



Improving the Precision of NDVI Estimates in Upland Cotton Field Trials

K.L. Hugie,* P.J. Bauer, K.C. Stone, E.M. Barnes, D.C. Jones, and B.T. Campbell

Core Ideas

- Estimates of NDVI in upland cotton were influenced by field heterogeneity.
- Row-column design and spatial analysis improved the precision of NDVI estimates.
- Extensive rank changes in genotype mean estimates were observed across models.

Controlling for experimental error attributable to field heterogeneity is important in high-throughput phenotyping studies that enable large numbers of genotypes to be evaluated across time and space. In the current study, we compared the efficacy of different experimental designs and spatial models in the analysis of canopy spectral reflectance data collected on upland cotton (*Gossypium hirsutum* L.). Canopy spectral reflectance, as measured by normalized difference vegetation index (NDVI), was measured at first bloom on three upland cotton performance trials conducted in Florence, SC, during 2014 and 2015. The relative efficiency and estimates of genotype effects were compared among randomized complete block, an α -lattice incomplete block, row-column incomplete block, nearest neighbor adjusted, and spatially correlated error models. The row-column model provided the greatest improvement in the precision of genotype effect estimates compared with the randomized complete block model. Genotype rankings based on NDVI varied substantially between the randomized complete block and alternative models, particularly at 5 and 10% selection intensities. These results suggest that the use of more complex experimental designs and spatial analyses should be routinely considered to minimize experimental error due to field heterogeneity and improve the precision and reliability of traits measured using high-throughput phenotyping systems. These findings also indicate that further research into the effects of field heterogeneity on the relationship between NDVI and lint yield in upland cotton is warranted.

Currently, the primary limitation on linking genotypic to phenotypic variation is the ability to measure relevant phenotypes in a precise, nondestructive, high-throughput capacity (Furbank and Tester, 2011). In response to the “phenotyping bottleneck,” there have been rapid developments in field-based high-throughput phenotyping (HTP) methods, chiefly focused on the application of imagery and proximal or remote sensing technologies (Andrade-Sanchez et al., 2014; Chapman et al., 2014; Crain et al., 2016). Studies have demonstrated the utility of phenotypic data collected using aerial and ground-based HTP systems across various stages of plant improvement programs, from the identification of novel genetic variation for complex traits (Honsdorf et al., 2014; Tanger et al., 2017; Yang et al., 2014) to the selection of superior breeding lines in preliminary performance trials (Rutkoski et al., 2016). In cotton (*Gossypium* spp.) specifically, genetic variation can be detected for spectral vegetation indices, such as the normalized difference vegetation index (NDVI), and canopy temperature, which correlate with lint yield (Pauli et al., 2016; Sharma and Ritchie, 2015).

Strategies to minimize experimental error in field trials, particularly that which is attributable to field heterogeneity, will become increasingly important as HTP systems enable crop improvement programs to evaluate increasing numbers of genotypes

Abbreviations: AIC, Akaike information criterion; AL, α -lattice incomplete block; aSED, average standard error of genotypic differences; DGPS, differential global positioning system; HTP, high-throughput phenotyping; LLRT, log-likelihood ratio chi-squared test; NDVI, normalized difference vegetation index; NNA, nearest neighbor adjustment; RC, row-column; RCB, randomized complete block; RE, relative efficiency; SP, spatially correlated errors.

K.L. Hugie, P.J. Bauer, K.C. Stone, and B.T. Campbell, USDA-ARS Coastal Plains Soil, Water, and Plant Research Center, 2611 W. Lucas St., Florence, SC 29501; E.M. Barnes and D.C. Jones, Cotton Incorporated, 6399 Weston Parkway, Cary, NC 27513.

Copyright © American Society of Agronomy and Crop Science Society of America. 5585 Guilford Rd., Madison, WI 53711 USA. This is an open access article distributed under the terms of the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)
Plant Phenome J. 1: 170009 (2017)
doi:10.2135/tppj2017.09.0009

Received 22 Sept. 2017

Accepted 15 Dec. 2017

*Corresponding author (kari.hugie@ars.usda.gov).

across time and space. Traditionally, a priori knowledge of spatial trends within a field and experimental design principles (i.e., replication, randomization, and blocking) are used to minimize experimental error due to field heterogeneity, improve discriminatory power between genotypes, and ultimately obtain precise and reliable measures of genotype performance. The randomized complete block (RCB) design is commonly used in cotton performance trials and is efficient when the variance within blocks is relatively small compared with the variance across blocks (Gusmão, 1986). However, it is well documented that environmental factors, such as differences in soil fertility, moisture availability, and topographic features within a field, often introduce considerable variability within blocks (Campbell and Bauer, 2007; Johnson et al., 2002; Liu et al., 2015). Incomplete block designs that reduce block size, such as α designs (Patterson and Williams, 1976) and row-column designs (John and Williams, 1998; Williams, 1986), provide additional control of experimental error and more precise estimates of genotype effects (Liu et al., 2015; Stroup et al., 1994).

The statistical models for the experimental designs outlined above assume constant variance among errors. However, experimental error introduced by field heterogeneity often persists despite a priori methods, resulting in spatial autocorrelation among errors. Violations in the assumption of constant variance can have considerable influence on the interpretation of performance trials and the selection of superior genotypes (Ball et al., 1993). In such cases, a posteriori statistical procedures can be used to account for spatial autocorrelation among residuals. Two of the most common analyses used to model field heterogeneity include nearest neighbor adjustment (NNA) using empirical data or systematically arranged check plots (Besag and Kempton, 1986; Kempton and Howes, 1981; Papadakis, 1937; Wilkinson et al., 1983) and spatially correlated error (SP) models (Zimmerman and Harville, 1991). Studies in multiple crops demonstrate that both procedures can improve trial efficiencies compared with RCB models by decreasing experimental error and can lead to more precise and reliable genotype estimates in the presence of field heterogeneity (Brownie et al., 1993; Cullis and Gleeson, 1991; Gilmour et al., 1997; Müller et al., 2010; Smith and Casler, 2004; Sripathi et al., 2017; Stroup et al., 1994; Yang et al., 2004). However, it is recommended that the most appropriate spatial analysis be determined on a trial-specific basis.

