

Utilization of Multiyear Plant Breeding Data to Better Predict Genotype Performance

Vivi N. Arief, Haile Desmae, Craig Hardner, Ian H. DeLacy, Arthur Gilmour, Jason K. Bull, and Kaye E. Basford*

ABSTRACT

Despite the availability of multiyear, multicycle, and multiphase data in plant breeding programs for annual crops, selection is often based on single-year, single-cycle, and single-phase data. As genotypes in the same fields are usually grown under the same management practice, data from these fields can and should be analyzed together. In Monsanto's North American maize (*Zea mays* L.) breeding program, this approach enables a spatial model to be fitted in each field, providing an estimate of spatial trend and a better estimate of residual variance in each field. Multiyear, multicycle analysis showed that the estimates of genotype \times year variance (V_{GY}) and genotype \times year \times location variance (V_{GYL}) were still the largest components of the estimated phenotypic variance. Analysis of any single-year subset of the data inflated the estimate of genotypic variance (V_G) by the size of the estimate of V_{GY} , resulting in potential bias in the estimates of genotype performance. These results demonstrate the advantage of a combined analysis of data across years and cycles to make selection decisions for genotype advancement.

V. N. Arief, I. H. DeLacy, and K.E. Basford, The Univ. of Queensland, School of Agriculture and Food Sciences, St. Lucia, QLD 4072, Brisbane, Australia; H. Desmae, International Crops Research Institute for Semi-Arid Tropics–West & Central Africa (ICRISAT– WCA), Bamako, Mali; C. Hardner, The Univ. of Queensland, Queensland Alliance for Agriculture and Food Innovation, Brisbane, QLD 4072, Australia; A. Gilmour, Statistical and ASReml Consultant, Cargo, NSW 2800, Australia; J. K. Bull, The Climate Corporation, 4 Cityplace Dr., St. Louis, MO 63141, USA. Received 15 Mar. 2018. Accepted 26 Nov. 2018. *Corresponding author (k.e.basford@uq.edu.au). Assigned to Associate Editor Marcio Resende Jr.

Abbreviations: BLUE, best linear unbiased estimate; BLUP, best linear unbiased prediction; MET, multi-environment trial; TPE, target population environments.

MULTI-ENVIRONMENT TRIALS (METs) are commonly used in plant breeding programs to evaluate the relative performance of genotypes for a target population of environments (TPE) (Basford and Cooper, 1998). Multi-environment trials are essential because of the presence of genotype \times environment interaction, since genotypes respond in different ways to different environmental conditions. This reduces the overall progress from selection in any one environment (Basford and Cooper, 1998) and decreases the response to selection for average genetic performance (Comstock and Moll, 1963) due to the difficulty in selecting the best-performing and most stable genotypes.

Multi-environment trial data are commonly analyzed using linear mixed models in either one-stage or two-stage analyses. A one-stage analysis is usually considered the gold standard, as it is more efficient than a two-stage analysis (Smith et al., 2001, 2005). However, it is computationally expensive (Möhring and Piepho, 2009) when applied to analyze many environments, especially when different types of models are needed to characterize

Published in Crop Sci. 59:1–11 (2019).
doi: 10.2135/cropsci2018.03.0182

© Crop Science Society of America | 5585 Guilford Rd., Madison, WI 53711 USA
This is an open access article distributed under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>).

variability for each individual environment (Piepho et al., 2012). In contrast, a two-stage analysis is computationally efficient (Möhring and Piepho, 2009) and can handle a larger amount of data or more complex models than a one-stage analysis. A two-stage analysis can be used to model any specific randomization layout and within-environment error for each individual environment in the first stage (Smith et al., 2001; Welham et al., 2010; Piepho et al., 2012), then use the adjusted genotype mean for the across-environment second-stage analysis. In the analysis of plant breeding MET data, a spatial model is usually fitted for each environment (Cullis and Gleeson, 1991; Cullis et al., 1998; Smith et al., 2001) to produce a spatially adjusted genotype mean in each environment. The spatially adjusted genotype means are then combined as the data for the second stage, across-environment analysis. The loss of efficiency in a two-stage analysis is reduced by weighting the adjusted means according to the (inverse) predicted accuracy of the value in the individual environment analysis (Cullis et al., 1996; Möhring and Piepho, 2009; Piepho et al., 2012).

It is a common practice in plant breeding programs to conduct multiphase selection to identify the elite genotypes from all those derived from each cycle of the breeding program. A cycle of breeding is a period from initial crossing to final selection phase. For annual crops, each cycle of breeding usually consists of two or three selection phases, each corresponding to 1 yr of multilocation field trials. These multilocation field trials are also used to test genotypes from multiple breeding cycles that were in the same or different selection phases. This practice results in a large number of genotypes being tested every year in any given location. Often, these genotypes are divided into sets depending on their breeding cycle, selection phase, genetic relationship in the breeding program, and the TPE for which they were bred. Some established genotypes and common checks are usually included in each trial to provide connection among sets, especially for sets that target the same TPE (Piepho et al., 2006). All genotypes in the same field are normally grown under the same field management, although they are from different sets, so an overall experimental design that minimizes the estimated error within a field should be used (Basford et al., 1996; Williams and John, 1996; Federer, 2005). Moreover, a spatial analysis for a field can be conducted using all available data in that field.

For annual crops, genotype performance is usually evaluated on a single phase and single cycle of data across locations, hence a single-year analysis. The results from the analysis of any single year's data have been shown to overestimate the genotypic (V_G) and genotype \times location (V_{GL}) variances, as they were confounded with the genotype \times year (V_{GY}) and genotype \times year \times location (V_{GYL}) variances, respectively (Nyquist and Baker, 1991; Holland and Nyquist,

2010; Arief et al., 2015). Moreover, as the same genotypes appear in different phases within and across cycles, it is logical to combine data across selection phases and across cycles to provide better estimates of the performance of the common genotypes. In the past, multiyear analysis was impractical due to the lack of computing power to handle large unbalanced data. However, a two-stage mixed model analysis is now enabling multiyear data analysis.

It has been shown that multiyear data analysis would provide better estimates of V_G and V_{GL} , thus better estimating genotype performance (Arief et al., 2015). In this study, we evaluate the advantage of using multiyear, multiphase data for selection through simulated data for one of Monsanto's North American maize (*Zea mays* L.) breeding programs.

MATERIALS AND METHODS

Maize Data

Data were obtained from METs from one of Monsanto's North American maize breeding programs. The hybrids, which from this point on are referred to as genotypes, were evaluated in three phases (P1 to P3) of testing in a breeding cycle. The best genotypes from P1 were advanced to P2 the following year, and the best genotypes from P2 were advanced to P3 the third year. In these METs, genotypes were grouped into sets. These sets corresponded to breeding cycles, selection phases, relative maturity zones, breeding objectives, or special purposes.

