Theor Appl Genet (2015) 128:1541–1554
DOI 10.1007/s00122-015-2530-0

ORIGINAL PAPER

Inter-block information: to recover or not to recover it?

Jens Möhring[1] · Emlyn Williams[2] · Hans-Peter Piepho[1]

Received: 22 September 2014 / Accepted: 27 April 2015 / Published online: 14 May 2015

Abstract

Key message **Comparing standard errors of treatment differences using fixed or random block effects with the approximation of Kackar and Harville helps in choosing the preferable assumption for blocks in the analysis of field experiments.**

Abstract Blocked designs are common in plant breeding field trials. Depending on the precision of variance estimates, recovery of inter-block information via random block effects may be worthwhile. A challenge in practice is to decide when recovery of information should be pursued. To investigate this question, a series of sugar beet trials laid out as α-designs were analysed assuming fixed or random block effects. Additionally, small trials laid out as α-designs or partially replicated designs were simulated and analysed assuming fixed or random block effects. Nine decision rules, including the Kackar–Harville adjustment, were used for choosing the better assumption regarding the block effects. In general, use of the Kackar–Harville adjustment works well and is recommended for partially replicated designs. For α-designs, using inter-block information is preferable for designs with four or more blocks.

Communicated by M. Bink.

✉ Hans-Peter Piepho
piepho@uni-hohenheim.de

[1] Biostatistics Unit, Institute for Crop Science, University of Hohenheim, Stuttgart, Germany

[2] Statistical Consulting Unit, School of Mathematical Sciences (MSI), The Australian National University, Canberra, ACT 0200, Australia

Introduction

To control field variability in plant breeding field trials, some kind of blocking is almost always used (John and Williams 1995). As the number of treatments is normally high, designs with incomplete blocks such as α-designs (Patterson and Williams 1976) or partially replicated (p-rep) designs (Cullis et al. 2006; Williams et al. 2011, 2014) are often preferred. The same incomplete blocking structure is often used in laboratory experiments, e.g., if not all samples can be measured on one microtiter plate or within a day. If it is plausible to assume that block effects are randomly drawn from some parent population of possible block effects or if there is an underlying randomization process to distribute treatments to blocks (Calinski and Kageyama 2000), taking blocks as a random factor allows the recovery of inter-block information (Yates 1940). For known variances, such a combined analysis weighting inter-block and intra-block treatment information is best. For better readability, we denote treatment information as simply as 'information' throughout the rest of the paper because our main aim is the estimation of treatment effects. Also, for brevity we denote an analysis assuming random block effects and hence recovering inter-block information as combined analysis. Weights in the combined analysis are based on estimated variances. Naïve estimates of standard errors for treatment effects from this combined analysis treat variance estimates as if they were known constants. The problem is that these weights will underestimate the true standard errors because errors in estimating the variances, and hence errors in the weights for the inter-block and intra-block information, are ignored (Mead et al. 2012). In particular, when the number of blocks is small and the block variance is large, estimates of the block variance may be so imprecise that weights are not determined with sufficient accuracy for recovery of inter-block information to be worthwhile (Casella 2008; Mead et al. 2012).

The problem in practice is to decide when it is worthwhile to recover the inter-block information. Clearly, the naïve estimates of precision are not a helpful guide in this context; it can be shown that these naïve estimated standard errors for the combined analysis will always be smaller than the estimated standard errors for intra-block analysis (Kenward and Roger 1997), even when recovery is not worthwhile. To avoid inefficient combined analysis, Mead (1988) and van Eeuwijk (1995) proposed that at least 10 degrees of freedom should be required for estimation of the block variance.

Based on the distributional properties of the combined treatment effect estimator for balanced incomplete block (BIB) designs, Graybill and Deal (1959) suggested that for sample size exceeding 10 an estimate from combined analysis is preferable. Noting that for BIB designs the relative intra- and inter-block information is split according to the ratio $\sigma_b^2 E/\sigma^2(1-E)$, where E is the efficiency factor and σ_b^2 and σ^2 are the block and error variances, respectively, Mead et al. (2012; Chapter 9) proposed that at least 4 % of the information is needed between blocks. So according to this rule, recovery is worthwhile when the efficiency factor is low (below 0.8) and the block-to-error variance ratio is less than 4. The efficiency factor E for a BIB design given by $(\lambda t)\big/(rk)$, where λ is the number of times two treatments occur together in a block, and t, r, and k are the number of treatments, replicates and plots per block, respectively. It is the ratio of the variance of a treatment difference for a BIB design and the variance of a difference for a randomized complete block design with the same number of experimental units and the same error variance (John and Williams 1995; Chapter 2). In practice this means that the inter-block analysis would be used only if block sizes are between two and four and the reduction of error variance due to blocking is small (Mead et al. 2012).

For other designs, the decision on recovery of inter-block information is more difficult (Mead et al. 2012). Kackar and Harville (1984) proposed a general adjustment of the standard error estimates based on a first-order Taylor expansion, which accounts for the errors in weights and usually leads to an inflation of the standard error estimates compared to naïve estimates from a combined analysis assuming known variances. This method was also used by Kenward and Roger (1997) in their procedure for small-sample inference on fixed effects in mixed models, which has become very popular after its implementation in some mixed model packages. The authors proposed to recover the inter-block information only when the Kackar–Harville adjusted standard errors of treatment differences (s.e.d.) are smaller than the corresponding estimated s.e.d. from an analysis with fixed block effects. While this approach has great practical appeal, its empirical performance does not seem to have been systematically investigated.

This paper pursues the approach of trying both analyses (with and without recovery) using a mixed model package with the residual maximum likelihood approach (REML; Patterson and Thompson 1971) as the estimation method for the variances and then picking one of the analyses based on a suitable decision rule designed to identify the smaller s.e.d. We will evaluate the decision rule using the Kackar–Harville adjustment and alternative decision rules. We considered a series of small single-trial designs from sugar beet and simulated single-trial resolvable incomplete block designs and partially replicated designs. The "Materials and methods" section starts with a description of the motivating series of sugar beet trials and then gives an overview of our simulation approach. Details concerning the model for simulating and analysing data for the investigated designs and nine different decision rules are given. Then we compare these rules based on the mean squared error of treatment differences (MSED) and the probability of correct model selection within the simulation. The paper ends with a discussion of results and our main conclusion.