Campbell and Bauer (2007) found that NNA led to more precise and reliable estimates of lint yield and recommended that the use of a posteriori methods should be regularly investigated in the analysis of cotton performance trials across the southeastern United States, a region characterized by highly variable soils. Similarly, Liu et al. (2015) analyzed yield data from cotton performance trials throughout Australian production regions and found that SP models increased the precision of genotype yield estimates by 76% on average compared with RCB. It can be expected that HTP traits, much like yield, are subject to environmental variability. For example, NDVI measures on cotton are known to be affected by soil fertility (Bronson et al., 2003) and soil moisture

availability (Li et al., 2001). However, the impact of field heterogeneity on the reliability and precision of HTP traits in cotton has not been reported to date. Thus, the objectives of this work were to: (i) evaluate the efficacy of different experimental designs and spatial analyses in the estimation of genotype effects for NDVI measured at first bloom in upland cotton; and (ii) measure their impact on selection for NDVI per se and lint yield.

Materials and Methods

Experimental Design

Canopy reflectance at first bloom was measured on three cotton performance trials conducted at the Pee Dee Research and Education Center near Florence, SC, in 2014 and 2015. All three trials were managed conventionally following standard cultural practices for rainfed cotton grown in the southeastern United States. Trial 1 was planted on 20 May 2014 on a mostly Goldsboro loamy sand (a fine-loamy, siliceous, subactive, thermic Aquic Paleudult), and Trials 2 and 3 were planted on 21 May 2015 on a mostly Norfolk loamy sand (a fine-loamy, kaolinitic, thermic Typic Kandudult). Each trial contained 288 entries arranged in an 8×36 α -lattice incomplete block design with two replications. Experimental units were single-row plots 10.7 m by 96.5 cm thinned to a density of 8 to 10 plants m^{-2} . Alternating rows of cv. PhytoGen 499WRF (PHY 499) were superimposed on the incomplete block designs, such that two rows of PHY 499 bordered every two rows of experimental plots. For example, if PHY 499 = P and experimental plots = X, rows were arranged in the following pattern: P-P-X-X-P-P-X-X-P-P. This design was implemented to evaluate the efficiency and practicality of planting a common cultivar check for use in calculating NNA to account for spatial variability. Each of the three trials had the same field dimensions, with 64 rows and 18 columns.

A historic rainfall event occurred in early October of 2015, with plots receiving 318 mm of precipitation during the course of 3 d, destroying the trials. Consequently, lint yield was not measured on Trials 2 and 3. Trial 1 was harvested on 1 Dec. 2014 with a spindle-type mechanical cotton picker, and the total seed cotton weight was recorded. A 25-boll sample was hand harvested from each plot prior to harvest and ginned on a 10-saw laboratory gin without lint cleaners. Lint percentage was determined by dividing the weight of the lint sample by the weight of the seed cotton sample, and lint yield was calculated by multiplying the lint percentage by the total seed cotton yield.

High-Throughput Phenotyping Platform and Measurement of Plant Canopy Spectral Reflectance

The HTP system used to collect proximally sensed spectral reflectance data was built on a high-clearance tractor sprayer (Lee Spider, LeeAgra). A toolbar supported by two hydraulically powered lift arms was attached to the front end of the tractor frame, providing support for a differential GPS (DGPS) receiver (Phoenix 200 DGPS receiver, Raven Industries) and spectral sensors. The DGPS receiver was placed on the center of the front-mounted

toolbar and programmed to position data at 5 Hz. The serial output from the DGPS receiver was streamed on an RS-232 network to a GeoScout GLS-400 datalogger (Holland Scientific) enabling geo-referencing of the spectral reflectance data with sub-meter accuracy via the National Marine Electronics Association GGA and RMC strings. An isolated secondary power source was retrofitted on the tractor to supply 12-V direct current power to the DGPS receiver, datalogger, and spectral sensors.

Four CropCircle ACS-430 active multispectral crop canopy sensors (Holland Scientific) were mounted on the toolbar, allowing canopy reflectance (ρ) to be collected on four adjacent rows simultaneously. The north-south positions (i.e., offsets) of each of the four sensors in relation to the DGPS receiver were programmed into the datalogger prior to data collection, thus eliminating the need to adjust for sensor offsets in geospatial post-processing steps. During data collection, the spectral sensors were positioned approximately 1.0 m above the tallest plants, and the speed of the HTP system was maintained near 1.2 m s⁻¹. Canopy reflectance was measured with the HTP system when approximately half of the genotypes were at first white bloom, corresponding to 4 July 2014 for Trial 1 and 21 July 2015 for Trials 2 and 3. The CropCircle ACS-430 spectral sensors integrate three optical measurement channels at 670 nm (red), 730 nm (red edge), and 780 nm (near infrared), and NDVI calculations were performed as

$$\text{NDVI}_{\text{red}} = \frac{\rho_{\text{NIR}} - \rho_{\text{red}}}{\rho_{\text{NIR}} + \rho_{\text{red}}} \quad [1]$$

Geospatial post-processing of the spectral reflectance data was performed in ArcGIS Version 10.3 software (ESRI) using methods modified from Wang et al. (2016) to establish plot boundaries. Initially, the georeferenced NDVI measurements and their quantities were used to obtain the four corners of each field. Next, the Fishnet function was used to generate a shapefile in which polygon boundaries were constructed around each set of four adjacent rows for each field column (i.e., “pass” boundaries), assigning unique pass identifiers and directionality of the HTP system to each polygon. More specifically, the Fishnet function creates a grid of adjacent rectangular cells based on the user-specified spatial extent of the grid, angle of rotation, number of rows and columns, as well as the height (i.e., plot length) and width (i.e., row width) of each cell. A point shapefile containing the centroids of each cell was also generated and used to trim the polygon borders so that they enclosed only the center 8.5 m of each set of four adjacent rows. The Intersect function was used to clip and append the polygon attributes with the canopy reflectance data. Lastly, unique plot identifiers were assigned to each data point based on sensor number and the directionality of the HTP system, and the mean NDVI for each plot was calculated using the MEANS procedure in SAS for Windows Version 9.4 (SAS Institute).

