Published December 13, 2018

RESEARCH

Can Spatial Modeling Substitute for Experimental Design in Agricultural Experiments?

Alejandra Borges, Agustín González-Reymundez, Oswaldo Ernst, Mónica Cadenazzi, José Terra, and Lucía Gutiérrez*

ABSTRACT

One of the most critical aspects of agricultural experimentation is the proper choice of experimental design to control field heterogeneity, especially for large experiments. However, even with complex experimental designs, spatial variability may not be properly controlled if it occurs at scales smaller than blocks. Therefore, modeling spatial variability can be beneficial, and some studies even propose spatial modeling instead of experimental design. Our goal was to evaluate the effects of experimental design, spatial modeling, and a combination of both under real field conditions using GIS and simulating experiments. Yield data from cultivars was simulated using real spatial variability from a large uniformity trial of 100 independent locations and different sizes of experiments for four experimental designs: completely randomized design (CRD), randomized complete block design (RCBD), α-lattice incomplete block design (ALPHA), and partially replicated design (PREP). Each realization was analyzed using different levels of spatial correction. Models were compared by precision, accuracy, and the recovery of superior genotypes. For moderate and large experiment sizes, ALPHA was the best experimental design in terms of precision and accuracy. In most situations, models that included spatial correlation were better than models with no spatial correlation, but they did not outperformed better experimental designs. Therefore, spatial modeling is not a substitute for good experimental design.

A. Borges, A. González-Reymundez, M. Cadenazzi, and L. Gutiérrez, Statistics Dep., Facultad de Agronomía, Univ. de la República, Garzón 780, Montevideo, Uruguay; A. González-Reymundez, current address, Dep. of Epidemiology and Biostatistics, Michigan State Univ. 909 Fee Rd., B601 West Fee Hall, East Lansing, 48823 MI; O. Ernst, Dep. of Crop Production, Facultad de Agronomía, Univ. de la República, Estación Experimental Mario Cassinoni, Paysandú, Uruguay; J. Terra, Estación Experimental Treinta y Tres, Instituto Nacional de Investigación Agropecuaria (INIA), Ruta 8, Km 281, Treinta y Tres 33000, Uruguay; L. Gutiérrez, Dep. of Agronomy, Univ. of Wisconsin–Madison, 1575 Linden Dr., Madison, WI 53706. Received 13 Mar. 2018. Accepted 3 Oct. 2018. *Corresponding author (gutierrezcha@wisc.edu). Assigned to Associate Editor Jode Edwards.

Abbreviations: ALPHA, α-lattice incomplete block experimental design; AR(1), spatially correlated error model with one-dimensional autoregressive process; Best_Gen, proportion of times the true 15% superior genotypes are recovered; COR, Pearson's correlation coefficient between true and estimated effects; CRD, completely randomized experimental design; EXP(2), spatially correlated error model with two-dimensional exponential process; MSEP, mean square error of prediction; NSC, no spatial correction model; PREP, partially replicated experimental design; $PREP_g$, partially replicated experimental design with fixed number of genotypes; $PREP_n$, partially replicated experimental design with fixed number of experimental units; RCBD, randomized complete block experimental design; YSD, yield standard deviation.

One of the most important aspects in any experiment in agricultural research is the proper choice of the experimental design and analysis model (Casler, 2015; Piepho et al., 2015). Well-designed experiments are based on the three principles proposed by R.A. Fisher (1935): randomization, replication, and local control. After Fisher, agricultural research has been based on these three principles that seek to control local

Published in Crop Sci. 59:44–53 (2019).
doi: 10.2135/cropsci2018.03.0177

variation, assuming independence between experimental units, and using replications to estimate experimental error and increase precision of mean estimates. On the other hand, spatial variation in environmental and soil factors is common under field conditions, so spatial autocorrelation is generally present in field experiments (Legendre, 1993; Grondona et al., 1996). Therefore, experiments can be designed to have local control of this spatial variability.

In agriculture, the most commonly used experimental designs are the completely randomized design (CRD) and the randomized complete block design (RCBD) (Piepho et al., 2015). In some cases, blocking is an effective way to control experimental error (Cochran and Cox, 1957; Casler, 2015), but it is not enough in agricultural situations where field heterogeneity and/or the size of the experiment is large (Brownie et al., 1993). Furthermore, designs including blocks without considering the real spatial variation among experimental units can strongly decrease the success of an experiment (Casler, 2015).

Some designs such as α-lattice incomplete block (ALPHA) experimental designs (Patterson and Williams, 1976) or row–column designs (Williams et al., 2006) were specifically created to control field heterogeneity in situations where a large number of treatments are evaluated, as well as other complex treatment structures (Williams et al., 2002). These designs are therefore better suited for a large number of treatments where spatial variability is high (Müller et al., 2010). However, they are not always used due to their complexity and practical difficulties, or simply because researchers are not used to choosing them (Casler, 2015). In cultivar evaluation trials, where the number of treatments is always high and there is an increasing need to distinguish cultivars by smaller differences in terms of yield (Casler and Undersander, 2000), better experimental designs and analysis models are needed to control spatial variation.

Spatial variability often occurs gradually, and sometimes it is not captured well enough by the experimental design (Grondona and Cressie, 1991). Even with strong local control, as in incomplete block designs, spatial variability may not be properly controlled if it occurs at smaller scales than the size of the sub-blocks (Grondona et al., 1996). Explicitly accounting for spatial variability can be achieved in various ways such as with different structures of correlation of the R matrix, trend analysis, or a combinations of both (Brownie et al., 1993; Casler and Undersander, 2000). Additionally, one- or two-dimensional spatial analysis can be implemented according to natural field heterogeneity (Gleeson and Cullis, 1987; Cullis and Gleeson, 1991; Gilmour et al., 1997; Qiao et al., 2000).

Whether it is better to control with experimental design or spatial analysis models is often debated. Müller et al. (2010) found that in barley (*Hordeum vulgare* L.) and sugar beet (*Beta vulgaris* L.) trials, the baseline model, which only includes a block and a replicate effect, showed the best fit according to Akaike information criterion for most cases. Spatial analyses were not necessarily better than the RCBD in all possible circumstances, particularly in cases where the spatial structure of the studied variable cannot be accurately characterized (Kravchenko et al., 2006). Uniformity trials have been conducted to determine the optimal experimental design in many cases. These trials were expensive to carry out, and complex designs were not evaluated (Koch and Rigney, 1951). Studies that compare different designs and analysis models using the same database with real field spatial variability have not been reported to our knowledge. Our goal was to evaluate the effect of experimental design, spatial modeling, and a combination of both under real field conditions using GIS information and simulating experiments. One of the questions of interest was whether spatial modeling can substitute for experimental design, or whether its advantage is marginal compared with a proper experimental design.