Yield data ($t\ ha^{-1}$) collected from the advancement trials for genotypes targeted for deployment in the Midwest (referred to as Midwest genotypes) were used for this study. The data consist of >10,000 unique genotypes tested in >1000 fields across hundreds of locations in several states for a 6-yr period. A field is where the genotypes were grown, whereas a location is a collection of nearby fields. These genotypes were obtained from eight breeding cycles, of which four cycles (C3–C6) had data for all three phases. There were common genotypes across years (Table 1, lower triangular values) corresponding to the selection phases (i.e., 3-yr selection cycle). However, as indicated in the paragraph above, there were also common genotypes across the different phases and cycles within a year and across years. Common genotypes across phases within a cycle were genotypes that were selected to be advanced to the next phase. Common genotypes across cycles were genotypes that were selected from the previous cycles or checks. As expected, the number of common genotypes across phases within cycles was greater than the number of common genotypes across cycles. The number of common locations was relatively consistent across years (Table 1, upper triangular values).

Data Analysis

The two-stage analysis approach (Smith et al., 2001, Welham et al., 2010) was used to estimate the variance components in the maize MET with mixed-model restricted maximum likelihood (Gilmour et al., 1997) implemented in ASReml (Gilmour et al., 2009).

The first-stage analysis was applied to all data in a field, including data for genotypes that were not targeted for the Midwest

Table 1. Number of genotypes tested in common (lower triangular values) and locations used in common (upper triangular values) across selection-phases, breeding cycles and years, and total number of unique genotypes, locations, and fields in each year, phase, and cycle in Monsanto's Advancement Trials.

Year (Y)†	Phase (P)‡	Cycle (C)§	Y1, P1, C1	Y1, P2, C2	Y1, P3, C3	Y1, P1, C3	Y2, P1, C3	Y2, P2, C3	Y2, P3, C3	Y2, P1, C4	Y3, P2, C4	Y3, P3, C4	Y3, P1, C5	Y4, P2, C5	Y4, P3, C5	Y4, P1, C6	Y5, P2, C6	Y5, P3, C6	Y5, P1, C7	Y6, P2, C7	Y6, P3, C7	Y6, P1, C8
Y1	P3	C1	29¶	84	8	8	50	80	56	58	77	59	50	56	60	48	49	39	37	37	38	
Y1	P2	C2	35#	27	17	17	24	28	22	24	25	24	22	21	21	22	19	21	21	17	18	
Y2	P3	C2	64	52	31	31	68	77	10	60	72	62	50	55	61	50	46	39	37	38		
Y1	P1	C3	51	32	40	31	31	43	33	36	42	37	32	32	34	32	29	30	26	27		
Y2	P2	C3	43	37	51	158	46	19	41	43	41	32	33	37	32	29	29	29	27	26		
Y3	P3	C3	37	36	70	35	75	26	20	82	82	13	52	61	62	49	50	43	36	39		
Y2	P1	C4	11	9	57	15	52	50	47	49	49	52	37	35	45	36	29	30	27	26		
Y3	P2	C4	13	12	15	18	48	61	189	56	56	12	51	42	55	44	33	37	31	33		
Y4	P3	C4	22	18	36	23	47	82	24	41	42	42	23	59	15	53	54	44	43	45		
Y3	P1	C5	10	9	11	8	16	68	25	70	59	43	56	43	65	45	35	37	33	35		
Y4	P2	C5	9	6	12	7	19	16	19	40	60	129	44	44	16	45	41	36	36	38		
Y5	P3	C5	10	8	12	5	13	17	13	21	47	41	75	33	10	59	8	8	44	45		
Y4	P1	C6	8	7	10	7	11	14	11	12	68	56	63	49	45	42	36	38	39			
Y5	P2	C6	6	4	5	5	11	8	5	6	12	23	22	55	268	48	48	10	51	50		
Y6	P3	C6	2	1	2	3	8	9	3	5	21	10	13	43	15	47	5	8	6			
Y5	P1	C7	6	5	7	3	9	9	7	9	11	10	11	46	14	68	42	46	47			
Y6	P2	C7	4	3	5	3	8	9	2	4	11	2	6	11	3	10	52	203	9			
Y6	P1	C8	2	1	3	3	7	5	3	4	7	4	5	9	4	6	51	10	66			
No. genotypes††			113	394	173	1106	655	199	1167	521	200	1649	543	192	1913	772	181	2108	720	2214		
No. locations‡‡			139	37	105	53	59	108	72	74	101	88	73	101	91	90	92	69	73	72		
No. fields§§			142	40	110	63	66	112	86	75	102	111	121	101	140	143	92	117	87	105		

† Year of testing. Each year contains multiple breeding cycles in multiple selection phases.

‡ Selection phase. For each breeding cycle, selection is conducted in three phases (P1–P3). Selected genotypes are advanced from P1 to P2 and P2 to P3.

§ Breeding cycle. Each cycle contains genotypes from the same crosses.

¶ No. of locations in common in the particular row and column of year, phase, and cycle.

No. of genotypes in common in the particular row and column of year, phase, and cycle.

†† No. of total unique genotypes in each year, phase, and cycle.

‡‡ Total no. of locations used to test genotypes from each year, phase, and cycle. Each location contains multiple fields.

§§ Total no. of fields used to test genotypes from each year, phase, and cycle. Field is where the genotypes are grown.

region (referred to as “other” genotypes) but were planted in the same fields as Midwest genotypes and assumed to be grown under the common agronomic conditions. In the second-stage analyses, only data for Midwest genotypes were analyzed.

First-Stage Analysis: Field Spatial Analysis

This analysis was conducted for each of the fields where genotypes targeted for the Midwest were grown. These fields varied in size and in field arrangement (Fig. 1). Some fields were also used to test other deployment regions for non-advancement trials (Fig. 1). Replications, if present, were placed within genotype sets (Fig. 1a, 1e, and 1f).

The heatmap for observed yield showed patterns that were associated with environmental factors such as ponding (Fig. 1a and 1b), the effects of relative maturity and set (Fig. 1c), or management factors such as usage of pivot irrigation (Fig. 1d). A spatial model (Cullis and Gleeson, 1991; Cullis et al., 1998; Smith et al., 2001) was fitted to take account of those patterns. An extra step was added to the first-stage analysis to cope with the limited number of replications. In the first step, genotype was fitted as a random effect to provide an estimate of the spatial trend. This step enables all genotypes to be used to estimate

the trend. In the second step, genotype was fitted as a fixed effect to obtain best linear unbiased estimates (BLUEs) that were required for the second-stage analysis (Smith et al., 2001), and the variance components were fixed to the values obtained from the first step.