Statistical Analyses

Lint yield for Trial 1 and NDVI for all three trials were analyzed separately using five linear mixed models in the MIXED

procedure in SAS. The following RCB model was used as the baseline model:

$$y_{ij} = \mu + \alpha_i + \beta_j + e_{ij} \quad [2]$$

where y_{ij} is the observation made on the i th genotype of the j th block, μ is the overall mean, α_i is the effect of the i th genotype, β_j is the effect of the j th complete block, and e_{ij} is the residual plot error associated with y_{ij} . Genotype effects were treated as fixed, while block and incomplete block, for the α -lattice (AL) model, effects were considered random. The efficiency of an alternative resolvable row-column (RC) experimental design was investigated in a post-blocking exercise (Ainsley et al., 1987), where artificial incomplete blocks were superimposed on the RCB design and analyzed as if they were true blocks. Thus, in the RC model, random row and column effects nested within complete blocks were added to the baseline RCB model. The NNA model included the terms in the RCB model plus the addition of a covariate. The NNA covariate was computed based on a modified Papadakis (1937) method using adjacent residuals from the bordering plots of PHY 499 according to

$$\rho_{kl} = \frac{1}{2}(\bar{e}_{k,l-1} + \bar{e}_{k,l+1}) \quad [3]$$

where ρ_{kl} is the covariate corresponding to the plot located in the k th row and l th column of each trial, and the two \bar{e}_{kl} values are the averages of the residuals of the cultivar check rows to the left and right of the kl th plot.

The RCB, AL, RC, and NNA models assume that errors are normally and independently distributed, with a zero mean and constant variance. Instead of assuming constant variance, SP models use heterogeneous covariance structures to account for spatial autocorrelation among errors. The SP models included the same effects of the baseline RCB model. However, the covariance among errors was modeled according to (Littell et al., 2006)

$$C(b) = \begin{cases} \sigma^2 & \text{if } b=0 \\ \sigma_{e_{(s)}, e_{(s+b)}} = \sigma^2 [f(b)] & \text{if } b>0 \end{cases} \quad [4]$$

where $C(b)$ is the covariance between two errors, denoted by $e_{(s)}$ and $e_{(s+b)}$, which are b units of distance apart, s is the spatial position of the ij th plot, and $f(b)$ is a function of the distance between the corresponding plots. The Universal Transverse Mercator coordinates corresponding to the plot centroids were used to define the distance (b), in meters, between plots. In contrast to NNA, data collected on the alternating check rows of PHY 499 were not included in the SP models. The VARIOGRAM procedure in SAS was used to visually inspect spatial autocorrelation and obtain estimates of covariance model parameters based on empirical semivariograms of the baseline RCB model fit residuals. Several covariance functions commonly used in the analyses of plant breeding trials were considered for the SP models, including three isotropic covariance models—power, Gaussian, and spherical models—and one anisotropic covariance model—the power anisotropic model (i.e., separable first-order autoregressive model, AR1 \times AR1). Residuals were also modeled with and

without a nugget effect, which is the semivariance at $h = 0$. Thus, eight different SP models were evaluated for each trial. Restricted maximum likelihood log-likelihood ratio chi-squared tests (LLRTs) were used to select the most suitable covariance function for the nested SP models, and Akaike information criterion (AIC) values were used as the selection criterion for non-nested models.

The alternative models were compared with the RCB model regarding the relative efficiency (RE) of the genotype effect estimates. The RE was calculated using the average standard error of pairwise genotypic differences (aSED): $RE = 100\% \times (\text{aSED of RCB})/(\text{aSED of the alternative model})$, such that an $RE > 100$ indicates an improvement in precision in relation to the RCB model (Qiao et al., 2000). Additionally, restricted maximum likelihood LLRTs and AIC values were used to compare the AL, RC, and SP models with the baseline RCB model. For all analyses, the first-order Kenward–Roger method (Kenward and Roger, 2009) was used to approximate denominator degrees of freedom in F tests

(Richter et al., 2015). Spearman rank correlations were computed based on the best linear unbiased estimates for genotype effects from each model to evaluate the impact of experimental designs and spatial analyses on selection. Additionally, the percentage of agreement between models at fixed selection intensities (i.e., the proportion of selected genotypes) of 5, 10, 25, and 50% was evaluated using the Czekanowski coefficient (D) (Campbell and Bauer, 2007; Qiao et al., 2000), calculated as $D = a/(a + c)$, where a is the number of lines selected by both the alternative and baseline model, and c is the number of lines selected by the baseline model only.

Results and Discussion

Diagnosis of Spatial Autocorrelation for NDVI

Field heterogeneity within each trial was initially investigated by plotting the two-row average NDVI of the alternating cultivar check, PHY 499 (Fig. 1). Clear spatial trends in NDVI measured on

