper, we use the AIC given by the *gamm4* package, which is based on marginal likelihood, therefore the AIC statistic used here is based on the marginal number of model parameters. Other information measures include the Bayesian information criterion BIC and the conditional AIC, see Saefken et al. (2018).

3.3. MEAN PAIRWISE SED

It is useful to have an overall measure of the precision of comparison of treatment means and one overall measure of precision is the average standard error of pairwise treatment differences (SED). Let $\text{Cov}(\hat{\mathbf{t}})$ be the $v \times v$ covariance matrix of a set of v estimated treatment means $\hat{\mathbf{t}}$, let $\text{trace}(\text{Cov}(\hat{\mathbf{t}}))$ be the sum of the diagonal elements and let $\text{sum}(\text{Cov}(\hat{\mathbf{t}}))$ be the grand sum of the elements of $\text{Cov}(\hat{\mathbf{t}})$. Then the average pairwise variance of all pairs of treatments can be found by a simple counting argument and the required mean SED is the square root of:

$$2 \times (v \times \text{trace}(\text{Cov}(\hat{\mathbf{t}})) - \text{sum}(\text{Cov}(\hat{\mathbf{t}})))/(v \times (v - 1)) \quad (4)$$

More usually, a linear model is parameterized by a constant intercept term μ followed by $v - 1$ treatment differences $\boldsymbol{\tau}$ and then the covariance matrix $\text{Cov}(\hat{\mathbf{t}})$ in Eq. (4) must be replaced by the covariance matrix $\text{Cov}(\hat{\boldsymbol{\tau}})$, as shown in the Appendix.

4. SIMULATED EXAMPLE OF A NESTED BLOCKS DESIGN

The following example examines multi-level nesting for three complete replicate blocks of 48 treatments using simulated data based on a study of a large number of cereal trials by Patterson and Hunter (1983).

4.1. SPATIAL DATA MODEL

Patterson and Hunter (1983) reviewed data from 244 UK cereal variety trials and examined the relationship between block size and precision. In most trials, plots were approximately 2 m wide by 20–25 m long and usually the plots of each replicate were arranged side by side. They assumed an empirical exponential variance function for the semi-variance ϕ_x of the difference between pairs of plots x plot units apart within the same replicate block, as shown in Eq. (5):

$$\phi_x = s^2 (1 - \lambda r^x) \tag{5}$$

r is a serial correlation coefficient, s^2 is the asymptotic semi-variance of widely separated plots and λ is a parameter between 0 and 1. Based on data from the UK cereal trials, Patterson and Hunter (1983) estimated the parameters as; $r = 0.942$; $\lambda = 0.725$; $s = 0.4572$.

4.2. SIMULATED DATA

Let $\mathbf{R}_{n \times n}(r)$ be an $n \times n$ correlation matrix where the (i, j) th element equals $r^{\text{abs}(i-j)}$ for $i = 1 \dots n, j = 1 \dots n$ and let $\mathbf{I}_{n \times n}$ be an $n \times n$ identity matrix. Define a correlation matrix $\mathbf{C}_{n \times n}(\lambda, r)$ such that:

$$\mathbf{C}_{n \times n}(\lambda, r) = (1 - \lambda) \times \mathbf{I}_{n \times n} + \lambda \times \mathbf{R}_{n \times n}(r) \tag{6}$$

Then a variance matrix for three separate replicate blocks of 48 plots with plots arranged side-by-side in linear arrays and with semi-variances based on Eq. (5), is given by the Kronecker product:

$$\mathbf{V} = s^2 \times (\mathbf{I}_{3 \times 3} \otimes \mathbf{C}_{48 \times 48}(0.725, 0.942)) \tag{7}$$

Let \mathbf{ZZ}' be a Choleski factorisation of \mathbf{V} and let $\boldsymbol{\epsilon}$ be a set of 144 random variables with mean zero and unit variance. Let \mathbf{T} be the treatment design indicator matrix of an optimized treatment factor \mathbf{T}_f . Then $\mathbf{y} = \mathbf{Tt} + \mathbf{Z}\boldsymbol{\epsilon}$ is a simulated data set with $\text{var}(\mathbf{y}) = \mathbf{V}$ and $\text{Cov}(\hat{\mathbf{t}}) = (\mathbf{T}'\mathbf{V}^{-1}\mathbf{T})^{-1}$, as required.

4.3. EXAMPLE NESTED BLOCKS DESIGN

Let \mathbf{N}_0 be a replicate blocks factor for 3 replicate blocks of 48 plots and let $\mathbf{N}_1, \mathbf{N}_2, \mathbf{N}_3$ and \mathbf{N}_4 be four nested blocks factors with 9, 18, 36 and 72 levels and blocks of size 16, 8, 4 and 2 respectively. Let $\mathbf{D}_f = (\mathbf{N}_0, \mathbf{N}_1, \mathbf{N}_2, \mathbf{N}_3, \mathbf{N}_4)$ be a data frame of block factors

Table 2. Mixed model analysis of 5000 simulations for each of the 16 factorial combinations of blocks N_1 , N_2 , N_3 , N_4 with replicate blocks N_0 and showing the number of model terms, the counts of minimum AIC simulations per model and the mean AIC and the mean SED for each model

Model	Blocks U	Model terms	Classified counts	Mean AIC	Mean SED
B1.1	N_0	50	27	239.0	0.306
B1.2	$N_0 + N_1$	51	350	220.6	0.263
B1.3	$N_0 + N_2$	51	1009	215.9	0.247
B1.4	$N_0 + N_3$	51	276	220.2	0.247
B1.5	$N_0 + N_4$	51	33	230.3	0.267
B1.6	$N_0 + N_1 + N_2$	52	552	215.2	0.246
B1.7	$N_0 + N_1 + N_3$	52	1093	214.8	0.242
B1.8	$N_0 + N_1 + N_4$	52	379	218.0	0.248
B1.9	$N_0 + N_2 + N_3$	52	447	215.2	0.240
B1.10	$N_0 + N_2 + N_4$	52	427	215.8	0.241
B1.11	$N_0 + N_3 + N_4$	52	34	221.2	0.244
B1.12	$N_0 + N_1 + N_2 + N_3$	53	98	214.5	0.238
B1.13	$N_0 + N_1 + N_2 + N_4$	53	174	215.1	0.239
B1.14	$N_0 + N_1 + N_3 + N_4$	53	78	215.8	0.239
B1.15	$N_0 + N_2 + N_3 + N_4$	53	19	216.3	0.237
B1.16	$N_0 + N_1 + N_2 + N_3 + N_4$	54	4	215.6	0.235

in the required order of fitting, and let T_f be a treatment factor for three replicates of 48 treatments. Then T_f can be optimised by the function in Eq. (2) with parameters T_f and D_f .

4.4. RANDOM BLOCKS ANALYSIS

A single optimized treatment design T_f was realized and 5000 simulated data vectors were generated for the assumed plot covariance matrix shown in Eq. (7). The choice of treatment effects is arbitrary and, in this study, all treatment effects are assumed null. Each simulation was analysed using the *gamm4* function shown in Eq. (3) with fixed effects T_f and with random effects U corresponding to each of the 16 possible block combination of the four block factors N_1, N_2, N_3, N_4 together with the replicate blocks factor N_0 . Table 2 shows the mean AIC and the mean SED averaged over all simulations for each model and also shows the number of times each model attained the minimum overall AIC for each simulation. The theoretical mean pairwise treatment SED for the optimized design was found from Eq. (4) using $\text{Cov}(\hat{\mathbf{t}})$ based on the treatment design indicator matrix \mathbf{T} and the variance matrix \mathbf{V} as discussed above. For this particular design realization, the theoretical mean pairwise treatment SED was found to be 0.2324.

