

Evaluation of Moving Grid Adjustment (MGA) Method in Field Variation Control

Jixiang Wu*, Johnie N. Jenkins, Jack C. McCarty

Genetics and Sustainable Agriculture Research Unit, USDA-ARS, Mississippi State, MS, USA

Email: *Jixiang.wu@usda.gov

How to cite this paper: Wu, J.X., Jenkins, J.N. and McCarty, J.C. (2024) Evaluation of Moving Grid Adjustment (MGA) Method in Field Variation Control. *Open Journal of Statistics*, 14, 450-466.

<https://doi.org/10.4236/ojs.2024.145019>

Received: August 29, 2024

Accepted: October 22, 2024

Published: October 25, 2024

Copyright © 2024 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Spatial variation is often encountered when large scale field trials are conducted which can result in biased estimation or prediction of treatment (*i.e.* genotype) values. An effective removal of spatial variation is needed to ensure unbiased estimation or prediction and thus increase the accuracy of field data evaluation. A moving grid adjustment (MGA) method, which was proposed by Technow, was evaluated through Monte Carlo simulation for its statistical properties regarding field spatial variation control. Our simulation results showed that the MGA method can effectively account for field spatial variation if it does exist; however, this method will not change phenotype results if field spatial variation does not exist. The MGA method was applied to a large-scale cotton field trial data set with two representative agronomic traits: lint yield (strong field spatial pattern) and lint percentage (no field spatial pattern). The results suggested that the MGA method was able to effectively separate the spatial variation including blocking effects from random error variation for lint yield while the adjusted data remained almost identical to the original phenotypic data. With application of the MGA method, the estimated variance for residuals was significantly reduced (62.2% decrease) for lint yield while more genetic variation (29.7% increase) was detected compared to the original data analysis subject to the conventional randomized complete block design analysis. On the other hand, the results were almost identical for lint percentage with and without the application of the MGA method. Therefore, the MGA method can be a useful addition to enhance data analysis when field spatial pattern exists.

Keywords

Spatial Variation, Linear Mixed Model Approach, Experimental Design, Moving Grid Adjustment, Crop Trial

1. Introduction

Field variation in a large field trial could cause a significant issue in statistical analysis and conclusions. Therefore, spatial variation control has become a crucial step to screen large-scale plant genotypes in a large field trial. Conventional blocking designs (like randomized complete block, split plots, Latin square, and incomplete block designs, etc.) [1] have been commonly used for field variation control since Fisher proposed his “The Arrangement of Field Experiments” [2]. Fisher’s experimental design principles (1926) provide some measure of protection against spatial variation, many statisticians agree that additional layers of spatial control are beneficial [3]-[5]. Interestingly, though the spatial analysis was proposed by Papadakis just 11 years after Fisher’s blocking design [6], its applications to field trial analysis have been unpopular compared to Fisher’s blocking designs. A possible reason is the difficulty of integrating spatial analysis into the conventional field trial data process. Therefore, investigating spatial variation control methods featured with efficiency and integrability will be helpful to enhance field trial analysis.

Fisher’s blocking design (1926) is ANOVA (analysis of variance) based thus simplicity is one of the major features. Under the principle of a linear model, Fisher’s blocking design can be extended to multi-dimension or hierarchical blocking designs. Examples include Latin square and alpha-lattice designs [1] [7]. An efficient blocking design should be assumed that within-block variability is minimized while between-block variability being maximized [8]. Because uniformity within a large block is unlikely in most field trials, conventional blocking designs would be rarely satisfactory. On the other hand, it is more likely that plot-to-plot variation within a block may be affected by competition between genotypes with the trial [9]-[11] by soil variation and fertility and previous land use [12] [13].

Removing spatial variation from a field trial has been a crucial goal for many statisticians [14]. Over 80 years ago Papadakis [6] believed that traditional blocking in field experiments might not adequately represent soil fertility patterns and he instead proposed adjusting the yield of each plot by the performance of the neighboring plots. According to Bartlett [15], the adjustment method suggested by Papadakis (1937) was approximately valid and sometimes useful in experiments where a block size is large. Unfortunately, Papadakis’s method did not receive as serious attention as Fisher’s blocking designs until the 1970’s [16]-[18]. The nearest neighbor adjustment (NNA) method was claimed to be a relatively simple yet effective method for accounting for within-block variation [18] [19]. Later, a random field approach [20] was proposed as an improvement on the NNA methods. The key point of using the NNA methods is to develop covariates to adjust the original data [21]. Stroup *et al.* (1994) compared RCB analysis with the NNA methods in a multi-environment winter-wheat trial data and concluded that the NNA approach was better as evidenced by a lower coefficient of variation and greater ability to distinguish the differences among cultivars. A recent study showed that modified Papadakis covariates incorporating spatial positions could enhance the adjustment efficiency [21]. However, such a modified covariate may

require more information that many data lack and thus it could reduce the application of this method.

Above mentioned covariates are based on the residuals obtained from a particular experimental design (equivalent to a particular statistical model). In other words, the calculation of these covariates is model-dependent. Recently, a moving grid adjustment (MGA) method was proposed by Technow [22] and a covariate is also required to adjust field spatial patterns. Compared to Papadakis's covariate, the MGA covariate has two noticeable differences. The first one is that the MGA covariate is based on phenotypic data rather than residuals, implying that the MGA adjustment can stand alone. The second one is that each MGA covariate has the flexibility to catch different directions of spatial patterns. In addition, the corresponding R package for the MGA method `mvngGrAd` is publicly available [22]. Numerical evaluation regarding the adjustment efficiency via Monte Carlo simulations will reveal more statistical properties under the conventional RCB design.

Therefore, there were two objectives to be targeted in this study. Our first objective was to numerically validate the efficacy of the MGA methods via simulated data generated from a commonly used RCB design with one-way and two-way spatial patterns. Correlation coefficients between adjusted and control data were used to determine the efficiency of the MGA method. The second objective was to apply the MGA method to analyze a cotton trial data set with emphasis on cotton yield (sensitive to field spatial pattern) and lint percentage (non-sensitive to field spatial pattern). The main purpose of this study is to provide a useful method to enhance breeding data analysis when spatial patterns are significant.