Ideally, a customized spatial model would be fitted for each field. However, due to the size of the data and number of trials, this approach was not practical. Moreover, it is a company requirement that the analysis must be automated. Therefore, the following spatial model was chosen as the base model for both steps in the first-stage analysis for each field:

$$y_{ipq} = \mu + \text{lin}(c) + \text{lin}(c).\text{lin}(r) + (b|s)_{jk} + g_i + (r|b|s)_{pjik} + (c|b|s)_{ajk} + \text{spl}(r) + \text{spl}(c) + \varepsilon_{ipq} : \text{AR1}(c) \otimes \text{AR1}(r)$$

where y_{ipq} is the grain yield for genotype i in row p and column q ; μ is the intercept; $\text{lin}(r)$ and $\text{lin}(c)$ are the fixed effect of linear function of rows and columns, with $\text{lin}(c).\text{lin}(r)$ being their interaction; $(b|s)_{jk}$ is the effect of block j within set k ; g_i is the effect of genotype i ; $(r|b|s)_{pjik}$ and $(c|b|s)_{ajk}$ are the random effects

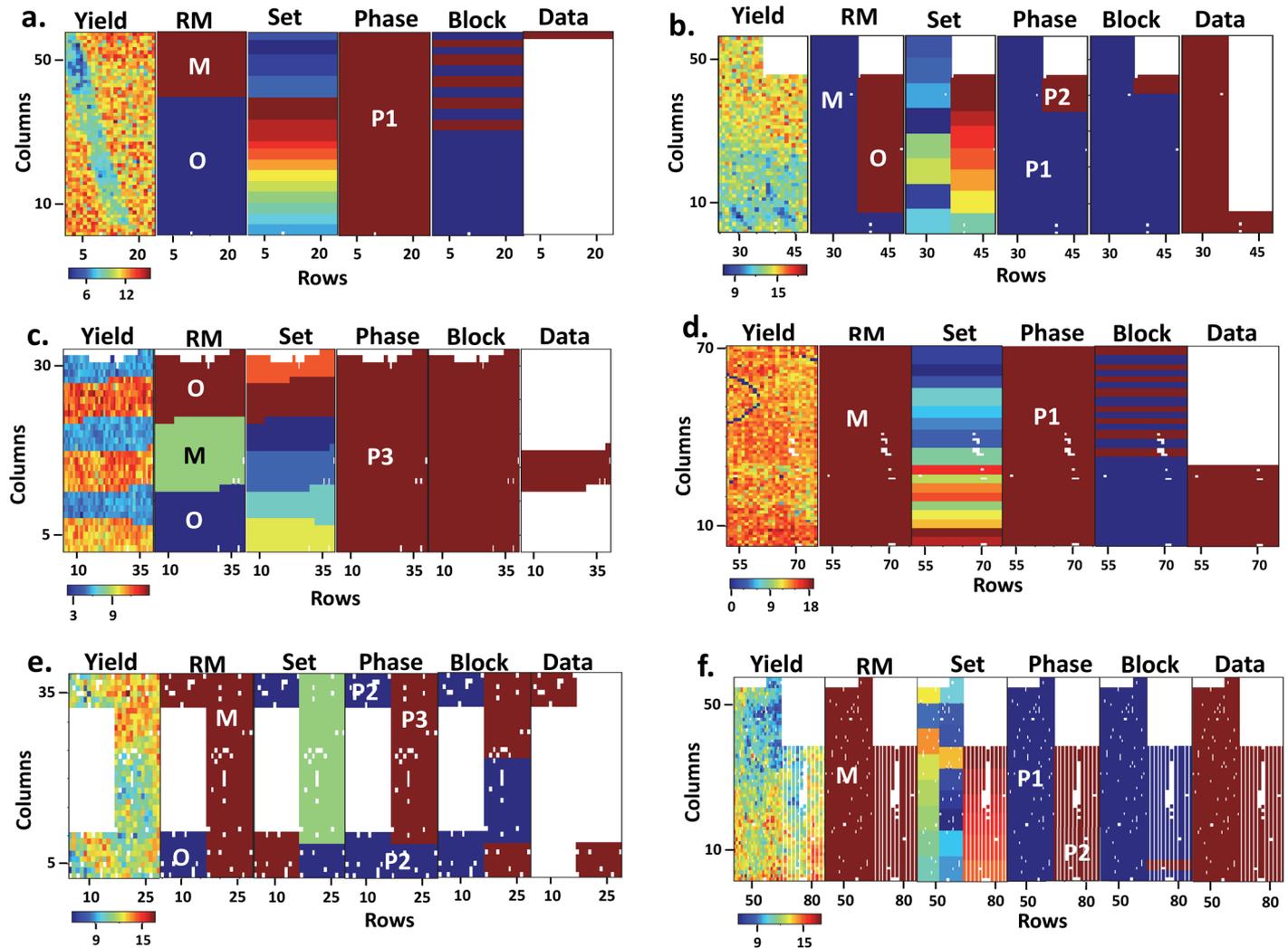


Fig. 1. Field layout for the six example fields in Monsanto’s Midwest Advancement Trials. Yield = yield data (t ha⁻¹), RM = relative maturity, M = genotypes targeted for Midwest region (Midwest genotypes), O = genotypes targeted for other regions (other genotypes), Set = genotype groups; Phase = selection phase (P1–P3), Block = replication, Data = data for Midwest genotypes.

of row p and column q , respectively, within block j and set k ; $\text{spl}(r)$ and $\text{spl}(c)$ are the spline effects of row r and column c , respectively; and ε_{ipq} is the residual effect modeled using first-order autoregression (AR1) for row r and column c . As noted above, the first-step analysis was used to obtain estimates of the variance components of the random spatial terms (i.e., rows and columns terms) when genotype was fitted as a random effect. In the second step, these estimates of variance components were used to obtain the genotype BLUEs and their respective weights by fitting the genotype as a fixed effect. The weights were calculated as the inverse of residual variance for each field multiplied by the replication number for each genotype and the pooled residual variance across fields (Cullis et al., 1996).

As trial sizes varied from field to field, this base spatial model was adjusted according to the following criteria:

1. No spatial model was fitted for a field with 10 or fewer replicated observations (i.e., when no. of nonmissing observations – no. of unique genotypes ≤ 10). For these fields, yield and mean yield were used as BLUEs, and the weights were set as 1 (as the residual variance for these fields was assumed to be equal to the estimated pooled variance and the number of replications was set to 1).
2. For fields where a spatial model was fitted:
 - a. The term $(b|s)_{jk}$ was only fitted when all genotypes were replicated;
 - b. The term $\text{lin}(c).\text{lin}(r)$ was only fitted when the number of rows and columns were both more than five;
 - c. The terms $\text{spl}(r)$, $(r|b|s)_{pjk}$, and $\text{AR1}(r)$ were only fitted when the number of rows was five or more; and
 - d. The terms $\text{spl}(c)$, $(c|b|s)_{qjk}$, and $\text{AR1}(c)$ were only fitted when the number of columns was five or more.

These first-stage analyses were conducted as an automated process using R software (R Development Core Team, 2008). This process was designed to report fields with convergence problems. This automated process also produced a series of plots for each field to display the results from the first-stage analysis. A variogram was produced to display the autoregression correlation of residuals, and a heatmap each was produced for centered yield data, genotype BLUEs, residual effects, and trend effects (calculated as the sum of all spatial terms in the model). These plots were used to evaluate the fit of the base spatial model and could be used by the breeder to get a good understanding of the fields being used for these trials.

Second-Stage Analysis: Across-Location Analysis

Across-location analyses were conducted using the BLUEs and weights for genotypes designated for Midwest advancement trials. Across-location analyses were conducted for the following data:

1. Single-year, single-phase data (18 analyses) to obtain variance estimates for each year and each phase.
2. Single-year, across-phase data (six analyses) to obtain variance estimates for each year.
3. Across-year, single-phase data (three analyses) to obtain variance estimates for each phase.

