Theor Appl Genet (2014) 127:1049–1060
DOI 10.1007/s00122-014-2278-y

ORIGINAL PAPER

Efficiency of augmented p-rep designs in multi-environmental trials

Jens Moehring · Emlyn R. Williams · Hans-Peter Piepho

Received: 5 September 2013 / Accepted: 26 January 2014 / Published online: 20 February 2014

Abstract
Key message **The paper shows that unreplicated designs in multi-environmental trials are most efficient. If replication per environment is needed then augmented p-rep designs outperform augmented and replicated designs in triticale and maize.**
Abstract In plant breeding, augmented designs with unreplicated entries are frequently used for early generation testing. With limited amount of seed, this design allows to use a maximum number of environments in multi-environmental trials (METs). Check plots enable the estimation of block effects, error variances and a connection of otherwise unconnected trials in METs. Cullis et al. (J Agri Biol Environ Stat 11:381–393, 2006) propose to replace check plots from a grid-plot design by plots of replicated entries leading to partially replicated (p-rep) designs. Williams et al. (Biom J 53:19–27, 2011) apply this idea to augmented designs (augmented p-rep designs). While p-rep designs are increasingly used in METs, a comparison of the efficiency of augmented p-rep designs and augmented designs in the range between replicated and unreplicated designs in METs is lacking. We simulated genetic effects and allocated them according to these four designs to plot yields of a triticale and a maize uniformity trial. The designs varied in the number of environments, but have a fixed number of entries and total plots. The error model and the assumption of fixed or random entry effects were varied in simulations. We extended our simulation for the triticale data by including correlated entry effects which are common in genomic selection. Results show an advantage of unreplicated and augmented p-rep designs and a preference for using random entry effects, especially in case of correlated effects reflecting relationships among entries. Spatial error models had minor advantages compared to purely randomization-based models.

Communicated by I. Mackay.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-014-2278-y) contains supplementary material, which is available to authorized users.

J. Moehring · H.-P. Piepho (✉)
Institute for Crop Science, Bioinformatics Unit, University of Hohenheim, Stuttgart, Germany
e-mail: piepho@uni-hohenheim.de

E. R. Williams
Statistical Consulting Unit, The Australian National University, Canberra, ACT 0200, Australia

Introduction

Multi-environmental trials (METs) are often used in plant breeding (Smith et al. 2001, 2005; Crossa et al. 2006; Piepho et al. 2008; Viana et al. 2010; Burgueno et al. 2011) to evaluate entries under varying environmental conditions in a diverse target region. If available land, seed or financial resources are limited, which is often the case in early generations (Besag and Kempton 1986; Smith et al. 2006; Federer and Crossa 2012), the number of environments can be increased by reducing the number of replicates per environment. Federer (1956) suggested an augmented design with replicated checks and unreplicated entries within an environment. He created a replicated design for checks and augmented blocks by adding unreplicated entries. An augmented design requires only seeds for one replicate per entry and environment (Federer and Crossa 2012) and replicated checks are used for local error control (Stringer and Cullis 2002). But this design entails the danger of misadjustment, if checks show a different error variance or

otherwise perform differently from unreplicated entries (Kempton 1984). Additionally, it allocates a high number of plots to checks, which are usually well-established varieties and which are normally not directly of interest for selection (Santos et al. 2002). Therefore, Cullis et al. (2006) proposed to replace replicated checks by unreplicated entries in a grid-plot design. They denoted this design as partially replicated (p-rep) design. Williams et al. (2011) applied this idea to augmented designs, where all entries are replicated in one of the used environments and the replicated entries of each environment are randomized according to a resolvable incomplete block design. The augmented p-rep design can be seen as one possible design lying between a replicated design with two replicates and an unreplicated design. Therefore, it is reasonable to expect the performance of augmented p-rep design to also be intermediate between a replicated and an unreplicated design. But it is unknown how well it performs, especially in comparison to augmented designs. Chandra (1994) pointed out that the preference of replicated and unreplicated designs depends on the reduction of the plot error variance after adjustment based on check plot observations and therefore, e.g., on the stage of selection, the crop (Kempton 1984) and on the number of available test environments. Cullis et al. (2006) showed the superiority of p-rep designs over three versions of grid-plot designs with replicated checks for single trial analysis only. Clarke and Stefanova (2011) evaluated single trial performance of augmented designs and grid-plot designs with varying numbers of checks and plots per checks. Additionally they replaced some unreplicated entries of an augmented design to replicate other entries. They showed that the latter design is preferable under the assumption of spatial error models and that there is an optimal number of plots for replicated entries (checks or entries of interest) in single trial designs. In contrast to these evaluations in single trial analysis, the idea of p-rep designs is most often used in the context of METs (Beeck et al. 2010; Hickey et al. 2011; Crawford et al. 2011) or it is adapted to multi-trait MET analysis (Smith et al. 2011) or multi-phase MET experiments (Smith et al. 2006; Butler et al. 2009). Up to now, a comparison of the efficiency of p-rep and augmented p-rep designs in the range between the replicated and the unreplicated design in the case of MET data is lacking.

The present paper assesses the performance of augmented p-rep designs in comparison to replicated and unreplicated designs for typical MET data in early generation testing of triticale and maize. Furthermore, the performance of the commonly used augmented design is investigated. This should help plant breeders to get a feeling of the relative performance of different designs. We studied a range of different scenarios using fixed or independent random entry effects and independent or spatially correlated errors. Furthermore, we extended the comparison of independent entry effects to correlated entry effects. In plant breeding, the use of models with correlated entry effects is common in pedigree-based best linear unbiased prediction (BLUP) (Piepho et al. 2008) or genomic selection (Jannink et al. 2010). Furthermore, these models allow separating error and genetic effects in unreplicated designs and, therefore, provide new opportunities in the analysis of such designs.

Materials and methods

This paper uses simulation to compare four field trial designs resulting in varying allocations of entries within a MET. Plot errors from a triticale and a maize uniformity trial were added to simulated entry effects according to one of four randomized designs. We first describe the uniformity trials. Later we describe the simulation of genetic effects, the designs and the methods used to analyze the data and evaluate the results.

Uniformity trial

We used two uniformity trials, one with triticale (variety: SW Talentro) and one with maize (variety: Companero, early, forage maize), both with 1,080 plots ordered in 30 columns and 36 rows. The plot size was 2×3.85 m in triticale and 1.5×6 m in maize with the longer side of the plots in the direction across columns. The trials were grown in 2007 and 2008 (harvest) at the experimental station Ihinger Hof of the University of Hohenheim (Germany). All plots within a trial were treated identically using best experimental practice.