Materials and methods

Motivating example: series of sugar beet plant breeding trials

The series comprised 285 trials from 21 locations. Within a trial one out of 26 subgroups each with 32 test genotypes and four checks was allocated to plots according to an α-design with two replicates of six incomplete blocks of size six. Separate analyses for corrected sugar yield were performed for each trial using random (combined analysis) or fixed (intra-block analysis) block effects. Both analyses used fixed effects for replicate and genotype. For each trial the size of variance estimates and the average estimated standard error were calculated. For the combined analysis we also calculated corrected standard errors using the Kenward–Roger approximation. To compare both models we calculated the average correlation of differences of test genotype estimates from different trials. For this purpose, we correlated least square mean differences of test genotypes from a single trial against mean differences across all other trials with the same subgroup of test genotypes. The second analysis corresponds to a two-stage analysis ignoring weights from the first stage. As the data are nearly balanced according to this subgroup of test genotypes, we expect that the correlation with a one-stage analysis is quite high (Möhring and Piepho 2009). We assume that a more precise estimate of test genotype differences results in a higher correlation.

Overview of simulation approach

On average, results from the motivating example show a clear preference for a combined analysis, but results vary from trial to trial. This raises the question whether there is a method to distinguish between trials where the combined analysis is better and trials where the intra-block analysis is better. We, therefore, started with a simulation taking the design and the median variance parameter values from the motivating example into account (scenario 1). Because results from this simulation are quite simple—most decision rules always select the preferred combined analysis—we varied our input variables and concentrated on cases, where the decision was less clear. Therefore, we performed additional simulation studies using different experimental designs (α-designs and p-rep designs). For each study we used a range of different scenarios. Scenarios varied in given number of blocks, block size, and the ratio of variance values. Data generation and analyses were done in SAS. Designs were created using the software package CycDesigN 5.1 (VSN International; http://www.vsni.co.uk/; Williams et al. 2014). Simulation studies comprised $n = 5000$ or $n = 1000$ datasets for each scenario. Each dataset was subjected to two analyses that differed only in the assumption regarding the block effects. The first model assumed block effects as fixed, and the second took block effects as random. For both analyses the MSED of all pairwise treatment differences was calculated (see section "Evaluation criteria" for details. Furthermore, for each dataset nine different rules were used to decide whether the analysis with fixed or random block effects should be used for the dataset. For each rule and each simulated scenario the resulting MSED was averaged across datasets. Additionally, the probability of making the right decision (i.e., selecting the more accurate of the two alternative analyses) was calculated as an evaluation criterion.

Data simulation

Observations for all datasets were simulated according to the following model:

$$y_{ijk} = \mu + r_j + b_{jk} + t_i + e_{ijk}, \tag{1}$$

where y_{ijk} is the response of the i-th treatment in the j-th replicate and k-th block, μ is the general intercept, r_j is the j-th replicate effect, b_{jk} is the k-th block effect within the j-th replicate, t_i is the i-th treatment effect, and e_{ijk} is the plot error corresponding to the response y_{ijk}. The design effects b_{jk} and e_{ijk} were drawn from normal distributions with homogeneous variances. For simplification, and without loss of generality, all treatment and replicate effects (t_i and r_j) were set to zero.

Experimental field designs and scenarios of simulation studies

Our simulation study started with a scenario which is quite similar to the motivating example (scenario 1). We then concentrated on worst-case scenarios where decision rules varied in their preferred analysis and the intra-block analysis can be a better option than the inter-block analysis. We generated a number of common α-designs and p-rep designs using CycDesigN 5.1 (Table 1). In our simulation studies we varied the following parameters: the number of blocks (scenario 2), the block sizes (scenario 3), and the block-to-error variance ratio (scenario 4). For the latter we varied the variance parameter values, but because only the ratio of block and error variance influences the relative performance of combined and intra-block analysis, we mainly just reported the ratio of block and error variances. For tables reporting MSED values we added the information about the error variance.

To concentrate our simulation on the most critical cases for combined analysis, we modified parameters in scenarios 2–4. As long as we were not varying the corresponding parameter value, we reduced the number of blocks to four. This is the minimum possible number of blocks. We used a block size of 12 because we expected that estimates for error variance are more stable in this case. Additionally, we assumed the highest block-to-error variance ratio found in our motivating example (we used a ratio of 5). The larger the ratio, the less benefit will accrue from using inter-block information (Mead et al. 2012). Additionally, as the absolute value of variances only affects the absolute value of the MSED, we varied the block variance parameter value and set the error variance to an arbitrary value of 40 for scenarios 2–5. The strategy of fixing the error variance and varying the block variance has the advantage that MSED values for different variance ratios are comparable. The efficiencies for our non-BIB designs reported in Table 1 were taken from CycDesigN 5.1. For the α-designs the number of treatments is half the number of experimental units. Therefore, designs also varied in the number of treatments and the number of plots. For p-rep designs we used 48 treatments and 72 experimental units. As this study aims at single-trial analysis, we used just one trial of each of the p-rep designs generated for four locations (Williams et al. 2011). The allocation of treatments to experimental units was randomized for each dataset.

As the block model for both designs is identical, we subjected each dataset to intra-block analysis and combined analysis. For combined analysis we adjusted the degrees of freedom and standard errors with and without the method of Kenward and Roger (1997). The replicate effect was taken as fixed throughout. Note that in

Table 1 Overview of considered designs and scenarios

Design	Scenario	Number of					Ratio of block-to-error variance	Efficiency factor
		Blocks	Experimental units per block	Obser-vations	Replicates	Treatments		
α-Design	1	12	6	72	2	36	0.3568[§]	0.7778
α-Design	2 a*	4	12	48	2	24	5	0.9200
α-Design	2 b	6	12	72	2	36	5	0.8974
α-Design	2 c	8	12	96	2	48	5	0.8868
α-Design	2 d	10	12	120	2	60	5	0.8765
α-Design	2 e	12	12	144	2	72	5	0.8667
α-Design	3 a	4	4	16	2	8	5	0.7778
α-Design	3 b	4	6	24	2	12	5	0.8462
α-Design	3 c	4	8	32	2	16	5	0.8824
α-Design	3 d	4	10	40	2	20	5	0.9048
α-Design	3 e*	4	12	48	2	24	5	0.9200
α-Design	4 a–f*	4	12	48	2	24	0.05–50	0.9200
p-rep	5 a	3	24	72	1.5	48	5	0.6425
p-rep	5 b	6	12	72	1.5	48	5	0.8409
p-rep	5 c	9	8	72	1.5	48	5	0.8991
p-rep	5 d	12	6	72	1.5	48	5	0.9292