2. Materials and Methods

2.1. Moving Grid Adjustment (MGA) Method

Assuming that a field trial is arranged in a (nearly) rectangular format, each plot (called grid in this study) is associated with its corresponding row and column indexes. With the MGA method [22], the covariate denoted as x_{ij} is calculated as follows

$$x_{ij} = \frac{\sum_{i'j'} y_{i'j'} I(y_{i'j'} \in G_{ij})}{\sum_{i'j'} I(y_{i'j'} \in G_{ij})} \quad (1)$$

The grid of the plot with row i and column j is denoted by G_{ij} and $I(\cdot)$ is an indicator function that takes the value 1 if a plot is included in G_{ij} or 0 if not. The observed phenotypic values from all plots which could be included in G_{ij} are denoted by $y_{i'j'}$. The covariate x_{ij} is taken as a measure of the growing conditions for the plot with row i and column j and will be used as a covariate to calculate an adjusted phenotypic value $y_{ij,adj}$ according to the following formula,

$$\hat{y}_{ij,adj} = y_{ij} - \hat{b}(x_{ij} - \bar{x}) \quad (2)$$

where \bar{x} is the mean of all x_{ij} and b is the regression coefficient, which can be estimated from the following simple linear regression model:

$$y_{ij} = a + bx_{ij} + e_{ij} \quad (3)$$

Since each adjusted data $y_{ij,adj}$ from Equation (2) shares the identical information as the original data y_{ij} does, it is very convenient to analyze the adjusted data $y_{ij,adj}$ subject to different statistical models.

2.2. Simulation Data

A field layout with 25 rows and 40 columns (1,000 plots in total) was used for our Monte Carlo simulation study. With these 1000 plots, an RCB design with four reps where 250 entries/treatments were randomly assigned. However, interested readers can use different sizes of field layouts (different row and column numbers) and/or different treatments for additional investigation. One hundred (100) simulation data sets were generated for each of the four settings provided in **Table 1**.

Table 1. Four variance components settings for simulation.

Parameter [†]	Setting 1 [‡]	Setting 2	Setting 3	Setting 4
V_G	50	50	50	50
V_R	0	25	50	0
V_C	0	25	0	50
V_e	25	25	25	25

[†]: V_G = variance component for genotype, V_R = variance component for row direction, V_C = variance component for column direction, and V_e = variance component for random error. [‡]: Setting 1: no field spatial pattern; Setting 2: spatial pattern exists in both directions; Setting 3: spatial pattern in row direction only; and Setting 4: spatial pattern in column direction only.

In our simulation study, we used the following procedures to generate simulation data:

Step 1: Using the variance components of setting 1 to generate a simulated data vector without spatial patterns (**Table 1**) as control data, namely y_0 .

Step 2: Using each set of variance components of V_R (variance for row direction) and V_C (variance for column direction) from each of the Settings 2 to 4 to generate row and column effects. Sort the simulated row and column effects to generate spatial effects in both row and/or column directions, namely **sp**.

Step 3: Using y_0 from Step 1 and **sp** from Step 2 to make pseudo-phenotypic data with spatial patterns, namely, $y_1 = y_0 + \mathbf{sp}$.

2.3. Actual Data Background

The data used in this study as demonstration/application were from our previous study [23]. The data included 188 cotton recombinant inbred (RI) (F8) lines developed by the single-hill procedure [24] from the cross of HS 46 (P_1) MARCABUCAG8US-1-8 (P_2) [25]. Originally, these 188 RI lines with their two parental lines (P_1 and P_2) and a check, 'Stoneville 474' (ST474) were planted at the Plant

Science Research Center, Mississippi State, MS in 1999. The experiment design was an RCB design with four replicates, where each replicate was in the column direction. Each block consisted of 205 two-row plots with 12 m in length, 0.97 m of between-row spacing, and approximately 10 cm of plant spacing. Within each block, two parental lines (P_1 and P_2) were repeated twice and the check ST474 was repeated 13 times. The entire field was almost a rectangular arrangement with 17 rows and 52 columns. Rows were numbered from south to north direction while columns were numbered from west to east direction. Data collected from each plot included the following traits: lint yield per hectare (LY, kg) and lint percentage (LP, %).

2.4. Data Processing

For our simulation study, both y_0 (data without spatial pattern, control) and y_1 (data with spatial pattern) were adjusted by the MGA method and the adjusted data are named by ay_0 and ay_1 accordingly. The following correlation coefficients were calculated: y_0 and y_1 , y_0 and ay_1 , and ay_0 and ay_1 . For simplification, we named these three correlation coefficients as r_{01} , ar_{01} , and aar_{01} , respectively. The correlation r_{01} indicates the impact of spatial patterns on the phenotype data. The correlation coefficient ar_{01} indicates the adjustment efficiency of the MGA method on the data with a spatial pattern. The correlation coefficient aar_{01} indicates similarity the adjusted data from both data with and without a spatial pattern.

For application, both lint yield and lint percentage were adjusted by the MGA method. Both original data and adjusted data were analyzed subject to CR (completely randomized) and RCB (randomized complete block) designs. A minimum norm quadratic unbiased estimation (MINQUE) method [26] [27] was used to estimate variance components subject to these two designs. In addition to variance component estimation, we estimated the residuals for both original and adjusted data subject to these two designs. The R library minque [28] was applied to estimate variance components and residuals.

Both simulated data and actual cotton data were adjusted by MGA with the R package mvngGrAd [22]. Heatmaps were generated from the R package desplot [29]. All R scripts including integrating the R functions from the cited packages were developed and customized by the senior author of this study. All data processing and graphics generating were conducted at the RStudio platform [30] [31].

3. Simulation Results

As addressed in Section 2, we define y_0 representing phenotypic data without a spatial pattern thus y_0 can be considered as the control group; y_1 representing phenotypic data with a spatial pattern for each of the four settings (Table 1). Both ay_0 and ay_1 are the adjusted data for y_0 and y_1 by the MGA method, respectively. For simplification, we only report three correlation coefficients as r_{01} , ar_{01} , and aar_{01} , where, r_{01} between y_0 and y_1 indicates the impact of spatial patterns on phenotypic data; ar_{01} is between y_0 and ay_1 indicating the adjustment efficiency of the MGA method on the data with spatial patterns; aar_{01} is between

ay_0 and ay_1 indicating the similarity between adjusted data from data with and without a spatial pattern. The correlation coefficients are summarized (minimum, maximum, mean values) from 100 simulations for each of the four settings and are provided in **Tables 2-5**, separately.

The high correlation coefficients, r_{01} , aar_{01} , and ar_{01} , showed that the data for both y_0 and y_1 and ay_0 and ay_1 are identical or nearly identical when there

Table 2. Correlation coefficient among simulated non-adjusted and adjusted data[†] for setting 1.

	Min	Max	Mean
r_{01}^{\ddagger}	1.0000	1.0000	1.0000
aar_{01}	1.0000	1.0000	1.0000
ar_{01}	0.9929	1.0000	0.9979

[†]: y_0 = simulated data without spatial pattern (control); y_1 = simulated data with spatial pattern; ay_0 = adjusted data for y_0 ; and ay_1 = adjusted data for y_1 . [‡]: r_{01} = correlation between y_0 and y_1 ; aar_{01} = correlation between ay_0 and ay_1 ; ar_{01} = correlation between y_0 and ay_1 .