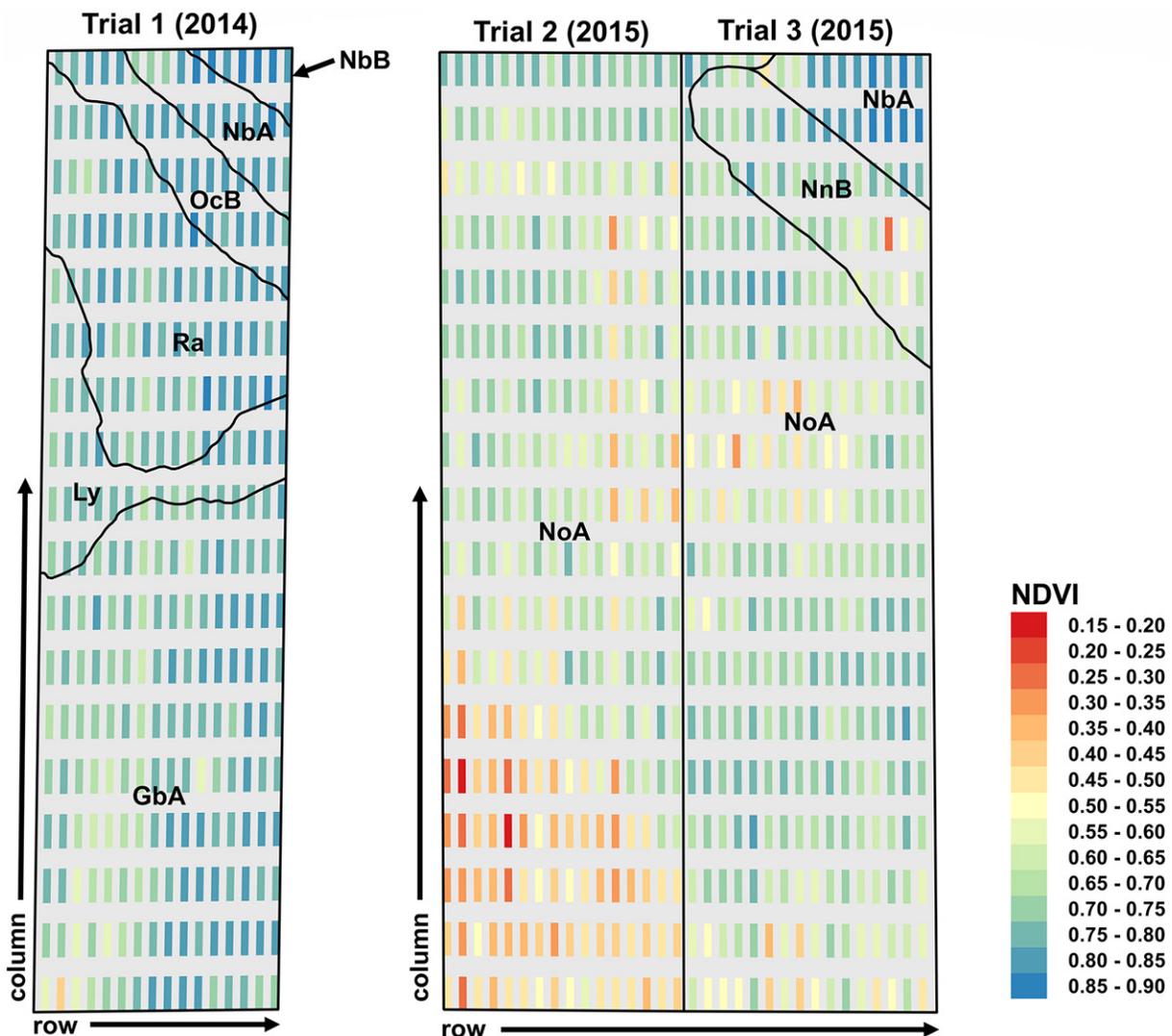


Fig. 1. Maps of soil type(s) and the two-row plot average of normalized difference vegetation index (NDVI) measured at first bloom on alternating cultivar check plots of cv. PhytoGen 499WRF for three trials grown from 2014 to 2015 in Florence, SC. GbA = Goldsboro loamy sand, 0 to 2% slopes; Ly = Lynchburg sandy loam; Ra = Rains sandy loam; OcB = Ocilla sand, 0 to 3% slopes; NbA = Noboco loamy sand, 0 to 1% slopes; NbB = Noboco loamy sand, 1 to 3% slopes; NoA = Norfolk loamy sand, 0 to 1% slopes; NnB = Norfolk loamy sand, 1 to 3% slopes.

PHY 499 were observed in Trials 2 and 3. In the early 1990s, the USDA–NRCS conducted a soil survey of the fields at the Pee Dee Research and Education Center using a 30.5-m grid sampling pattern, allowing for the superimposition of soil type(s) within each field over NDVI. However, variation in soil type did not correspond to the spatial trends observed for NDVI, except for the Noboco loamy sand (NbA) soil in Trial 3. Considering that the trials were grown under rainfed conditions, variability in soil moisture availability resulting from field topography and lateral water movement may have contributed to variation in NDVI. Residuals from the RCB model were plotted against row and column positions, and the spatial trends of NDVI observed in the experimental rows were similar to those apparent in the check plots of PHY 499 (data not shown).

Empirical (isotropic) semivariograms based on the RCB model fit residuals were used to diagnose spatial autocorrelation in NDVI among experimental plots (Fig. 2). A positive relationship between semivariance and distance was observed within all three

trials, indicating spatial dependence among residuals. The distance at which the semivariance stabilized (i.e., the range) ranged from approximately 25 to 75 m. The semivariograms also indicated substantial nugget effects (i.e., semivariance at 0 m) across all three trials, which is often indicative of random variation (Gilmour et al., 1997).

Model Comparisons for NDVI

Among the covariance functions evaluated for the spatial models, the isotropic power model with nugget effect was the most suitable SP model for Trials 1 and 3, while the power anisotropic (i.e., AR1 × AR1) model with nugget effect was the most suitable SP model for Trial 2. The power isotropic and anisotropic covariance structures have proven effective in modeling spatial dependence for yield measures across various crops, including peanut (*Arachis hypogaea* L.) (Casanoves et al., 2005), ryegrass (*Lolium* spp.) (Sripathi et al., 2017), wheat (*Triticum aestivum* L.) (Qiao et al., 2000), grapevine (*Vitis vinifera* L.) (Gonçalves et al., 2007), and upland cotton (Liu et al., 2015).