4. Across-year, single-cycle data for cycles with complete phases (i.e., C3 to C6, four analyses) to obtain variance estimates for each cycle.
5. Across-year, multicycle data for cycles with complete phases (i.e., data for C3 to C6, one analysis) to obtain variance estimates for each cycle. The analysis of this dataset provided estimates of genotypic variance for each cycle with a pooled estimate of genotype \times year, genotype \times location, and genotype \times year \times location, respectively.
6. All across-year data (two analyses) to obtain an overall estimate of genotypic variance and an estimate for each selection phase with a pooled estimate of genotype \times year, genotype \times location, and genotype \times year \times location, respectively.

The estimate of genotypic variance from any single-year data (i.e., Datasets 1 and 2) was expected to be inflated by the genotype \times year variance. Similarly, the estimate of genotype \times location variance from any single year was expected to be inflated by the genotype \times year \times location variance. The size of these confounded effects were evaluated by comparing the variance components from single-year analysis with the variance components from multiyear analysis (i.e., Datasets 3 to 4).

The models used to obtain the estimate of variance components were as follows:

1. $z_{it(jk)} \odot w_{it(jk)} = \mu + e_{i(jk)} + g_i + (\text{gl})_{ij} + \bar{\varepsilon}_{it(jk)}$ applied to single-year, single-stage data (Dataset 1) and single-year data (Dataset 2) to obtain an estimate of genotypic variance and genotype \times location variance for each phase and each year for Dataset 1, and for each year across stages for Dataset 2.
2. $z_{it(jk)} \odot w_{it(jk)} = \mu + e_{i(jk)} + g_i + (\text{gl})_{ij} + (\text{gy})_{ik} + (\text{gly})_{ijk} + \bar{\varepsilon}_{it(jk)}$ applied to single-phase across-year data (Dataset 3) and single-cycle data for cycles with complete phases (Dataset 4) to obtain an estimate of genotypic variance for each phase for Dataset 3, and for each cycle for cycles with complete phases for Dataset 4. This model was also applied to all data (Dataset 6) to obtain overall genotypic variance across years, stages, and cycles. In addition to genotypic variance, this model also provided variance estimates of genotype \times location, genotype \times year, and genotype \times year \times location, respectively.
3. $z_{it(jk)} \odot w_{it(jk)} = \mu + e_{i(jk)} + g_i | f_n + (\text{gl})_{ij} + (\text{gy})_{ik} + (\text{gly})_{ijk} + \bar{\varepsilon}_{it(jk)}$ applied to multicycle data (Dataset 5) to obtain genotypic variance for each cycle with pooled variance for genotype \times year, genotype \times location, and genotype \times year \times location variance, respectively. This model was used to remove the confounded effect of genotype \times year variance in the estimate of genotypic variance for each cycle.
4. $z_{it(jk)} \odot w_{it(jk)} = \mu + e_{i(jk)} + g_i | s_m + (\text{gl})_{ij} + (\text{gy})_{ik} + (\text{gly})_{ijk} + \bar{\varepsilon}_{it(jk)}$ applied to all data (Dataset 6) to obtain genotypic variance for each phase with pooled variance for genotype \times year, genotype \times location, and genotype \times year \times location, respectively.

location, respectively. This model was used to remove the confounded effect of genotype \times year variance in the estimate of genotypic variance for each phase;

where $z_{it(jk)} \odot w_{it(jk)}$ is the weighted BLUE from the first-stage analysis for genotype i in field t within location j and year k ; $e_{it(jk)}$ is the effect of environments defined as fields t within location j and year k ; g_i is the effect of genotype i ; $(gl)_{ij}$ is the effect of genotype i in location j ; $(gy)_{ik}$ is the effect of genotype i in year k ; $(gly)_{ijk}$ is the effect of genotype i in location j and year k ; $g_i|f_n$ is the effect of genotype i within cycle n ; $g_i|s_m$ is the effect of genotype i within phase m ; and $\bar{e}_{it(jk)}$ is the pooled residual. In a case when all fields in the datasets had <10 replicated observations, the weights were set to 1 and the pooled residual was set as the pooled residual from overall data.

The variance components obtained from these analyses were presented as a percentage of phenotypic variance (V_p). Phenotypic variance was calculated as the sum of genotypic variance (V_G), variance of genotype \times year interaction (V_{GY}), variance of genotype \times location interaction (V_{GL}), variance of genotype \times year \times location (V_{GYL}), and residual variance (V_R). For Models 3 and 4, V_p was calculated using an average V_G across cycles and phases, respectively.

Simulated Multi-environment Trial Data

Simulated data were used to evaluate the use of single-year and multiyear data for selection. The phenotypic data were calculated as the sum of genotypic (G), genotype \times year (GY), genotype \times location interaction (GL), genotype \times year \times location (GYL), and residual (R) effects. These effects were drawn from normal distributions with zero mean and variance of V_G , V_{GY} , V_{GL} , V_{GYL} , and V_R , respectively. These variance components were estimated using all data (Dataset 6) analyzed with Model 4 and expressed as the percentage of V_p (i.e., $V_p = 100\%$). As V_G was truncated due to selection (Cochran, 1951), the values of G were generated based on Stage 1's V_G and retained throughout the phases within a cycle (Arief et al., 2013). The values of GY, GL, GYL, and the residual were generated for each phase and each cycle assuming selection does not affect these variances.

For each cycle, a fixed percentage of top-performing genotypes were advanced from phase to phase. The residual variance was estimated from two replications of 20 checks. The phenotypic value for these checks was randomly generated in a similar way to the genotypes. These checks were not included in selection, and two of these checks were replaced every year. Simulated data were generated for 20, 40, and 60 locations for Phases 1, 2, and 3, respectively.

Three selection scenarios were evaluated. In all these scenarios, selection was always made within phase and cycle with no reselection (i.e., nonselected genotypes could not be reselected in later phases or later cycles). There was also no common genotypes across cycles, except for checks. In the first scenario, selection was based on genotype performance calculated from single-year, single-phase data, as in current practice. This scenario was appropriate for analyzing Dataset 1. In the second scenario, selection was based on genotype performance calculated from single-cycle data (i.e., three consecutive years

of data). This scenario was appropriate for analyzing Dataset 4. In the third scenario, selection was based on genotype performance calculated from all available data from 1 to 8 yr in the 8-yr period. This scenario was appropriate for analyzing Dataset 6. Each scenario was evaluated using 500 simulated datasets.

Three criteria were calculated for each simulation scenario:

1. The ratio of V_G to V_p (i.e., per-plot heritability) to evaluate the accuracy of variance component estimation. This value was compared with the expected ratio calculated from the parameter variance components.
2. The correlation between the estimated genotype performances and their parameter G values to evaluate the accuracy in estimating genotype performance. This correlation was calculated based on the performance of the current-year Phase 3 genotypes and their G value.
3. The proportion of common selected genotypes between the two selection methods. A set of 100 genotypes was selected either based on the estimated performance or based on the parameter G values. The number of genotypes that were selected in both methods was calculated and divided by 100. This proportion of common genotypes was used to evaluate how similar the two selection methods were in ranking the genotypes. As in the second criterion, selection was based on the performance of the current-year Phase 3 genotypes.