4.5. INTERPRETATION

Table 2 shows sub-models based on all possible factorial combinations of the nested block factors and counts of the number of times each model gave the best overall fit, based on AIC, for each simulation. The sub-model counts show the relative importance of the various block sizes and depths of nesting for fitting the individual simulations. Model selection is

an important methodology for fitting parsimonious and repeatable models to treatment data, Burnham and Anderson (2002), and based on the classification by minimum AIC, the full additive blocks model B1.16 would hardly ever be selected as the best model for data based on the spatial model in Eq. (5). However, as a mixed model analysis automatically finds a best fitting set of model coefficients, possibly zero, and as the full blocks model B1.16 has a mean AIC close to the minimum of all the models and a pairwise treatment SED of 0.2355, which is quite close to its theoretical value of 0.2324, it is not obvious whether selection of a more parsimonious blocks model would be useful or worthwhile.

4.6. HGAM ANALYSIS

For comparison, the same set of simulations was fitted by the *gamm4* function shown in Eq. (3) with fixed effects $\mathbf{T}_f + \mathbf{S}$ where \mathbf{S} is fitted by G2 or G3 from Table 1 and the random effects model \mathbf{U} is fitted by the replicate blocks factor \mathbf{N}_0 . Smoother G1, which assumes a single common smoother for all levels of the grouping factor, is unrealistic for a design with separate complete randomized blocks and has not been fitted. Pedersen et al. (2019) found that with a random grouping factor, it was important to specify $m = 1$ instead of $m = 2$ for group-level smoothers to avoid instabilities therefore a choice of $m = 1$ is fixed for the G2 individual block level smoothers. For the remaining smoothers, the choice of m is data dependent and Table 3 shows choices of $m = 1$ or $m = 2$ for these smoothers. A range of values of $k \geq 10$ was tested (not shown) but increasing k beyond 10 made little difference therefore $k = 10$ has been used throughout. Table 3 shows the mean estimated degrees of freedom (EDF) per model averaged over all simulations, the number of marginal fitted terms per model, the mean marginal AIC averaged over all simulations and the mean pairwise marginal SED averaged over all simulations. Pairwise SED's were calculated by extracting the treatments covariance matrix from the *gamm4* model for each simulation and using the method shown in Eq. (4).

4.7. INTERPRETATION

Following Pedersen et al. (2019), the models in Table 3 can be compared by AIC and the best model based on AIC was either H1.1 or H1.3. The mean SED for both models was about 0.231, which is slightly less than the theoretical value of 0.2324. There seems

Table 3. Hierarchical GAM models with complete replicate blocks \mathbf{N}_0 and two choices of m for the order of the penalized derivatives of the global smoother in G2 or the separate smoothers in G3

Blocks U	Smoothing model S	Model terms	Mean EDF	Mean AIC	Mean SED
H1.1 \mathbf{N}_0	$s(\mathbf{plots}, m = 1) + s(\mathbf{plots}, \text{by} = \mathbf{N}_0, m = 1)$	54	61.0	209.6	0.231
H1.2 \mathbf{N}_0	$s(\mathbf{plots}, m = 2) + s(\mathbf{plots}, \text{by} = \mathbf{N}_0, m = 1)$	55	61.1	213.4	0.231
H1.3 \mathbf{N}_0	$s(\mathbf{plots}, \text{by} = \mathbf{N}_0, m = 1)$	53	61.3	208.2	0.231
H1.4 \mathbf{N}_0	$s(\mathbf{plots}, \text{by} = \mathbf{N}_0, m = 2)$	56	57.3	219.0	0.236

The tabulation shows the number of fitted model terms, the mean EDF, the mean AIC and the mean SED averaged over all simulations. A basis dimension of $k = 10$ was used for all smoothers

little reason to fit a common global smoother over the three separate independent complete blocks therefore on theoretical grounds H1.3 should be a more appropriate choice of model than H1.1.

4.8. COMPARISON OF MIXED MODEL VERSUS HGAM ANALYSIS

The mean SED based on the full nested blocks model, B1.16, was 0.235 compared with a mean SED of 0.231 from H1.3, the best fitting HGAM model. The theoretical mean SED value was 0.2324 so both models appear to fit the data reasonably well. Comparison of model complexity can be based on either the number of fitted model terms or on the effective degrees of freedom (EDF). The complexity of B1.16 and H1.3 based on the number of model terms, 54 and 53 respectively, appears similar. The *lme4* package does not provide an EDF for mixed models but Wood (2017): Section 2.4.6. suggests that the EDF of a mixed model should be less than the total number of model terms. In that case, and based on a comparison of EDF's, H1.3 appears substantially more complex than B1.16.

The *blocksdesign* algorithm does not generally find a unique global optimum therefore the results may differ slightly for different optimizations. The simulation model was tested a number of times using different optimizations (not shown) but no evidence was found that the particular treatment optimization had any significant impact on the conclusions of the study.

5. SIMULATED EXAMPLE OF NESTING WITHIN A SET OF CROSSED BLOCKS

Crossed row-and-column block designs are useful for 2-dimensional spatial arrays of plots, especially if there is no obvious preferred direction of blocking. Patterson and Hunter (1983) did not study field variability in two dimensions but spatial models for field and ecological experiments have been extensively studied, Cullis and Gleeson (1991), Dutilleul (1993), Eccleston and Chan (1998) and Hu and Spilke (2009). One of the simplest 2-dimensional spatial models, and the model used here, is a separable first-order autoregressive model $AR1 \otimes AR1$ where each plot effect is the product of two separable autoregressive processes, one for rows and one for columns. There are many other possible models, for example Rodriguez-Alvarez et al. (2016) discuss random spatial components for 2-dimensional field trials based on penalized splines but these are beyond the scope of this paper.

5.1. SIMULATED DATA

For designs with long narrow plots arranged side-by-side in rows and lengthwise in columns, spatial correlation is likely to be less within columns than within rows and it will be assumed here that the semi-variance of the difference between pairs of plots separated by x plot units in the same column is $s^2(1 - \lambda r^x)$ with $r = 0.5$ and $\lambda = 1$ which gives

$C_{3 \times 3}(1, 0.5)$ for the within-column correlation matrix. Equation (7) then generalizes to:

$$V = s^2 \times (C_{3 \times 3}(1, 0.5) \otimes C_{48 \times 48}(0.725, 0.942)) \tag{8}$$

Let ZZ' be a Choleski factorisation of V and let ϵ be a set of 144 random variables with mean zero and unit variance. Let T be the treatment design indicator matrix of an optimized treatment factor T_f . Then $y = Tt + Z\epsilon$ is a simulated data set with $var(y) = V$ and $Cov(\hat{t}) = (T'V^{-1}T)^{-1}$, as required.

5.2. EXAMPLE CROSSED BLOCKS DESIGN

Let R_0 be a replicate blocks factor for 3 replicate rows of 48 plots and let Q_1, Q_2, Q_3, Q_4 and Q_5 be five hierarchically nested column block factors with 3, 6, 12, 24 and 48 levels, respectively, where the column blocks are assumed to be fully crossed with R_0 . Let T_f be a treatment factor for 3 replicates of 48 treatments and let $D_f = (R_0, Q_1, Q_2, Q_3, Q_4, Q_5)$. Then T_f can be optimised by the function in Eq. (2) with parameters T_f and D_f , as discussed in the previous example. The weighting option for crossed blocks interactions has a default of 0.5 but can be reset to any other value between 0 and 1 to give less or more importance to the factorial block interactions, as required.