To characterize the spatial distribution of errors for the triticale uniformity trial we separated small- and large-scale error effects (Zimmerman and Harville 1991; Gilmour et al. 1997) by first fitting fixed row and column effects and then looking for additional small-scale spatial error structure. We calculated an empirical semivariogram to describe the spatial dependency of plots within columns (Journel and Huijbregts 1978), again after correcting for fixed effects of rows and columns. It should be mentioned here that the same semivariogram is produced if only row effects are previously subtracted, because column effects are automatically excluded by considering only plots within columns. Additionally, we used residual maximum likelihood (REML) (Patterson and Thompson 1971) to directly estimate the semivariance depending on the distance d between plots within columns. We used a model including both fixed effects for row and a variance–covariance structure V to estimate error variances depending on the distance. The variance of the column effect is not included because it is confounded with V and will have no influence on the

final graph. We modeled a joint variance–covariance structure (V) for all plots within a column. We assumed that this matrix is structured as a Toeplitz (diagonal-constant) matrix (see Eq. 1). The structure requires n different parameters, one for the diagonal of V and $n-1$ for the off-diagonals, where n is the number of rows within a column.

$$V = \begin{pmatrix} \sigma_1^2 & \sigma_2 & \dots & \sigma_n \\ \sigma_2 & \sigma_1^2 & \cdots & \sigma_{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_n & \sigma_{n-1} & \cdots & \sigma_1^2 \end{pmatrix} \quad (1)$$

We calculated the semivariance $\gamma(d)$ for a distance d by subtracting the estimated covariance σ_{d+1} from the estimated variance σ_1^2.

$$\gamma(d) = \sigma_1^2 - \sigma_{d+1}.$$

We did an analogous calculation for exploring the variances within a row by exchanging rows and columns. We further analyzed the triticale uniformity trial using independent or one- or two-dimensional spatial error models and fitting fixed or random effects for column, row, incomplete block, row and column or no block effects. Furthermore, we evaluated the optimal block size by comparing the AIC for models with fixed row and column effects and random effects for incomplete blocks with varying block sizes. Incomplete blocks were arranged in either rows or columns. For example, for blocks within rows and a block size of three each row is divided into 12 blocks.

Simulation of genotypic effects

We simulated two types of genetic effects. First, we drew 180 independent entry effects from a normal distribution with constant variance σ_g^2. Additionally, for triticale we drew entry effects from a multivariate normal distribution with a marker-based variance–covariance structure $Z_g Z_g' \sigma_u^2$, where Z_g is the design matrix to match entry effects g to marker effects u ($g = Z_g u$) and σ_u^2 is the variance of marker effects u. This variance–covariance structure corresponds to the ridge regression BLUP (RR-BLUP) model (Piepho 2009). We determined Z_g using DArT data from a worldwide set of 161 spring- and winter-triticale genotypes characterized by the triticale-specific DArT array used in Badea et al. (2011) and Alheit et al. (2012). To simulate 180 genotypes, we added marker data from 19 doubled haploid (DH) lines of the population of DH lines used for breeding triticale in Hohenheim (Germany). These lines were also characterized by the same DArT array. For entry-by-environment effects of each environment, we assumed the same variance–covariance structure as for the entry main effects. Entry-by-environment effects of different environment were assumed to be independent. We scaled entry and entry-by-environment effects according to long-term variance components published in Laidig et al. (2008). Because breeding trials are often performed within one year, so that entry-by-year and entry-by-year-by-environment interaction effects cannot be dissected from entry main and entry-by-environment effects, respectively, we also used the sum of both corresponding variance components. Therefore, we used two variance component ratios for each crop to simulate entry and entry-by-environment effects: for triticale we used the ratios 10.9:3.6 and 15.6:15.3, in maize we used the variance component ratios 38.7:15.6 and 47.3:33.4. Because calculation time is high especially if entry or error effects are correlated, we only performed 600 simulations for triticale in Table 3 and 135 simulations for Table 4. Out of the 600 simulations, in 537 all 24 models converged, therefore we limit our presentation of results to these. In maize, we used 500 simulations. To demonstrate that the shown results are stable, we performed an analysis of variance with the simulation-specific evaluation criteria (see "Evaluation criteria"). We used a letter display to show the significances of evaluation criteria differences between different designs.

Design

We compared four different designs. They are identical concerning the total number of plots and the number of entries. Therefore, all designs test a fixed set of entries using a fixed number of plots. The designs vary in the allocation of entries and the subdivision of the uniformity trial into environments, and thus the number of plots per entry and the size and especially the number of environments vary. Specifically, we divided the uniformity trial into three, five or six environments depending on whether an augmented p-rep design (Fig. 1a), an augmented design (Fig. 1b) (both with five environments), a fully replicated (three environments) or an unreplicated design (six environments) was chosen. The environments were further split into one or two replicates, depending on the design, and several blocks of size 12. Three blocks make up a column, and 5–10 columns constitute an environment. To create the augmented p-rep design, we used a beta version of a module for generating augmented p-rep designs in CycDesigN (VSN International; http://www.vsni.co.uk/). The chosen design was restricted to replicate each entry twice in one of all possible environments. Simulating five environments, this procedure entailed testing each entry on six plots. For the augmented design, we replaced two of four plots of replicated entries per block by plots of the first entry, which was then treated as a check. This resulted generally in two check plots per block and a design with 37 check plots and plots for 179 unreplicated entries. For the replicated and unreplicated designs, we used the same α-design (Patterson and Williams 1976) with six replicates. In the unreplicated