RESULTS

First-Stage Analysis of Real Data

The spatial model managed to model the spatial trends in all 81% of the fields where the model was fitted. The spatial model was not fitted for the remaining 19% of the overall total fields, as these fields had <10 repeated observations. There were no convergence problems with this spatial model. This ease of calculation was likely due to the two-step approach, where in the first step genotype was fitted as random, hence all plots were used to estimate the trend. The terms used to model the trend (i.e., spline terms for row and column, random row and column, and replication) generally captured most of the pattern in the field and removed that trend from the genotype effects (e.g., Fig. 2b, 2e, and 2f). The AR1 model fitted for rows and columns also adequately modeled the residual correlation (Fig. 2b, c, and d). However, as one base spatial model was used in all fields, there were some fields where the patterns were still confounded with residuals (Fig. 2a and 2d). In a few cases, the field trends were confounded with set effects (Fig. 2c).

Estimates of Variance Components

All variance components, except for environment variance (V_E), were expressed as a percentage of V_p . Environment variance was expressed as a percentage of total variance (i.e., $V_p + V_E$) and varied from 57 to 83% across datasets, with an average of 73%. The V_R was the largest

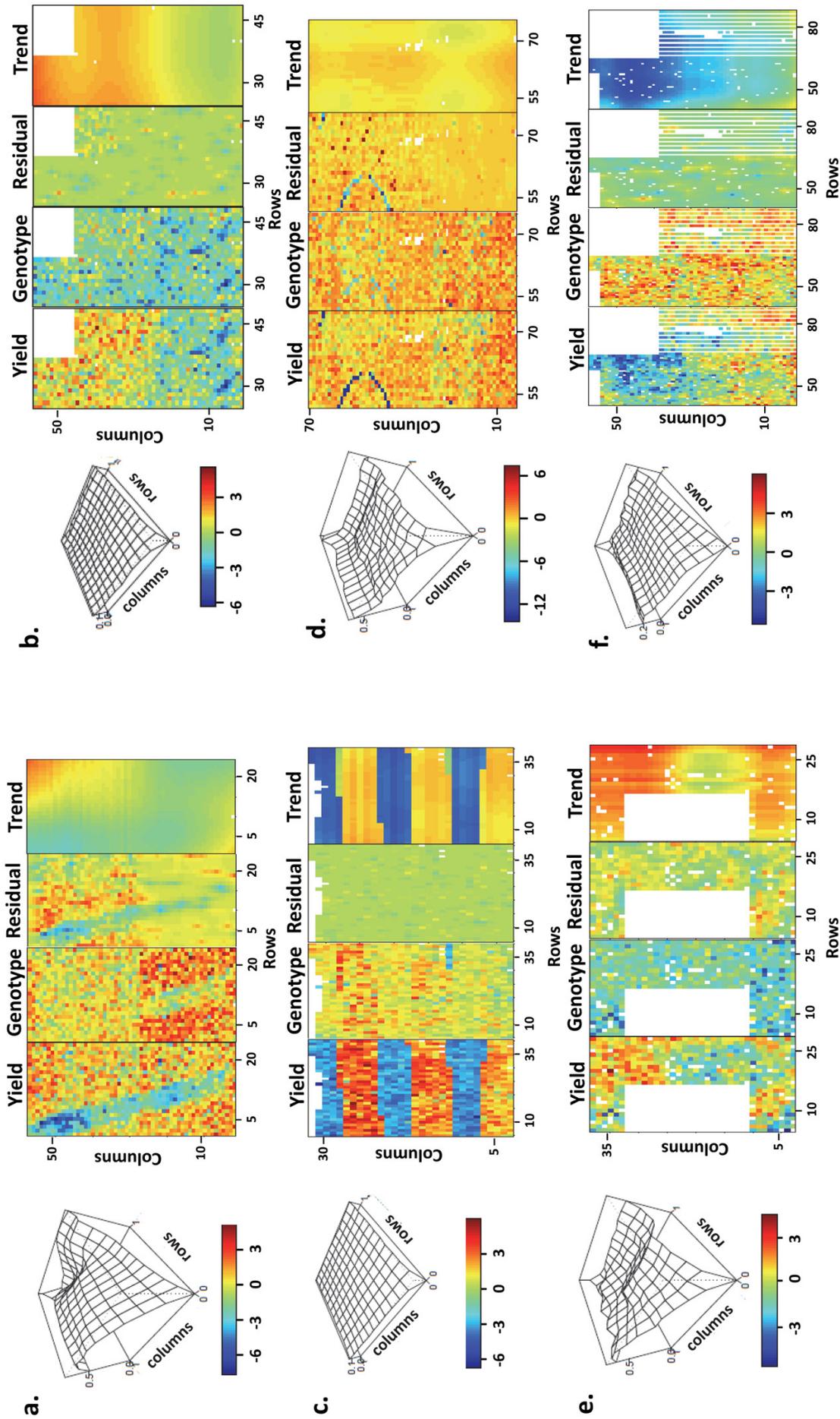


Fig. 2. Results from the first-stage analysis for the six example fields. Each field has five plots: a variogram to display the autoregression (AR1) correlation model for the residual, a heatmap for centered yield data (t ha⁻¹) (Yield), a heatmap for genotype effects (best linear unbiased estimates [BLUEs]) (Genotype), a heatmap for residual effects (Residual), and a heatmap for the sum of block, column, row, linear row, linear column, spline column, and spline row effects (Trend). The color intensity from blue to red indicates the range of the values from small to large.

component of V_p and was relatively consistent across datasets (mean = 58%, minimum = 47%, and maximum = 66%).

The V_G from single-year analysis (Fig. 3a and 3b) was always greater than the one estimated from across-year analysis (Fig. 3c–3e, black bars). The V_G from any single-year analysis (Fig. 3a and 3b) was roughly equivalent to the sum of V_G and V_{GY} from across-year analysis (Fig. 3c–3e; black and white bars, respectively). The same pattern was observed for V_{GL} , where V_{GL} from any single-year analysis (Fig. 3h and 3i) was roughly equivalent to the sum of V_{GL} and V_{GYL} from across-year analysis (Fig. 3j–3n; black and white bars, respectively). For any multiyear analysis, V_{GL} was always the smallest component of V_p (Fig. 3j–3n, black bars).

The V_G from single-phase analysis showed an increasing pattern across phases (Fig. 3c), whereas V_G estimated from the analysis of all data showed decreasing pattern from P1 to P3 (Fig. 3g). This discrepancy and the decreasing V_{GY} across phases from single-phase analysis (Fig. 3j) indicated the possibility of some confounding effects between V_G and V_{GY} , especially as there were not many common genotypes across cycles (Table 1). In contrast, the estimates of V_G from single-cycle analysis

(Fig. 3d) were consistent with the estimates of V_G from multicycle analysis (Fig. 3e). The two models (Models 2 and 4) fitted to all data produced similar estimates of V_{GY} , V_{GL} , and V_{GYL} (Fig. 3m and 3n), with V_G from Model 2 (Fig. 3f) similar to the average V_G across phases from Model 4 (Fig. 3g).