MATERIALS AND METHODS

General Approach

Yield data from cultivars were simulated using real field variability, genotypic effects, and a number of experimental designs in different locations within a large field. Each realization was analyzed using a series of models with different levels of correction for spatial variability. Evaluation criteria for comparing designs and analysis models included precision, accuracy, and the recovery of superior genotypes.

Wheat Yield Data and Spatial Variability from Yield Monitors

A field of ∼64 ha was sown with the wheat (*Triticum aestivum* L.) cultivar 'Nogal' (USDA-ARS, 1992) on 20 June 2008 at a density of 120 kg ha^{-1} of seed. The field was harvested in rectangular plots of ∼15-m × 5-m area to obtain 1445 yield plots (kg ha^{-1}). Each plot has geographic coordinates X (east–west) and Y (north–south) and elevation (m). An empiric variogram was calculated with the values from the yield monitor and a Matern variogram ($\kappa = 1$) was fitted. Due to the original inclination of the plots with respect to the X and Y coordinates, kriged values were used as the baseline field heterogeneity. Yield maps were created using the 'sp' package (Pebesma and Bivand, 2005) of R statistical software (R Core Team, 2016).

Genotypic Effects

Three experiment sizes were evaluated: (i) small size with 45 experimental units, (ii) moderate size with 150 experimental units, and (iii) large size with 600 experimental units. For the small experiment size, 15 genotypic effects for the simulation process were obtained from yield mean data of 15 wheat cultivars from the National Cultivar Trial Networks from 3-yr evaluation trials (2007–2009) (INIA, 2017). For moderate and large experiment size, genotypic effects (G_i) were simulated from a Gaussian distribution, assuming that $G_i \sim N(0, \sigma^2_G)$, where σ^2_G was the genotypic variance of

the 15 genotypes obtained from the National Cultivar Trial Network. For moderate size, 50 genotypic effects were simulated for CRD, RCBD, and ALPHA designs, whereas 200 genotypic effects were simulated for large-size experiments. We used two strategies for partially replicated experimental (PREP) designs: (i) fixing the number of experimental units (i.e., 150 and 600 experimental units for the moderate and large size respectively) so that 108 and 428 genotypic effects were evaluated and simulated for the medium and large experiment size, respectively; or (ii) fixing the number of genotypes evaluated. Therefore, 50 and 200 genotypes were evaluated and simulated for the moderate and large experiment sizes, respectively, using fewer experimental units than the other experimental designs (i.e., 70 experimental units for the medium-size experiment instead of 150, and 280 experimental units for the large-size experiment instead of 600), as described below.

Experimental Designs

The genotypes (treatments) were assigned to experimental units (plots of ~15-m × 5-m area) in one of four experimental designs: CRD, RCBD, ALPHA, and PREP.

The model used for the CRD experimental design with three replications was:

$$Y_{ij} = \mu + G_i + \varepsilon_{ij}$$

where Y_{ij} is the observed yield for the ith genotype in the jth replicate, μ is the overall mean, G_i is the ith genotypic effect, and ε_{ij} are the residual errors associated with the observation Y_{ij}, $\varepsilon_{ij} \sim N(0, \sigma^2_\varepsilon)$, where σ^2_ε is the error variance, and the covariance $COV(\varepsilon_{ij}, \varepsilon_{ij'})$ was modeled according to the spatial corrections described below.

The model used for the RCBD with three replications was:

$$Y_{ij} = \mu + G_i + \beta_j + \varepsilon_{ij}$$

where Y_{ij} is the observed yield for the ith genotype in the jth block, and β_j is the jth block effect. The blocks were considered fixed effects and were located in the east–west direction.

An ALPHA design with three complete replications and a number of incomplete blocks or sub-block was used. The number of experimental units per incomplete block was three for the small size, five for the moderate size, and five [$ALPHA_{(s = 5)}$] or ten [$ALPHA_{(s = 10)}$] for the large experiment size. The model used for the ALPHA experiment with three full replicates was:

$$Y_{ij} = \mu + G_i + \beta_j + \gamma_{k(j)} + \varepsilon_{ijk}$$

where Y_{ijk} is the observed yield for the ith genotype, jth replication, and kth incomplete block, β_j is the jth complete replication effect, and $\gamma_{k(j)}$ is the kth incomplete block effect nested on jth replication, and ε_{ijk} are the residual errors associated with the observation Y_{ijk}, $\varepsilon_{ijk} \sim N(0, \sigma^2_\varepsilon)$, and the covariance $COV(\varepsilon_{ijk}, \varepsilon_{ijk'})$ was modeled according to the spatial corrections described below. Incomplete blocks were considered random factors nested in each complete replication, assuming $\gamma_{k(j)} \sim N(0, \sigma^2_s)$, where σ^2_s is the variance of the sub-blocks. Complete replications were located in the east–west direction, as in the RCBD design, and incomplete blocks were orthogonal to the complete replications.

A PREP design where repeated genotypes were randomized in a RCBD design with three replications was used. The model used for the PREP was:

$$Y_{ij} = \mu + G_i + \beta_j + \varepsilon_{ij}$$

where Y_{ij} is the observed yield of the ith genotype and jth replication. The model for G_i was $G_i = g_i + t_i$, where g_i is the effect of the ith nonreplicated genotypic effect with $i = 1, \ldots, n_g$ (where n_g is the number of nonreplicated genotypes), and t_i is the ith replicated genotypic effect (test line) with $i = n_g + 1, \ldots, n_g + n_c$ (where n_c is the number of replicated genotypes).

In the case that the number of experimental units was fixed ($PREP_n$), 20% of the genotypes were replicated, whereas the remaining genotypes were unreplicated. In the case that the number of genotypes was fixed ($PREP_g$), 15% of the genotypes were replicated, containing either 70 or 280 plots for moderate and large experiment size respectively (47% of the number of plots in the CRD, RCBD, and ALPHA designs).

Locations

One hundred randomly selected locations within the field were used to conduct all simulations. The selected location was used as the upper left corner of each experimental design, and the shape was always rectangular with longer east–west dimension for all locations to follow the direction of smallest variance. The shape of all locations was the same for a given experimental size. For each experimental size, the same experimental units were used for all experimental designs, except for $PREP_g$, where fewer experimental units were used. Each location was characterized by their yield SD (YSD), expressed as tons per hectare. For each experimental size, the mean YSD of all 100 locations was calculated. Locations were then classified as either with high or low variability (Fig. 1) according to this criterion: locations with YSD lower than the mean YSD were considered as low-variability locations, and those with greater YSD than the mean YSD were considered as high-variability locations. Results are reported as averages for all high- and low-variability locations for each experimental size.