Table 3. Correlation coefficient among simulated non-adjusted and adjusted data[†] for setting 2.

	Min	Max	Mean
r_{01}^{\ddagger}	0.7488	0.8048	0.7793
aar_{01}	0.9522	0.9824	0.9694
ar_{01}	0.9676	0.9820	0.9753

[†]: y_0 = simulated data without spatial pattern (control); y_1 = simulated data with spatial pattern; ay_0 = adjusted data for y_0 ; and ay_1 = adjusted data for y_1 . [‡]: r_{01} = correlation between y_0 and y_1 ; aar_{01} = correlation between ay_0 and ay_1 ; ar_{01} = correlation between y_0 and ay_1 .

Table 4. Correlation coefficient among simulated non-adjusted and adjusted data[†] for setting 3.

	Min	Max	Mean
r_{01}^{\ddagger}	0.7506	0.8026	0.7786
aar_{01}	0.9581	0.9946	0.9785
ar_{01}	0.9716	0.9849	0.9801

[†]: y_0 = simulated data without spatial pattern (control); y_1 = simulated data with spatial pattern; ay_0 = adjusted data for y_0 ; and ay_1 = adjusted data for y_1 . [‡]: r_{01} = correlation between y_0 and y_1 ; aar_{01} = correlation between ay_0 and ay_1 ; ar_{01} = correlation between y_0 and ay_1 .

Table 5. Correlation coefficient among simulated non-adjusted and adjusted data[†] for setting 4.

	Min	Max	Mean
r_{01}^{\ddagger}	0.7412	0.8012	0.7692
aar_{01}	0.9414	0.9865	0.9708
ar_{01}	0.9595	0.9796	0.9727

[†]: y_0 = simulated data without spatial pattern (control); y_1 = simulated data with spatial pattern; ay_0 = adjusted data for y_0 ; and ay_1 = adjusted data for y_1 . [‡]: r_{01} = correlation between y_0 and y_1 ; aar_{01} = correlation between ay_0 and ay_1 ; ar_{01} = correlation between y_0 and ay_1 .

is no spatial pattern ($sp = 0$) (setting 1, **Table 2**). When a spatial pattern exists in both directions (setting 2), the mean correlation coefficient (r_{01}) between y_0 and y_1 was 0.779 while the mean correlation coefficients between ay_0 and ay_1 (aar_{01}) or ay_1 and y_0 (ar_{01}) were 0.969 and 0.975, respectively (**Table 3**). These two correlation coefficients were significantly higher than that between y_0 and y_1 . Similar results can be observed for settings 3 and 4 (**Table 4** and **Table 5**).

Figures 1-3 showed the heatmaps for one set of simulated data of y_0 , sp , y_1 ,

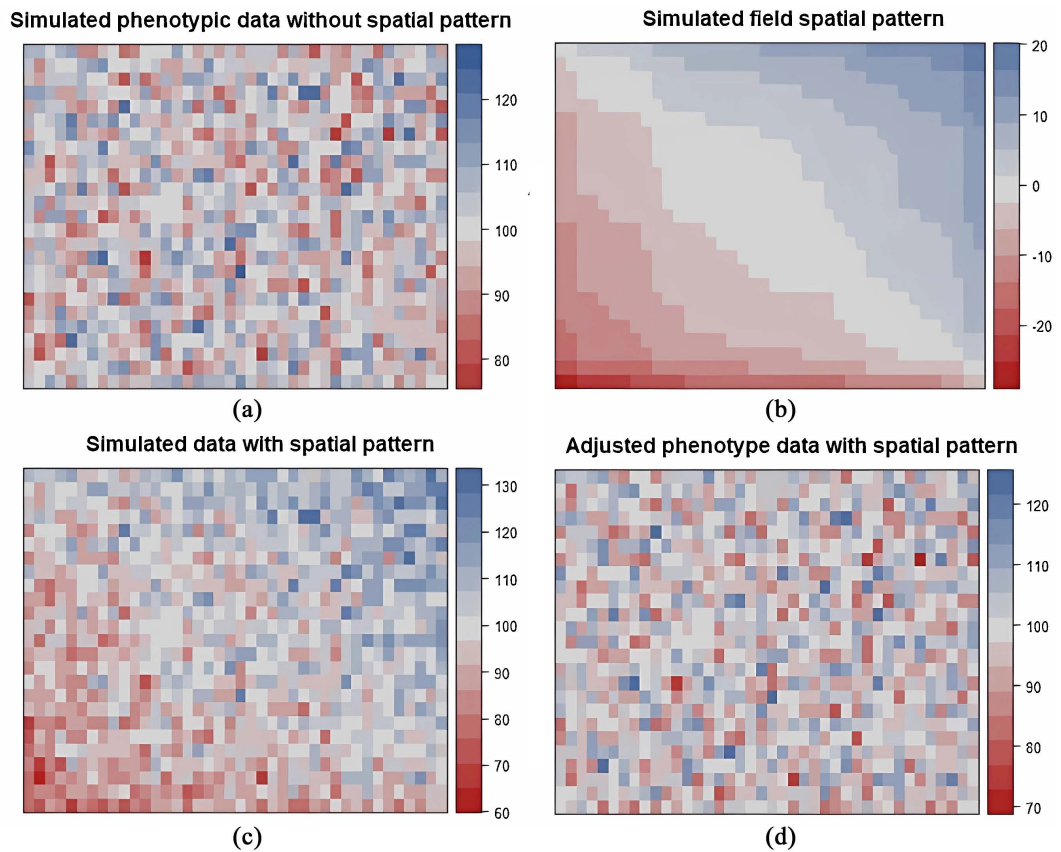


Figure 1. Heat maps for simulated data without spatial pattern (a), simulated spatial pattern in two directions (b), and simulated data with spatial pattern ((c) = (a) + (b)), and adjusted data with MGA method (d).