Simulation Results

The simulation results showed that across-year analysis provided better estimates of the variance components than single-year analysis (i.e., single-year, single-phase, and 1-yr data; Fig. 4a). The V_G/V_p ratio from single-year data was higher than the expected ratio of 0.1 and closer to the $(V_G + V_{GY})/V_p$ ratio of 0.17. The correlation of genotypic performance (i.e., best linear unbiased prediction [BLUP]) and the genotype parameters (G) from multiyear analysis was also higher than those from single-year analysis (Fig. 4b). As expected, the proportion of common top 100 genotypes showed a similar pattern as the correlation (Fig. 4c). Given those three criteria, the results from single-cycle analysis were similar to the ones from multiyear analysis with three or more years (Fig. 4). In general, there was no improvement in all three criteria after 3 yr.

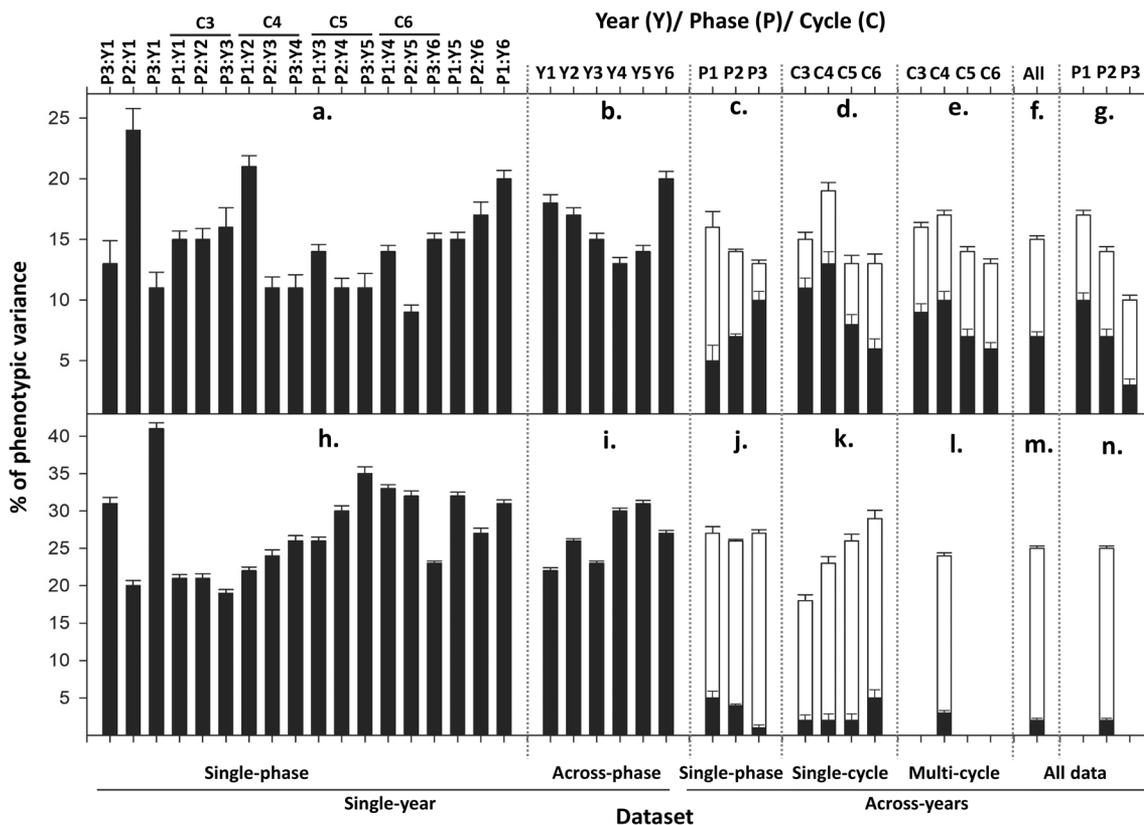


Fig. 3. Variance components expressed as percentage of phenotypic variance from the analysis of Monsanto's Midwest Advancement Trials. The top row contains the genotypic (V_G , black) and genotype \times year (V_{GY} , white), and the bottom row contains the genotype \times location (V_{GL} , black) and genotype \times year \times location (V_{GYL} , white) variances. These variance components were estimated from (a, h) single-year, single-cycle, and single-phase data; (b, i) single-year, across-phase data; (c, j) across-year, single-phase data; (d, k) across-year, single-cycle data; (e, l) across-year, multicycle data; (f, m) across-year data with overall V_G , V_{GY} , V_{GL} , and V_{GYL} ; and (g, n) across-year data with estimated V_G for each cycle and overall V_{GY} , V_{GL} , and V_{GYL} .