Simulation Procedure

Yield was obtained for each plot according to the procedure below. First, treatments were assigned to plots according to one of the three experimental designs described above. Second, yield of each plot was obtained using the equation:

$$Y_{ij} = G_i + \varepsilon_{ij} + \delta_{ij}$$

where Y_{ij} is the yield plot data simulated corresponding to the ith cultivar and the jth replication, G_i is the ith genotypic effect corresponding to the randomly assigned treatment to the plot, ε_{ij} is the field experimental error that represents the spatial heterogeneity of the field and was obtained from the yield monitor, and δ_{ij} is a repeatability error. To avoid using a deterministic model, the δ_{ij} values were assumed as independent random variables with $\delta_{ij} \sim N(0, \sigma^2_r)$, where σ^2_r is a random noise or repeatability. We conducted the simulations using two values for σ^2_r. First, we used a value of $\sigma^2_r = 0.07$ that represents 5% of the total field heterogeneity and a yield heritability of ~0.5 when a simple experimental design is used. The second value, $\sigma^2_r = 0.2$, targeted a lower yield heritability.

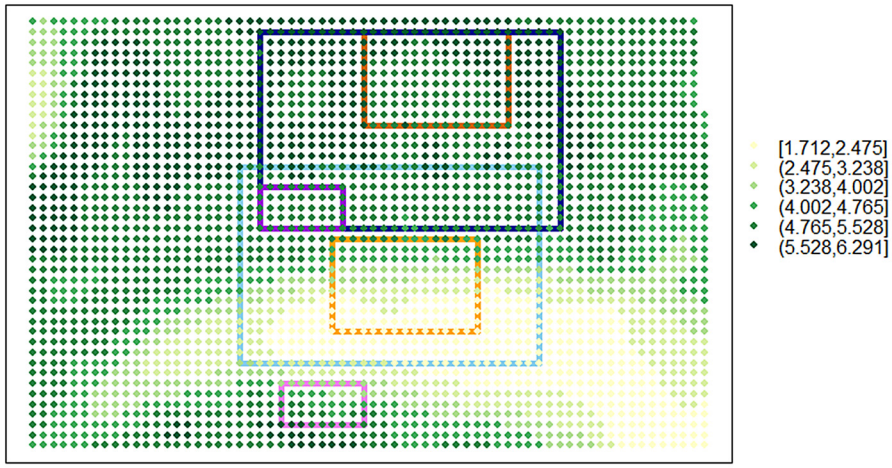

Fig. 1. Wheat yield map (tons ha^{-1}), according to east–west oriented coordinates (*Y* coordinates) and north–south oriented coordinates (*X* coordinates). Darker colors indicate higher yield values. Localization of the experimental units for two contrasting locations and three experimental design sizes are shown: for small experiment size (15 genotypes), light violet lines represent high-variability location and dark violet lines represent for low-variability ones; for moderate experiment size (50 genotypes), light orange lines represent high-variability locations and dark orange lines represent low-variability ones; for large experiment size (200 genotypes), light blue lines represent high-variability locations and dark blue lines represent low-variability locations. Two locations out of 100 field locations that were evaluated for each experimental design size are shown.

For each experimental design and location, 1000 simulations were run, performing an independent randomization for each simulation. The 'agricolae' package (de Mendiburu, 2012) of R (R Core Team, 2016) was used for experimental design randomization. The simulations and statistical analysis of the data were performed with the 'nlme' package (Pinheiro et al., 2013) and personal code of R (R Core Team, 2016).

Analysis Models

Each vector of phenotypic yield was analyzed according to the following three models:

1. No spatial correction (NSC), where experimental errors are assumed independent (uncorrelated), $\varepsilon_{ijk} \sim N(0, \sigma^2_\varepsilon)$, where $\text{COV}(\varepsilon_{ij}, \varepsilon_{ij'}) = 0 \; \forall \; \varepsilon_{ij} \neq \varepsilon_{ij'}$.
2. A spatially correlated error model with one-dimensional autoregressive process [AR(1)], $\varepsilon_{ijk} \sim N(0, \sigma^2_\varepsilon)$, where $\text{COV}(\varepsilon_{ij}, \varepsilon_{ij'}) = \sigma^2_\varepsilon \rho^k$. The correlation function corresponding to an autoregressive model of order 1 decreases in absolute value with every unit of distance within columns: $h(k, \rho) = \rho^k$, $k = 0, 1, ...$, where ρ is the correlation parameter to be estimated, and k is the distance unit between rows (i.e., the direction of maximum variance in our uniformity trial).
3. Spatially correlated error model with two-dimensional exponential spatial correlation structure [EXP(2)] , $\varepsilon_{ijk} \sim N(0, \sigma^2_\varepsilon)$, where $\text{COV}(\varepsilon_{ij}, \varepsilon_{ij'})$ was modeled according to an isotropic exponential model. With d being the range, the correlation between two observations at r distance was $\exp(-r/d)$. The EXP(2) model was fitted for rows and columns with and without a nugget variance.

Estimation Method and Evaluation Criteria for Model Comparison

The residual maximum likelihood method was used to estimate parameters. Models were compared by precision and accuracy statistics and by their ability to recover superior genotypes. For all statistics, we calculated the mean and the SD over the 1000 realizations of the simulation for each experiment size, type of location, experimental design, and spatial correction.

The SE of the difference between cultivar means (SED) was used as a precision statistic to compare models. The lower the value, the better the precision.

The recovery of the best genotypes (Best_Gen), the Pearson's correlation coefficient between true and estimated genotypic effects (COR), and the mean square error of prediction (MSEP) were used as an accuracy statistic to compare models. The Best_Gen was calculated as the proportion of times 15% of the true superior genotypes were recovered. The MSEP was calculated according to Gauch et al. (2003) as follows:

$$\text{MSEP} = \Sigma(X_n - Y_n)2/N$$

where X_n and Y_n are the model-based and true genotypic values, and N is the number of genotypes, where summation is over $n = 1, 2, \ldots, N$.

RESULTS

The experimental designs and spatial correction models performed similarly for the small experiment size in both low- and high-variability locations, and with low and high yield heritability (Supplemental Tables S1 and S2).

For moderate experiment size in high-variability conditions, the ALPHA design had the best performance for most statistics (Table 1), although RCBD with AR(1) performed similar to the ALPHA with NSC. For CRD and RCBD, only the AR(1) model was better than the NSC (Table 1). Therefore, there is some performance compensation by

spatial modeling, in the case of AR(1). The PREP had a poor performance over all designs, and spatial corrections did not improve the performance. Even with $PREP_n$ with the same number of experimental units as the others experimental designs, performance was poorer.

For moderate experiment size with low variability, no differences in the performance of experimental design or spatial corrections were found among CRD, RCBD, and ALPHA for most statistics (Table 1). The PREP design in these circumstances underperformed for all statistics in both cases ($PREP_g$ and $PREP_n$, Table 1).