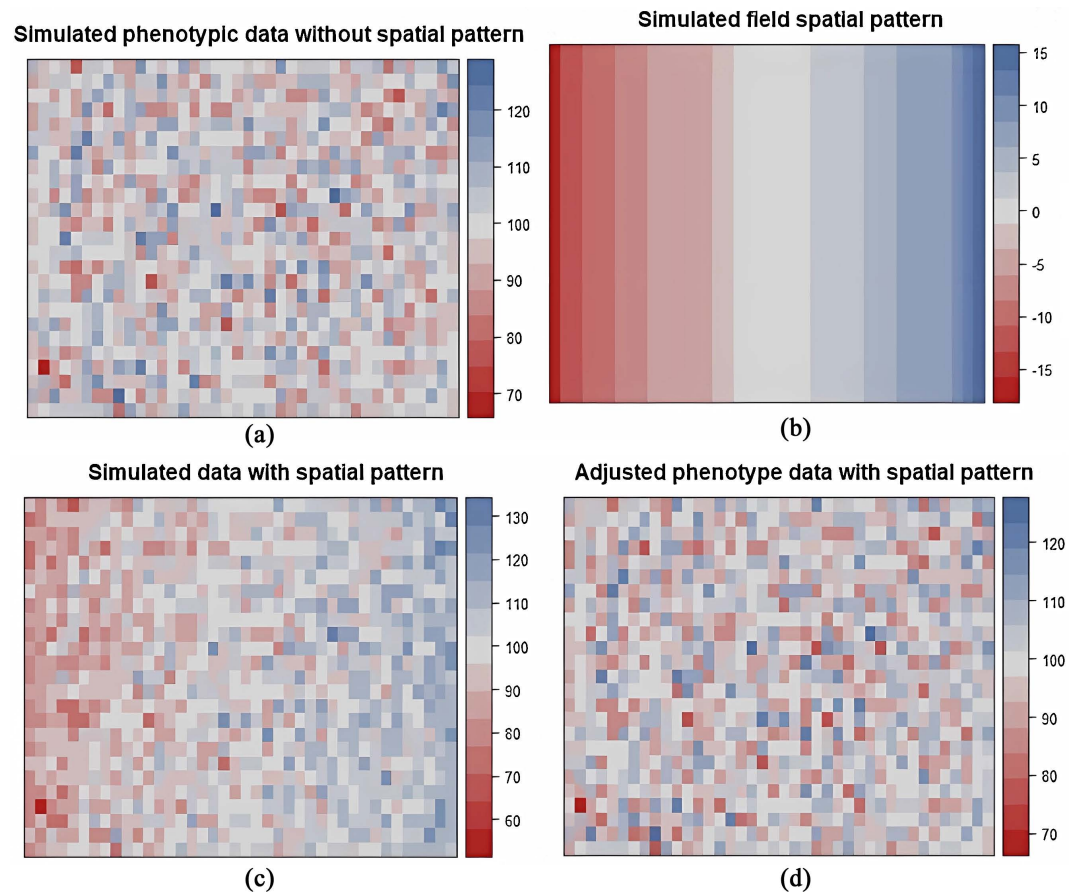


Figure 2. Heat maps for simulated data without spatial pattern (a), simulated spatial pattern in row direction (b), and simulated data with spatial pattern ((c) = (a) + (b)), and adjusted data with MGA method (d).

and ay_1 for settings 2 to 4, respectively. These three figures showed that the heatmap for the adjusted data ay_1 is very similar to that for the control data y_0 . The heatmaps from these three different settings are consistent with the correlation analysis. Therefore, according to the simulation results (Tables 2-5 and Figures 1-3), we can conclude that 1) the MGA method can effectively remove field spatial variation if it does exist and 2) the MGA method will not change phenotype results if field spatial variation does not exist.

4. Application

4.1. Heatmaps for Observed Cotton Lint Yield and Lint Percentage

The heatmaps for observed lint yield and lint percentage data are provided in Figure 4(a) and Figure 5(a) (Figure 4 and Figure 5). Lint yield showed that spatial patterns were associated with soil conditions in both column (west-east) and row (north-south) directions (Figure 4(a)). The heatmap showed that lint yield gradually increased from west to east as indicated in column direction and gradually increased from south to north as indicated in the row direction (Figure 4(a)). In other words, more cotton yielded in the NE (north-east) area compared to the SW (south-west) area. Such a spatial pattern likely was associated with the

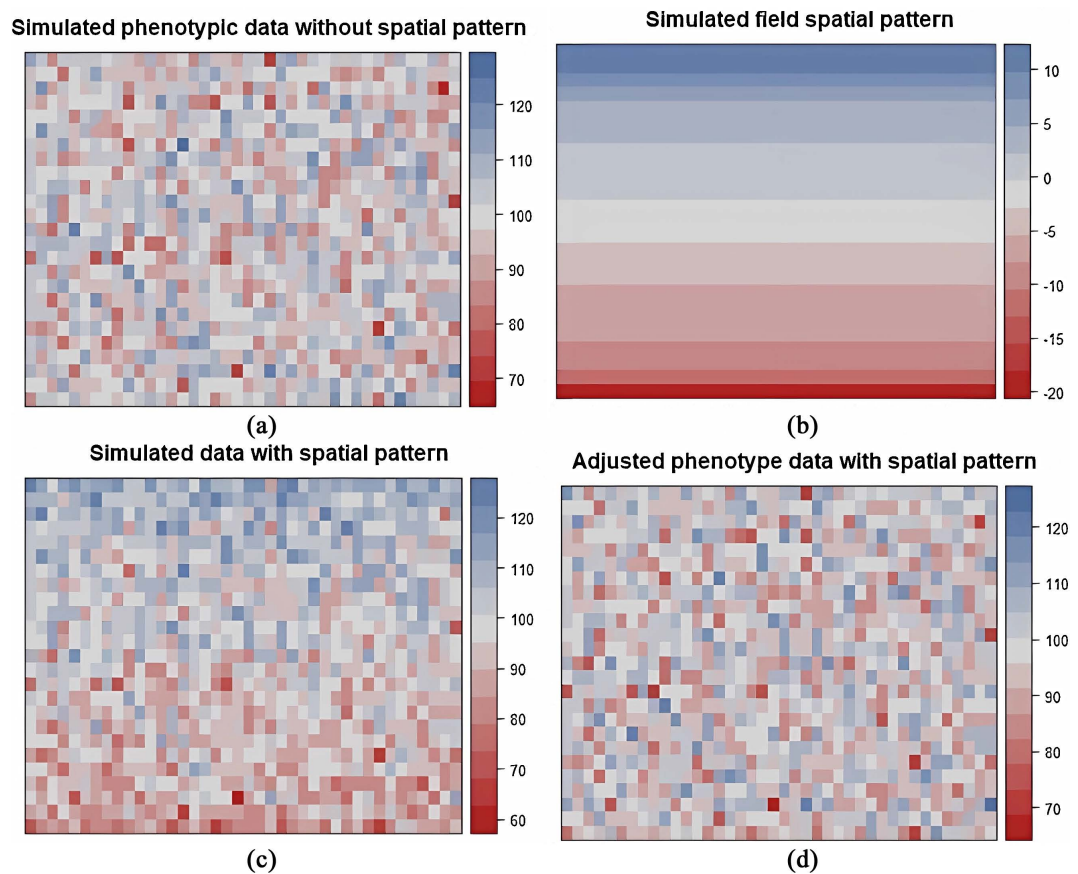


Figure 3. Heat maps for simulated data without spatial pattern (a), simulated spatial pattern in column direction (b), and simulated data with spatial pattern ((c) = (a) + (b)), and adjusted data with MGA method (d).

gradual changes in soil conditions including different soil types, moisture levels, and nutrient components. The heatmap for the residuals for the original yield data being analyzed subject to CR (completely randomized) design also showed a similar pattern as the phenotypic data (Figure 4(b)). It suggested that lint yield was sensitive to the field condition as shown in the heatmaps on both phenotypic data and estimated residuals. On the other hand, regarding lint percentage, no visible field spatial variation trend was observed for the phenotypic data and estimated residuals subject to a CR design analysis (Figure 5(a) and Figure 5(b)), suggesting that this trait was not sensitive to field conditions and adjustment is not needed.