5.3. RANDOM BLOCKS ANALYSIS

The possible combinations of the five nested columns Q_1, Q_2, Q_3, Q_4, Q_5 , including possible interactions with R_0 , is too large for tabulation. Table 4 shows just three fitted models, B2.1, which is the basic additive row-and-column model with three rows and 48 columns, B2.2, which is an additive row-and column model with a full set of nested column blocks and B2.3, which is an interactive row-and-column model with a full set of nested column blocks including all possible interactions between rows and nested column blocks. Note that the intersection blocks between R_0 and Q_5 are single plot blocks therefore the $R_0 \times Q_5$ interactions cannot be estimated and must be assumed null.

As in the previous example, there were 5000 simulations all based on the same treatment design and Table 4 shows the minimum AIC counts per model, the mean AIC per model

Table 4. Mixed model analysis of 5000 simulations of three rows R_0 and five sets of nested columns Q_1, Q_2, Q_3, Q_4, Q_5 comprising B2.1, a basic additive row-and-column model, B2.2, an additive row-and-column model with a full set of nested columns and B2.3, a fully interactive row-and-column model with a full set of nested columns

Model	Blocks U	Model terms	Classified counts	Mean AIC	Mean SED
B2.1	$R_0 + Q_5$	51	731	225.1	0.256
B2.2	$R_0 + Q_1 + Q_2 + Q_3 + Q_4 + Q_5$	55	911	218.3	0.242
B2.3	$R_0 \times (Q_1 + Q_2 + Q_3 + Q_4) + Q_5$	59	3358	210.7	0.206

The tabulation shows the number of model terms, the counts of minimum AIC simulations per model, the mean marginal AIC and the mean SED for each model

and the mean SED per model for the full set of simulations. The theoretical mean pairwise treatment SED for the optimized design was found from Eq. (4) using $\text{Cov}(\hat{\mathbf{t}})$ based on the treatment design indicator matrix \mathbf{T} and the variance matrix \mathbf{V} , as discussed in the previous example. For this example, the theoretical mean pairwise treatment SED was found to be 0.2043.

5.4. INTERPRETATION

The comparisons show that the mean AIC decreased substantially with increasing model complexity and that there was a corresponding improvement in the mean SED over the three models. B2.3 was the best overall model for about 66% of the simulations with B2.2 best for about 18% and B2.1 best for the rest. The mean SED of the best fitting model, B2.3, was 0.206 which is within about 1.3% of the theoretical value of 0.2043. These results show that for most of the simulations, the full interactive model, B2.3, was required for a good model fit. As with the previous example, it is not obvious if selection of a parsimonious block model using a criterion such as the AIC would be useful or worthwhile but further study of the utility, or otherwise, of model selection could be of value.

5.5. HGAM ANALYSIS

The same set of simulations was also fitted by a HGAM model with smoothers fitted within complete replicate rows \mathbf{R}_0 using the *gamm4* function shown in Eq. (3). For this study, two alternative random blocks models were fitted, either simple replicate row blocks \mathbf{R}_0 , as in the previous example, or a basic additive row-and-and-column blocks model, $\mathbf{R}_0 + \mathbf{Q}_5$. Table 5 shows eight combinations of two random blocks models, \mathbf{R}_0 or $\mathbf{R}_0 + \mathbf{Q}_5$, two smoothing models, G2 or G3, and two choices of $m = 1$ or $m = 2$ for the global smoother in G2 or the separate smoothers in G3. A preliminary analysis showed that G1 gave a very poor fit to the data therefore G1 is not considered further here. For each model

Table 5. HGAM models G2 or G3 with two choices of random blocks, \mathbf{R}_0 or $\mathbf{R}_0 + \mathbf{Q}_5$, and two choices of m for the order of the penalized derivatives of the global smoother in G2 or the separate smoothers in G3 and with basis dimension of $k = 10$ for each smoother

Blocks U	Smoothing model S	Model terms	Mean EDF	Mean AIC	Mean SED
H2.1 \mathbf{R}_0	$s(\mathbf{plots}, m = 1) + s(\mathbf{plots}, \text{by} = \mathbf{R}_0, m = 1)$	54	59.8	204.6	0.227
H2.2 \mathbf{R}_0	$s(\mathbf{plots}, m = 2) + s(\mathbf{plots}, \text{by} = \mathbf{R}_0, m = 1)$	55	59.5	209.2	0.228
H2.3 \mathbf{R}_0	$s(\mathbf{plots}, \text{by} = \mathbf{R}_0, m = 1)$	53	61.4	206.6	0.232
H2.4 \mathbf{R}_0	$s(\mathbf{plots}, \text{by} = \mathbf{R}_0, m = 2)$	56	57.4	217.3	0.237
H2.5 $\mathbf{R}_0 + \mathbf{Q}_5$	$s(\mathbf{plots}, m = 1) + s(\mathbf{plots}, \text{by} = \mathbf{R}_0, m = 1)$	55	60.8	198.0	0.204
H2.6 $\mathbf{R}_0 + \mathbf{Q}_5$	$s(\mathbf{plots}, m = 2) + s(\mathbf{plots}, \text{by} = \mathbf{R}_0, m = 1)$	56	60.3	201.8	0.204
H2.7 $\mathbf{R}_0 + \mathbf{Q}_5$	$s(\mathbf{plots}, \text{by} = \mathbf{R}_0, m = 1)$	54	61.4	198.6	0.206
H2.8 $\mathbf{R}_0 + \mathbf{Q}_5$	$s(\mathbf{plots}, \text{by} = \mathbf{R}_0, m = 2)$	57	57.2	209.0	0.210

The analysis shows the number of model terms, the mean EDF, the mean AIC and the mean SED averaged over simulations

combination, the table shows the number of marginal model terms, the mean EDF, the mean AIC and the mean SED, calculated as described in the previous example.

5.6. INTERPRETATION

Comparison of the mean AIC and the mean SED of the models with a single random block factor \mathbf{R}_0 versus the models with two additive random blocks factors $\mathbf{R}_0 + \mathbf{Q}_5$ shows clear evidence of a substantial reduction in both the mean AIC and the mean SED due to the addition of \mathbf{Q}_5 . The best overall model was either H2.5 or H2.7, consistent with the previous example, and the mean SED of 0.204 for H2.5 was slightly smaller than the theoretical value of 0.2046. The mean SED for models without random block factor \mathbf{Q}_5 was at least 14% larger than the theoretical value which shows that \mathbf{Q}_5 is essential for a good model fit. Smoothing alone is not sufficient to account for trends between columns and the best model requires a combination of smoothing to account for smooth trends within rows plus a random blocks model to account for residual column block effects. The residual random variability between columns makes the recovery of inter-column treatment information by a mixed model analysis essential.

5.7. COMPARISON OF NESTED BLOCKS VERSUS HGAM ANALYSIS

From Table 4, the mean SED of the full row-and-column blocks model including interacting row and column effects, B2.3, was 0.206 whereas from Table 5, the best fitting HGAM model, H2.5, had an SED of 0.204. The theoretical mean pairwise treatment SED for the design was 0.2043 therefore there is very little to choose between the fit of the two models. The full row-and-column blocks model required 59 model terms whereas the best fitting HGAM model required 55 marginal model terms or a mean EDF of 60.8. As discussed previously, more work is needed on the comparison of model complexity of mixed models versus HGAM's based on the numbers of fitted model parameters.