For large experiment size under high variability, the results were similar to those of the moderate size, but the advantages of a more complex experimental design were more noticeable. The ALPHA design outperformed all other experimental designs, especially when it was combined with AR(1) or EXP(2) using either 5 or 10 experimental units per incomplete block (Table 2, Fig. 2). The AR(1) spatial correction improved model performance of CRD and RCBD but did not outperform experimental design in general (Table 2), although RCBD with AR(1) obtained similar values to $ALPHA_{(s = 10)}$ with no spatial correction (Fig. 2). For this size of experiment, the performance of the $PREP_g$ was closer to the simple designs for most of the statistics (Table 2).

For large experiment size with low variability, the differences among experimental designs and spatial corrections were small, although larger than for moderate size designs. The $ALPHA_{(s = 10)}$ experimental design with some spatial correction (Table 2, Fig. 2) obtained the best

Table 1. Mean and SD (in parentheses) of the recovery of 15% of superior genotypes (Best_Gen), the SE for mean differences (SED), the Pearson's correlation coefficient between observed and estimated genotypic effects (COR), and the mean square error of prediction (MSEP), for four experimental designs and three spatial corrections in two types of field locations (high and low spatial heterogeneity). This table shows results for the moderate experiment size (50 genotypes) and high heritability.

Variability	YSD†	Locations	Design‡	Model§	Best_Gen	SED	COR	MSEP
	tons ha^{-1}	no.				tons ha^{-1}		tons ha^{-1}
High	0.80	46	CRD	NSC	0.45 (0.15)	0.49 (0.02)	0.63 (0.08)	0.25 (0.04)
			CRD	AR(1)	0.49 (0.15)	0.38 (0.03)	0.68 (0.07)	0.18 (0.03)
			CRD	EXP(2)	0.45 (0.15)	0.49 (0.02)	0.64 (0.07)	0.29 (0.06)
			RCBD	NSC	0.46 (0.15)	0.47 (0.02)	0.65 (0.07)	0.32 (0.04)
			RCBD	AR(1)	0.52 (0.15)	0.33 (0.03)	0.72 (0.06)	0.24 (0.03)
			RCBD	EXP(2)	0.46 (0.15)	0.47 (0.02)	0.65 (0.07)	0.34 (0.06)
			ALPHA	NSC	0.55 (0.14)	0.36 (0.02)	0.75 (0.05)	0.21 (0.03)
			ALPHA	AR(1)	0.64 (0.13)	0.29 (0.01)	0.85 (0.04)	0.16 (0.03)
			ALPHA	EXP(2)	0.64 (0.13)	0.29 (0.01)	0.85 (0.04)	0.16 (0.03)
			$PREP_g$	NSC	0.38 (0.15)	0.60 (0.10)	0.53 (0.10)	0.52 (0.16)
			$PREP_g$	AR(1)	0.39 (0.15)	0.56 (0.09)	0.55 (0.10)	0.49 (0.18)
			$PREP_g$	EXP(2)	0.40 (0.15)	0.57 (0.09)	0.56 (0.10)	0.47 (0.15)
			$PREP_n$	NSC	0.25 (0.14)	0.62 (0.07)	0.48 (0.07)	0.54 (0.09)
			$PREP_n$	AR(1)	0.26 (0.14)	0.58 (0.06)	0.50 (0.07)	0.49 (0.10)
			$PREP_n$	EXP(2)	0.26 (0.14)	0.58 (0.06)	0.51 (0.07)	0.48 (0.09)
Low	0.25	53	CRD	NSC	0.68 (0.13)	0.21 (0.01)	0.88 (0.03)	0.05 (0.01)
			CRD	AR(1)	0.68 (0.13)	0.21 (0.01)	0.88 (0.03)	0.05 (0.01)
			CRD	EXP(2)	0.69 (0.13)	0.21 (0.01)	0.88 (0.02)	0.05 (0.01)
			RCBD	NSC	0.69 (0.12)	0.20 (0.01)	0.89 (0.02)	0.06 (0.01)
			RCBD	AR(1)	0.69 (0.13)	0.19 (0.01)	0.89 (0.02)	0.06 (0.01)
			RCBD	EXP(2)	0.69 (0.12)	0.21 (0.27)	0.89 (0.02)	0.06 (0.01)
			ALPHA	NSC	0.70 (0.12)	0.19 (0.01)	0.90 (0.02)	0.06 (0.01)
			ALPHA	AR(1)	0.71 (0.12)	0.18 (0.01)	0.90 (0.02)	0.06 (0.01)
			ALPHA	EXP(2)	0.71 (0.12)	0.18 (0.01)	0.90 (0.02)	0.06 (0.01)
			$PREP_g$	NSC	0.56 (0.14)	0.30 (0.05)	0.76 (0.05)	0.13 (0.04)
			$PREP_g$	AR(1)	0.55 (0.14)	0.29 (0.05)	0.75 (0.06)	0.14 (0.04)
			$PREP_g$	EXP(2)	0.56 (0.14)	0.29 (0.05)	0.76 (0.05)	0.13 (0.04)
			$PREP_n$	NSC	0.45 (0.15)	0.31 (0.04)	0.73 (0.04)	0.13 (0.02)
			$PREP_n$	AR(1)	0.45 (0.15)	0.30 (0.04)	0.73 (0.04)	0.13 (0.03)
			$PREP_n$	EXP(2)	0.46 (0.15)	0.30 (0.04)	0.74 (0.04)	0.13 (0.02)

† YSD, yield SD.

‡ CRD, completely randomized design; RCBD, randomized complete block design; ALPHA, incomplete blocks and α-lattice design; $PREP_g$, partially replicated design with 50 genotypes (this experiment preserved the number of genotypes and thus used fewer experimental units); $PREP_n$, partially replicated design with 108 genotypes (this experiment preserved the number of experimental units and thus evaluated more genotypes).

§ NSC, no spatial correction model; AR(1), spatial correlated error model with one-dimensional autoregressive process; EXP(2), spatial correlated error model with two-dimensional exponential spatial correlation structure without a nugget variance.

results. The PREP design had no advantages in these conditions either.

Similar relative performance of experimental designs and spatial corrections was observed when a low heritability was simulated for both, medium (Supplemental Table S3) and large experiment sizes (Supplemental Table S4). In both cases, all the experiments and spatial corrections performed poorer than with high heritability.

The use of a nugget variance in the EXP(2) spatial correction increased the mean and the variability in the mean square error estimation of the CRD and RCBD experiments and did not improve any other statistic of any experimental design (data not shown); therefore, it is not further discussed.

DISCUSSION

Experiment size and field heterogeneity clearly affected the performance of the experimental designs and spatial correction. With large experiment sizes under high-variability conditions, the choice of the experimental design is essential to obtain good results in terms of precision and accuracy in the estimation of genotypic effects.