* Scenarios 2 a, 3 e and 4 e are identical

§ The ratio is 812 to 2276 which are the median values from the motivating example

Table 2 Decision rules of selecting combined analysis

Rule	Condition for assuming random block effects (combined analysis)
Rule 1	Never
Rule 2	Always
Rule 3	If corrected average s.e.d. of combined analysis is smaller
Rule 4	If block variance estimate is positive
Rule 5	If block variance estimate is positive and if corrected average s.e.d. of combined analysis is smaller
Rule 6	If there are at least than ten degrees of freedom for estimation the block variance
Rule 7	If the efficiency factor is smaller than 0.8 and the block-to-error variance ratio is smaller than 4
Rule 8	If the simulated probability for using the combined analysis is larger than 0.5
Rule 9	If the simulated average MSED of the combined analysis is the smallest one

s.e.d. standard error of treatment differences, *MSED* mean square error of treatment differences

the α-designs, replicates are complete, so no inter-replicate information exists and hence no information could be recovered by taking replicates as random. In case of p-rep designs, there are no complete replicates and recovering inter-replicate treatment information raises the same question for replicates, which was discussed for incomplete blocks. We simulated no replicate effects and ignored the inter-replicate information in both models, and we think that this does not affect the relative performance of our two models with respect to the assumption made for blocks.

Decision rules

We used nine different rules for deciding whether an intra-block analysis or a combined analysis should be performed (Table 2). *Rule 1* and *Rule 2* are invariant to the analysed dataset. *Rule 3* to *Rule 7* are based on readily computable criteria used in the literature, and *Rule 8* and *Rule 9* are based on a simulation approach.

Rule 1 ignores inter-block information throughout. *Rule 2* always recovers the inter-block information. *Rule 3* to *Rule 9* in principle allow both analyses. *Rule 3* uses the Kackar and

Harville adjustment for s.e.d. estimation. *Rule 6* uses the decision rule of Mead (1988) and van Eeuwijk (1995). *Rule 7* uses for all designs the decision rule that Mead et al. (2012) proposed for a BIB design. *Rule 4* and *Rule 5* are extensions to *Rule 2* and *Rule 3*, respectively, requiring that in cases of a zero block variance estimate the user switches to an intra-block analysis. The motivation for these rules is the following: To our knowledge most mixed model software packages constrain very small variance estimates to a boundary value of zero. This is also true for the package we used (SAS, PROC MIXED). In our simulation we assume a block variance to be zero whenever the program constrained the variance to zero. Constraining a block variance estimate to zero effectively removes block effects from the model and therefore results in the smallest possible estimated average s.e.d. (Mead et al. 2012); furthermore, a zero estimate for the variance may indicate that a model that allows negative variances is more appropriate (Nelder 1954). As it is uncommon to have no block effects, our analysis reverts to fixed effects for blocks as a backup against underestimation of the error term and an invalid analysis. *Rule 8* and *Rule 9* are based on a simulation as described in the next sub-section.

Simulation for Rule 8 and Rule 9

The basic idea of *Rule 8* and *Rule 9* is to use information from a given dataset to (1) simulate comparable data with known true treatment differences, (2) analyse them with models assuming fixed or random block effects and (3) select the better model. Therefore, for a given dataset, variance estimates are computed using model (1). These variance estimates are then used for a further simulation, where in each simulation run data are created based on model (1), taking block effects as random. The same two analyses with either fixed or random block effects are then performed. For each simulation run and each analysis, the MSED is calculated. If for a simulation run the block variance is estimated to be zero, both rules use the MSED of the intra-block analysis. These MSED values are averaged across simulation runs for a dataset. *Rule 9* takes the model for the analysis with the smaller average MSED. Additionally, we counted the number of simulation runs, where each analysis was selected based on the MSED. *Rule 8* decides to take the model for the analysis which is selected the most often.

Note that here we use these two rules on data which are themselves simulated, so overall, our study involves two levels of simulation: (1) an outer simulation that generates datasets for a given design and scenario and (2) an inner simulation when applying *Rule 8* and *Rule 9* to each dataset from level (1).

Evaluation criteria

Based on the decisions taken using *Rule 1* to *Rule 9*, the MSED and whether the truly better model was selected are assessed for each dataset. As we simulated no treatment effects, the MSED is the average of the squared estimated treatment differences. The MSED values per dataset were then averaged across datasets for each design and scenario. By "truly better model" we mean that for a given dataset the MSED for that model is smaller than for the other model. This does not imply that the better model is the best possible model, as there may be yet better models than the ones considered for analysis. Smaller average MSED values and a higher probability indicate better rules for deciding whether it is worthwhile to recover inter-block information.

Results

Motivating example

Using combined analysis, the block-to-error variance ratio varied between 0 and about 5 with a median ratio of 0.39. About 17 and 87 % of the ratios were smaller than 0.05 and 1.25, respectively. The block variance estimates were more variable (values between 0 and 13,951 with a median of 812) than the error variance estimates (values between 474 and 10,702 with a median of 2276). In 39 trials the block variance estimate was constrained to zero. As each trial comprised 12 blocks there were 10 degrees of freedom for estimating the block variance. Aside from the variability of the ratio of variance estimates and the small number of blocks, the estimated standard error of treatment differences for the combined analysis with and without using the Kacker–Harville approximation was always smaller than for the intra-block analysis for all 285 trials. The average correlation of differences of treatment estimates from different trials was 0.4130 for the combined analysis and 0.4022 for the intra-block analysis, but for 111 out of 285 trials the intra-block analysis shows the higher correlation. No obvious effect of the size of variances or their ratio was detectable. Both the average correlation and the estimated standard error tended to favour the assumption of random block effects across all trials. This is in accordance with results from using the same design and the median values for block and error variance parameters in a simulation study (scenario 1). In 79.41 % of the simulation runs a combined analysis was better. The MSED for non-zero block variance estimates was 2672.64 compared to 2958.56 for the intra-block analysis (5000 simulation runs). All decision rules except of *Rule 1* selected the combined analysis in all simulation runs.