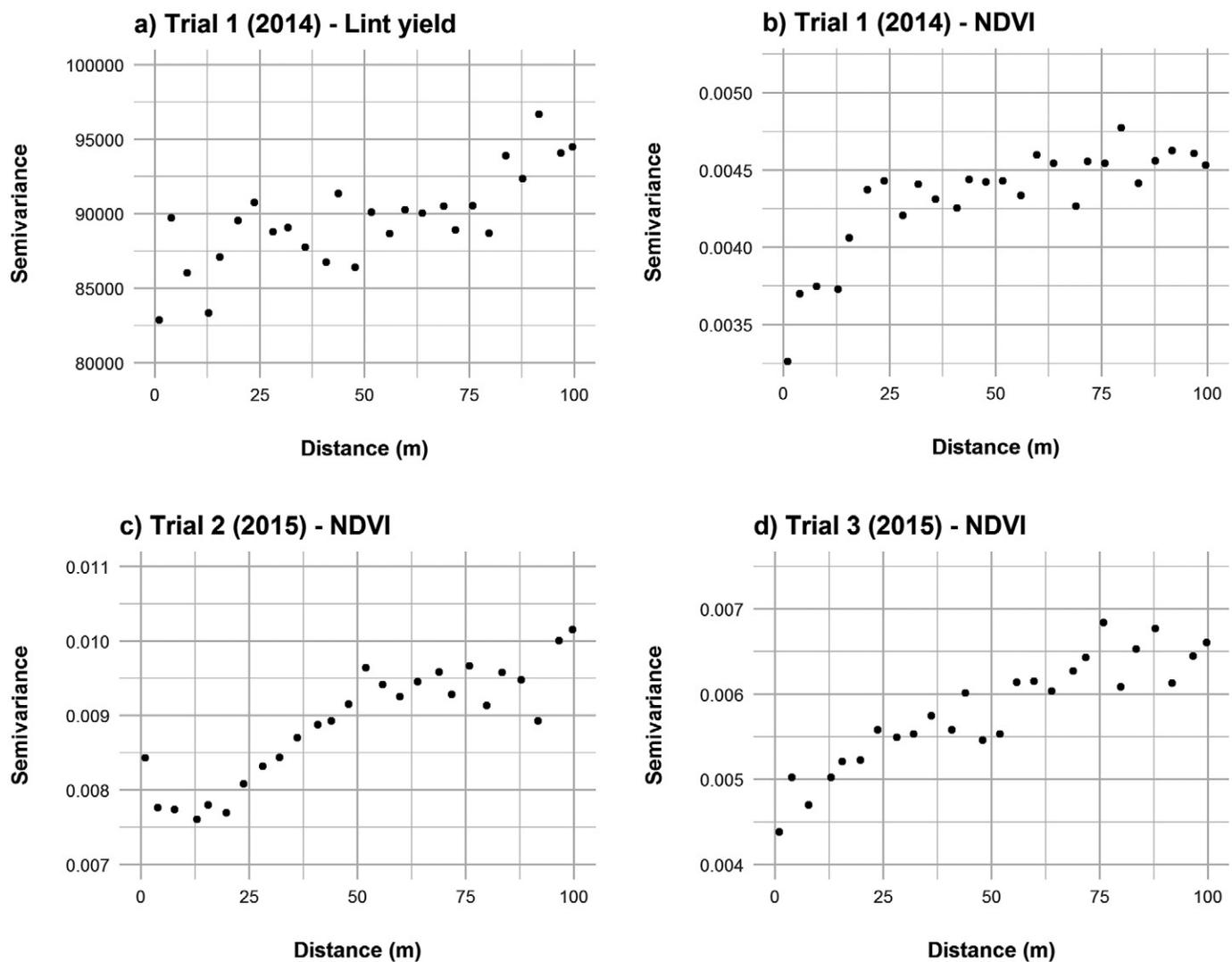


Fig. 2. Empirical (isotropic) semivariograms based on residuals from the randomized complete block model for (a) lint yield and (b) normalized difference vegetation index (NDVI) for Trial 1, (c) NDVI for Trial 2, and (d) NDVI for Trial 3. Trials were conducted from 2014 to 2015 in Florence, SC.

Complete blocks within each of the three trials were arranged parallel to the column direction. The spatial trends in NDVI suggest that the RCB design was not effective at minimizing within-block heterogeneity, particularly for Trials 2 and 3 (Fig. 1). Expectedly, the AL, RC, and SP models were superior to the baseline RCB model based on the model selection criteria (i.e., LLRTs and AIC values) and the precision of genotype effect estimates as measured by the RE (Table 1). The SP model was superior to the AL model for all three trials according to AIC values and the RE. The addition of a covariate derived from the check rows of PHY 499 in the NNA model also improved the precision of genotype effect estimates compared with the RCB, AL, and SP models for Trials 2 and 3. However, the RC model provided the greatest improvement in precision of NDVI estimates compared with the RCB model, ranging from a 20 to 66% reduction of aSED in Trials 1 and 2, respectively. Although post-blocking is not recommended for routine analysis, it is frequently used to investigate the efficiency of alternative experimental designs (Qiao et al., 2000; Robinson et al., 1988). Similar gains in efficiency using RC designs compared with RCB have been reported in cotton yield trials (Liu et al., 2015; Williams and Luckett, 1988). Genotypic

Table 1. Comparison of α -lattice incomplete block (AL), row-column incomplete block (RC), nearest neighbor adjusted (NNA), and spatially correlated errors (SP) models with the randomized complete block (RCB) model for normalized difference vegetation index (NDVI) measured at first bloom on three cotton performance trials conducted at Florence, SC, from 2014 to 2015.

Model	<i>P</i> value†	AIC‡	LLRT§	RE¶ %
Trial 1, 2014				
RCB	<0.0001	-343.9	-	-
AL	<0.0006	-351.9	10.0***	102.7
RC	<0.0001	-406.1	66.2***	119.9
NNA	<0.0001	-	-	100.1
SP	<0.0002	-380.0	40.1***	111.1
Trial 2, 2015				
RCB	0.4444	-99.9	-	-
AL	0.1719	-174.6	76.7***	123.3
RC	0.2993	-295.7	199.8***	165.9
NNA	0.1945	-	-	135.3
SP	0.0615	-242.0	148.1***	132.2
Trial 3, 2015				
RCB	0.2230	-256.6	-	-
AL	0.1657	-292.1	37.5***	111.5
RC	0.0314	-384.5	129.9***	140.6
NNA	0.1829	-	-	125.2
SP	0.2263	-318.7	66.1***	115.5

*** Significant at the 0.001 probability level.

† *P* value of fixed effect of genotype.

‡ Akaike information criterion. The smaller the AIC, the better the model fit.