The ALPHA design had the best performance across design and spatial correction. This design has been shown

Table 2. Mean and SD (in parentheses) of the recovery of 15% of superior genotypes (Best_Gen), the SE for mean differences (SED), the Pearson's correlation coefficient between observed and estimated genotypic effects (COR), and the mean square error of prediction (MSEP) for four experimental designs and three spatial corrections in two types of field locations (high and low spatial heterogeneity). This table shows results for the large experiment size (200 genotypes) and high heritability.

Variability	YSD†	Locations	Design‡	Model§	Best_Gen	SED	COR	MSEP
	tons ha^{-1}	no.				tons ha^{-1}		tons ha^{-1}
High	1.10	58	CRD	NSC	0.40 (0.07)	0.66 (0.01)	0.51 (0.05)	0.43 (0.04)
			CRD	AR(1)	0.45 (0.07)	0.43 (0.02)	0.60 (0.04)	0.27 (0.02)
			CRD	EXP(2)	0.40 (0.07)	0.65 (0.01)	0.51 (0.05)	0.43 (0.04)
			RCBD	NSC	0.40 (0.07)	0.64 (0.01)	0.52 (0.05)	0.52 (0.04)
			RCBD	AR(1)	0.50 (0.07)	0.36 (0.02)	0.67 (0.03)	0.31 (0.02)
			RCBD	EXP(2)	0.40 (0.07)	0.64 (0.01)	0.52 (0.05)	0.52 (0.04)
			$ALPHA_{(s = 5)}$	NSC	0.52 (0.07)	0.35 (0.01)	0.71 (0.03)	0.23 (0.02)
			$ALPHA_{(s = 5)}$	AR(1)	0.60 (0.07)	0.28 (0.01)	0.80 (0.03)	0.18 (0.02)
			$ALPHA_{(s = 5)}$	EXP(2)	0.60 (0.07)	0.28 (0.01)	0.80 (0.03)	0.18 (0.02)
			$ALPHA_{(s = 10)}$	NSC	0.47 (0.07)	0.49 (0.01)	0.63 (0.04)	0.35 (0.02)
			$ALPHA_{(s = 10)}$	AR(1)	0.68 (0.06)	0.27 (0.01)	0.87 (0.02)	0.19 (0.02)
			$ALPHA_{(s = 10)}$	EXP(2)	0.68 (0.06)	0.27 (0.01)	0.87 (0.02)	0.19 (0.02)
			$PREP_g$	NSC	0.39 (0.07)	0.76 (0.06)	0.46 (0.05)	0.65 (0.07)
			$PREP_g$	AR(1)	0.44 (0.07)	0.59 (0.04)	0.57 (0.05)	0.36 (0.07)
			$PREP_g$	EXP(2)	0.44 (0.07)	0.59 (0.04)	0.57 (0.05)	0.36 (0.07)
			$PREP_n$	NSC	0.26 (0.07)	0.81 (0.04)	0.40 (0.04)	0.80 (0.06)
			$PREP_n$	AR(1)	0.31 (0.07)	0.60 (0.02)	0.52 (0.04)	0.47 (0.06)
			$PREP_n$	EXP(2)	0.31 (0.07)	0.60 (0.02)	0.52 (0.04)	0.47 (0.06)
Low	0.46	35	CRD	NSC	0.63 (0.06)	0.31 (0.01)	0.78 (0.02)	0.10 (0.01)
			CRD	AR(1)	0.63 (0.06)	0.28 (0.01)	0.79 (0.02)	0.10 (0.01)
			CRD	EXP(2)	0.63 (0.06)	0.30 (0.01)	0.79 (0.02)	0.10 (0.01)
			RCBD	NSC	0.63 (0.06)	0.30 (0.01)	0.79 (0.02)	0.11 (0.01)
			RCBD	AR(1)	0.63 (0.06)	0.27 (0.01)	0.80 (0.02)	0.10 (0.01)
			RCBD	EXP(2)	0.63 (0.06)	0.30 (0.01)	0.79 (0.02)	0.11 (0.01)
			$ALPHA_{(s = 5)}$	NSC	0.64 (0.06)	0.24 (0.01)	0.83 (0.02)	0.07 (0.01)
			$ALPHA_{(s = 5)}$	AR(1)	0.66 (0.06)	0.21 (0.01)	0.85 (0.02)	0.06 (0.01)
			$ALPHA_{(s = 5)}$	EXP(2)	0.66 (0.06)	0.21 (0.01)	0.85 (0.02)	0.06 (0.01)
			$ALPHA_{(s = 10)}$	NSC	0.66 (0.06)	0.27 (0.01)	0.83 (0.02)	0.09 (0.01)
			$ALPHA_{(s = 10)}$	AR(1)	0.71 (0.06)	0.20 (0.01)	0.88 (0.01)	0.05 (0.01)
			$ALPHA_{(s = 10)}$	EXP(2)	0.71 (0.06)	0.20 (0.01)	0.88 (0.01)	0.05 (0.01)
			$PREP_g$	NSC	0.55 (0.07)	0.41 (0.04)	0.68 (0.03)	0.20 (0.02)
			$PREP_g$	AR(1)	0.56 (07)	0.37 (0.03)	0.71 (0.04)	0.17 (0.03)
			$PREP_g$	EXP(2)	0.56 (0.07)	0.37 (0.03)	0.71 (0.04)	0.17 (0.03)
			$PREP_n$	NSC	0.45 (0.07)	0.44 (0.03)	0.63 (0.03)	0.22 (0.02)
			$PREP_n$	AR(1)	0.46 (0.07)	0.38 (0.02)	0.66 (0.03)	0.18 (0.02)
			$PREP_n$	EXP(2)	0.46 (0.07)	0.38 (0.02)	0.66 (0.03)	0.18 (0.02)

† YSD, yield SD.

‡ CRD, completely randomized design; RCBD, randomized complete block design; $ALPHA_{(s = 5)}$, incomplete blocks and α-lattice design with a block size of 5; $ALPHA_{(s = 10)}$, incomplete blocks and α-lattice design with a block size of 10; $PREP_g$, partially replicated design with 200 genotypes (this experiment preserved the number of genotypes and thus used fewer experimental units); $PREP_n$, partially replicated design with 428 genotypes (this experiment preserved the number of experimental units and thus evaluated more genotypes).

§ NSC, no spatial correction model; AR(1), spatial correlated error model with one-dimensional autoregressive process; EXP(2), spatial correlated error model with two-dimensional exponential spatial correlation structure without a nugget variance.