Simulation results

Results from our simulation (scenarios 2–5) are given for two separate cases: (1) when the block variance estimate is constrained to zero and (2) when it is larger than zero. The rationale for this separation is that constraining the block variance estimate to zero in a combined analysis de facto drops the block effect from the model. As it is uncommon to have no real block effects, we assume that an estimate constrained to zero only reflects estimation error. Using an intra-block analysis can be seen as a fall-back option in these situations that makes sure the block effects stay in the model.

Results for α-designs

The probability that an intra-block analysis is best is identical to the probability that *Rule 1* selects the better model. In general, this probability decreases with increasing number of blocks (Table 3), decreasing block size (Table 9), and decreasing block-to-error variance ratio (Table 4). For a given scenario and a given number of blocks, *Rule 1* to *Rule 6* lead to the same decision for all simulated data sets (Tables 5, 10, 11). The probability for selecting the intra-block analysis increases with decreasing number of blocks (scenarios 2) for *Rule 3*, *Rule 5*, *Rule 6*, *Rule 8,* and *Rule 9* (Table 5). For increasing block-to-error variance ratio (scenario 4), the probability of selecting intra-block analysis increases for *Rule 8* and *Rule 9* (Table 10). Furthermore, for increasing block size (scenario 3), the probability of selecting intra-block analysis increases for *Rule 7* to *Rule 9* (Table 10). Therefore, in principle all decision rules which potentially allow the selection of both models switch to a combined analysis, if this analysis tends to be better. If the truly better model varies between simulation runs, this does not affect the probabilities that a combined analysis is

Table 3 Probability of selecting the better model for nine decision rules and scenario 2 (varying number of blocks) depending on the value of the block variance estimate (zero or positive)

Scenario	Estimate of block variance	Number of		Probabilities for the nine different rules								
		Simulations	Blocks	1	2	3	4	5	6	7	8	9
2 a	Positive	4829	4	0.4813	0.5187	0.4813	0.5187	0.4813	0.4813	0.4813	0.5250	0.4999
	0	171	4	0.6024	0.3976	0.3976	0.6024	0.6024	0.6024	0.6024	0.6024	0.6024
2 b	Positive	4987	6	0.4738	0.5262	0.4738	0.5262	0.4738	0.4738	0.4738	0.5242	0.5185
	0	13	6	0.77	0.23	0.23	0.477	0.77	0.77	0.77	0.77	0.77
2 c	Positive	5000	8	0.4496	0.5504	0.5504	0.5504	0.5504	0.4496	0.4496	0.5496	0.5374
2 d	Positive	5000	12	0.4125	0.5875	0.5875	0.5875	0.5875	0.5875	0.4125	0.5875	0.5867
2 e	Positive	5000	24	0.3796	0.6204	0.6204	0.6204	0.6204	0.6204	0.3796	0.6204	0.6204

All scenarios used an α-design with block size of 12, and a block-to-error variance ration of 5. Table 5 also looks at scenario 2, but considers the probability of selecting an intra-block analysis

Table 4 Probability of selecting the better model for nine decision rules and scenario 4 (varying block-to-error variance ratio) depending on the value of the block variance estimate (zero or positive)

Scenario	Estimate of block variance	Ratio of block-to-error variance	Number of simulations	Probabilities for the nine different rules								
				1	2	3	4	5	6	7	8	9
4 a	Positive	0.05	497	0.2034	0.7966	0.2034	0.7966	0.2034	0.2034	0.2034	0.7923	0.7666
	0	0.05	503	0.4053	0.5947	0.5947	0.4053	0.4053	0.4053	0.4053	0.4053	0.4053
4 b	Positive	0.125	571	0.2680	0.7320	0.7320	0.2680	0.2680	0.2680	0.2680	0.7303	0.6953
	0	0.125	429	0.4918	0.5082	0.5082	0.4918	0.4918	0.4918	0.4918	0.4918	0.4918
4 c	Positive	0.5	797	0.3676	0.6324	0.3676	0.6324	0.3676	0.3676	0.3676	0.6198	0.4780
	0	0.5	203	0.5369	0.4631	0.4631	0.5369	0.5369	0.5369	0.5369	0.5369	0.5369
4 d	Positive	1.25	886	0.4413	0.5587	0.4413	0.5587	0.4413	0.4413	0.4413	0.5485	0.4480
	0	1.25	114	0.6316	0.3684	0.3684	0.6316	0.6316	0.6316	0.6316	0.6316	0.6316
4 e	Positive	5	4829	0.4813	0.5187	0.4813	0.5187	0.4813	0.5187	0.4813	0.5250	0.4999
	0	5	171	0.6024	0.3976	0.3976	0.6024	0.6024	0.3976	0.6024	0.6024	0.6024
4 f	Positive	50	999	0.4955	0.5045	0.4955	0.5045	0.4955	0.4955	0.4955	0.5145	0.5065
	0	50	1	1	0	0	1	1	1	1	1	1

All scenarios used an α-design with four blocks of size of 12. Table 11 also looks at scenario 4, but considers the probability of selecting the truly better method

Table 5 Probability of selecting intra-block analysis for nine decision rules and scenario 2 (varying number of blocks) depending on the value of the block variance estimate (zero or positive)

Scenario	Estimate of block variance	Number of		Probabilities for the nine different rules§								
		Simulations	Blocks	1	2	3	4	5	6	7	8	9
2 a	Positive	4829	4	1	0	1	0	1	1	1	0.1814	0.8233
	0	171	4	1	0	0	1	1	1	1	1	1
2 b	Positive	4987	6	1	0	1	0	1	1	1	0.0040	0.0578
	0	13	6	1	0	0	1	1	1	1	1	1
2 c	Positive	5000	8	1	0	0	0	0	1	1	0.0032	0.0454
2 d	Positive	5000	12	1	0	0	0	0	0	1	0	0.0009
2 e	Positive	5000	24	1	0	0	0	0	0	1	0	0

All scenarios used an α-design with block size of 12, and a block-to-error variance ratio of 5

§ These are exact 0s and 1s

Table 6 MSED for nine decision rules and scenario 2 (varying number of blocks) depending on the value of the block variance estimate (zero or positive)