§ Restricted maximum likelihood log-likelihood ratio chi-squared test (LLRT) statistics comparing alternative model to the RCB model.

¶ Relative efficiency of the alternative model compared with the RCB model based on the average standard error of pairwise genotypic differences (aSED).

effects for NDVI were significant ($\alpha = 0.05$) in Trial 1 in 2014 across all models (Table 1). Conversely, no significant differences in NDVI were detected among genotypes evaluated in Trial 2, and significant genotypic effects were detected in Trial 3 only when analyzed with the RC model.

Selection Based on NDVI

Genotype NDVI estimates from the AL, RC, NNA, and SP models were compared with estimates from the RCB model based on Spearman rank correlation coefficients and the percentage of agreement, *D*, between models in the selection of superior lines. Correlations between NDVI estimates were generally lower in Trials 2 and 3, where extensive spatial trends were present (Table 2). Among the two trials in 2015, correlations between NDVI estimates from the SP and AL models were the highest, while correlations between estimates from the RC model compared with the NNA and RCB models were the lowest.

Adjustments of genotype NDVI estimates from the alternative models varied substantially from the RCB model, resulting in differences between genotype rankings, particularly at selection intensities $\leq 10\%$ (Table 3). With the exception of NDVI estimates from the AL and NNA models in Trial 1, the concurrence of genotypes ranked in the top 5 and 10% from the RCB and alternative models did not exceed 53%. Qiao et al. (2000) evaluated different incomplete block experimental designs and SP models in the analysis of yield performance trials in wheat and selected a preferred model based on LLRTs, the RE, and variogram diagnostics. Similar to the current study, they observed that the percentage of agreement between statistical models decreased at higher selection

Table 2. Spearman rank correlations for mean genotype effect estimates using the randomized complete block (RCB), α -lattice incomplete block (AL), row-column incomplete block (RC), nearest neighbor adjusted (NNA), and spatially correlated errors (SP) models for normalized difference vegetation index (NDVI) measured at first bloom on three cotton performance trials conducted at Florence, SC, from 2014 to 2015.

Model	RCB	AL	RC	NNA
Trial 1, 2014				
AL	0.97			
RC	0.88	0.90		
NNA	1.00	0.96	0.88	
SP	0.88	0.95	0.86	0.88
Trial 2, 2015				
AL	0.79			
RC	0.59	0.72		
NNA	0.72	0.84	0.67	
SP	0.74	0.94	0.73	0.88
Trial 3, 2015				
AL	0.86			
RC	0.64	0.70		
NNA	0.71	0.84	0.58	
SP	0.80	0.96	0.73	0.89

Table 3. Agreement in selection proportion for the top genotypes at selection intensities of 5, 10, 25, and 50% based on the Czekanowski coefficient (D) for comparison of four alternative models, including α -lattice incomplete block (AL), row-column incomplete block (RC), nearest neighbor adjusted (NNA), and spatially correlated errors (SP) models with the baseline randomized complete block (RCB) model for normalized vegetation difference index (NDVI) measured at first bloom on three cotton performance trials conducted at Florence, SC, from 2014 to 2015.

Model	Agreement in selection			
	5%	10%	25%	50%
----- % agreement -----				
Trial 1, 2014				
AL	0.80	0.72	0.82	0.90
RC	0.33	0.52	0.79	0.82
NNA	0.87	0.90	0.96	0.97
SP	0.53	0.55	0.71	0.84
Trial 2, 2015				
AL	0.40	0.46	0.59	0.79
RC	0.33	0.25	0.49	0.69
NNA	0.00	0.29	0.54	0.77
SP	0.20	0.36	0.52	0.78
Trial 3, 2015				
AL	0.53	0.52	0.71	0.85
RC	0.20	0.31	0.47	0.72
NNA	0.27	0.41	0.51	0.76
SP	0.47	0.38	0.62	0.77

intensities, and the greatest divergence in genotype rankings was observed between the preferred model and the RCB model. Still, the percentages of agreement between NDVI estimates reported in this study are substantially lower on average than those from studies evaluating similar spatial analyses on yield performance (Campbell and Bauer, 2007; Qiao et al., 2000; Sripathi et al., 2017). These results highlight the importance of using appropriate experimental designs and statistical analyses, and the sensitivity of HTP traits, such as NDVI, to field heterogeneity.

The effects of field heterogeneity on the relationship between NDVI at first bloom and lint yield is of equal, if not greater, importance than the impact on estimates of NDVI per se. Maps of lint yield based on the PHY 499 check rows and residuals from the RCB model for Trial 1 (2014) revealed similar spatial trends to NDVI (data not shown). However, spatial dependence among lint yield residuals was not as strong, based on the empirical semivariogram, as that of NDVI (Fig. 2). Comparisons between models based on LLRTs, AIC, and the RE suggested that the RC model was the preferred model for the analysis of lint yield for Trial 1. Thus, genotype yield estimates from the RC model were used to evaluate the relationship between NDVI and lint yield.

Genotype yield estimates from the RC model were regressed onto the genotype NDVI estimates of all five models for Trial 1, and the proportion of top yielding genotypes selected based on NDVI at different selection intensities was calculated (Table 4). There was

Table 4. Adjusted r^2 and proportion of top yielding genotypes from a single cotton performance trial (Trial 1) conducted at Florence, SC, in 2014 selected at selection intensities of 5, 10, 25, and 50% based on genotype effect estimates for normalized difference vegetation index (NDVI) measured at first bloom using randomized complete block (RCB), α -lattice incomplete block (AL), row-column incomplete block (RC), nearest neighbor adjusted (NNA), and spatially correlated errors (SP) models. Genotype effect estimates for lint yield were calculated using the RC model.