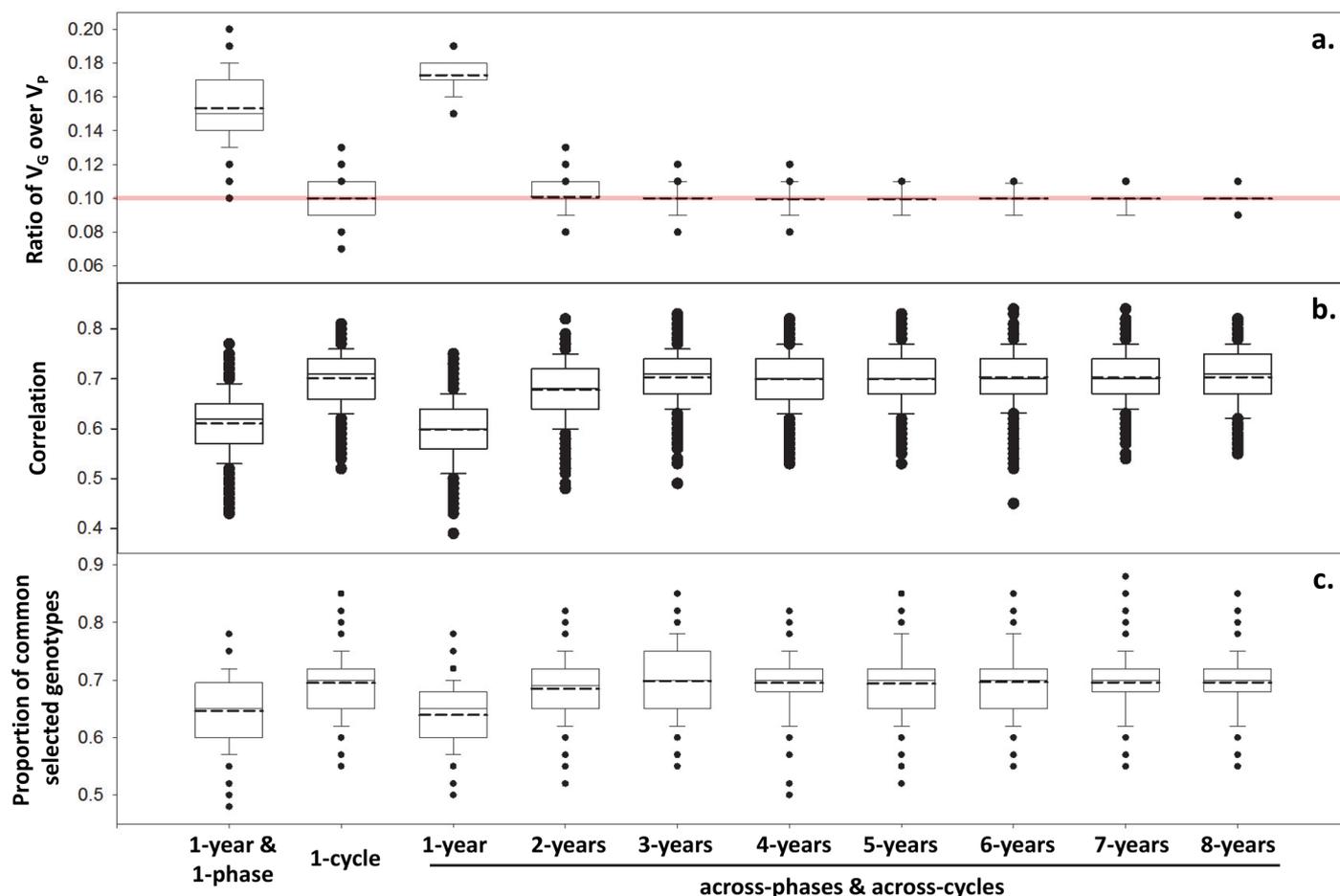


Fig. 4. Simulation results. (a) Ratio of genotypic (V_G) and phenotypic variance (V_P) (i.e., per-plot heritability). The red line indicates the expected ratio. (b) Correlation between the true genotype values (i.e., G) and their performance (i.e., best linear unbiased prediction [BLUP]) for all the current-year Stage 3 genotypes. (c) Proportion of the common top 100 genotypes selected in the current-year Stage 3 trial based on BLUP and G values. This proportion was calculated as the number of the same genotypes in the top 100 selected genotypes based on BLUP and G values divided by 100. Dotted lines in the boxplot indicate the mean.

DISCUSSION

The maize data used for our analysis were unbalanced, with the design and size varying across fields. Under such a scenario, a two-stage analysis approach is commonly used, whereby genotype means are first estimated separately for each field and then combined to form a data array for an overall mixed model analysis (Frensham et al., 1997). A weighted two-stage analysis approach has been found to perform adequately for variety prediction (Smith et al., 2001; Welham et al., 2010; Piepho et al., 2012). The first-stage analysis for each individual environment offers a convenient analysis that accounts for any specific randomization layout and within-environment error modeling (Piepho et al., 2012)—for example, a fitting spatial model for each environment.

We argue that the spatial model should use data for all genotypes under the same management in that field regardless of their designated set. This approach offers several advantages. First, it provides a better estimate of genotype performance, especially for genotypes that were repeated in a field. Second, it provides better estimates of variance

components. Third, it enables a spatial model to be fitted for small trials. For example, in this study, spatial analysis could not be conducted for the Midwest advancement trials only, as these trials were relatively small (Fig. 1a–1e).

While using all available data of a field increases the ability to fit spatial model, most fields have a small number of repeated observations. This created convergence problems when a spatial model was fitted. Therefore, in this study, an additional step was added to the first stage of the two-stage approach. In this additional step, a genotype was fitted as a random effect to obtain estimates of variance components for the spatial terms. These variance components were then used in the second step where genotype was fitted as a fixed effect. This second step is necessary to obtain a genotype BLUE, rather than a genotype BLUP, for the second-stage analysis (Smith et al., 2001). This two-step procedure enables the spatial model to be fitted in fields with a small number of repeated observations and eliminate the convergence problem. Moreover, a better estimate of spatial variability is expected from this two-step procedure, as information from all genotypes was used.

Automatization of data analysis is required for any large breeding program. Due to the large number of fields, a customized model for each field is not practical. Thus, an implementation of a one-model-fits-all approach is necessary. In this study, one base spatial model was fitted for all fields, with some adjustment to accommodate different field sizes. A series of heatmaps produced for each field can provide a quick tool to evaluate the fit of this one-model-fits-all in capturing the pattern in each field. In general, the base spatial model fitted for all fields managed to capture the spatial trend in the fields (Fig. 2), except for a few fields where some residual structure was still observed (Fig. 2a and 2d) or for special-purpose trials where treatment effects were confounded with field trends (Fig. 2c). However, due to the large number of fields used in the second-stage analysis, the differences between a one-model-fits-all and customized models are likely to be negligible (Qiao et al., 2004).

A single-year analysis invariably overestimates V_G and V_{GL} components (Fig. 3), as these estimates are confounded with the V_{GY} (Nyquist and Baker, 1991; Holland and Nyquist, 2010; Arief et al., 2015). Multiyear analysis provides an estimate for the V_{GY} and V_{GYL} components, and therefore better estimates of V_G and V_{GL} . The accuracy of V_{GY} and V_{GYL} estimates are affected by the number of genotype connections across years, as indicated by the differences between V_G estimated from single-phase data and all data (Fig. 3a). The V_G was expected to decrease from P1 to P3 due to selection (Cochran, 1951). However, analysis of single-phase data showed an opposite pattern: V_G increased from P1 to P3 (Fig. 3c, black bars). These results were likely to be an artifact of a poorly estimated V_{GY} , as V_{GY} was decreasing from P1 to P3 (Fig. 3j, black bars). The lack of genotype connections across years within stages caused V_G to be partially confounded with V_{GY} , resulting in an overestimated V_G and an underestimated V_{GY} . In contrast, V_G estimated from all data showed the expected decrease across phases (Fig. 3g). There were more genotype connections across years in all data than in single-phase data, and that helps in the estimate of V_{GY} . When the genotype connections across years were sufficient, similar estimates of V_G for each cycle were obtained from both single- and multicycle data (Fig. 3d and 3e).

The ratios of variance components from all data were consistent with the ones from Texas commercial maize (Barrero Farfan et al., 2013). However, the ratio of V_{GL} from this study was smaller than the one in other crops (Cullis et al., 1996; Butler et al., 2000; de la Vega et al., 2007; Arief et al., 2015). The genotypes in these trials were targeted for the Midwest and were only tested in locations assigned to these maturity zones. As the genotypes were highly adapted and the test locations have high similarity, small estimates of

V_{GL} were expected. It is interesting to note that the estimated V_{GL} values (Fig. 3j–3n, black bars) were relatively smaller than the V_{GY} values (Fig. 3c–3g, white bars). These results could indicate the effectiveness of relative maturity zone in partitioning TPEs for maize in North America. However, V_{GYL} is still the largest component of the phenotypic variance (Fig. 3c–3n, white bars), so multiyear data analysis would provide an advantage over single-year data analysis.