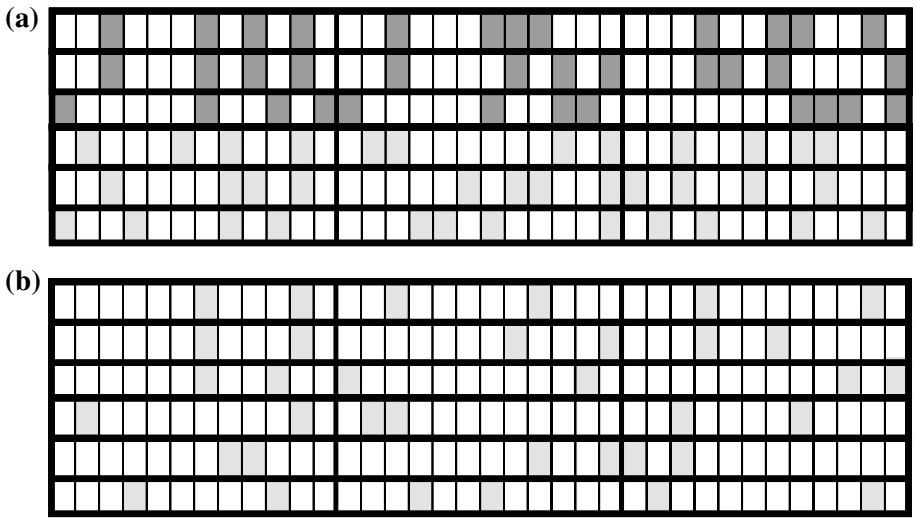

Fig. 1 Allocation of plots to replicated entries (replicate 1 *dark gray*, replicate 2 in *light gray*) in augmented p-rep designs (**a**) and allocation of check plots (*light gray*) in an augmented design (**b**) to 18 blocks of size 12 within one environment. In both layouts, the two *boldfaced vertical lines* divide each row of 36 plots into three blocks of 12 plots each, referred to in this legend as field block for clarity. The 36 replicated entries of the augmented p-rep design are allocated to the 72 *gray* plots (four *gray* plots per field block) using an α-design with two replicates and nine blocks of size four per replicate. In both designs, plots for unreplicated entries are shown in *white*

design, we took each replicate of the α-design as a different environment, whereas in the replicated design we merged two neighboring replicates to form the α-design for an environment.

Analysis

We analyzed the data according to the field design and the model of simulated entry effects using a mixed model approach. For datasets including entry effects with the marker-based variance–covariance structure, we also analyzed the data assuming independent entry effects with constant variance. Additionally, for both independent and correlated entry effects we performed analysis assuming fixed entry effects. Table 1 gives an overview of the models used for simulating and analyzing entry effects in the data. Furthermore, we used models with independent errors as well as one-dimensional first-order autoregressive (AR1) (Gilmour et al. 1997) and linear variance (LV) (Williams 1986) structures with and without nugget for plot errors within a block.

The full model accounting for replicate and incomplete block effects is given by

$$L \cdot R + L \cdot R \cdot B + G + G \cdot L, \tag{2}$$

where L, R, B, and G denote environment, replicate, incomplete block, and entry, respectively. Environment effects and replicate effects nested within environment, both represented by $(L{\cdot}R)$ are taken as fixed, while the block effect nested within replicate and environment and the entry-by-environment interaction $(G{\cdot}L)$ are taken as random. The entry main effect (G) is taken as fixed or as random. Crossing of two factors is denoted by a dot between the corresponding two factors. In our models, we took effects for environment and replicate as fixed, thus ignoring inter-environment and inter-replicate information (Piepho and

Table 1 Overview of six models used to generate genetic effects and to analyze the simulated data. Each row in the table below the header line corresponds to one model

Simulated entry and entry-environmental effects are	Entry effects analyzed as			Entry-by-environment effects analyzed as	
	Fixed	Random			
		Independent	Correlated, using markers	Independent	Correlated, using markers
Independent	✓			✓	
Independent		✓		✓	
Correlated, using markers	✓			✓	
Correlated, using markers		✓		✓	
Correlated, using markers	✓				✓
Correlated, using markers			✓		✓

All analysis models were combined with several spatial models

Möhring 2006). Additionally, we assumed homogeneous variances for blocks and errors across environments. This is in contrast to the common procedure in analysis of METs, where heterogeneous variances are assumed (Smith et al. 2001; Möhring and Piepho 2009). But in our case, different simulated environments were based on plot errors of the same uniformity trial. Therefore, we assumed homogeneous variances for both block and error variances. For random entry main effects it was assumed that effects follow a normal distribution with variance–covariance matrix $I_g\sigma_g^2$ or $Z_gZ'_g\sigma_u^2$, where I_g is an identity matrix of size g and Z_g is the design matrix to match entry g to marker effects u ($g = Z_gu$). The latter assumption corresponds to the ridge regression BLUP (RR-BLUP) model (Piepho 2009), in which $Z_gZ'_g$ is proportional to the assumed marker-based variances–covariances matrix between genotypes. This model is easy to implement and usually performs equally well as other models assuming other distributions for marker effects [e.g., a t-distribution (BayesA), or a mixture of normal distributions (BayesB); (Guo et al. 2012)]. For random entry-by-environment interaction effects, the Kronecker product of I_l and $I_g\sigma_g^2$ or $Z_gZ'_g\sigma_u^2$ is assumed, where I_l is an identity matrix of size l corresponding to the number of environments. For the unreplicated design, there is a single replicate per environment, so R is dropped from Eq. (2). Additionally, for the unreplicated design and assuming either fixed entry effects or independent random entry effects, the entry-by-environment interaction and an independent error are totally confounded; therefore the interaction is dropped from the model. All analyses are performed using SAS version 9.3 (SAS Institute 2011).

Evaluation criteria

For all models, we calculated the mean square error of a difference (MSED) between estimated and true simulated effect. Additionally, for fixed entry effects, we calculated the mean standard error of differences (s.e.d.), where the s.e.d. were estimated based on the fitted model. These values are denoted as *pre* and *emp*, respectively, in several papers (Besag and Kempton 1986; Zimmerman and Harville 1991; Wu et al. 1998; Wu and Dutilleul 1999). If the fitted model is valid, the square of the s.e.d. had the same expectation as the MSED. We further calculated the mean of the true simulated effects for the estimated best 18 entries as a measure of the realized selection gain. Taking the best 18 entries corresponds to a selection fraction of 10 %.

Results

We first report the main results for the analysis of the plot data of the triticale uniformity trial. This shows which error structure can be expected to fit well in our simulations. A more detailed report is found in the Online resource. Later, we show results from our simulations.