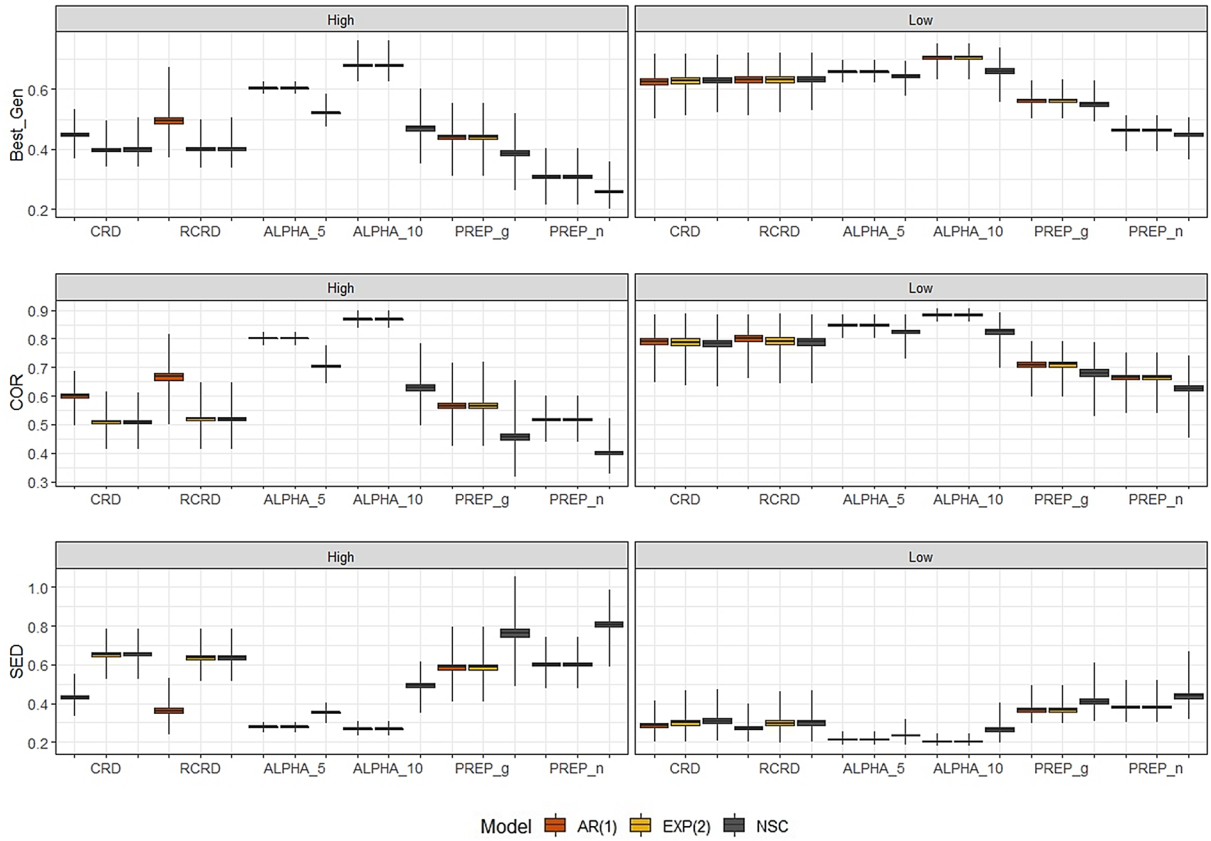

Fig. 2. Recovery of the best genotypes (Best_Gen), Pearson's correlation coefficient between true and estimated genotypic effects (COR), and the SE of the difference between cultivar means (SED) for six experimental designs: completely randomized design (CRD), randomized complete block design (RCBD) incomplete blocks and α-lattice design with block size of 5 (ALPHA_5), incomplete blocks and α-lattice design with block size of 10 (ALPHA_10), partially replicated design with 200 genotypes ($PREP_g$, this experiment preserved the number of genotypes and thus used fewer experimental units), and a partially replicated design with 428 genotypes (PREPn, this experiment preserved the number of experimental units and thus evaluated more genotypes). Each experimental design was analyzed with three spatial models: no spatial correction model (NSC), spatial correlated error model with one-dimensional autoregressive process [AR(1)], spatial correlated error model with two-dimensional exponential spatial correlation structure [EXP(2)], in high- (left) and low-variability (right) locations.

as superior to RCBD (Masood et al., 2008; Gonçalves et al., 2010) due to stronger local control. The experimental units' layout and blocking orientation in our study might have favored CRD designs over all other experimental designs, with RCBD being the most unfavorable design. Experimental designs that exerted some local control on the spatial variability, like ALPHA or row–column designs (Müller et al., 2010; Sripathi et al., 2017), generally increase the power of the ANOVA (Legendre et al., 2004). Additionally, we found that the ALPHA design was the best in its ability to recover superior genotypes in terms of yield and the accuracy measured by the correlation between true and expected genotypic effects. Although we expect unbiased estimations in all models due to the randomization process (Fisher, 1935), we expect models taking into account the underlying variability in the field to be more efficient (Lehmann and Casella, 1998). The larger variability not controlled with some experimental designs possibly induced more noise in the estimation of the genotypic effect, and therefore smaller values of Best_Gen and COR were obtained. The low efficiency of simpler designs in controlling field heterogeneity should reduce the accuracy in recovering superior genotypes. Stroup et al. (1994) suggested that the presence of spatial correlation in the data inflates the estimation of experimental error and could lead to inaccurate estimates of treatment means using classical RCBD analysis.

Even with low field heterogeneity present, the ALPHA design was slightly better when a large number of genotypes are used, mainly with some spatial correction. The efficiency of an experimental design depends on its complexity, and it is regulated by experiment size and the nature of the spatial variability. The larger the number of treatments, the smaller the efficiency of simpler designs in general (Casler, 2015).

Model performance improves when combining well-designed experiments with spatial adjustments (Gilmour et al., 1997; Qiao et al., 2000; Williams et al., 2006), suggesting that both, design information and spatial modeling should be considered. We found that under high

field heterogeneity, spatial models improved the performance of all experimental designs. On the other hand, spatial modeling did not outperform experimental design in any of our situations. Therefore, spatial modeling should be considered as a supplemental strategy rather than an alternative to experimental design, as was also suggested by Richter and Kroschewski (2012) and Piepho et al. (2015). Under high field variability, CRD or RCBD with spatial corrections were similar to ALPHA without spatial correction for precision.

One approach that was beyond the scope of this paper but is worth mentioning is the use of spatial designs. Our paper focused on answering the question as to whether spatial analysis can be a substitute for a good classical experimental design, and we found that there is no good substitute for a well-planned experimental design. However, we did not evaluate the use of a priori spatial information in the optimization of classical experimental designs, which have been called spatial designs (Williams et al., 2006). There is a growing interest in the literature for these spatial designs (Piepho and Williams, 2010; Williams and Piepho, 2013; Piepho et al., 2016). However, one of the most challenging aspects of this approach is demonstrating the validity of the restrictions in the randomization and the presence of error variance bias in these designs (Williams and Piepho, 2018). Furthermore, even if soil spatial variation was previously well characterized and experiments are designed accordingly, there is still variation that cannot be predicted a priori (Cooper et al., 2014), making it even more challenging to produce good randomizations.