The simulation model was tested several times using different treatment optimizations (not shown) and although the outcome of the analysis was slightly different each time, the overall interpretation remained the same and there was no evidence that the particular treatment optimization had any significant impact on the conclusions of the study.

6. A LARGE ROW-AND-COLUMN FIELD TRIAL EXAMPLE

Durban et al. (2003) discussed an experiment with two replicates of 272 spring barley varieties arranged in an array of 34 columns (east–west) and 16 rows (north–south), subject to the constraint that rows 1–8 contained one complete set of treatment replicates and rows 9–16 contained the other. They showed that a conventional row-and-column model was inadequate due to residual trends within rows and they sought to improve model efficiency by fitting a 2-dimensional GAM loess smoothing model, Hastie (2020) and Hastie and Tibshirani (1986, 1987). In this paper, the data is re-analysed both by a mixed model random blocks analysis assuming four additional sets of nested column blocks superimposed on the original design and by an HGAM smoothing model with smoothers nested within rows. The treatment

design will not be optimal for the additional nested column blocks but, nevertheless, should provide insight into the utility of multi-level block designs for field trials.

The data was obtained from the *agridat* package, Wright (2020), and is available in the *blocksdesign* package, see supplementary material.

6.1. SUPERIMPOSED COLUMN BLOCKS

Let \mathbf{R}_0 represent the original replication factor with 2 levels, let \mathbf{R}_1 represent the original rows factor with 16 levels and let \mathbf{Q}_5 represent the original columns factor with 34 levels fully crossed with \mathbf{R}_0 and \mathbf{R}_1 . Let \mathbf{Q}_1 , \mathbf{Q}_2 , \mathbf{Q}_3 , \mathbf{Q}_4 represent four additional column block factors with 2, 4, 8 and 16 levels, respectively, all fully crossed with rows. Unlike the simulated row-and-column example discussed previously, not all of the nested column blocks can be of equal width and two column blocks from each of \mathbf{Q}_2 , \mathbf{Q}_3 and \mathbf{Q}_4 must be an extra plot wide. It will be assumed here that the extra wide column blocks are always located down the two outside edges of the design. Some further discussion of a design layout that avoids this issue is given later.

6.2. RANDOM BLOCKS ANALYSIS

As was done in the previous example, we analyse three random blocks models for the observed data, B3.1, the original row-and-column blocks model with additive row and column effects, B3.2, a full additive model for a full set of rows and columns including the extra superimposed columns and B3.3, a full interactive model for the full set of rows and columns including the extra superimposed columns. Table 6 shows the number of model terms, the model AIC statistic and the model pairwise SED for these three models.

6.3. INTERPRETATION

The AIC statistic shows significant improvement in model fit for B3.2 relative to B3.1 and for B3.3 relative to B3.2, which is good evidence that inclusion of the extra superimposed columns improves model fit. The mean pairwise SED of B3.1, the original row-and-column design, was .291 which is approximately 12% larger than the mean pairwise SED of B3.3, the interactive full row and column model including the extra column blocks, which was 0.260. The mean pairwise SED for model B3.2, which included the extra column blocks but

Table 6. Mixed model analysis of spring barley variety trial yield data showing the AIC and SED statistics for B3.1, the original additive row and column blocks model, B3.2, a fully additive row and column blocks model including the extra superimposed column blocks and B3.3, a fully interactive row and column blocks model including the extra superimposed column blocks

Model	Random Blocks	Model terms	Model AIC	Mean SED
B3.1	$\mathbf{R}_0 + \mathbf{R}_1 + \mathbf{Q}_5$	276	887.2	0.291
B3.2	$\mathbf{R}_0 + \mathbf{R}_1 + \mathbf{Q}_1 + \mathbf{Q}_2 + \mathbf{Q}_3 + \mathbf{Q}_4 + \mathbf{Q}_5$	280	874.7	0.285
B3.3	$(\mathbf{R}_0 + \mathbf{R}_1) \times (\mathbf{Q}_1 + \mathbf{Q}_2 + \mathbf{Q}_3 + \mathbf{Q}_4) + \mathbf{Q}_5$	288	864.1	0.260

Table 7. HGAM models G2 or G3 with two choices of random blocks \mathbf{U} and two choices of m for the order of the penalized derivatives of the global smoother in G2 or the separate smoothers in G3 and with a basis dimension of $k = 10$ for each smoother

	Blocks \mathbf{U}	Smoother \mathbf{S}	Model terms	EDF	AIC	SED
H3.1	$\mathbf{R}_0 + \mathbf{R}_1$	$s(\text{plots}, m = 1) + s(\text{plots}, \text{by} = \mathbf{R}_1, m = 1)$	292	305.9	846.4	0.249
H3.2	$\mathbf{R}_0 + \mathbf{R}_1$	$s(\text{plots}, m = 2) + s(\text{plots}, \text{by} = \mathbf{R}_1, m = 1)$	293	305.7	854.3	0.250
H3.3	$\mathbf{R}_0 + \mathbf{R}_1$	$s(\text{plots}, \text{by} = \mathbf{R}_1, m = 1)$	291	329.9	901.1	0.272
H3.4	$\mathbf{R}_0 + \mathbf{R}_1$	$s(\text{plots}, \text{by} = \mathbf{R}_1, m = 2)$	307	319.4	948.0	0.271
H3.5	$\mathbf{R}_0 + \mathbf{R}_1 + \mathbf{Q}_5$	$s(\text{plots}, m = 1) + s(\text{plots}, \text{by} = \mathbf{R}_1, m = 1)$	293	305.9	847.4	0.248
H3.6	$\mathbf{R}_0 + \mathbf{R}_1 + \mathbf{Q}_5$	$s(\text{plots}, m = 2) + s(\text{plots}, \text{by} = \mathbf{R}_1, m = 1)$	294	307.8	854.4	0.248
H3.7	$\mathbf{R}_0 + \mathbf{R}_1 + \mathbf{Q}_5$	$s(\text{plots}, \text{by} = \mathbf{R}_1, m = 1)$	292	308.2	876.8	0.257
H3.8	$\mathbf{R}_0 + \mathbf{R}_1 + \mathbf{Q}_5$	$s(\text{plots}, \text{by} = \mathbf{R}_1, m = 2)$	308	310.2	928.4	0.258

The analysis shows the number of fitted model terms, the EDF, the AIC and the SED for the observed data

assumed additivity of row and column effects, was 0.285 which is only a small improvement over the original row and column model. These results show that a highly structured row-and-column design with nested column blocks and proper allowance for row-by-column interaction effects gives a better design for modelling spatial positional effects.

6.4. HGAM MODELS

As discussed in the previous examples, a HGAM with suitable random blocks and suitable smoothers can be used to fit local trends within block levels and Table 7 shows eight HGAM smoothing models for trends fitted within the rows of the barley variety trial data. These are essentially the same eight smoothing models that were discussed in the previous row-and-column example (see Table 5), except that the replicate blocks factor \mathbf{R}_0 has been included as an extra factor in the random row blocks model.

6.5. INTERPRETATION

Unlike the previous crossed blocks example, the effects of the inclusion or exclusion of the crossed blocks factor \mathbf{Q}_5 in the blocks model depends on whether the smoothing model includes a global smoother or not. \mathbf{Q}_5 had little effect on H3.5 and H3.6 versus H3.1 and H3.2, all of which had global smoothers, whereas it gave a clear improvement to H3.7 and H3.8 versus H3.3 and H3.4, none of which had global smoothers. Overall, the best fitting models from Table 7 were H3.1 and H3.5 and, in the remainder of this analysis, H3.1 will be used as the best fitting HGAM model.