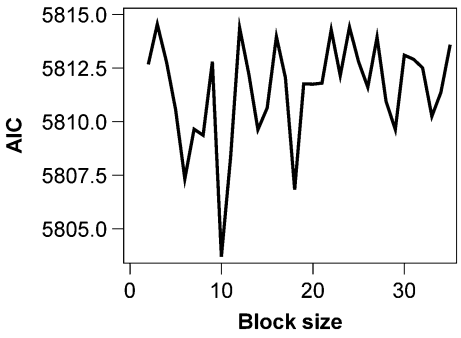

Fig. 2 AIC depending on block size using models with fixed column and row effects and a random effect for blocks. Blocks are arranged in direction of columns with a maximum block size equal to the number of plots per column (36)

Uniformity trial

The triticale uniformity trial can be characterized by an average yield level of 53.8 dt ha^{-1} (standard error 5.3 dt ha^{-1}). As our aim is to compare field designs requiring some blocking structure, we first searched for an adequate block size. We fitted models including fixed effects for row and column and random effects for incomplete blocks with varying block size. Figure 2 shows that the minimal AIC and, therefore, optimal block size within columns is 10. Presumably, the spiky plot just represents sampling variation. With 36 plots per column, a block size of 10 results in blocks including plots of different columns. We therefore choose a block size of 12, which divides a column into three blocks. Additionally, in models without blocks, fitting row effects are less important (see results in Table S1 in Online resource). This indicates that blocking in the direction of plots with common long sides is preferable, which is in accordance with the way cereal breeders normally choose blocks.

Figure 3 shows the empirical and model-based semivariance as related to the distance between plots within a column. In computing the empirical semivariance, row and column effects are first estimated and eliminated. The model-based approach fits both the semivariance and fixed row effects in one model using the Toeplitz structure described in (1). Both approaches describe the variability within the field and result in nearly identical curves. Both curves show an increasing semivariance with increasing distance and a clear non-zero semivariance for small distances. This indicates some spatial variability and the need for a nugget effect. The lower slope of the semivariance

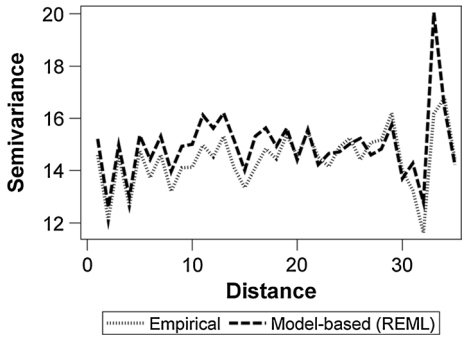

Fig. 3 Empirical and model-based semivariance in dependence of the distance between plots within a column. For the empirical semivariogram fixed row effects are previously removed. For the model-based semivariogram fixed row effects are included in the model

within columns compared to the variance within rows (data not shown) shows that incomplete blocks should be arranged in direction of columns.

In Table 2, AIC (Akaike 1974) values are given for some models used for analyzing the uniformity trial. Again, further models were used and their AIC values are given in Table S1 in the Online resource. While both the use of blocks and the use of spatial error models within blocks decreased the AIC, the best model fit is obtained using a row–column model and a two-dimensional spatial linear variance error model for the complete trial (Williams et al. 2006).

Our results suggest the following:

1. If some kind of incomplete blocking is required, e.g., if augmented or augmented p-rep designs are of interest, then relatively large blocks within columns are preferred. For practical reasons, we chose a block size of 12 plots.
2. Two-dimensional spatial error models outperform one-dimensional spatial error models.
3. A nugget effect most often improved model fit.
4. There is a slight preference for using a linear variance or a one-dimensional autoregressive error structure, both with nugget, compared to other possible error structures.

As our interest is in comparing block designs (augmented design, augmented p-rep design), we analyzed our simulated data using models for designs with incomplete blocks and a block size of 12 and, thus, with either independent or one-dimensional spatial error models within a block. We concentrated on linear variance and the often-used first-order autoregressive model.

Results from simulation

We will discuss the results of our simulation in four parts: (1) the influence of assuming random or fixed entry effects in the analysis, (2) the influence of correlated entry effects as compared to independent entry effects in simulating and analyzing data, (3) the influence of spatial error models, and finally, (4) the influence of the four field designs. Because of most often comparable results for both crops and all variance component ratios, we only report results from one crop and one variance component ratio. The used crop and variance component ratio in question is given in the table or figure.

1. When entry effects simulated as independent were analyzed, no clear preference of assuming fixed or random entry effects was found for correlation between true and estimated entry effects and the realized selection gain (Tables 3, 5). In case of simulating and analyzing correlated entry effects, there was an advantage for BLUP compared to best linear unbiased estimation (BLUE). This is expected, as in this case entries are correlated and, therefore, the value of any one entry depends partly on the value of every other correlated entry. Additionally, and not unexpectedly, because BLUP maximizes the probability of correct ranking of entries for known variance components (Searle et al. 1992), BLUP outperformed BLUE for the evaluation criteria MSED, although in the simulations variance components were estimated from the data.
2. A comparison of correlations between different approaches for simulating entry effects is not directly possible, because they are based on differently simulated true effects. Nevertheless, the correlation between true and estimated entry effects using either inde-

Table 2 AIC values for models with varying block structure and error model; for two-dimensional spatial error structures all observations are correlated, for one-dimensional error structure the correlation was limited to observations within a block

Number of dimensions of spatial error model	Random design effects in the baseline model	Added spatial error model			
		None	AR1	AR1 + nugget	LV + nugget
2	None	6,642.69	6,305.33	6,056.85	6,057.15
1	Blocks within column	6,312.46	6,314.16	6,316.15	6,305.36
2	Column + row	6,120.83	6,102.05	6,054.85	6,051.14

AR1 first-order autoregressive, *LV* linear variance Williams 1986)

Table 3 Four evaluation criteria for using different models for simulating and analyzing data