Model	r^2	Top yielding genotypes selected based on NDVI			
		5%	10%	25%	50%
----- % top yielding genotypes -----					
RCB	0.23	0.13	0.28	0.47	0.67
AL	0.24	0.07	0.28	0.44	0.65
RC	0.27	0.20	0.34	0.50	0.63
NNA	0.23	0.20	0.28	0.47	0.65
SP	0.24	0.07	0.28	0.44	0.66

a moderate, but significant, positive relationship between lint yield and NDVI, with r^2 estimates ranging from 0.23 to 0.27. Despite genotype rank changes in NDVI among the different statistical models, similar proportions of the top yielding lines were selected based on NDVI at 10, 25, and 50% selection intensities independent of the model. However, based on selection for genotypes in the top 5% for NDVI, the RC and NNA models captured the greatest proportion of top yielding lines (20%) among the five models. These results, in combination with the findings of Sharma and Ritchie (2015) and Pauli et al. (2016), suggest that NDVI may be used to eliminate poor-performing genotypes before harvest. However, the impact of field heterogeneity on the relationship between NDVI and lint yield needs further study.

Conclusions

High-throughput phenotyping systems are becoming more routinely used in cotton improvement programs, allowing breeders to measure traits indicative of plant health and development across increased space and time. This study highlights the importance of using appropriate experimental designs and supplemental spatial analyses to account for variation in NDVI introduced by field heterogeneity. These considerations likely extend to other common HTP traits, such as canopy temperature and plant height. The greatest improvement in the precision of genotype estimates was achieved through post-blocking, and the results suggest that RC experimental designs may be preferable not only to the RCB but also the AL design. Furthermore, resolvable RC designs can be optimized to accommodate spatial correlation with or without a priori knowledge of the underlying spatial structure (Williams and Piepho, 2013).

Adjustments for field heterogeneity using a common cultivar check is not recommended for replicated trials, considering that the alternative methods presented provide comparable, if not improved, control of experimental error without the additional

land and labor requirements. However, numerous studies have demonstrated that NNA based on adjacent experimental plots can effectively capture spatial variability in field trials (Campbell and Bauer, 2007; Sripathi et al., 2017; Stroup et al., 1994). The efficacy of different a posteriori spatial analyses is reliant on various experimental design factors. For example, previous research has demonstrated that highly rectangular field designs, like those characteristic of the current study, are especially prone to within-block heterogeneity (Liu et al., 2015; Sripathi et al., 2017). The findings of this study suggest that spatial models can help improve the reliability and precision of NDVI estimates in the presence of field heterogeneity, particularly when blocking is ineffective. Although beyond the scope of the current study, Velazco et al. (2017) recently described a method incorporating two-dimensional regression spline modeling that may warrant future investigation as an alternative to the sequential evaluation of different SP models.

Ultimately, these results demonstrate that failure to account for field heterogeneity can have sizeable consequences on selection for HTP traits per se, such as NDVI. Although lint yield data were limited in this study, the results suggest that further research into the impact of field heterogeneity on the relationship between NDVI and lint yield in upland cotton is warranted.

Acknowledgments

This research project was supported by funding from CRIS no. 6082-21000-007-00D of the USDA and a grant from Cotton Incorporated. Special thanks to Kendreal Wingate, Ernie Strickland, and summer students for technical assistance. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the USDA.

References

- Ainsley, A.E., L.J. Paterson, and H.D. Patterson. 1987. A method for predicting the efficiency of incomplete-block trials. *Biometrics* 43:55–59. doi:10.2307/2531948
- Andrade-Sanchez, P., M.A. Gore, J.T. Heun, K.R. Thorp, A.E. Carmo-Silva, A.N. French, et al. 2014. Development and evaluation of a field-based high-throughput phenotyping platform. *Funct. Plant Biol.* 41:68–79. doi:10.1071/FP13126
- Ball, S.T., D.J. Mulla, and C.F. Konzak. 1993. Spatial heterogeneity affects variety trial interpretation. *Crop Sci.* 33:931–935. doi:10.2135/cropsci1993.0011183X003300050011x
- Besag, J., and R. Kempton. 1986. Statistical analysis of field experiments using neighbouring plots. *Biometrics* 42:231–251. doi:10.2307/2531047
- Bronson, K.F., T.T. Chua, J.D. Booker, J.W. Keeling, and R.J. Lascano. 2003. In-season nitrogen status sensing in irrigated cotton. *Soil Sci. Soc. Am. J.* 67:1439–1448. doi:10.2136/sssaj2003.1439
- 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
- Campbell, B.T., and P.J. Bauer. 2007. Improving the precision of cotton performance trials conducted on highly variable soils of the southeastern USA Coastal Plain. *Plant Breed.* 126:622–627. doi:10.1111/j.1439-0523.2007.01397.x
- Casanoves, F., R. Macchiavelli, and M. Balzarini. 2005. Error variation in multi-environment peanut trials: Within-trial spatial correlation and between-trial heterogeneity. *Crop Sci.* 45:1927–1933. doi:10.2135/cropsci2004.0547
- Crain, J.L., Y. Wei, J. Barker III, S.M. Thompson, P.D. Alderman, M. Reynolds, et al. 2016. Development and deployment of a portable field phenotyping platform. *Crop Sci.* 56:965–975. doi:10.2135/cropsci2015.05.0290
- Chapman, S.C., T. Merz, A. Chan, P. Jackway, S. Hrabar, M.F. Dreccer, et al. 2014. Pheno-copter: A low-altitude, autonomous remote-sensing robotic helicopter for high-throughput field-based phenotyping. *Agronomy* 4:279–301. doi:10.3390/agronomy4020279
- 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
- Furbank, R.T., and M. Tester. 2011. Phenomics: Technologies to relieve the phenotyping bottleneck. *Trends Plant Sci.* 16:635–644. doi:10.1016/j.tplants.2011.09.005
- 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
- Gonçalves, E., A. St. Aubyn, and A. Martins. 2007. Mixed spatial models for data analysis of yield on large grapevine selection field trials. *Theor. Appl. Genet.* 115:653–663.
- Gusmão, L. 1986. Inadequacy of blocking in cultivar yield trials. *Theor. Appl. Genet.* 72:98–104. doi:10.1007/BF00261462
- Honsdorf, N., T.J. March, B. Berger, M. Tester, and L. Pillen. 2014. High-throughput phenotyping to detect drought tolerance QTL in wild barley introgression lines. *PLoS One* 9:e97047. doi:10.1371/journal.pone.0097047
- John, J.A., and E.R. Williams. 1998. *t*-Latinized designs. *Aust. N.Z. J. Stat.* 40:111–118. doi:10.1111/1467-842X.00012
- Johnson, R.M., R.G. Downer, J.M. Bradow, P.J. Bauer, and E.J. Sadler. 2002. Variability in cotton fiber yield, fiber quality, and soil properties in a southeastern Coastal Plain. *Agron. J.* 94:1305–1316. doi:10.2134/agronj2002.1305
- Kempton, R.A., and C.W. Howes. 1981. The use of neighbouring plot values in the analysis of variety trials. *J. R. Stat. Soc., Ser. A* 30:59–70.
- Kenward, M.G., and J.H. Roger. 2009. An improved approximation to the precision of fixed effects from restricted maximum likelihood. *Comput. Stat. Data Anal.* 53:2583–2595. doi:10.1016/j.csda.2008.12.013
- Li, H., R.J. Lascano, E.M. Barnes, J. Booker, L.T. Wilson, K.F. Bronson, and E. Segarra. 2001. Multispectral reflectance of cotton related to plant growth, soil water and texture, and site elevation. *Agron. J.* 93:1327–1337. doi:10.2134/agronj2001.1327
- Littell, R.C., G.A. Milliken, W.W. Stroup, R.D. Wolfinger, and O. Schabenberger. 2006. *SAS for mixed models*. 2nd ed. SAS Inst., Cary, NC.
- Liu, S.M., G.A. Constable, B.R. Cullis, W.N. Stiller, and P.E. Reid. 2015. Benefit of spatial analysis for furrow irrigated cotton breeding trials. *Euphytica* 201:253–264.
- 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
- Papadakis, J.S. 1937. Méthode statistique pour des expériences sur champ. *Bull. Inst. Amélior. Plant. Thessalonique* 23.
- 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
- Pauli, D., P. Andrade-Sanchez, A.E. Carmo-Silva, E. Gazave, A.N. French, J. Heun, et al. 2016. Field-based high-throughput plant phenotyping reveals the temporal patterns of quantitative trait loci associated with stress-responsive traits in cotton. *G3: Genes, Genomes, Genet.* 6:865–879. doi:10.1534/g3.115.023515
- 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
- Richter, C., B. Kroschewski, H.-P. Piepho, and J. Spilke. 2015. Treatment comparisons in agricultural field trials accounting for spatial variation. *J. Agric. Sci.* 153:1187–1207. doi:10.1017/S0021859614000823
- Robinson, D.L., C.D. Kershaw, and R.P. Ellis. 1988. An investigation of two-dimensional yield variability in breeders' small plot barley trials. *J. Agric. Sci.* 111:419–426. doi:10.1017/S0021859600083581
- Rutkoski, J., J. Poland, S. Mondal, E. Autrique, L.G. Pérez, J. Crossa, et al. 2016. Canopy temperature and vegetation indices from high-throughput phenotyping improve accuracy of pedigree and genomic selection for grain yield in wheat. *G3: Genes, Genomes, Genet.* 6:2799–2808.