Scenario	Estimate of block variance	Number of		Average MSED for the nine different rules								
		Simulations	Blocks	1	2	3	4	5	6	7	8	9
2 a	Positive	4829	4	43.78	43.99	43.78	43.99	43.78	43.78	43.78	44.00	44.01
	0	171	4	43.92	45.96	45.96	43.92	43.92	43.92	43.92	43.92	43.92
2 b	Positive	4987	6	44.56	44.53	44.56	44.53	44.56	44.56	44.56	44.53	44.54
2 c	Positive	5000	8	45.17	45.10	45.10	45.10	45.10	45.17	45.17	45.10	45.11
2 d	Positive	5000	12	45.50	45.35	45.35	45.35	45.35	45.35	45.50	45.35	45.35
2 e	Positive	5000	24	46.07	45.88	45.88	45.88	45.88	45.88	46.07	45.88	45.88

All scenarios used an α-design with block size of 12, and a block-to-error variance ratio of 5. For simplicity, we dropped the presentation of MSED for simulation runs with zero block variance estimates for six blocks because the number of these simulations is very low

selected (results not shown). For larger numbers of blocks, an intra-block analysis is sub-optimal (*Rule 1* and *Rule 7* for scenario 2 e). Using an intra-block analysis seems to be better if the block variance estimate is zero (scenarios 2a, 3 and 4). Additionally, *Rule 8* shows higher probabilities of selecting the better model and smaller MSED values than *Rule 9* (Tables 3, 6). Therefore, *Rule 4*, *Rule 5* or *Rule 8* optimizes the probability of selecting the truly better model. We expect that a decreasing block variance and an increasing number of blocks tend to favour a combined analysis. Therefore, intra-block analysis is assumed to be best for designs with four blocks, a large block size of 12, and a relatively high block-to-error variance ratio. *Rule 4* maximizes the probability to select the better model for non-zero block variance (Tables 3, 4, 9). *Rule 5* and *Rule 8* have the advantage that they potentially allow to select both models, the combined analysis and the intra-block analysis, and hence fare a bit better in borderline cases. Evaluating the MSED shows comparable results. But the differences between decision rules are smaller because average MSED values for the analyses with and without recovery are more similar. For scenarios with more than four blocks, the combined analysis shows both higher probabilities of selecting the better model compared to intra-block analysis and the smallest average MSED values. For four blocks and a block-to-error variance ratio of 5 or 50, the intra-block analysis has the smallest MSED (Table 12) but a smaller probability of selecting the better model compared to the combined analysis. MSED values for decisions rules which potentially can select both models are often intermediate between the MSED values of intra-block or combined analysis or identical to one of them.

Results for p-rep designs

For p-rep designs the same tendencies for increasing block-to-error variance ratio and increasing the number of blocks were observed (data not shown). If there is a sufficient number of blocks and treatments per block, the combined analysis is preferred. We, therefore, just present scenarios with a high block-to-error variance ratio of 5. Changing the number of blocks for a given set of 48 treatments and a given number of 72 experimental units produces less clear results (Tables 7, 13). As the block size varies, the number of blocks and the number of replicated treatments per block vary too. While for positive block variance estimates and scenarios 5 a–c the combined analysis (*Rule 4*) shows both higher probabilities of selecting the better model and smaller MSED values, for scenario 5 d the intra-block analysis shows better values of the evaluation criteria (Tables 7, 13). Again, *Rule 3*, *Rule 5*, *Rule 8,* and *Rule 9* switch to a combined analysis if the designs get larger (data not shown) or if the probability that this analysis is truly best increases (this probability is equal to the one given for *Rule 2*; Table 8). In contrast to our findings for the α-design, *Rule 4* is not always the best decision rule and, e.g., *Rule 5* or *Rule 8* can provide better decisions, as both rules potentially allow to switch to an intra-block analysis.

Discussion

Almost all field experiments involve some kind of blocking. As such block effects should be considered in the analysis, the decision on whether block effects are taken as fixed or random in the analysis is crucial. For BIB designs, Graybill and Deal (1959) suggested that inter-block information should always be used if $r \cdot t - b - t + 1 \geq 18$ and $b - t = 9$ or $b - t \geq 10$. Seshadri (1963) and Stein (1966) stated that $t \geq 9$ or $t \geq 4$ is required. Brown and Cohen (1974) showed that for four and more blocks inter-block information should be used. Our simulation results for α-designs (which are not BIB designs) show a comparable result when averaged. The results of our simulation study indicate that four blocks are sufficient for using inter-block information. "Sufficient" means that on average across a large number of experiments with the same design and with the same variance–covariance structure, the MSED is smaller when inter-block information is used. This criterion should not be confused with the criterion "good" used, e.g., in Bhattacharya (1998), where "good" means that the chosen model is best for each single simulation run. So our aim was to select a rule to find the best expected model for a new experiment.

For large designs it is obvious that using inter-block information is likely to increase precision of treatment estimates and their standard errors (Kenward and Roger 1997). For smaller designs, Casella (2008) argued that a linear combination of inter-block and intra-block estimates may have higher variance due to uncertainty in the weight estimates. Stein (1956) proposed that a reasonable number of degrees of freedom for the block variance is required. Mead (1988) and van Eeuwijk (1995) proposed that 10 degrees of freedom are required. All authors assume that for small sample sizes or small designs, imprecise variance estimates may result in poorer treatment effect estimates from combined analysis compared to intra-block analysis. Mead et al. (2012) proposed that inter-block information should not be used unless there is enough inter-block information. But not recovering inter-block information is equivalent to using extreme values for dispersion parameters (i.e., an infinite block variance)

Table 7 Probability of selecting the right model using scenario 5 (different number of blocks and block sizes with a fixed number of treatments) depending on the value of the block variance estimate (zero or positive)

Scenario	Estimate of block variance	Number of		Simulations	Probabilities for the nine different rules								
		Blocks	Replicated treatments/block		1	2	3	4	5	6	7	8	9
5 a	Positive	6	8	994	0.4125	0.5875	0.4054	0.5875	0.4054	0.4125	0.4125	0.5845	0.4678
	0	6	8	6	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
5 b	Positive	8	6	998	0.4058	0.5942	0.5271	0.5942	0.5271	0.4058	0.4058	0.5932	0.5721
	0	8	6	2	1	0	0	1	1	1	1	1	1
5 c	Positive	12	4	1000	0.3660	0.6340	0.5890	0.6340	0.5890	0.6340	0.3660	0.6340	0.6340
5 d	Positive	24	2	991	0.5550	0.4450	0.4884	0.4450	0.4884	0.4450	0.5106	0.5116	0.5156
	0	24	2	9	1	0	0.8889	1	1	0	1	1	1