4.2. Heatmaps for Adjusted Cotton Lint Yield and Lint Percentage

Both lint yield and lint percentage data were adjusted by the MGA method. Similarly, the heatmaps for adjusted lint yield and adjusted lint percentage data are provided in Figure 4 and Figure 5. Adjusted data for lint yield showed that spatial patterns were removed and the heatmap appeared more uniform in both column (west-east) and row (north-south) directions (Figure 4(c)). The heatmap for the estimated residuals from the adjusted lint yield data subject to CR design analysis also appeared much more uniform compared to that from the original lint yield data (Figure 4(d)). As expected, the heatmaps for both the adjusted lint

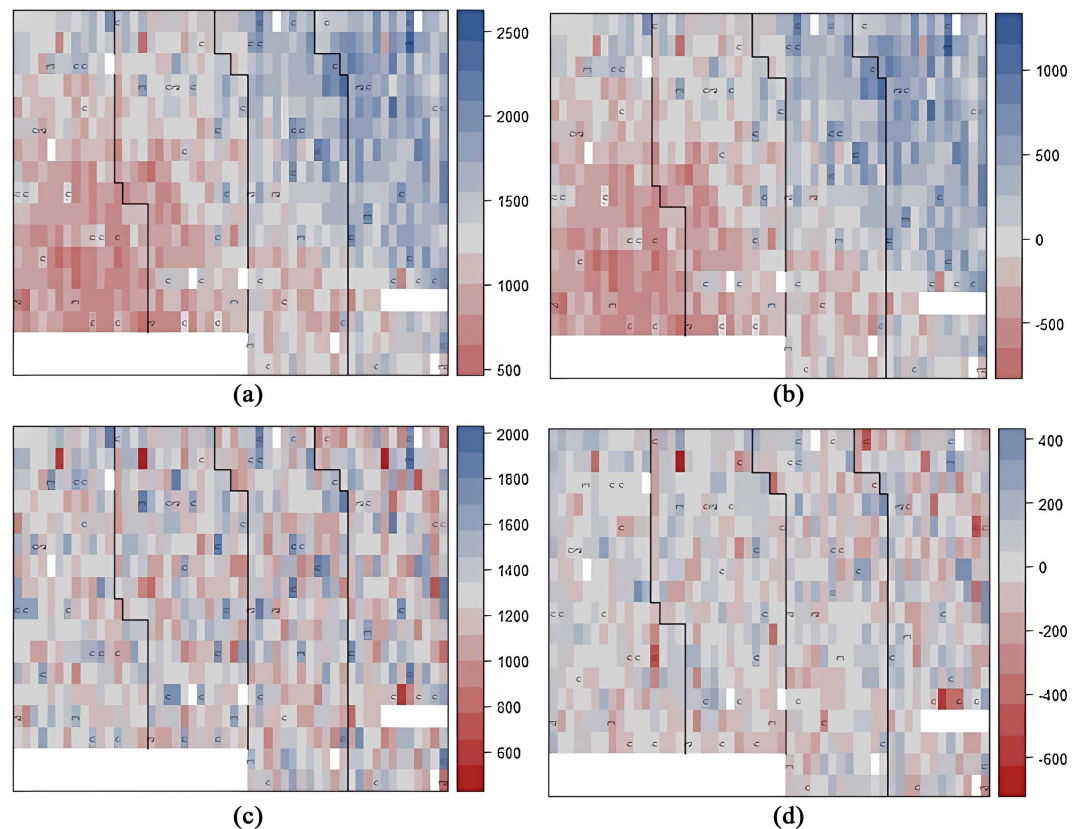


Figure 4. Heat maps for the original lint yield data (a), residuals subject to CR (completely randomized) design from the original data (b), and adjusted lint yield data (c), and residuals subject to CR (completely randomized) design from the adjusted data (d).

percentage and the corresponding estimated residuals were similar to those for the original lint percentage data and the corresponding residuals subject to CR design analysis (**Figure 5(c)** and **Figure 5(d)**). It suggested that the MGA method would not change the phenotypic data if spatial patterns do not exist.

4.3. Variance Components Estimations Subject to CR and RCB Designs

The heatmaps for the estimated residuals subject to CR design for both agronomic traits were presented above. It is also important to numerically compare the results from the original and adjusted data subject to an RCB (randomized complete block) design. Variance components subject to two experimental designs for both original and adjusted data were estimated by linear mixed model approach [27] with minque package (Wu, 2014). The estimated variance components are provided in **Table 6** and **Table 7** for lint yield and lint percentage, respectively.

Results in **Table 6** showed that genotypic effects for lint yield were not significant subject to a CR design analysis because of the dominant variance due to the residuals. However, significant genotypic effects were detected subject to RCB design analysis. The results indicated that the RCB design used for the trial was able to remove the variation due to the column direction. The results agreed with

the field spatial pattern detected in this study (Figure 4(b)). With the adjusted lint yield data, the estimated genotypic and residual variance components were