Evaluation criterion	Model assumed for			Design			
	Simulation of entry and entry-by-environment effects	Analysis of		Replicated	Augmented	Augmented p-rep	Unreplicated
		Entry effects	Entry-by-environment effects				
Correlation	Independent	Fixed	Independent	0.8481[b]	0.8434[a]	0.8588[c]	0.8651[d]
Correlation	Independent	Independent	Independent	0.8482[b]	0.8438[a]	0.8590[c]	0.8652[d]
Correlation	Correlated	Fixed	Independent	0.7955[b]	0.7830[a]	0.8038[c]	0.8103[d]
Correlation	Correlated	Independent	Independent	0.7957[b]	0.7846[a]	0.8047[c]	0.8105[d]
Correlation	Correlated	Fixed	Correlated	0.7956[b]	0.7831[a]	0.8045[c]	0.8104[d]
Correlation	Correlated	Correlated	Correlated	0.8373[a]	0.8377[a]	0.8500[b]	0.8556[c]
Selection gain	Independent	Fixed	Independent	4.739[a]	4.750[a]	4.817[a]	4.826[a]
Selection gain	Independent	Independent	Independent	4.734[a]	4.752[a]	4.821[a]	4.827[a]
Selection gain	Correlated	Fixed	Independent	3.504[a]	3.472[a]	3.542[a]	3.566[a]
Selection gain	Correlated	Independent	Independent	3.505[a]	3.480[a]	3.553[a]	3.567[a]
Selection gain	Correlated	Fixed	Correlated	3.505[ab]	3.470[a]	3.540[bc]	3.568[cd]
Selection gain	Correlated	Correlated	Correlated	3.662[ab]	3.660[a]	3.710[bc]	3.733[cd]
MSED	Independent	Fixed	Independent	4.279[b]	4.433[a]	3.897[c]	3.694[d]
MSED	Independent	Independent	Independent	3.115[b]	3.192[a]	2.908[c]	2.792[d]
MSED	Correlated	Fixed	Independent	5.473[b]	5.805[a]	5.253[bc]	5.113[c]
MSED	Correlated	Independent	Independent	4.133[b]	4.232[a]	4.017[bc]	3.967[c]
MSED	Correlated	Fixed	Correlated	5.469[b]	5.914[a]	5.234[bc]	5.109[c]
MSED	Correlated	Correlated	Correlated	2.334[a]	2.314[a]	2.173[a]	2.063[a]
S.e.d.	Independent	Fixed	Independent	2.923[b]	2.954[a]	2.763[c]	2.713[d]
S.e.d.	Correlated	Fixed	Independent	2.766[b]	2.875[a]	2.665[c]	2.629[d]
S.e.d.	Correlated	Fixed	Independent	2.762[b]	2.859[a]	2.658[c]	2.627[d]

The evaluation criteria are the correlation of simulated and estimated entry effects, the mean square error of a difference between simulated and estimated entry effects (MSED), model-based estimated standard error of a difference between simulated and estimated entry effects (s.e.d.) and selection gain measured as average true value of the 18 best-estimated entries

Variance component ratio entry to entry-by-environment: 10.9:3.6, $n = 537$

Values in a row followed by a common letter are not significantly different by a t test ($\alpha = 0.05$)

pendent or correlated entry effects in simulation and analysis was quite similar, because we scaled them using long-term variance components. The correlation decreased if correlated entry effects were simulated but fixed or independent entry main effects were used in analysis.

3. In contrast to the clear advantage of BLUP with correlated entry effects, the advantage of spatial error models for estimating entry main effects was limited (Table 3). Additionally, the computing time increased and the probability of convergence was reduced, especially for models assuming fixed entry main effects and an autoregressive error structure.
4. In Tables 3, 4 and 5, a clear ranking of the four designs for both crops with both variance component ratios is observed. For the highest ratio of entry and entry-by-environment variance (triticale with a ratio of 10.9:3.6), the replicated design and the augmented design showed a similar performance. Therefore, seeing the entries twice in each of three locations (in total six observations) in the replicated design approximately compensated the advantage of seeing entries once in each of five locations in the augmented design. For the other ratios, the relative performance of the augmented design increased with decreasing ratio. For all ratios, both designs were worse than the augmented p-rep design and the unreplicated design. The unreplicated design showed the best performance according to all evaluation criteria and for all simulated crops and all variance component ratios. The augmented p-rep design showed a clearly better performance than the augmented design. While the evaluation criteria of the augmented p-rep design were slightly better than expected from the linear interpolation between replicated and unreplicated design, the augmented design was always worse (results not shown).

Table 4 Correlation between simulated and estimated entry effects for different spatial error models and different designs

Spatial error model	Entry main effect in the analysis	Design			
		Replicated	Augmented	Augmented p-rep	Unreplicated
Independent	Fixed	0.7895[a] (135)	0.8189[b] (135)	0.8284[c] (135)	0.8424[d] (135)
First-order autoregressive plus nugget	Fixed	0.7924[a] (67)	0.8185[b] (67)	0.8284[c] (75)	0.8405[d] (77)
Linear variance plus nugget	Fixed	0.7883[a] (105)	0.8192[b] (110)	0.8276[c] (108)	0.8430[d] (130)
Independent	Independent	0.8046[a] (135)	0.8311[b] (135)	0.8398[c] (135)	0.8524[d] (135)
First-order autoregressive plus nugget	Independent	0.8044[a] (135)	0.8318[b] (135)	0.8396[c] (132)	0.8520[d] (135)
Linear variance plus nugget	Independent	0.8050[a] (128)	0.8313[b] (127)	0.8397[c] (127)	0.8525[d] (134)

Additionally, the number of converging simulation runs is given in parentheses

Data are simulated with independent entry and entry-by-environment effects, entry-by-environment effects in the analysis are taken as independent

Variance component ratio entry to entry-by-environment: 15.6:15.3; $n = 135$

Values in a row followed by a common letter are not significantly different by a t test ($\alpha = 0.05$)

Table 5 Estimated correlation for using different models for analyzing triticale and maize data with varying variance component ratios, $n = 537$ for ratio 10.9:3.6, $n = 135$ for ratio 15.6:15.3, and $n = 500$ for maize

Crop	Analysis of entry effects	Variance component ratio	Design			
			Replicated	Augmented	Augmented p-rep	Unreplicated
Triticale	Fixed	10.9:3.6	0.8481[b]	0.8434[a]	0.8588[c]	0.8651[d]
Triticale	Independent	10.9:3.6	0.8482[b]	0.8438[a]	0.8590[c]	0.8652[d]
Maize	Fixed	38.7:15.6	0.8440[a]	0.8437[a]	0.8588[b]	0.8651[c]
Maize	Independent	38.7:15.6	0.8440[a]	0.8441[a]	0.8588[b]	0.8652[c]
Maize	Fixed	42.3:33.4	0.8282[a]	0.8414[b]	0.8526[c]	0.8645[d]
Maize	Independent	42.3:33.4	0.8282[a]	0.8419[b]	0.8526[c]	0.8646[d]
Triticale	Fixed	15.6:15.3	0.7895[a]	0.8189[b]	0.8284[c]	0.8424[d]
Triticale	Independent	15.6:15.3	0.8046[a]	0.8311[b]	0.8398[c]	0.8524[d]