6.6. COMPARISON OF NESTED BLOCKS MODELS WITH HGAM MODELS

The mean pairwise SED for H3.1 was about 0.249 compared with about 0.260 for B3.3, which suggests that the best fitting HGAM model gave an improvement in precision of about 4% compared with the best fitting multi-level blocks model. However, H3.1 required 292 fitted (marginal) terms and about 305.9 EDF compared with the 288 model terms for

B3.3, which suggests that the HGAM model is more complex and requires more model parameters than the mixed model.

It is not possible to compare B3.3 and H3.1 directly using AIC because they have different numbers of fixed model effects. In principle, maximum likelihood ratio tests can be used to provide a general AIC for comparing model fit (see Pinheiro and Bates 2000) but as we use REML to estimate model coefficients and smoothing parameters and residual variances, it does not seem consistent to use ML to compare model fit. Instead, we prefer to use graphical and other informal methods to examine model fit and to compare the outputs of the different models.

6.7. RESIDUAL COPLOTS

Coplots are graphical plots that fit locally-weighted polynomial regression lines to data within the levels of a conditioning variable, Cleveland (1979 and 1993). Figures 1 and 2 show coplots of genotype adjusted residuals for B3.3, the best fitting row-and-column blocks model, and H3.1, the best fitting HGAM model, all plotted against plot position for each row. The coplots show residual trends within rows after adjusting for row and genotype effects where the 16 panels taken in order from bottom left to top right represent the individual rows running from rows 1–16.

Both models seem to account for most of the residual trends in the different rows. However, comparison of Fig. 1 with Fig. 2 does show there are some small differences in fit between the two models with some rows being fitted better by one model and some by the other. Overall, it does not seem practicable to choose between the models on the basis of their residual plots.

6.8. WIREFRAME PLOTS OF FIELD VARIABILITY

A wireframe plot is a graph that can be used to visualize a 3-dimensional surface such as the trend surface of plot yields in a field trial. Figures 3 and 4 show wireframe plots based on the lattice package, Sarkar (2008), for the predicted plot yields of a nominal genotype as a function of plot position for the best mixed model B3.3 and the best HGAM model H3.1.

Comparison of the two plots shows that the shape of the fitted surface is very similar for both plots except that Fig. 3 shows discrete block effects whereas Fig. 4 shows smoothed trends within individual rows. As the true surface is unknown, it is impossible to judge which is the better model but as both models give a similarly shaped trend surface, both models seem to fit the data equally well. In this analysis, the smoothing model works well because there are smooth trends within rows but in other situations, positional effects within rows might be patchier, in which case the nested blocks model might be more useful. It is important to note that there is very little good evidence of smooth trends between rows, which helps explain why a 2-dimensional smoothed surface, as discussed by Durban et al. (2003), is unsuitable for this data and why a random model for row effects is more appropriate than a smoothing model.

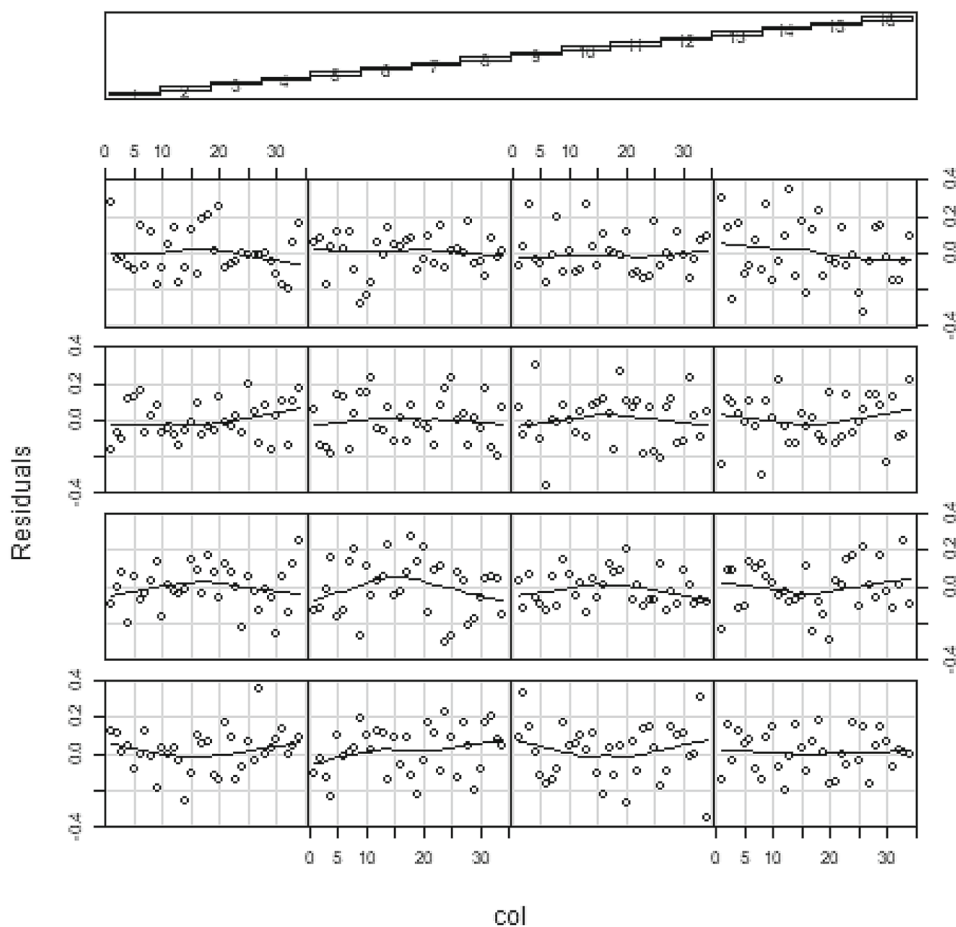


Figure 1. Residuals for mixed model B3.3.

6.9. COMPARISONS OF MODEL PREDICTIONS

Model predictions are important for comparing alternative models since, if different models give different predictions, not all the models can be valid and a further investigation might be needed. Table 8 compares the correlations between the raw treatment means and the predictions of B3.3, the best mixed model and H3.1, the best HGAM model.

The correlations between the raw means and the predicted yields of B3.3 and H3.1 are 0.8456 and 0.8462, respectively, which shows that both models give predictions substantially different from the raw means. The correlation of 0.9721 between B3.3 and H3.1 shows that the predictions of these two models are quite strongly correlated. Figure 5 shows a scatter plot of the fitted values for H3.1 versus B3.3, which shows a strong linear relationship consistent with a linear correlation model. The two models produce highly correlated predictions but there is still sufficient scatter to change the order of ranking of the highest yielding genotypes, which could be critically important for genotype selection.