The simulation results indicated that, under the assumption of constant V_G , V_{GY} , V_{GL} , V_{GYL} , and V_R across years and cycles, multiyear analysis provides better estimates of variance components and genotype performance, as well as increased accuracy of selection (Fig. 4). The simulation results also showed no advantage of using >3 yr of data (Fig. 4b and 4c). This was due to the current simulation setting where there were no overlapping genotypes across cycles. No improvement after 3 yr of data was also due to the assumption that spatial variations had been equally accounted for in all three simulation scenarios. For the real data, although the number of overlapping genotypes across cycles was minimal (Table 1), multiyear data could provide an advantage over single-cycle data by providing better estimates of spatial variation. However, it is clear that the analysis of multiyear data always provides better estimates of genotype performance than the analysis of single-year data.

CONCLUSION

The combined across-year analysis enables the use of information from all years to improve estimates of genotype performance. It enables a spatial model to be fitted for each field and separation of the effects of genotype and genotype \times year interaction in the estimation of variance components across fields. In simulated data, this resulted in a better correlation between estimated performance and the parameter genotypic values. As the simulation contained no overlapping cycles, there was no advantage from analyzing >3 yr of data (containing one complete cycle). In practice, the analysis of more years and across-cycle data can provide better estimates of genotype performance than the analysis of single-cycle data. Thus, in general, the combined analysis of data across years and cycles is better for making selection decisions for genotype advancement.

Conflict of Interest

The authors declare that there is no conflict of interest.

References

- Arief, V., I. DeLacy, T. Tian, J. Crossa, H.-J. Braun, K. Basford, and M. Dieters. 2013. Evaluation of two-times testing for the CIMMYT's international wheat nurseries. Wheat Breeding Assembly, Brisbane.

- Arief, V.N., I.H. DeLacy, J. Crossa, T. Payne, R. Singh, H.-J. Braun, et al. 2015. Evaluating testing strategies for plant breeding field trials: Redesigning a CIMMYT international wheat nursery to provide extra genotype connection across cycles. *Crop Sci.* 55:164–177. doi:10.2135/cropsci2014.06.0415
- Barrero Farfan, I.D., S.C. Murray, S. Labar, and D. Pietsch. 2013. A multi-environment trial analysis shows slight grain yield improvement in Texas commercial maize. *Field Crops Res.* 149:167–176. doi:10.1016/j.fcr.2013.04.017
- Basford, K.E., and M. Cooper. 1998. Genotype \times environment interactions and some considerations of their implications for wheat breeding in Australia. *Aust. J. Agric. Res.* 49:153–174. doi:10.1071/A97035
- Basford, K.E., E.R. Williams, B.R. Cullis, and A. Gilmour. 1996. Experimental design and analysis for variety trials. In: M. Cooper and G.L. Hammer, editors, *Plant adaptation and crop improvement*. CAB Int., Wallingford, UK. p. 125–138.
- Butler, D.G., R.J. Redden, I.H. DeLacy, and T. Usher. 2000. Analysis of line \times environment interactions for yield in navy beans. 1. Component of variance. *Aust. J. Agric. Res.* 51:597–605. doi:10.1071/AR97135
- Cochran, W.G. 1951. Improvement by means of selection. In: J. Neyman, editor, *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, CA. 31 July–12 Aug. 1950. Univ. California Press, Berkeley. p. 449–470.
- Comstock, R.E., and R.H. Moll. 1963. Genotype-by-environment interactions. In: W.D. Hanson and H.F. Robinson, editors, *Statistical genetics and plant breeding*. Natl. Acad. Sci., Natl. Res. Council, Washington, DC. p. 164–194.
- 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., B. Gogel, A. Verbyla, and R. Thompson. 1998. Spatial analysis of multi-environment early generation variety trials. *Biometrics* 54:1–18. doi:10.2307/2533991
- Cullis, B.R., F.M. Thomson, J.A. Fisher, A.R. Gilmour, and R. Thompson. 1996. Analysis of the NSW wheat variety database. II. Variance component estimation. *Theor. Appl. Genet.* 92:28–39. doi:10.1007/BF00222948
- de la Vega, A.J., I.H. DeLacy, and S.C. Chapman. 2007. Progress over 20 years of sunflower breeding in central Argentina. *Field Crops Res.* 100:61–72. doi:10.1016/j.fcr.2006.05.012
- Federer, W.T. 2005. Augmented split block experiment design. *Agron. J.* 97:578–586. doi:10.2134/agronj2005.0578
- Frensham, A., B. Cullis, and A. Verbyla. 1997. Genotype by environment variance heterogeneity in a two-stage analysis. *Biometrics* 53:1373–1383. doi:10.2307/2533504
- Gilmour, A., 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
- Gilmour, A.R., B.J. Gogel, B.R. Cullis, S.J. Welham, and R. Thompson. 2009. *ASReml user guide*, Release 3.0, VSN Int., Hemel Hempstead, UK.
- Holland, J.B., and W.E. Nyquist. 2010. Estimating and interpreting heritability for plant breeding: An update. In: J. Janick, editor, *Plant breeding reviews*. Vol. 22. John Wiley & Sons, Oxford, UK. p. 9–112. doi:10.1002/9780470650202.ch2
- Möhrling, J., and H.-P. Piepho. 2009. Comparison of weighting in two-stage analysis in plant breeding trials. *Crop Sci.* 49:1977–1988. doi:10.2135/cropsci2009.02.0083
- Nyquist, W.E., and R.J. Baker. 1991. Estimation of heritability and prediction of selection response in plant populations. *Crit. Rev. Plant Sci.* 10:235–322. doi:10.1080/07352689109382313
- Piepho, H.P., A. Büchse, and B. Truberg. 2006. On the use of multiple lattice designs and α -designs in plant breeding trials. *Plant Breed.* 125:523–528. doi:10.1111/j.1439-0523.2006.01267.x
- Piepho, H.-P., J. Möhrling, T. Schulz-Streeck, and J.O. Ogutu. 2012. A stage-wise approach for the analysis of multi-environment trials. *Biom. J.* 54:844–860. doi:10.1002/bimj.201100219
- Qiao, C.G., K.E. Basford, I.H. DeLacy, and M. Cooper. 2004. Advantage of single-trial models for response to selection in wheat breeding multi-environment trials. *Theor. Appl. Genet.* 108:1256–1264. doi:10.1007/s00122-003-1541-4
- R Development Core Team. 2008. *R: A language and environment for statistical computing*. R Found. Stat. Comput., Vienna.
- Smith, A., B. Cullis, and R. Thompson. 2001. Analysis variety by environments data using multiplicative mixed models and adjustments for spatial field trend. *Biometrics* 57:1138–1147. doi:10.1111/j.0006-341X.2001.01138.x
- Smith, A.B., B.R. Cullis, and R. Thompson. 2005. The analysis of crop cultivar breeding and evaluation trials: An overview of current mixed model approaches. *J. Agric. Sci.* 143:449–462. doi:10.1017/S0021859605005587
- Welham, S.J., B.J. Gogel, A.B. Smith, R. Thompson, and B.R. Cullis. 2010. A comparison of analysis methods for late-stage variety evaluation trials. *Aust. N. Z. J. Stat.* 52:125–149. doi:10.1111/j.1467-842X.2010.00570.x
- Williams, E.R., and J.A. John. 1996. Row-column factorial designs for use in agricultural field trials. *J. R. Stat. Soc. Ser. C Appl. Stat.* 45:39–46. doi:10.2307/2986221