The best spatial model seems to be, most likely, case specific (Cullis and Gleeson, 1991; Richter and Kroschewski, 2012; Moehring et al., 2014). We found that the spatial correlation structure modeled with AR(1) seemed to be more suitable to describe the spatial pattern present in our uniformity trial, where the strongest spatial variability gradient was given in the direction of the columns (Fig. 1). We used the direction of maximum spatial variability known a priori to model the AR(1) process. This could theoretically favor the AR(1) model. However, because in the absence of prior information on spatial variability it is a common practice to evaluate the spatial processes in both directions, we do not believe our results are biased. The exponential structure in two dimensions gave the same weights to spatial variability in rows and columns that did not correspond with the pattern observed in the field at most of the locations. However, a two-dimensional AR(1) process was better than a one-dimensional one in several studies (Cullis and Gleeson, 1991; Moehring et al., 2014) because it does not assume isotropy. One of the limitations of our study is that only an isotropic spatial correction model was considered. Given that our uniformity trial had higher variability in the north–south direction than in the east–west direction, an anisotropic model might have provided better spatial control. In that situation, an AR(1) × AR(1) model might outperform the AR(1) and EXP(2) models (Moehring et al., 2014).

Because of the nature of PREP, a choice between fixing the number of treatments (and therefore having fewer experimental units) or fixing the number of experimental units (and having more treatments) had to be made. We evaluated both strategies. First, we compared all experimental designs for a fixed set of treatments; therefore, $PREP_g$ used fewer experimental units and was at a disadvantage compared with the other designs. Under these circumstances, $PREP_g$ did not perform as well as the other experimental designs. With a moderate size of experiment, it was the worst design. However, under large experiment sizes, the differences with simpler designs were smaller, while $PREP_g$ used fewer resources. Second, we evaluated the performance of the PREP experimental designs, fixing the resources by using the same number of experimental units and therefore evaluating a larger number of treatments ($PREP_n$). This situation was not better in terms of precision and accuracy than the $PREP_g$. Other studies did find that for a fixed experiment size, PREP designs are efficient and can outperform replicated designs in multiple-environment experiments (Cullis et al., 2006, Moehring et al., 2014). It can be expected that in PREP designs, some precision and accuracy are sacrificed because there are fewer replications. Endelman et al. (2014) pointed out that in preliminary yield trials in multiple locations, allocating additional plots per entry increases accuracy. What we are not considering in this work is the impact of evaluating larger population sizes with more genotypes. In our $PREP_n$, 428 genotypes were evaluated, vs. the 200 that other designs evaluated using the same number of experimental units. Moehring et al. (2014) found that an augmented design outperformed replicated and classical augmented designs in terms of prediction accuracy, providing a better sample of the genotype × environment interaction. Another important consideration is the allocation of repeated plots. There are different ways in which this can be accomplished, mainly by increasing the number of repeated genotypes, or increasing the number of replications. However, the most important decision to make is the total number of repeated plots used and not the number of repeated genotypes, because that number determines the total degrees of freedom that affect design performance (Clarke and Stefanova, 2011). Optimizing this might slightly improve the performance of our PREP. The use of larger population sizes would have other advantages not reflected in our study, such as larger population sizes for genomic studies (i.e., mapping or genomic selection). Larger population sizes would result in larger selection intensities that could increase the selection response in breeding populations.

Trait heritability did not change the conclusions of this work (Supplemental Tables S2–S4). The relative performance of the experimental designs and spatial correction models was the same regardless of the heritability. The experiments and spatial models all performed more poorly with lower heritability. Therefore, we believe that our conclusions can be generalized to a large number of situations.

In summary, under high field heterogeneity and with a large number of treatments, spatial modeling without local control does not outperform local control with proper experimental designs. Once you have chosen the proper experimental design, spatial modeling can further improve its performance, as also seen in Qiao et al. (2000). This is especially the case in moderate- to large-sized experiments and under both high and low field variability. Spatial patterns using precision agriculture technology could therefore be used to better design experiments and to find an adequate spatial correction for each experiment (Cooper et al., 2014).

Conflict of Interest

The authors declare that there is no conflict of interest.

Supplemental Data Available

Supplemental material is available online for this article.

Acknowledgments

A. Borges was funded by a young scientist award from the Comisión Sectorial de Investigación Científica (CSIC-ID 714) to conduct this study. Funding was also provided by a research grant from the Fondo María Viñas (FMV_2009_1_2496). We would like to thank two anonymous reviewers for their careful reading and constructive comments.

References

Brownie, C., D.T. Bowman, and J.W. Burton. 1993. Estimating spatial variation in analysis of data from yield trials: A comparison of methods. Agron. J. 85:1244–1253. doi:10.2134/agronj1993.00021962008500060028x

Casler, M.D. 2015. Fundamentals of experimental design: Guidelines for designing successful experiments. Agron. J. 107:692–705. doi:10.2134/agronj2013.0114

Casler, M.D., and D.J. Undersander. 2000. Forage yield precision, experimental design, and cultivar mean separation for alfalfa cultivar trials. Agron. J. 92:1064–1071. doi:10.2134/agronj2000.9261064x

Clarke, G.P., and K. Stefanova. 2011. Optimal design for early-generation plant-breeding trials with unreplicated or partially replicated test lines. Aust. N. Z. J. Stat. 53:461–480. doi:10.1111/j.1467-842X.2011.00642.x

Cochran, W.G., and G.M. Cox. 1957. Experimental designs. 2nd ed. John Wiley & Sons, New York.

Cooper, M., C.D. Messina, D. Podlich, L.R. Totir, A. Baumgarten, N.J. Hausmann, et al. 2014. Predicting the future of plant breeding: Complementing empirical evaluation with genetic prediction. Crop Pasture Sci. 65:311–336. doi:10.1071/CP14007

Cullis, B.R., and A.C. Gleeson. 1991. Spatial analysis of field experiments-an extension to two dimensions. Biometrics 47:1449–1460. doi:10.2307/2532398

Cullis, B.R., A.B. Smith, and N.E. Coombes. 2006. On the design of early generation variety trials with correlated data. J. Agric. Biol. Environ. Stat. 11:381–393. doi:10.1198/108571106X154443

de Mendiburu, F. 2012. Agricolae: Statistical procedures for agricultural research. R package version 1.1–3. R Found. Stat. Comput., Vienna. http://CRAN.R-project.org/package=agricolae (accessed 13 Nov. 2018).

Endelman, J.B., G.N. Atlin, Y. Beyene, K. Semagn, X. Zhang, M.E. Sorrells, and J.L. Jannink. 2014. Optimal design of preliminary yield trials with genome-wide markers. Crop Sci. 54:48–59. doi:10.2135/cropsci2013.03.0154

Fisher, R.A. 1935. The design of experiments. Oliver and Boyd, Oxford, UK.