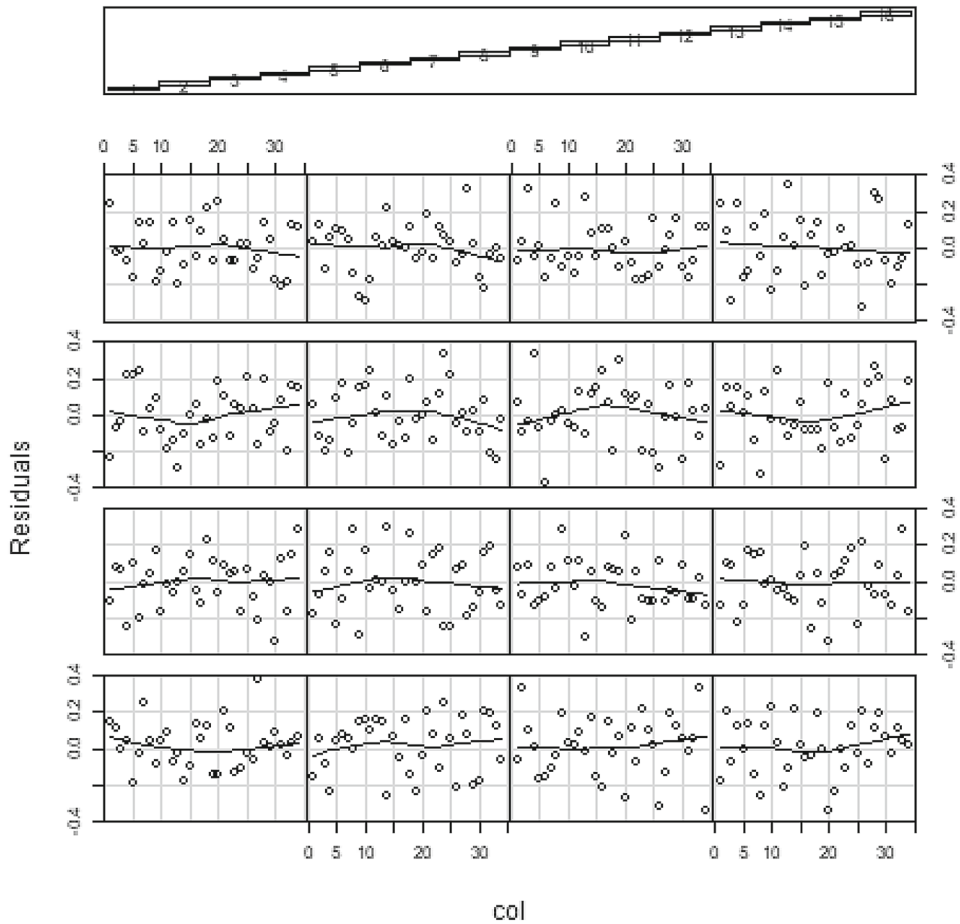


Figure 2. Residuals for HGAM model H3.1.

For example, G112 was in 4th place in the raw means, 53rd place in the B3.3 means and 81st place in the H3.1 means. Examination of the plot positions and yields of the two replicates of genotype G112 show that replicate 1, with a yield of 6.67 tonnes per hectare, was in row 1, column 34 whereas replicate 2, with a yield of 5.91 tons per hectare, was in row 14, column 19. Thus, replicate 1 of G112 was located on a very high yielding corner plot, which means that the correct choice of model could be of critical importance for this genotype.

6.10. COMMENT ON DESIGN LAYOUT

A difficulty with the original spring barley trial design was that the design had 16 rows and 34 columns, which meant that not all of the nested column blocks could be of equal width. An alternative layout with 17 rows and 32 columns allows all nested column blocks to be of equal width but requires eight and one-half rows per complete replicate. An example in the supplementary material shows how this can be done by using *blocksdesign* to create one

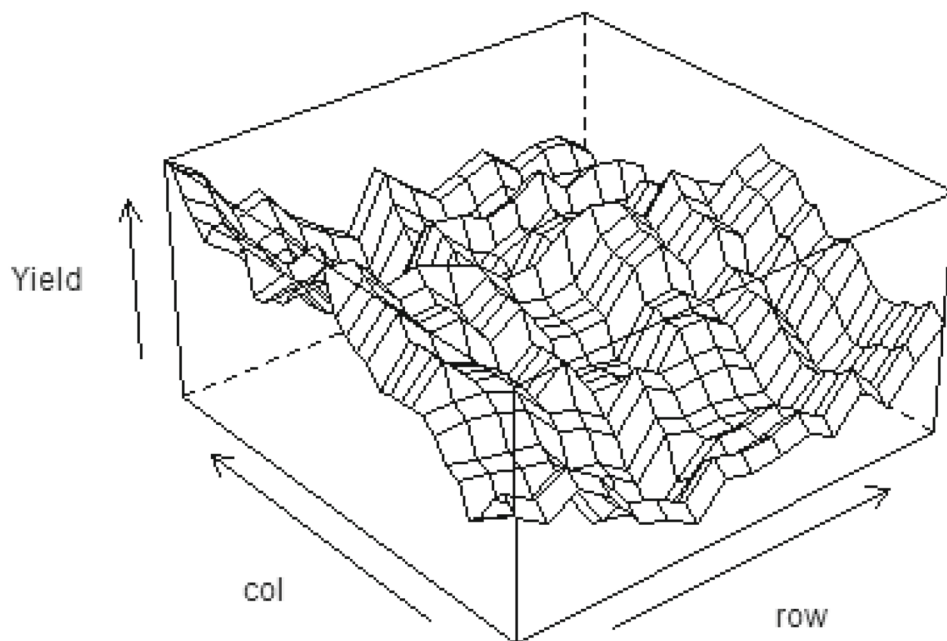


Figure 3. Predicted yield surface for mixed model B3.3.

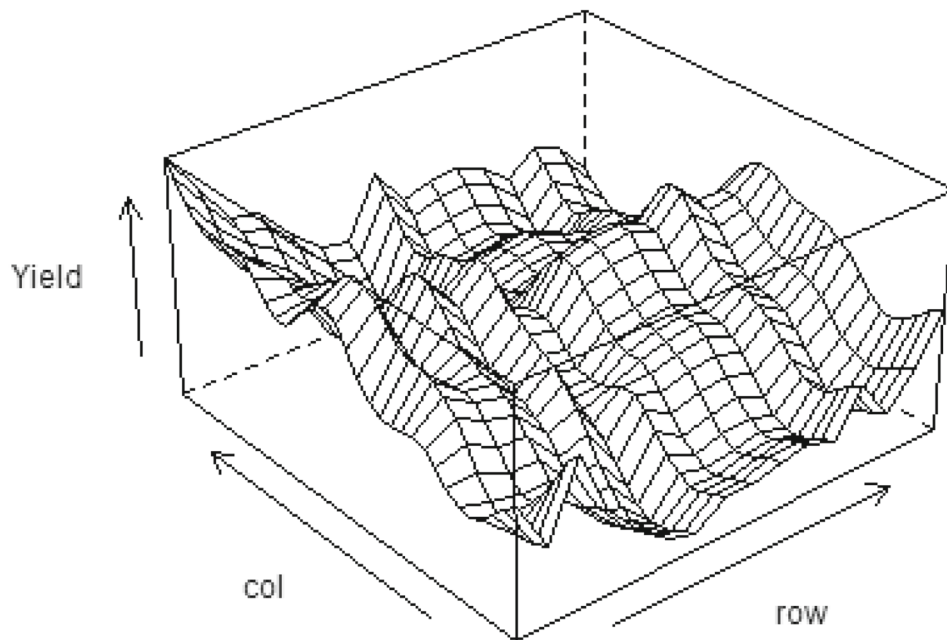


Figure 4. Predicted yield surface for HGAM model H3.1.

complete replicate from rows 1–8 plus 16 plots from row 9 with the other complete replicate from rows 10–17 plus the remaining plots of row 9. Usually, there is some flexibility in the

Table 8. Correlations between the raw treatment means, the predicted treatment means of the mixed model B3.3 and the predicted treatment means of the HGAM model H3.1