All scenarios used a p-rep design with a block-to-error variance ratio of 5

Table 8 Probability of selecting intra-block analysis for nine decision rules and scenario 5 (different number of blocks and block sizes with a fixed number of treatments depending on the value of the block variance estimate (zero or positive)

Scenario	Estimate of block variance	Number of			Probabilities for the nine different rules								
		Simulations	Blocks	Replicated treatments/block	1	2	3	4	5	6	7	8	9
5 a	Positive	994	6	8	1	0	0.9708	0	0.9708	1	1	0.0070	0.4054
	0	6	6	8	1	0	0	1	1	1	1	1	1
5 b	Positive	998	8	6	1	0	0.2876	0	0.2876	1	1	0.0010	0.0441
	0	2	8	6	1	0	0	1	1	1	1	1	1
5 c	Positive	1000	12	4	1	0	0.1670	0	0.1670	0	1	0	0
5 d	Positive	991	24	2	1	0	0.7316	0	0.7316	0	0.6428	0.6155	0.6458
	0	9	24	2	1	0	0.8889	1	1	0	1	1	1

All scenarios used a p-rep design with block-to-error variance ratio of 5

and is worse than a combined analysis (Robinson 1991). Our analyses showed that assuming random block effects on average increased the precision for nearly all simulated designs even with extreme variance ratios.

Simulated designs

We focussed on small α-designs with four or more blocks because they are commonly used in plant breeding. We further simulated p-rep designs (Williams et al. 2011, 2014). The use of partially replicated designs in multi-environmental experiments is increasing (Smith et al. 2006; Butler et al. 2009; Beeck et al. 2010; Hickey et al. 2011; Crawford et al. 2011; Longin et al. 2013; Rebetzke et al. 2014) because in series of trials with the same total number of plots they are more efficient than replicated designs or augmented designs (Cullis et al. 2006; Möhring et al. 2014). In p-rep designs, only replicated treatments give information about block effects, if treatment effects are assumed as fixed. Therefore, block variance estimates are expected to be less precise. In our experience, most designs used in plant breeding tend to be at least as large as the ones studied in this paper. As the decision for small designs is critical, we focussed here on these smaller and more critical design sizes. For larger designs, inter-block information should always be used. There are many alternative designs, such as augmented designs (Federer 1956) and other resolvable or non-resolvable block and row-column designs with constant or variable block size (Shrikhande 1951; John and Williams 1995). We conjecture that for these designs the favourable rules *Rule 5* and *Rule 8* may also give good suggestions.

Series of trials

Our simulations concentrated on single-trial analysis. In plant breeding, single-trial analysis is common. If the focus is on the analysis of a series, it is also important to fit variance parameters for blocks and error to explain the variability in each trial. As block and error models can differ between trials, it is useful in the joint analysis of a series of trials to estimate variances per trial (Möhring and Piepho 2009; Piepho et al. 2012). If a two-stage approach is used for analysing a series of trials, single-trial analysis forms the first stage. Furthermore, for α-designs, treatments in trials are equally replicated and form complete replicates. Therefore, no inter-trial information exists and the replicate effect can be taken as fixed. Treatments in p-rep designs are not equally replicated per trial, but are usually equally replicated across the whole experiment. To ensure this property, p-rep designs are usually created as a series of trials. Again it is important to model the variability for each trial separately. Therefore, analysis of a p-rep design as a series of trials should fit trial-specific variances. In summary, we may assume a very similar performance of intra-block and combined analysis for a series of trials, if trial-specific variances are fitted in a joint analysis over trials as compared with analysis of a single-trial. For this reason our study was restricted to single-trial analyses.

Influence of the block-to-error variance ratio

The value of block variance in comparison to the residual error variance influences the amount of information that can be recovered. If the value of the true block variance goes to infinity (or is large compared to other variances) no inter-block information can be recovered and estimates from combined analysis and intra-block analysis are identical. Therefore, recovery of information improves the estimation of treatment effects when block effects are small (Weerakkody 1992). This potential benefit cannot usually be fully realized, however, because variance estimates are estimated and not known. In our simulations, we varied the block-to-error variance ratio and found that both the probability that inter-block information should be used and the probability of a decision for using inter-block information given by our rules increased with increasing ratio. Therefore, the critical case for assuming random block effects is the case with high block-to-error variance ratio. Throughout this paper we, therefore, most often simulated a block-to-error variance ratio of 5, which is relatively high for practical plant breeding trials. Hence, our choice tended to be slightly conservative in the sense that if our suggestion favours the use of inter-block information, this is probably also true for most field trials having smaller ratios.

Distribution of block effects

Our simulations were based on block effects, which were simulated as normally distributed random effects with trial-specific variances. Therefore, an analysis assuming normally distributed random block effects (so *Rule 2*) will be best for large designs or known true variances, as data were simulated with this model. Our results furthermore show that *Rule 4* is best for smaller α-designs and unknown variances. In principle, other distributions for block effects are possible. We have not investigated this point, but we assume that the assumption of an approximate normal distribution of block effects is reasonable for most practical purposes.

Conclusion

For α-designs and a non-zero block variance estimate, the use of inter-block information (R*ule 4*) can be suggested for all simulated designs with four blocks or more. Furthermore, for most p-rep designs *Rule 4* is best. For all simulated p-rep designs trying both analyses and picking the best model using the Kackar-Harville adjustment for non-zero block variance estimates (*Rule 5*) generally works well, too. It can result in too many intra-block analyses for small designs to be performed, but it can also identify cases where an intra-block analysis is preferred. *Rule 8* shows no clear advantage and the same tendency compared to *Rule 5*. Due to the easy access to this procedure via the Kenward-Roger approximation with standard statistical software packages, we recommend *Rule 5* when deciding whether block effects should be taken as fixed or as random, despite its minor imperfections for small designs.

Author contribution statement JM worked out the concept of the paper (with HPP), performed all simulations and analyses, and he wrote the paper. ERW was involved in the concept of the paper and helped with editing the paper. HPP worked out the concept of the paper (with JM) and helped with editing the paper.

Acknowledgments JM and HPP were supported by DFG grant PI 377/13-1. Strube Research is thanked for providing their dataset. We thank two anonymous reviewers for their helpful hints, which enhanced the quality of this paper.