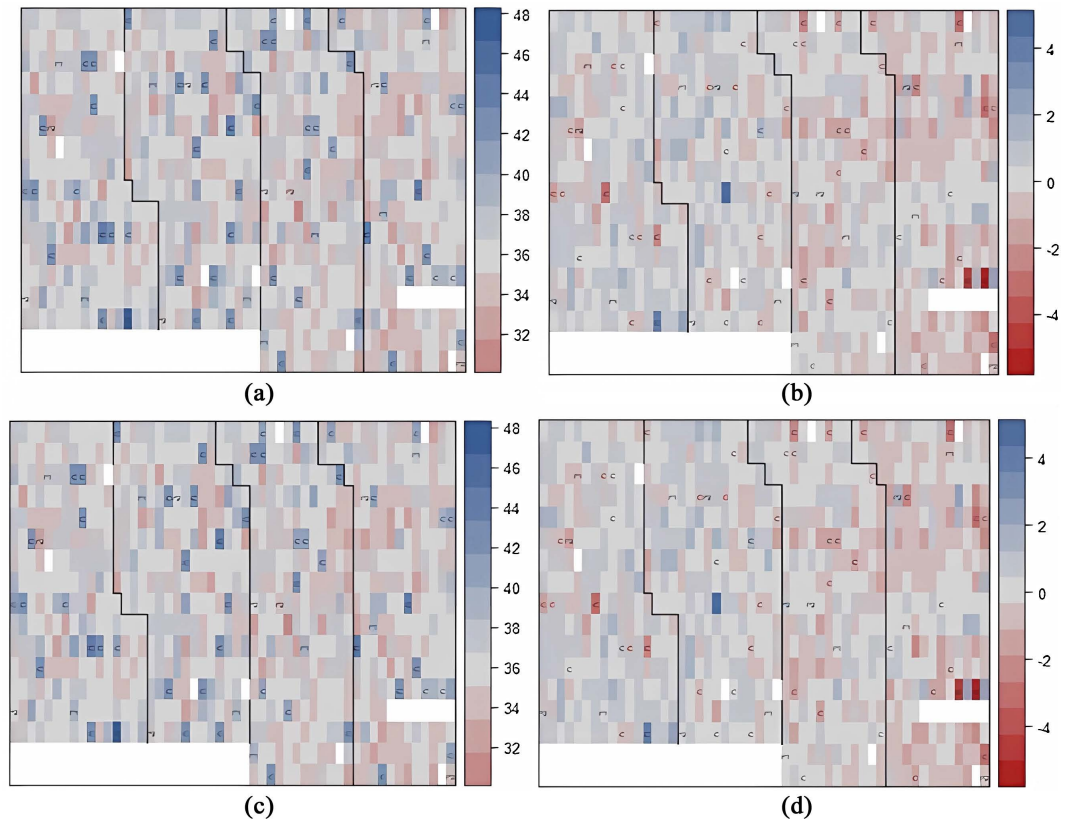


Figure 5. Heat maps for the original lint percentage data (a), residuals subject to CR (completely randomized) design from the original data (b), and adjusted lint percentage data (c), and residuals subject to CR (completely randomized design from the adjusted data (d).

Table 6. Variance components for original and adjusted lint yield subject to two experimental designs[†].

Variance component [‡]	CR design			
	Original data		Adjusted data	
V_G	-3.05	NS	15,426.61	**
V_e	117,378.91	**	21,458.93	**
	RCB design			
	Original data		Adjusted data	
V_G	11,885.25	**	15,420.95	**
V_B	61,987.00	NS	-0.02	NS
V_e	56,801.03	**	21,481.87	**

[†]: CR design = completely randomized design; RCB design = randomized complete block design. [‡]: V_G = variance component for genotype effects; V_B = variance component for block effects; and V_e = variance component for residuals.

Table 7. Variance components for original and adjusted lint percentage subject to two experimental designs[†].

Variance component [‡]	CR design			
	Original data		Adjusted data	
V_G	1.92	**	1.93	**
V_e	0.77	**	0.75	**
	RCB design			
	Original data		Adjusted data	
V_G	1.95	**	1.96	**
V_B	0.11	NS	0.10	NS
V_e	0.66	**	0.66	**

[†]: CR design = completely randomized design; RCB design = randomized complete block design. [‡]: V_G = variance component for genotype effects; V_B = variance component for block effects; and V_e = variance component for residuals.

nearly identical between CR and RCB designs. In addition, block effects which were significant in the original data for lint yield but no significant block effects were detected in the adjusted data (Table 6). Block effects were trivial and non-significant for lint percentage. As expected, both genotypic variances and residual variances were comparable between CR and RCB experimental designs. In addition, these estimated variance components from adjusted data were comparable to those from the original data.

Based on the results from lint yield and lint percentage, we can conclude that 1) the spatial pattern would be visible if it does exist; 2) spatial pattern including block effects can be removed with the MGA method; and 3) results would remain the same if a spatial pattern doesn't exist.

5. Discussion

It has been a common practice that large-scale entries are evaluated for either early screening of breeding lines or phenotyping for association mapping study. However, field spatial variation is a common phenomenon in a field trial, especially for a large field trial. Such field spatial variation could greatly impact the conclusions such as inflation/deflation of differences among entries being evaluated. Field spatial variation for a large trial normally results in inflated residual variance and also can be visualized via heatmap. Effectively removing field variation is highly needed and can be indicated by reduced variance and/or depatterned heatmap for residuals.

There are two major ways to control field spatial variation. One major way is to apply various experimental designs such as conventional block designs and augmented experimental designs [1] [32]-[34]. If field conditions are known prior to a field trial being executed, appropriate blocking designs would help control field

variation well for a small block size. However, the assumption of within-block uniformity is easily violated when block size is large due to many uncontrollable factors [8], including the row-column incomplete block designs. Among various blocking designs, row-column designs (sometimes called rectangular design) [35]-[37] were more efficient than the conventional RCB designs for different trials and experiment sizes [38]. With our limited number of field trial data processing, we also noticed that the rectangular provided equally well or better field variation control compared to the other blocking designs [39]-[41]. Without spatial analysis, the rectangular design is a recommended option over the conventional RCB designs. The other major approach is to apply spatial statistical analysis which is NNA (near-neighbor adjustment) based. Though NNA based spatial analysis was initially proposed in 1937 (Papadakis 1937), the application of NNA based method has been very few compared to Fisher's based blocking designs.

The efficiency of the MGA method was numerically evaluated by Monte Carlo simulation. The simulation results showed that the MGA method can be used to effectively remove a field spatial pattern in one-way or two-way direction as evidenced by the heatmaps (Figures 1-3) and high coefficient of correlation between control data and adjusted data (Tables 2-5). The application of MGA method to an actual cotton trial data also showed the similar conclusion (Figure 4 & Figure 5; Table 6 & Table 7). Our results also suggested that the MGA method can effectively remove block effects (Table 6). In one of our previous studies, we applied several augmented designs to analyze the same lint yield and lint percentage data. We found that the rectangular blocking design (row-column) was better than other blocking designs for lint yield because both row and column variations were significantly captured. Similar results were found among different experimental designs for lint percentage because no significant row or column effects were detected [39].