Gauch, H.G., J.T.G. Hwang, and G.W. Fick. 2003. Model evaluation by comparison of model-based predictions and measured values. Agron. J. 95:1442–1446. doi:10.2134/agronj2003.1442

Gilmour, A.R., B.R. Cullis, and A.P. Verbyla. 1997. Accounting for natural and extraneous variation in the analysis of field experiments. J. Agric. Biol. Environ. Stat. 2:269–293. doi:10.2307/1400446

Gleeson, A.C., and B.R. Cullis. 1987. Residual maximum likelihood (REML) estimation of a neighbour model for field experiments. Biometrics 43:277–287. doi:10.2307/2531812

Gonçalves, E., A. St Aubyn, and A. Martins. 2010. Experimental designs for evaluation of genetic variability and selection of ancient grapevine varieties: A simulation study. Heredity 104:552–562. doi:10.1038/hdy.2009.153

Grondona, M.O., and N. Cressie. 1991. Using spatial considerations in the analysis of experiments. Technometrics 33:381–392. doi:10.1080/00401706.1991.10484867

Grondona, M.O., J. Crossa, P.N. Fox, and W.H. Pfeiffer. 1996. Analysis of variety yield trials using two-dimensional separable ARIMA processes. Biometrics 52:763–770. doi:10.2307/2532916

INIA. 2017. Evaluación de cultivares. Inst. Nacl. Invest. Agropecu. http://www.inia.uy/productos-y-servicios/servicios-t%C3%A9cnicos/Evaluación-de-Cultivares (accessed 13 Nov. 2018).

Koch, E., and J. Rigney. 1951. A method of estimating optimum plot size from experimental data. Agron. J. 43:17–21. doi:10.2134/agronj1951.00021962004300010005x

Kravchenko, A.N., G.R. Robertson, S.S. Snap, and A.J.M. Smucker. 2006. Using Information about spatial variability to improve estimates of total soil carbon. Agron. J. 98:823–829. doi:10.2134/agronj2005.0305

Legendre, P. 1993. Spatial autocorrelation: Trouble or new paradigm? Ecology 74:1659–1673. doi:10.2307/1939924

Legendre, P., M.R.T. Dale, M.J. Fortin, P. Casgrain, and J. Gurevitch. 2004. Effects of spatial structures on the results of field experiments. Ecology 85:3202–3214. doi:10.1890/03-0677

Lehmann, E.L., and G. Casella. 1998. Theory of point estimation. 2nd ed. Springer, New York.

Masood, A.M., K. Farooq, Y. Mujahid, and M.Z. Anwar. 2008. Improvement in precision of agricultural field experiments through design and analysis. Pak. J. Life Social Sci. 6:89–91.

Moehring, J., E.R. Williams, and H.P. Piepho. 2014. Efficiency of augmented p-rep designs in multi-environmental trials. Theor. Appl. Genet. 127:1049–1060. doi:10.1007/s00122-014-2278-y

Müller, B.U., K. Kleinknecht, J. Möhring, and H.P. Piepho. 2010. Comparison of spatial models for sugar beet and barley trials. Crop Sci. 50:794–802. doi:10.2135/cropsci2009.03.0153

Patterson, H.D., and E.R. Williams. 1976. A new class of resolvable incomplete block designs. Biometrika 63:83–92. doi:10.1093/biomet/63.1.83

Pebesma, E.J., and R.S. Bivand. 2005. Classes and methods for spatial data in R. R News 5(2):9–13. https://cran.r-project.org/doc/Rnews/Rnews_2005-2.pdf (accessed 3 Nov. 2018).

Piepho, H.P., and E.R. Williams. 2010. Linear variance models for plant breeding trials. Plant Breed. 129:1–8.

Piepho, H.P., E.R. Williams, and V. Michel. 2015. Beyond Latin squares: A brief tour of row-column designs. Agron. J. 107:2263–2270. doi:10.2134/agronj15.0144

Piepho, H.P., E.R. Williams, and V. Michel. 2016. Nonresolvable row-column designs with an even distribution of treatment replications. J. Agric. Biol. Environ. Stat. 21:227–242. doi:10.1007/s13253-015-0241-2

Pinheiro, J., D. Bates, S. DebRoy, D. Sarkar, S. Heisterkamp, B. Van Willigen, R Core Development Team. 2013. nlme: Linear and nonlinear mixed effects models. R package version 3.1-109. R Found. Stat. Comput., Vienna.

Qiao, C.G., K.E. Basford, I.H. DeLacy, and M. Cooper. 2000. Evaluation of experimental designs and spatial analyses in wheat breeding trials. Theor. Appl. Genet. 100:9–16. doi:10.1007/s001220050002

R Core Team. 2016. R: A language and environment for statistical computing. R Found. Stat. Comput., Vienna.

Richter, C., and B. Kroschewski. 2012. Geostatistical models in agricultural field experiments: Investigations based on uniformity trials. Agron. J. 104:91–105. doi:10.2134/agronj2011.0100

Sripathi, R., P. Conaghan, D. Grogan, and M.D. Casler. 2017. Field design factors affecting the precision or ryegrass forage yield estimation. Agron. J. 109:858–869. doi:10.2134/agronj2016.07.0397

Stroup, W.W., P.S. Baenzinger, and D.K. Mulitze. 1994. Removing spatial variation from wheat yield trials: A comparison of methods. Crop Sci. 34:62–66. doi:10.2135/cropsci1994.0011183X003400010011x

USDA-ARS. 1992. Germplasm Resources Information Network (GRIN) database. *Triticum aestivum* L. subsp. *aestivum*. POACEAE 'Nogal'. Natl. Germplasm Resour. Lab., Natl. Genet. Resour. Progr., Beltsville, MD. https://npgsweb.ars-grin.gov/gringlobal/accessiondetail.aspx?id=1457618 (accessed 1 Oct. 1992).

Williams, E.R., J.A. John, and D. Whitaker. 2006. Construction of resolvable spatial row–column designs. Biometrics 62:103–108. doi:10.1111/j.1541-0420.2005.00393.x

Williams, E.R., A.C. Matheson, and C.E. Harwood. 2002. Experimental design and analysis for tree improvement. CSIRO Publ., Clayton, VIC.

Williams, E.R., and H.P. Piepho. 2013. A comparison of spatial designs for field variety trials. Aust. N. Z. J. Stat. 55:253–258. doi:10.1111/anzs.12034

Williams, E.R., and H.P. Piepho. 2018. An evaluation of error variance bias in spatial designs. J. Agric. Biol. Environ. Stat. 23:83–91. doi:10.1007/s13253-017-0309-2