Values in a row followed by a common letter are not significantly different by a t test ($\alpha = 0.05$)

Discussion

This paper studies the efficiency of augmented p-rep designs in the range between replicated and unreplicated designs, especially compared to an augmented design. While unreplicated designs fit best for both crops and all simulated variance component ratios, the augmented p-rep design is just slightly inferior. The augmented design and the replicated design have a lower efficiency. We chose simulations to study the efficiency of these four designs for a range of different conditions. The disadvantage of simulation is that its results depend on the validity and usefulness of the underlying assumption for performing the simulation. As described, our simulations were mainly based on five inputs: a fixed total number of plots and entries, long-term variance components, marker data from a broad breeding population, plot errors of a uniformity trial and finally the chosen models for simulating and analyzing the data. If variance components or the amount of marker-based correlation between entries change, e.g., if a closer or more homogeneous or diverse breeding population is used, this will affect the results of our simulations. To account for this uncertainty, we varied our variance components ratio from about 3:1 (10.9:3.6) to 1:1 (15.6:15.3; entry to entry-by-environment).

Taking plot errors of uniformity trials allow conclusions about the performance of different spatial models without directly assuming a special model for simulation. Uniformity trials are mentioned in Fisher (1926) for proving the validity of estimation in field experiments. They are frequently used to study the variation within a field (Wiebke 1935), and to find the optimal plot size, the optimal plot arrangements or the best spatial model (Williams 1986; Richter and Kroschewski 2012). A more critical point is the assumed model for both simulating and analyzing data. We chose an additive model, thus we assumed that the performance of an entry in a special environment is the sum of the genetic effect, the environmental effect and the entry-by-environment effect. This can easily be implemented in standard statistical software. But there exist other models

for describing the effect of entry and entry-environmental effects, e.g., multiplicative models (Gollob 1968; Gabriel 1978). Additionally, we simulated normally distributed entry effects, thus assuming random entry effects. Therefore, a slight preference for BLUP in our simulations can be expected. Our results, showing an advantage of BLUP, are comparable to empirical results found, e.g., in Hill and Rosenberger (1985), Stroup and Mulitze (1991), Stanek et al. (1999) and Kleinknecht et al. (2013).

Fixing the total number of plots and entries

Our designs used a fixed number of plots and entries, but varied in the number of environments. As testing entries normally result in costs for managing plots and fixed costs per environment, the costs for our four designs are not exactly identical, though according to collaborating breeders the variable cost per plot by far dominates the fixed cost per environment. An alternative is to fix the costs. This requires an unequal number of plots assigning more plots to designs with a smaller number of environments. Therefore, fixing the cost will indirectly improve our evaluation criteria for the replicated design and deteriorate them for the unreplicated design compared to the augmented or augmented p-rep design. We ignored varying costs in our simulation. For the case of fixed number of the total number of plots and entries, Talbot et al. (1984) showed a preference for maximizing the number of environments while decreasing number of replicates per environment, which is in accordance with our findings. But the unreplicated design also has some disadvantages. Because entries are unreplicated per environment, a single-environment analysis is not possible. However, many breeders require a speedy analysis even of single environments for quick and early selection decisions, therefore it is often crucial that individual trials can be analyzed separately without having to wait for data from all environments to come in. Furthermore, no two-stage analysis is possible if entry effects are assumed as fixed or if both a random and uncorrelated error and entry effect are assumed. Additionally, in these cases a separation of entry-by-environment and error effects is impossible, which has two main consequences. On the one hand, no test for entry-by-environment interaction is possible, but in plant breeding this can usually be accepted as the existence of entry-by-environment interaction is expected a priori. On the other hand, if error variances are heterogeneous, this confounding inflates all error variance estimates by adding the entry-by-environment interaction variance and finally results in a different weighting of entry-by-environment means. A different weighting of entry-by-environment means probably decreases the correlation between true and estimated entry effects, although in analyses of plant breeding METs, Möhring and Piepho (2009) demonstrated that ignoring weights often lead to acceptable results. Using models with spatially correlated errors without nugget or using random and correlated entry effects allows an analysis of single environments. While the assumption of no nugget is often problematic, correlated entry effects are common in genomic selection. Therefore, the use of genomic selection offsets some of the disadvantages of unreplicated designs.

Fixed or random environmental effects

There is an ongoing discussion whether the factor 'environment' in the analysis of plant breeding METs should be assumed as random or fixed (Smith et al. 2001). Field trials are normally conducted at fields of experimental stations, and therefore are not drawn at random from the population of all fields within a region, where the crop can potentially be grown. Nevertheless, it is often assumed that fields at experimental stations are representative for a region. Clearly, random sampling of environments from a target region is desirable for an unbiased estimate of entry means in the target.

Assuming that environments are a random sample, taking environmental main effects as fixed but entry-by-environmental effects as random in the analysis ignores inter-environment information, but the loss of information is normally small (Piepho and Möhring 2006) because environment main effects are often large compared to entry effects. Taking them as random requires the estimation of a variance component, and therefore about 10 degrees of freedom for an adequate precision of this variance component (Mead et al. 2012). In our case there are only three to six environments, therefore we took environment main effects as fixed in our analyses.

Reasons for using an incomplete block design

Williams and Luckett (1988) concluded that their uniformity trials in barley and cotton support a two-dimensional blocking structure with row and column effects or a two-dimensional spatial error model (Cullis and Gleeson 1991). This is in accordance with our findings where a model including row and column effects and two-dimensional spatial errors fits best for the uniformity trial data. Nevertheless, using a block effect and one-dimensional spatial error models also improved the model fit compared to a model with just an independent error term. Furthermore, using incomplete blocks with long thin plots within a block is common in cereal breeding (Patterson and Hunter 1983). Therefore, we decided to use an incomplete block design in our simulations.

Stage-wise analysis

Especially in routine analysis of large series of METs, it is common to use a stage-wise approach, where adjusted means are estimated for each trial or environment (Möhring and Piepho 2009; Welham et al. 2010; Piepho et al. 2012). These means are then summarized in the second stage to estimate means across trials or environments. In genomic selection, a further stage for estimating genetic breeding values can be added (Schulz-Streeck et al. 2013). The advantage of stage-wise analysis is the option to easily account for specifics of the design and error structure for each environment in the first stage (Mathews et al. 2008). Additionally, it normally speeds up calculation. Unfortunately, a stage-wise approach for the unreplicated design is impossible if error effects are assumed as independent or as correlated including a nugget and entry effects are assumed as fixed or as random independent effects. We therefore used single-stage analysis assuming homogeneous block and error variances for our simulations.