Conflict of interest The authors declare that they have no conflict of interest.

Appendix

See Tables 9, 10, 11, 12 and 13.

Table 9 Probability of selecting the better model for nine decision rules and scenario 3 (varying number of plots per block) depending on the value of the block variance estimate (zero or positive)

Scenario	Estimate of block variance	Block size	Number of simulations	Probabilities for the nine different rules								
				1	2	3	4	5	6	7	8	9
3 a	Positive	4	4572	0.4563	0.5437	0.4563	0.5437	0.4563	0.4563	0.5763	0.5221	0.4175
	0	4	428	0.6238	0.3762	0.3762	0.6238	0.6238	0.6238	0.6238	0.6238	0.6238
3 b	Positive	6	4693	0.4673	0.5327	0.4673	0.5327	0.4673	0.4673	0.4693	0.4673	0.4341
	0	6	307	0.6319	0.3681	0.3681	0.6319	0.6319	0.6319	0.6319	0.6319	0.6319
3 c	Positive	8	4766	0.4824	0.5176	0.4824	0.5176	0.4824	0.4824	0.4824	0.4824	0.4478
	0	8	234	0.6325	0.3675	0.3675	0.6325	0.6325	0.6325	0.6325	0.6325	0.6325
3 d	Positive	10	4818	0.4819	0.5181	0.4819	0.5181	0.4819	0.4819	0.4819	0.5052	0.4498
	0	10	182	0.6703	0.3297	0.3297	0.6703	0.6703	0.6703	0.6703	0.6703	0.6703
3 e	Positive	12	4829	0.4813	0.5187	0.4813	0.5187	0.4813	0.5187	0.4813	0.5250	0.4999
	0	12	171	0.6024	0.3976	0.3976	0.6024	0.6024	0.3976	0.6024	0.6024	0.6024

All scenarios used an α-design with four blocks, and a block-to-error variance ratio of 5

Table 10 Probability of selecting intra-block analysis for nine decision rules and scenario 3 (varying number of plots per block) depending on the value of the block variance estimate (zero or positive)

Sceanrio	Estimate of block variance	Block size	Number of simulations	Probabilities for the nine different rules[§]								
				1	2	3	4	5	6	7	8	9
3 a	Positive	4	4572	1	0	1	0	1	1	0.4517	0.0755	0.6866
	0	4	428	1	0	0	1	1	1	1	1	1
3 b	Positive	6	4693	1	0	1	0	1	1	1	0.1048	0.7166
	0	6	307	1	0	0	1	1	1	1	1	1
3 c	Positive	8	4766	1	0	1	0	1	1	1	0.1244	0.7673
	0	8	234	1	0	0	1	1	1	1	1	1
3 d	Positive	10	4818	1	0	1	0	1	1	1	0.1598	0.8014
	0	10	182	1	0	0	1	1	1	1	1	1
3 e	Positive	12	4829	1	0	1	0	1	0	1	0.1814	0.8233
	0	12	171	1	0	0	1	1	0	1	1	1

All scenarios used an α-design with four blocks and a block-to-error variance ratio of 5

[§] These are exact 0s and 1s

Table 11 Probability of selecting intra-block analysis for nine decision rules and scenario 4 (varying block-to-error variance ratio) depending on the value of the block variance estimate (zero or positive)

Scenario	Estimate of block variance	Ratio of block-to-error variance	Number of simulations	Probabilities for the nine different rules[§]								
				1	2	3	4	5	6	7	8	9
4 a	Positive	0.05	497	1	0	1	0	1	1	1	0.0043	0.0385
	0	0.05	503	1	0	0	1	1	1	1	1	1
4 b	Positive	0.125	571	1	0	1	0	1	1	1	0.0053	0.0543
	0	0.125	429	1	0	0	1	1	1	1	1	1
4 c	Positive	0.5	797	1	0	1	0	1	1	1	0.0125	0.2673
	0	0.5	203	1	0	0	1	1	1	1	1	1
4 d	Positive	0.125	886	1	0	1	0	1	1	1	0.0350	0.4966
	0	0.125	114	1	0	0	1	1	1	1	1	1
4 e	Positive	5	4843	1	0	1	0	1	1	1	0.1814	0.8233
	0	5	157	1	0	0	1	1	1	1	1	1
4 f	Positive	50	999	1	0	1	0	1	1	1	0.4304	0.9289
	0	50	1	1	0	0	1	1	1	1	1	1

All scenarios used an α-design with four blocks of size of 12. Table 3 also looks at scenario 4, but considers the probability of selecting the truly better method

§ These are exact 0s and 1s

Table 12 MSED for nine decision rules and scenario 4 (varying block-to-error variance ratio) depending on the value of the block variance estimate (zero or positive)

Scenario	Estimate of block variance	Ratio of block-to-error variance	Number of simulations	Average MSED for the nine different rules								
				1	2	3	4	5	6	7	8	9
4 a	Positive	0.05	497	44.71	41.58	44.71	41.58	44.71	44.71	44.71	41.59	41.69
	0	0.05	503	41.29	40.60	40.60	41.29	41.29	41.29	41.29	41.29	41.29
4 b	Positive	0.125	571	44.65	42.64	44.65	42.64	44.65	44.65	44.65	42.65	42.76
	0	0.125	429	41.35	40.98	40.98	41.35	41.35	41.35	41.35	41.35	41.35
4 c	Positive	0.5	797	43.83	43.03	43.83	43.03	43.83	43.83	43.83	43.06	43.39
	0	0.5	203	42.39	43.26	43.26	42.39	42.39	42.39	42.39	42.39	42.39
4 d	Positive	1.25	886	43.58	43.55	43.58	43.55	43.58	43.58	43.58	43.57	43.78
	0	1.25	114	42.38	43.95	43.95	42.38	42.38	42.38	42.38	42.38	42.38
4 e	Positive	5	4829	43.78	43.99	43.78	43.99	43.78	43.78	43.78	44.00	44.01
	0	5	171	43.92	45.96	45.96	43.92	43.92	43.92	43.92	43.92	43.92
4 f	Positive	50	999	43.89	43.93	43.89	43.93	43.89	43.89	43.89	43.91	43.88
	0	50	1	42.95	48.35	48.35	42.95	42.95	42.95	42.95	42.95	42.95