The estimated genotype and residual variance components for lint yield from the adjusted data by the MGA method in this study (Table 6) were compared with the estimated variance components subject to the rectangular (row-column) model reported in one of our previous studies [39] (Table 4). Interestingly, the results were comparable between two studies (15,421 vs 16,400 and 21,482 vs 20,400, equivalent to 6.0% and 5.3% in difference for genotype and residual variance components, respectively). The results suggest that both the MGA method and the row-column augmented blocking design can effectively remove the two-way field spatial variation and thus they both function equally well regarding this data. However, it will be interesting to compare these two methods with more complicated field spatial patterns.

Data adjustment by the MGA can stand alone. Therefore, this adjustment method has a great flexibility to integrate with other statistical analyses including various genetic model analyses and association mapping study [28] [42]-[44]. Many crop trials are repeated over locations or years. It could be statistically challenging to analyze multi-environment data with a GGE (genotype and genotype

× environment interaction) model simultaneously when field spatial patterns vary among environments. However, spatial analysis model is usually fitted for each environment [45]-[48] to produce a spatially adjusted genotype means for each environment. With application of the MGA method and other spatial analysis methods, the spatially adjusted data are then combined for the second stage for an across-environment analysis [49] [50].

6. Conclusion

In conclusion, both our simulation study and actual data analysis showed that the MGA method has a desirable adjustment efficiency when a field spatial pattern exists in single or both directions. In addition, because the R package for the MGA method has been publicly available [22], it will be flexible to integrate with other statistical methods under the RStudio platform [30] for more complicated statistical analysis. Therefore, this method can be a great addition to enhance field trial data analyses with potential field variation.

Disclaimer

Mention of trade names or commercial products in this publication is solely for the purposes of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. USDA is an equal opportunity provider and employer.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Kuehl, R.O. (2000) Design of Experiments: Statistical Principles of Research Design and Analysis, Duxbury Press.
- [2] Fisher, R.A. (1926) The Arrangement of Field Experiments. *Journal of the Ministry of Agriculture*, **33**, 503-515.
- [3] Hoefler, R., González-Barrios, P., Bhatta, M., Nunes, J.A.R., Berro, I., Nalin, R.S., et al. (2020) Do Spatial Designs Outperform Classic Experimental Designs? *Journal of Agricultural, Biological and Environmental Statistics*, **25**, 523-552. <https://doi.org/10.1007/s13253-020-00406-2>
- [4] Cullis, B.R., Smith, A.B. and Coombes, N.E. (2006) On the Design of Early Generation Variety Trials with Correlated Data. *Journal of Agricultural, Biological, and Environmental Statistics*, **11**, 381-393. <https://doi.org/10.1198/108571106x154443>
- [5] Borges, A., González-Reymundez, A., Ernst, O., Cadenazzi, M., Terra, J. and Gutiérrez, L. (2019) Can Spatial Modeling Substitute for Experimental Design in Agricultural Experiments? *Crop Science*, **59**, 44-53. <https://doi.org/10.2135/cropsci2018.03.0177>
- [6] Papadakis, J.S. (1937) Methode Statistique Pour des Experiences Sur Champ. Bulletin, 23.
- [7] Patterson, H.D. and Williams, E.R. (1976) A New Class of Resolvable Incomplete

- Block Designs. *Biometrika*, **63**, 83-92. <https://doi.org/10.1093/biomet/63.1.83>
- [8] Gusmão, L. (1986) Inadequacy of Blocking in Cultivar Yield Trials. *Theoretical and Applied Genetics*, **72**, 98-104. <https://doi.org/10.1007/bf00261462>
- [9] Jensen, N.F. and Federer, W.T. (1964) Adjacent Row Competition in Wheat. *Crop Science*, **4**, 641-645. <https://doi.org/10.2135/cropsci1964.0011183x000400060027x>
- [10] Kempton, R.A. (1982) Adjustment for Competition between Varieties in Plant Breeding Trials. *The Journal of Agricultural Science*, **98**, 599-611. <https://doi.org/10.1017/s0021859600054381>
- [11] Kempton, R.A. and Lockwood, G. (1984) Inter-Plot Competition in Variety Trials of Field Beans (*Vicia faba* L.). *The Journal of Agricultural Science*, **103**, 293-302. <https://doi.org/10.1017/s0021859600047249>
- [12] Pearce, S.C. (1978) The Control of Environmental Variation in Some West Indian Maize Experiments. *Tropical Agriculture (Trinidad)*, **55**, 97-108.
- [13] Pearce, S.C. (1980) Randomized Blocks and Some Alternatives: A Study in Tropical Conditions. *Tropical Agriculture (Trinidad)*, **57**, 1-10.
- [14] Stroup, W.W., Baenziger, P.S. and Mutilze, D.K. (1994) Removing Spatial Variation from Wheat Yield Trials: A Comparison of Methods. *Crop Science*, **34**, 62-66. <https://doi.org/10.2135/cropsci1994.0011183x003400010011x>
- [15] Bartlett, M.S. (1938) The Approximate Recovery of Information from Replicated Field Experiments with Large Blocks. *The Journal of Agricultural Science*, **28**, 418-427. <https://doi.org/10.1017/s0021859600050875>
- [16] Bartlett, M.S. (1978) Nearest Neighbour Models in the Analysis of Field Experiments. *Journal of the Royal Statistical Society Series B. Statistical Methodology*, **40**, 147-158. <https://doi.org/10.1111/j.2517-6161.1978.tb01657.x>
- [17] Bartlett, M.S. (1988) Stochastic Models and Field Trials. *Journal of Applied Probability*, **25**, 79-89. <https://doi.org/10.2307/3214148>
- [18] Wilkinson, G.N., Eckert, S.R., Hancock, T.W. and Mayo, O. (1983) Nearest Neighbour (Nn) Analysis of Field Experiments. *Journal of the Royal Statistical Society Series B. Statistical Methodology*, **45**, 151-178. <https://doi.org/10.1111/j.2517-6161.1983.tb01240.x>
- [19] Besag, J. and Kempton, R. (1986) Statistical Analysis of Field Experiments Using Neighbouring Plots. *Biometrics*, **42**, 231-251. <https://doi.org/10.2307/2531047>
- [20] Zimmerman, D.L. and Harville, D.A. (1991) A Random Field Approach to the Analysis of Field-Plot Experiments and Other Spatial Experiments. *Biometrics*, **47**, 223-239. <https://doi.org/10.2307/2532508>
- [21] Taye, G. and Njuho, P. (2007) An Improvement on the Papadakis Covariate to Account for Spatial Variation. *Journal of Agricultural, Biological, and Environmental Statistics*, **12**, 397-413. <https://doi.org/10.1198/108571107x227946>
- [22] Technow, F. (n.d.) R Package Mvnggrad: Moving Grid Adjustment in Plant Breeding Field Trials. 0.1.6 ed2023.
- [23] Wu, J., Jenkins, J.N., McCarty, J.C. and Zhu, J. (2004) Genetic Association of Yield with Its Component Traits in a Recombinant Inbred Line Population of Cotton. *Euphytica*, **140**, 171-179. <https://doi.org/10.1007/s10681-004-2897-5>
- [24] Fehr, W. (1987) Principles of Cultivar Development: Volume 1, Theory and Technique. Macmillan Publishing Co.
- [25] Shappley, Z.W., Jenkins, J.N., Watson, C.E., Kahler, A.L. and Meredith, W.R. (1996) Establishment of Molecular Markers and Linkage Groups in Two F2 Populations of