For correlated entry effects, which are common in pedigree-based analysis or genomic selection, we tried out a two-stage analysis for all designs. Such an analysis assumes environmental-specific block and error variances and, therefore, allows a weighting of environments by their precision. As in the case of spatial statistics, the probability of convergence decreases, but results from two-stage analysis vary only slightly from our results using a one-stage approach (data not shown). Larger differences can be expected, if there is a real heterogeneity of variances in the data. In our case we used a single uniformity trial; therefore, no heterogeneity of variances was expected.

Number of replicated plots

As discussed above, the augmented p-rep design can be seen as a design intermediate between a replicated and an unreplicated design. If the percentage of replicated entries is reduced, the design becomes similar to the unreplicated design, whereas if the number is increased, the design becomes more similar to the replicated design. In our case, we replicated 20 % of the entries in each environment. For the settings studied in this paper, this corresponds to using 17 % of the plots for checks in augmented designs. Fisher (1926) proposed to use the square root of the number of entries as the number of check plots in augmented designs. Thus, about 14 plots or 8 % of the plots should be assigned to checks. For p-rep designs the number of additional plots is often higher: Beeck et al. (2010) replicated 20–60 % of the entries, Payne (2006) 19 % and Hickey et al. (2011) 70–90 %. 20 % plots for checks are also common in augmented designs (Mathews et al. 2008; Kehel et al. 2010).

Number of check varieties

In our experience from Germany, plant breeders most often conduct early generation trials with two to four check varieties, which are replicated with equal or varying frequencies. We simulated the most extreme case of using just one check variety for all replicated plots. The other extreme is to replicate each check twice, which maximizes the number of check varieties. The latter allocation is identical with a p-rep design. Therefore it can be expected, that the efficiency of augmented designs increases as the number of checks increases and the number of plots per check decreases. Additionally, breeders usually do not consider checks for selection. Because we fixed the number of entries and the number of plots used in our simulations, we used all entries for calculation of evaluation criteria. Therefore, the first entry is always included, irrespectively of whether it was used as check (augmented design) or normal entry (otherwise). Excluding the check variety from calculation of evaluation criteria would further decrease the relative efficiency of the augmented design, because this design has the highest number of plots of the check variety. Additionally, excluding this entry leads to an unequal number of data points used in different designs.

Probability of convergence

We observed convergence problems in analysis, especially when we fitted fixed entry effects and assumed an autoregressive error structure. This is in accordance to Robbins et al. (2012), Clarke and Stefanova (2011) and Piepho et al. (2013). They mentioned convergence problems when using autoregressive variance models with several fixed and random effects or if the autocorrelation parameter was near unity. In our case of fitting spatial error structures within blocks, an autocorrelation near unity implies fit of identical spatial effects for all plots within a block. This is identical to fitting a block effect across the range of plots assumed to be spatially correlated (Piepho et al. 2008). Therefore, the autocorrelation parameter can be confounded with other effects in the model.

Spatial error models

We detected minor effects of the chosen spatial model on the estimated main entry effects. All simulations were based on just two uniformity trials. Using other uniformity trials may show an advantage of spatial models, which is reported, e.g., in Wu et al. (1998) or Müller et al. (2012). But our result shows that using a randomization-based approach for modeling MET as baseline model and using spatial error structures as add-on option (Williams

1986; Stefanova et al. 2009; Beeck et al. 2010) can be advantageous.

Conclusion

Unreplicated designs showed the best values of all evaluation criteria in all simulated scenarios. Nevertheless, augmented p-rep designs can be advantageous, e.g., if there is (1) interest in results of single trial analysis or (2) a stage-wise analysis with weighting of individual environments according to their precision is preferred. Both designs clearly outperform replicated and classical augmented designs.

Acknowledgments Jens Möhring was supported by DFG Grant PI 377/13-1. We thank two anonymous reviewers for their helpful comments. The DArT marker data were created within research project 0315414A, funded by the Federal Ministry of Education and Research (BMBF), Germany.

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

References

Akaike H (1974) New look at the statistical model identification. IEEE Trans Automat Contr AC 19:716–723

Alheit KV, Maurer HP, Reif JC, Matthew RT, Hahn V, Weissmann EA, Würschum T (2012) Genome-wide evaluation of genetic diversity and linkage disequilibrium in winter and spring triticale (X *Triticosecale Wittmack*). BMC Genom 13:235

Badea A, Eudes F, Salmon D, Tuvesson S, Vrolijk A, Larsson C-T, Caig V, Huttner E, Kilian A, Laroche A (2011) Development and assessment of DArT markers in triticale. Theor Appl Genet 122:1547–1560

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

Besag J, Kempton R (1986) Statistical analysis of field experiments using neighbouring plots. Biometrics 42:231–251

Burgueno J, Crossa J, Cotes JM, Vicente FS, Das B (2011) Prediction assessment of linear mixed models for multienvironment trials. Crop Sci 51:944–954

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

Chandra S (1994) Efficiency of check plot designs in unreplicated field trials. Theor Appl Genet 88:618–620

Clarke GPY, Stefanova KT (2011) Optimal design for early-generation plant-breeding trials with unreplicated or partially replicated test lines. Aust NZ J Stat 53:461–480

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

Crossa J, Burgueno J, Cornelius PL, McLaren G, Trethowan R, Krishnamachari A (2006) Modelling genotype × environment interaction using additive genetic covariances of relatives for predicting breeding values of wheat genotypes. Crop Sci 46:1722–1733

Cullis BR, Gleeson AC (1991) Spatial analysis of field experiments—an extension to two dimensions. Biometrics 47:1449–1460

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

Federer WT, Crossa J (2012) I.4 screening experimental designs for quantitative trait loci, association mapping, genotype-by environment interaction, and other investigations. Front Physiol 3:156

Fisher RA (1926) The arrangement of field experiments. J Min Agric G Br 33:503–513

Gabriel KR (1978) Least squares approximation of matrices by additive and multiplicative model. J R Statist S Ser B 40: 186–196

Gilmour AR, Cullis BR, Verbyla AP (1997) Accounting for natural and extraneous variation in the analysis of field experiments. J Agric Biol Environ Stat 2:269–293