- Sharma, B., and G.L. Ritchie. 2015. High-throughput phenotyping of cotton in multiple irrigation environments. *Crop Sci.* 55:958–969. doi:10.2135/cropsci2014.04.0310
- Smith, K.F., and M.D. Casler. 2004. Spatial analysis of forage grass trials across locations, years, and harvests. *Crop Sci.* 44:56–62. doi:10.2135/cropsci2004.5600
- Sripathi, R., P. Conaghan, D. Grogan, and M.D. Casler. 2017. Spatial variability effects on precision and power of forage yield estimation. *Crop Sci.* 57:1383–1393. doi:10.2135/cropsci2016.08.0645
- Stroup, W.W., P.S. Baenziger, 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
- Tanger, P., S. Klassen, J.P. Mojica, J.T. Lovell, B.T. Moyers, M. Baraoidan, et al. 2017. Field-based high throughput phenotyping rapidly identifies genomic regions controlling yield components in rice. *Sci. Rep.* 7:42839. doi:10.1038/srep42839
- Velazco, J.G., M.X. Rodríguez-Ailvarez, M.P. Boer, D.R. Jordan, P.H.C. Eilers, M. Malosetti, and F.A. van Eeuwijk. 2017. Modelling spatial trends in sorghum breeding field trials using a two-dimensional P-spline mixed model. *Theor. Appl. Genet.* 130:1375–1392. doi:10.1007/s00122-017-2894-4
- Wang, X., K.R. Thorp, J.W. White, A.N. French, and J.A. Poland. 2016. Approaches for geospatial processing of field-based high-throughput plant phenomics data from ground vehicle platforms. *Trans. ASABE* 59:1053–1067. doi:10.13031/trans.59.11502
- Wilkinson, G.N., S.R. Eckert, T.W. Hancock, and O. Mayo. 1983. Nearest neighbour (NN) analysis of field experiments. *J. R. Stat. Soc., Ser. A* 45:151–211.
- Williams, E.R. 1986. Row and column designs with contiguous replicates. *Aust. N.Z. J. Stat.* 28:154–163. doi:10.1111/j.1467-842X.1986.tb00594.x
- Williams, E.R., and D.J. Lockett. 1988. The use of uniformity data in the design and analysis of cotton and barley variety trials. *Aust. J. Agric. Res.* 39:339–350. doi:10.1071/AR9880339
- 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
- Yang, R.-C., T.Z. Ye, S.F. Blade, and M. Bandara. 2004. Efficiency of spatial analyses of field pea variety trials. *Crop Sci.* 44:49–55. doi:10.2135/cropsci2004.4900
- Yang, W., Z. Guo, C. Huang, L. Duan, G. Chen, N. Jiang, et al. 2014. Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. *Nat. Commun.* 5:5087. doi:10.1038/ncomms6087
- Zimmerman, D.L., and D.A. Harville. 1991. A random field approach to the analysis of field-plot experiments and other spatial experiments. *Biometrics* 47:223–239. doi:10.2307/2532508