All scenarios used an α-design with four blocks of size of 12

Table 13 MSED for using different number of blocks and block sizes with a fixed number of treatments (scenario 5) depending on the value of the block variance estimate (zero or positive)

Scenario	Estimate of block variance	Number of			Average MSED for the nine different rules								
		Blocks	Replicated treat-ments/block	Simulations	1	2	3	4	5	6	7	8	9
5 a	Positive	6	8	994	72.05	71.67	72.12	71.67	72.12	72.05	72.05	71.68	72.08
5 b	Positive	8	6	998	77.02	76.44	76.66	76.44	76.66	77.02	77.02	76.44	76.50
5 c	Positive	12	4	1000	90.33	88.18	88.69	88.18	88.69	88.18	90.33	88.18	88.18
5 d	Positive	24	2	991	154.91	167.59	160.82	167.59	160.82	167.59	161.53	161.38	160.81

All scenarios used a p-rep design with a block-to-error variance ratio of 5. For simplicity, we dropped the presentation of MSED for simulation runs with zero block variance estimates for all scenarios because the number of these simulations is very low

References

Beeck CP, Cowling WA, Smith AB, Cullis BR (2010) Analysis of yield and oil from series of canola breeding trials, part I. Fitting factor analytic mixed models with pedigree information. Genome 53:992–1001

Bhattacharya CG (1998) Goodness of Yates-Rao procedure for recovery of inter-block information. Indian J Statist 60:134–144

Brown LD, Cohen A (1974) Point and confidence estimation of a common mean and recovery of interblock information. Ann Statist 2:963–976

Butler DG, Tan MK, Cullis BR (2009) Improving the accuracy of selection for late maturity α-amylase in wheat using multi-phase designs. Crop Pasture Sci 60:1202–1208

Caliński T, Kageyama S (2000) Block designs: a randomization approach. Volume I: analysis. Lecture notes in statistics 150. Springer, New York

Casella G (2008) Statistical design. Springer, New York

Crawford AC, Stefanova K, Lambe W, McLean R, Wilson R, Barclay I, Francki MG (2011) Functional relationships of phytoene synthase 1 alleles on chromosome 7A controlling flour colour variation in selected Australian wheat genotypes. Theor Appl Genet 123:95–108

Cullis BR, Smith AB, Coombes NE (2006) On the design of early generation variety trials with correlated data. J Agric Biol Environ Stat 11:381–393

Federer WT (1956) Augmented (or hoonuiaku) designs. Hawaii Plant Rec 55:191–208

Graybill FA, Deal RB (1959) Combining unbiased estimation. Biometrics 15:543–550

Hickey LT, Lawson W, Platz GJ, Dieters M, Arief VN, Germán S, Fletcher S, Park RF, Singh D, Pereyra S, Franckowiak J (2011) Mapping Rph20: a gene conferring adult plant resistance to Puccinia hordei in barley. Theor Appl Genet 123:55–68

John JA, Williams ER (1995) Cyclic and computer generated designs. Chapman and Hall, London

Kackar AN, Harville DA (1984) Approximation for standard errors of estimators of fixed and random effects in mixed linear models. J Amer Statist Assoc 79:853–862

Kenward MG, Roger JH (1997) Small-sample inference for fixed effects from restricted maximum likelihood. Biometrics 53:983–997

Longin CFH, Gowda M, Mühleisen J, Ebmeyer E, Kazman E, Schachschneider R, Schacht J, Kirchhoff M, Zhao Y, Reif JC (2013) Hybrid wheat: quantitative genetic parameters and consequences for the design of breeding programs. Theor Appl Genet 126:2791–2801

Mead R (1988) The design of experiments. Statistical principles for practical application. Cambridge University Press, Cambridge

Mead R, Gilmour SG, Mead A (2012) Statistical principles for the design of experiments: applications to real experiments. Cambridge series in statistical and probabilistic mathematics. Cambridge University Press, Cambridge

Möhring J, Piepho HP (2009) Comparison of weighting in two-stage analyses of series of experiments. Crop Sci 49:1977–1988

Möhring J, Williams ER, Piepho HP (2014) Efficiency of augmented p-rep designs in multi-environmental trials. Theor Appl Genet 127:1047–1060

Nelder JA (1954) The interpretation of negative components of variance. Biometrika 41:544–548

Patterson HD, Thompson R (1971) Recovery of inter-block information when block sizes are unequal. Biometrika 58:545–554

Patterson HD, Williams ER (1976) A new class of resolvable incomplete block designs. Biometrika 63:83–92

Piepho HP, Möhring J, Schulz-Streeck T, Ogutu JO (2012) A stagewise approach for the analysis of multi-environment trials. Biom J 54:844–860

Rebetzke GJ, Verbyla AP, Verbyla KL, Morell MK, Cavanagh CR (2014) Use of a large multiparent wheat mapping population in genomic dissection of coleoptile and seedling growth. Plant Biotechnol J 12:219–230

Robinson GK (1991) That BLUP is a good thing: estimation of random effects. Statist Sci 6:15–32

Seshadri V (1963) Constructing uniformly better estimators. J Amer Statist Assoc 58:172–175

Shrikhande SS (1951) Designs for two-way elimination of heterogeneity. Ann Math Stat 22:235–247

Smith AB, Lim P, Cullis BR (2006) The design and analysis of multiphase plant breeding experiments. J Agric Sci 144:393–409

Stein C (1956) Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. Proc. Third Berkeley Symp. Math. Statist. Prob. 1, University of California Press, Berkeley, pp 197–206

Stein C (1966) An approach to the recovery of inter-block information in balanced incomplete block designs. In: Research Papers in Statistics. Neymann Festschrift FN, David (ed) John Wiley, New York, pp 351–366

Van Eeuwijk FA (1995) Linear and bilinear models for the analysis of multi-environment trials, part I. An inventory of models. Euphytica 84:1–7

Weerakkody GJ (1992) A note on the recovery of inter-block information in balanced incomplete block designs. Comm Stat Theory Methods 21:1125–1136

Williams ER, Piepho HP, Whitaker D (2011) Augmented p-rep designs. Biom J 53:19–27

Williams ER, John JA, Whitaker D (2014) Construction of more flexible and efficient p-rep designs. Aust N Z J Stat 56:89–96

Yates F (1940) The recovery of inter-block information in balanced incomplete block designs. Ann Eugen 10:317–325