Gollob HF (1968) A statistical model that combines features of factor analysis and analysis of variance techniques. Psychometrika 33:73–115

Guo Z, Tucker DM, Lu J, Kishore V, Gay G (2012) Evaluation of genome-wide selection efficiency in maize nested association mapping populations. Theor Appl Genet 124:261–275

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

Hill RR, Rosenberger JL (1985) Methods of combining data from germplams evaluation trials. Crop Sci 25:467–470

Jannink J-L, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. Brief Funct Genomics 9:166–177

Journel AG, Huijbregts CJ (1978) Mining geostatistics. Academic Press, London

Kehel Z, Habash DZ, Gezan SA, Welham SJ, Nachit MM (2010) Estimation of spatial trend and automatic model selection in augmented designs. Agron J 102:1542–1552

Kempton RA (1984) The design and analysis of unreplicated field trials. Vortr Pflanzenzuecht 7:219–242

Kleinknecht K, Möhring J, Singh KP, Zaidi PH, Atlin GN, Piepho H-P (2013) Comparison of the performance of BLUE and BLUP of genotype effects from zoned Indian maize data. Crop Sci 53:1384–1391

Laidig F, Drobek T, Meyer U (2008) Genotypic and environmental variability of yield for cultivars from 30 different crops in German official variety trials. Plant Breed 127:451–457

Mathews KL, Marosetti M, Chapman S, McIntyre L, Reynolds M, Shorter R, van Eeuwijk FA (2008) Multi-environment QTL mixed models for drought stress adaptation in wheat. Theor Appl Genet 117:1077–1091

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, New York

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

Müller BU, Schützenmeister A, Piepho H-P (2012) Arrangement of check plots in augmented block designs when spatial analysis is used. Plant Breed 129:581–589

Patterson HD, Hunter EA (1983) The efficiency of incomplete block designs in national list and recommended list cereal variety trials. J Agric Sci 101:427–433

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
Payne RW (2006) New and traditional methods for the analysis of unreplicated experiments. Crop Sci 46:2476–2481
Piepho H-P (2009) Ridge regression and extensions for genomewide selection in maize. Crop Sci 49:1165–1176
Piepho H-P, Möhring J (2006) Selection in cultivar trials—is it ignorable? Crop Sci 46:192–201
Piepho H-P, Möhring J, Melchinger AE, Büchse A (2008) BLUP for phenotypic selection in plant breeding and variety testing. Euphytica 161:209–228
Piepho H-P, Möhring J, Schulz-Streeck T, Ogutu JO (2012) A stage-wise approach for the analysis of multi-environment trials. Biom J 54:844–860
Piepho H-P, Möhring J, Williams ER (2013) Why randomize agricultural experiments? J Agron Crop Sci 199:374–383
Richter C, Kroschewski B (2012) Geostatistical models in agricultural field experiments: investigations based on uniformity trials. Agron J 104:91–105
Robbins KR, Backlund JE, Schnelle KD (2012) Spatial correlations of unreplicated trials using a two-dimensional spline. Crop Sci 52:1138–1144
Santos AH, Bearzoti E, Ferreira DF, da Silva Filho JL (2002) Simulation of mixed models in augmented design. Sci Agric 59:483–489
SAS Institute Inc. (2011) Base SAS® 9.3 Procedures Guide. Cary, NC: SAS Institute Inc
Schulz-Streeck T, Ogutu JO, Piepho H-P (2013) Comparisons of single-stage and two-stage approaches to genomic selection. Theor Appl Genet 126:69–82
Searle SR, Casella G, McCulloch CE (1992) Variance components. Wiley, New York, pp 268–269
Smith AB, Cullis BR, Gilmour AR (2001) Analyzing variety by environment trials using multiplicative mixed models and adjustments for spatial field trend. Biometrics 57:1138–1147
Smith AB, Cullis BR, Thompson R (2005) The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. J Agric Sci 143:449–462
Smith AB, Lim P, Cullis BR (2006) The design and analysis of multiphase plant breeding experiments. J Agric Sci 144:393–409
Smith AB, Thompson R, Butler DG, Cullis BR (2011) The design and analysis of variety trials using mixtures of composite and individual plot samples. J R Stat S Ser C (Appl Stat) 60:437–455
Stanek EJ III, Well A, Ockene I (1999) Why not routinely use best linear unbiased predictors (BLUPs) as estimates of cholesterol, per cent fat from kcal and physical activity? Stat Med 18:2943–2959
Stefanova KT, Smith AB, Cullis BR (2009) Enhanced diagnostics for the spatial analysis of field trials. J Agric Biol Environ Stat 14:392–410
Stringer JK, Cullis BR (2002) Application of spatial analysis techniques to adjust for fertility trends and identify interplot competition in early stage sugarcane selection trials. Aust J Agric Res 53:911–918
Stroup WW, Mulitze DK (1991) Nearest neighbor adjusted best linear unbiased prediction. Am Stat 45:194–200
Talbot M (1984) Yield variability of crop varieties in the UK. J Agric Sci Cambridge 102:315–321
Viana JMS, Sobreira FM, De Resende MDV, Faria VR (2010) Multitrait BLUP in half-sib selection of annual crops. Plant Breed 129:599–604
Welham SJ, Gogel BJ, Smith AB, Thompson R, Cullis BR (2010) A comparison of analysis methods for late-stage variety evaluation trials. Aust NZ J Stat 52:125–149
Wiebe GA (1935) Variation and correlation in grain yield among 1500 wheat nursery plots. J Agric Res 50:331–357
Williams ER (1986) Neighbour analysis of uniformity data. Aust J Stat 28:182–191
Williams ER, Luckett DJ (1988) The use of uniformity data in the design and analysis of cotton and barley variety trials. Aust J Agric Res 39:339–350
Williams ER, John JA, Whitaker D (2006) Construction of resolvable spatial row–column designs. Biometrics 62:103–108
Williams ER, Piepho H-P, Whitaker D (2011) Augmented p-rep designs. Biom J 53:19–27
Wu T, Dutilleul P (1999) Validity and efficiency of neighbor analyses in comparison with classical complete and incomplete block analyses of field experiments. Agron J 91:721–731
Wu T, Mather DE, Dutilleul P (1998) Application of geostatistical and neighbor analyses to data from plant breeding trial. Crop Sci 38:1533–1545
Zimmerman DL, Harville DA (1991) A random field approach to the analysis of field-plot experiments and other spatial experiments. Biometrics 47:223–239