- Upland Cotton. *Theoretical and Applied Genetics*, **92**, 915-919.
<https://doi.org/10.1007/bf00224030>
- [26] Rao, C.R. (1971) Estimation of Variance and Covariance Components—Minque Theory. *Journal of Multivariate Analysis*, **1**, 257-275.
[https://doi.org/10.1016/0047-259x\(71\)90001-7](https://doi.org/10.1016/0047-259x(71)90001-7)
- [27] Zhu, J. (1989) Estimation of Genetic Variance Components in the General Mixed Model, Ph.D. Dissertation, North Carolina State.
- [28] Wu, J. (2019) minque: Various Linear Mixed Model Analyses. 2nd Version.
<https://cran.r-project.org/web/packages/minque/index.html>
- [29] Wright, K. (2023) desplot: Plotting Field Plans for Agricultural Experiments. 1.10 Version.
- [30] RStudio Team (2022) Rstudio: Integrated Development for R. RStudio, Inc.
- [31] R Core Team (2023) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.
- [32] Zystro, J., Colley, M. and Dawson, J. (2019) Alternative Experimental Designs for Plant Breeding. In: Goldman, I., Ed., *Plant Breeding Reviews*, John Wiley & Sons, Inc. 87-117.
- [33] Federer, W.T. (2002) Construction and Analysis of an Augmented Lattice Square Design. *Biometrical Journal*, **44**, 251-257.
[https://doi.org/10.1002/1521-4036\(200203\)44:2<251::aid-bimj251>3.0.co;2-n](https://doi.org/10.1002/1521-4036(200203)44:2<251::aid-bimj251>3.0.co;2-n)
- [34] Federer, W.T. (2005) Augmented Split Block Experiment Design. *Agronomy Journal*, **97**, 578-586. <https://doi.org/10.2134/agronj2005.0578>
- [35] John, J.A. and Eccleston, J.A. (1986) Row-Column α -Designs. *Biometrika*, **73**, 301-306. <https://doi.org/10.1093/biomet/73.2.301>
- [36] Williams, E.R., John, J.A. and Whitaker, D. (2005) Construction of Resolvable Spatial Row-Column Designs. *Biometrics*, **62**, 103-108.
<https://doi.org/10.1111/j.1541-0420.2005.00393.x>
- [37] Williams, E.R. and John, J.A. (1989) Construction of Row and Column Designs with Contiguous Replicates. *Applied Statistics*, **38**, 149-154.
<https://doi.org/10.2307/2347689>
- [38] Williams, E.R. and Piepho, H.P. (2013) A Comparison of Spatial Designs for Field Variety Trials. *Australian & New Zealand Journal of Statistics*, **55**, 253-258.
<https://doi.org/10.1111/anzs.12034>
- [39] Bondalapati, K.D., Jenkins, J.N., McCarty, J.C. and Wu, J. (2015) Field Experimental Design Comparisons to Detect Field Effects Associated with Agronomic Traits in Upland Cotton. *Euphytica*, **206**, 747-757. <https://doi.org/10.1007/s10681-015-1512-2>
- [40] Adhikari, S. N., Wu, J. and Caffè, M. (Year) Comparing Linear Mixed Models for Preliminary Yield Trials That Follow Augmented Experimental Designs. *Conference on Applied Statistics in Agriculture*, 14-22. <https://doi.org/10.4148/2475-7772.1472>
- [41] Wu, J., Bondalapati, K., Glover, K., Berzonsky, W., Jenkins, J.N. and McCarty, J.C. (2012) Genetic Analysis without Replications: Model Evaluation and Application in Spring Wheat. *Euphytica*, **190**, 447-458. <https://doi.org/10.1007/s10681-012-0835-5>
- [42] Zhu, J. (1998) Genetic Models and Analytical Methods. China Agricultural Press.
- [43] Wu, J., Jenkins, J.N. and McCarty, J.C. (2014) qgtools: Tools for Quantitative Genetics Data Analyses. 1st Version.
<https://cran.r-project.org/web/packages/qgtools/index.html>
- [44] Yu, J., Holland, J.B., McMullen, M.D. and Buckler, E.S. (2008) Genetic Design and

- Statistical Power of Nested Association Mapping in Maize. *Genetics*, **178**, 539-551. <https://doi.org/10.1534/genetics.107.074245>
- [45] Cullis, B.R. and Gleeson, A.C. (1991) Spatial Analysis of Field Experiments—An Extension to Two Dimensions. *Biometrics*, **47**, 1449-1460. <https://doi.org/10.2307/2532398>
- [46] Cullis, B., Gogel, B., Verbyla, A. and Thompson, R. (1998) Spatial Analysis of Multi-Environment Early Generation Variety Trials. *Biometrics*, **54**, 1-18. <https://doi.org/10.2307/2533991>
- [47] Smith, A., Cullis, B. and Thompson, R. (2001) Analyzing Variety by Environment Data Using Multiplicative Mixed Models and Adjustments for Spatial Field Trend. *Biometrics*, **57**, 1138-1147. <https://doi.org/10.1111/j.0006-341x.2001.01138.x>
- [48] Arief, V.N., Desmae, H., Hardner, C., DeLacy, I.H., Gilmour, A., Bull, J.K., *et al.* (2019) Utilization of Multiyear Plant Breeding Data to Better Predict Genotype Performance. *Crop Science*, **59**, 480-490. <https://doi.org/10.2135/cropsci2018.03.0182>
- [49] Cullis, B.R., Thomson, F.M., Fisher, J.A., Gilmour, A.R. and Thompson, R. (1996) The Analysis of the NSW Wheat Variety Database. I. Modelling Trial Error Variance. *Theoretical and Applied Genetics*, **92**, 21-27. <https://doi.org/10.1007/bf00222947>
- [50] Piepho, H.P., Ogutu, J.O., Schulz - Streeck, T., Estaghirou, B., Gordillo, A. and Technow, F. (2012) Efficient Computation of Ridge - regression Best Linear Unbiased Prediction in Genomic Selection in Plant Breeding. *Crop Science*, **52**, 1093-1104. <https://doi.org/10.2135/cropsci2011.11.0592>