	Raw	B3.3	H3.1
Raw	1	0.8456	0.8462
B3.3	0.8456	1	0.9721
H3.1	0.8462	0.9721	1

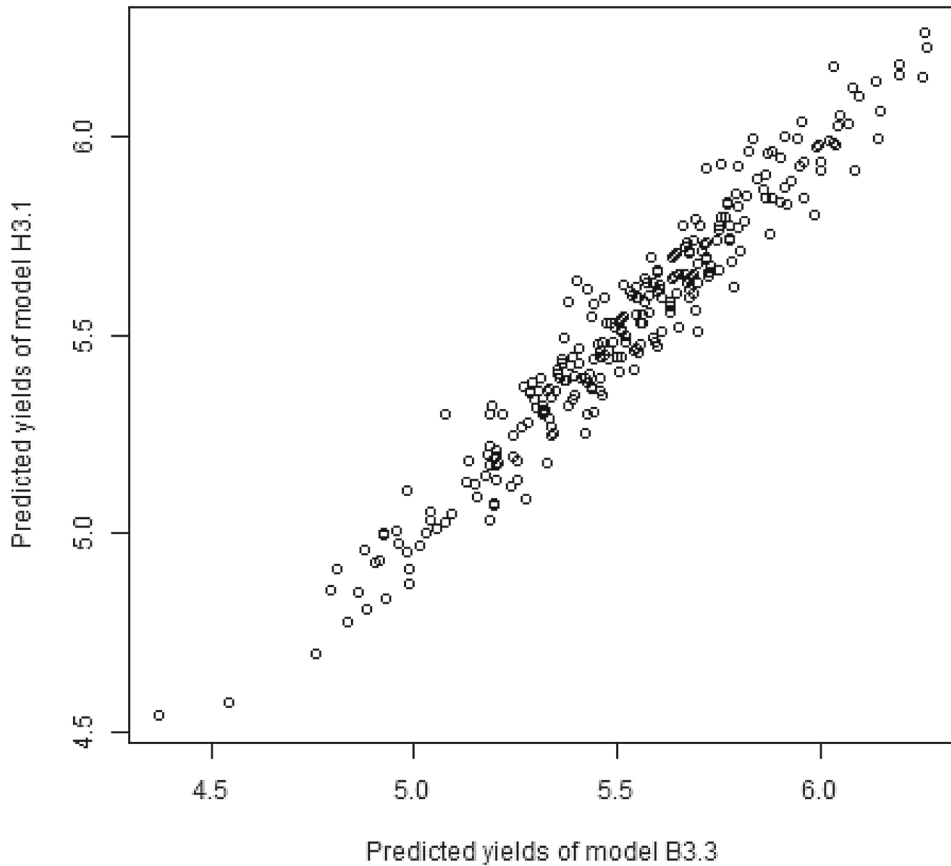


Figure 5. Predicted yield from H3.1 versus B3.3.

actual plot layout and, wherever possible, it is best to ensure equal block sizes in the same level of nesting.

7. DISCUSSION

The advantages of simplicity, flexibility and robustness make block designs the designs of choice for many agricultural experiments. Cultural operations such as planting or harvesting might occur over several days and it is often useful for the main replicate blocks to form

management units so that all cultural operations on the same replicate block are carried out at the same time Bailey (2008): Chapter 4. Incomplete block designs with a single level of nesting are widely used for the control of variability within replicate blocks in large field trials. Unlike replicate blocks, however, there is usually no prior information about the likely effects of incomplete blocks and therefore no information to guide the choice of size of blocks. Furthermore, there is no good reason to assume that any single size of block will be sufficient to account for variability over the wide range of scales of measurement in large replicate blocks.

Modern software has the capability to deal with very complex blocking systems involving crossed and nested block designs and there is no reason why modern experimental design should be restricted to designs with a single level of nesting. Modern software can fit very general block models either by fitting all the available block structures simultaneously or by using appropriate software to search for the best sub-model that fits the data. This methodology provides a very flexible and general approach to block modelling but the models must be based on structures that were envisaged and built into the design at the design stage. It would be highly risky to make *post hoc* searches for block models based on block structures that were not envisaged as part of the original design.

Our analysis has compared mixed models and HGAM models and has shown that a mixed model can be very effective even when the data is generated by a spatial correlation process. However, our row-and-column examples have also shown that model fit can sometimes be improved by including both random block effects and HGAM smoothers in the same analysis. In our examples, the HGAM row smoothers accounted for smooth plots-within-rows trend components while the random column block terms accounted for residual random column block components. Including both random blocks and spatial smoothers in the same model can be expected to fit smooth trends both within and between blocks while simultaneously accounting for block deviations from smoothness. This approach can be applied to any field block design and further reinforces the argument for a highly structured block design at the design stage to provide maximum freedom for model fitting at the analysis stage.

Nesting is a recursive process that starts with a minimum number of block constraints in the top level of blocking and then adds additional constraints at each level of nesting. Usually this process gives good efficiency at each level of nesting except that as the block sizes become very small the blocks become more constrained and then there may be some small loss of efficiency due to the constraints of blocking. A referee has suggested that, in some situations, a nested design could be built by starting with the smallest blocks and then grouping them into larger blocks, as required. However, constraints would still need to be added to ensure good properties for the larger blocks so it is not clear if this is a useful or practicable approach.

One issue that has not been fully addressed in this paper is whether an analysis should include every block design structure or whether model selection should be used to find a subset of block structures that are adequate for the data. There is usually no interest in the block structures themselves so the actual block model is irrelevant provided that it ensures maximum precision on the treatment estimates. Throughout this paper we have fitted models with a full set of block factors and have relied on *lmer()* to provide proper weighting for

the various random block components. As far as we can judge, this process works well, although *lmer()* does provides warnings when models fail to converge or are 'over-fitted'. We recognize that more work could be done to clarify model selection methods for multi-level block designs.

ACKNOWLEDGEMENTS

I would like to thank Prof. Dr. Hans-Peter Piepho of the University of Hohenheim and David Rügamer of the Ludwig-Maximilians-University Munich (LMU) for useful help and advice. I would also like to acknowledge the help and advice of three anonymous referees.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

[Received September 2019. Accepted September 2020. Published Online October 2020.]

APPENDIX

Let $\mathbf{1}$ be a $(v-1)$ dimensional vector of unit elements, let \mathbf{I} be a $(v-1) \times (v-1)$ identity matrix and let $\mathbf{U} = (-\mathbf{1}, \mathbf{I})$. Let $\hat{\mathbf{t}}$ be a vector of estimated treatment means and let \mathbf{C} be the covariance matrix of $\hat{\mathbf{t}}$. Then $\hat{\boldsymbol{\tau}} = \mathbf{U}\hat{\mathbf{t}}$ is a vector of estimated treatment differences and \mathbf{UCU}' is the covariance matrix of $\hat{\boldsymbol{\tau}}$. Assuming a suitable partition of \mathbf{C} , the matrix \mathbf{UCU}' can be expanded to give:

$$\text{Cov}(\hat{\boldsymbol{\tau}}) = \mathbf{1}\mathbf{1}'\mathbf{C}_{11} - \mathbf{C}_{21}\mathbf{1}' - \mathbf{1}\mathbf{C}_{12} + \mathbf{C}_{22} \quad (\text{A1})$$

It then follows that $v \times \text{trace}(\text{Cov}(\hat{\boldsymbol{\tau}})) - \text{sum}(\text{Cov}(\hat{\boldsymbol{\tau}})) = v \times \text{trace}(\text{Cov}(\hat{\mathbf{t}})) - \text{sum}(\text{Cov}(\hat{\mathbf{t}}))$, as required.

REFERENCES

- Atkinson, A.C., Donev, A.N. and Tobias, R. D. (2007). "Optimum Experimental Designs, with SAS". Oxford, Oxford University Press.
- Bailey, R. A. (1999). "Choosing designs for nested blocks". *Listy Biometryczne-Biometrical Letters*, 36, 85–126.
- Bailey R. A. (2008). "Design of Comparative Experiments (Cambridge Series in Statistical and Probabilistic Mathematics)". Cambridge University Press.
- Bates, D., Maechler, M., Bolker, B. and Walker, S. (2015). "Fitting Linear Mixed-Effects Models Using lme4". *Journal of Statistical Software*, 67, 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Burnham, K. P. and Anderson, D. R. (2002). "Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach" (2nd ed.). Springer.

- Cleveland, W. S. (1979). "Robust Locally-Weighted Regression and Smoothing Scatterplots". *Journal of the American Statistical Association*, 74, 829–836.
- Cleveland, W. S. (1993). "Visualizing Data". New Jersey: Hobart Press.
- Cullis, B. and Gleeson, A. (1991). "Spatial analysis of field experiments—an extension to two dimensions". *Biometrics*, 47, 1449–1460.
- Cullis, B., Smith, A. and Coombes, N. (2006). "On the Design of Early Generation Variety Trials with Correlated Data". *Journal of Agricultural, Biological, and Environmental Statistics*, 11, 381–393.
- Dean, A., Morris, M., Stufken, J. and Bingham, D. (eds.) (2015). "Handbook of Design and Analysis of Experiments". Boca Raton: CRC Press.
- Dutilleul, P. (1993). "Spatial Heterogeneity and the Design of Ecological Field Experiments". *Ecology*, 74, 1646–1658.
- Durban, M., Hackett, C., McNicol, J., Newton, A., Thomas, W. and Currie, I. (2003). "The practical use of semi-parametric models in field trials". *Journal of Agricultural, Biological, and Environmental Statistics*, 8, 48–66.
- Eccleston, J. and Chan, B. (1998). "Design Algorithms for Correlated Data". *COMPSTAT Proceedings in Computational Statistics, 13th Symposium*, Editors Payne, R and Green P, 41–52.
- Edmondson, R. N. (2020). "blocksdesign: Nested and Crossed Block Designs for Factorial, Fractional Factorial and Unstructured Treatment Sets". R package version 4.4.
- Gilmour, A. R., Cullis, B. R. and Verbyla, A. P. (1997). "Accounting for Natural and Extraneous Variation in the Analysis of Field Experiments". *Journal of Agricultural, Biological, and Environmental Statistics*, 2, 269–293.
- Gilmour, S. G. and Trinca, L. A. (2006). "Response Surface Methodology and Related Topics" (ed. A. I. Khuri), Chapter 2, pp. 19–46. New York: World Scientific Publishers.
- Goos, P. (2002). "The optimal design of blocked and split-plot experiments". New York: Springer.
- Goos, P. and Vandebroek, M. (2001). "Optimal response surface designs in the presence of random block effects". *Computational Statistics and Data Analysis*, 37, 433–453.
- Goos, P. and Donev, A. N. (2006). "Blocking response surface designs". *Computational Statistics and Data Analysis*, 51, 1075–1088.
- Hastie, T. (2020). "gam: Generalized Additive Models". R package version 1.20.
- Hastie, T. J. and Tibshirani, R. J. (1986). "Generalized Additive Models". *Statistical Science*, 1, 297–318.
- Hastie, T. J. and Tibshirani, R. J. (1987). "Generalized Additive Models: Some Applications". *Journal of the American Statistical Association*, 82, 371–386.
- Hu, X. and Spilke, J. (2009). "Comparison of various spatial models for the analysis of cultivar trials". *New Zealand Journal of Agricultural Research*, 52, 277–287.
- John, J. A. and Williams, E. R. (1995). "Cyclic and Computer Generated Designs", 2nd Edition. London: Chapman and Hall.
- Jones, B., Allen-Moyer, K. and Goos, P. (2020). "A-optimal versus D-optimal design of screening experiments". *Journal of Quality Technology*. Published online: 21 May 2020.
- Mitchell, T. J. (1974). "An algorithm for the Construction of D-Optimal Experimental Designs". *Technometrics*, 16, 203–210.
- Muller, S., Sealy, J. L. and Welsh, A. H. (2013). "Model Selection in Linear Mixed Models". *Statistical Science*, 28, 135–167.
- Patterson, H. D. and Hunter, E. A. (1983). "The efficiency of incomplete block designs in National List and Recommended List cereal variety trials". *Journal of Agricultural Science, Cambridge*, 101, 427–433.
- Pedersen, E. J., Miller, D. L., Simpson, G. L. and Ross, N. (2019). "Hierarchical generalized additive models in ecology: an introduction with mgcv". *PeerJ* 7: e6876. 10.7717/peerj.6876
- Piepho, H. P., Buchse, A. and Emrich, K. (2003). "A Hitchhiker's Guide to Mixed Models for Randomized Experiments". *Journal of Agronomy and Crop Science*, 189, 310–322.
- Piepho, H. P., Williams, E. R., and Michel, V. (2015). "Beyond Latin Squares: A Brief Tour of Row-Column Designs". *Agronomy Journal*, 107, 2263–2270
- Pinheiro, J. C. and Bates, D. M. (2000). "Mixed-Effects Models in S and S-PLUS". Springer, Berlin.

- R Core Team (2020). "R: A language and environment for statistical computing". R Foundation for Statistical Computing, Vienna, Austria.
- Rodriguez-Alvarez, M. X., Boer, M. P., van Eeuwijk, F. A. and Eilers, P. H. C. (2016). "Spatial Models for Field Trials". [arXiv:1607.08255](https://arxiv.org/abs/1607.08255) [stat.ME].
- Rodriguez-Alvarez, M. X., Boer, M.P., van Eeuwijk, F. A. and Eilers, P.H.C. (2018). "Correcting for spatial heterogeneity in plant breeding experiments with P-splines". *Spatial Statistics*, 23, 52–71.
- Saefken, B., Ruegamer, D., Kneib, T. and Greven, S. (2018). "Conditional Model Selection in Mixed-Effects Models with cAIC4". [arXiv:1803.05664](https://arxiv.org/abs/1803.05664)
- Sarkar, D. (2008). "Lattice: Multivariate Data Visualization with R". New York: Springer. ISBN 978-0-387-75968-5.
- Shah, K. R. and Sinha, B. K. (2012). "Theory of Optimal Designs". New York: Springer.
- Wilkinson, G.N., Eckert, R. S., Hancock, T.W. and Mayo, O. (1983). "Nearest Neighbour (NN) Analysis of Field Experiments", *Journal of the Royal Statistical Society. Series B*, 45, 151–211.
- Wolfinger, R. (1993). "Covariance structure selection in general mixed models". *Communications in Statistics - Simulation and Computation*, 22, 1079–1106
- Wood, S. N. and Scheipl, F. (2020). "gamm4: Generalized Additive Mixed Models using 'mgcv' and 'lme4' ". R package version 0.2-6.
- Wood, S. N. (2017). "Generalized Additive Models: An Introduction with R", 2nd Edition. Chapman and Hall/CRC Press.
- Wright, K. (2020). "agridat: Agricultural Datasets". R package version 1.17
- Zuur, A., Ieno, E. N., Walker, N., Saveliev, A. A. and Smith, G. M. (2009). "Mixed Effects Models and Extensions in Ecology with R". New York: Springer

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.