Using soybean historical field trial data to study genotype by environment variation and identify mega-environments with the integration of genetic and non-genetic factors

Matheus D Krause[1], Kaio O G Dias[2], Asheesh K Singh[1], and William D Beavis[1,*]

[1]*Department of Agronomy, Iowa State University, Ames, IA 50011, USA*

[2]*Department of General Biology, Federal University of Viçosa, Viçosa, Brazil*

[*]*Corresponding author: wdbeavis@iastate.edu*

Saturday 15th April, 2023

1 Highlights

- Mega-environments can be identified with phenotypic, geographic, and meteorological data.
- Reliable estimates of variances can be obtained with proper analyses of historical data.
- Genotype by location was more important than genotype by year variation for seed yield.
- The trend in genotype by environment variances was captured in probability distributions.

2 Abstract

Soybean (*Glycine max* (L.) Merr.) provides plant-based protein for global food production and is extensively bred to create cultivars with greater productivity in distinct environments. Plant breeders evaluate new soybean genotypes using multi-environment trials (MET). The application of MET assumes that trial locations provide representative environmental conditions that cultivars are likely to encounter when grown by farmers. In addition, MET are important to depict the patterns of genotype by environment interactions (GEI). To evaluate GEI for soybean seed yield and identify mega-environments (ME), a retrospective analysis of 39,006 data points from experimental soybean genotypes evaluated in preliminary and uniform field trials conducted by public plant breeders from 1989-2019 was considered. ME were identified from phenotypic information from the annual trials, geographic, soil, and meteorological records at the trial locations. Results indicate that yield variation was mostly explained by location and location by year interactions. The static portion of the GEI represented 26.30% of the total yield variance. Estimates of variance components derived from linear mixed models demonstrated that the phenotypic variation due to genotype by location interaction effects was greater than genotype by year interaction effects. A trend analysis indicated a two-fold increase in the genotypic variance between 1989-1995 and 1996-2019. Furthermore, the heterogeneous estimates of genotypic, genotype by location, genotype by year, and genotype by location by year variances, were encapsulated by distinct proba-

bility distributions. The observed target population of environments can be divided into at least two and at most three ME, thereby suggesting improvements in the response to selection can be achieved when selecting directly for clustered (i.e., regions, ME) versus selecting across regions. Clusters obtained using phenotypic data, latitude, and soil variables plus elevation, were the most effective. In addition, we published the R package SoyURT which contains the data sets used in this work.

3 Keywords

Soybean; Genotype by Environment Interaction; Multi-environmental trials; Target Population of Environments; Mega-environments

4 Introduction

The terms genotype (G) and phenotype (P) were first coined by Johannsen (1911) after the rediscovery of Mendel's work. Since then, the understanding of the mapping function that links G to P has been an on-going research interest (Pigliucci 2001, p. 2). The mapping of G to P for most quantitatively expressed traits is complicated by the differential response of genotype(s) to different environments, i.e. genotype by environment interactions (GEI), wherein phenotypic variation is shaped by G, Environment (E), and GEI (Tabery 2008; Sprague and Federer 1951). The GEI component will increase P variance and reduces heritability estimates, complicating breeding decisions and lowering response to selection. Additionally, it leads to variation in the adaptation and plasticity of genotypes' response to targeted agroecological zones (Mackay *et al.* 2019; Cooper and DeLacy 1994; Haldane 1947). Hence, the GEI is particularly important to breeders as they explore the responsiveness of varieties for specific or broad environments.

Plant breeders evaluate candidate genotypes in multi-environment trials (MET) to reveal GEI patterns (Oakey *et al.* 2016; Smith *et al.* 2001b). MET utilize locations sampled from a target population of environments (TPE) representing farm production environments, i.e. the growing

conditions that an experimental genotype is expected to encounter as a cultivar grown by farmers (Bustos-Korts *et al.* 2021). Hence, a TPE comprises many environments, spatially across agroecological zones and temporally over the years (Crespo-Herrera *et al.* 2021). The manifestation of GEI in a TPE has two components, the "static" environmental characteristics such as soil, longitude, latitude, and "non-static" seasonal environmental characteristics such as weather and management practices (Cullis *et al.* 2000). If GEI is large and associated with consistent sub-groupings of environments within the TPE, greater gains from selection should be achieved by subdividing locations into Mega-Environments (ME) (Crespo-Herrera *et al.* 2021; Yan 2016; Atlin *et al.* 2000a).

According to CIMMYT (1989, p. 58), "ME are broad, not necessarily contiguous areas, defined by similar biotic and abiotic stresses, cropping system requirements …". Another definition is a group of environments that share the same winning genotypes (Kang 2020; Gauch and Zobel 1997), or that within ME there is minimal crossover interaction among the genotypes across environments (Singh *et al.* 2021; Smith *et al.* 2021). It should be noted that the terms subregion, region/regional, subdivision, clusters, zones, agroclimatic, ecogeographic, and ME are sometimes interchangeably used in the literature. One way of exploring GEI is to divide the TPE into ME, and to select within ME (Yan 2016). Some studies have investigated strategies to subdivide the TPE in maize (Windhausen *et al.* 2012), barley (Atlin *et al.* 2000b), wheat (George and Lundy 2019; Bustos-Korts 2017), sorghum (da Silva *et al.* 2021), alfalfa (Annicchiarico 2021), rice (Krishnamurthy *et al.* 2017), oat (Yan *et al.* 2010), and soybean (Zdziarski *et al.* 2019; Yan and Rajcan 2002).

There are several methods for dividing the TPE into ME. For example, the genotype main effect plus GEI (GGE) biplots (Yan *et al.* 2000) on soybean MET data was used by Zdziarski *et al.* (2019) to identify two ME in Midwestern Brazil with contrasting altitudes, levels of fertilizer, and incidence of soybean cyst nematode profiles. da Silva *et al.* (2021) and Krishnamurthy *et al.* (2017) also took advantage of GGE biplots to pinpoint ME for pre-commercial sorghum hybrids in Brazil and rice genotypes in India, respectively. For wheat, Crespo-Herrera *et al.*

(2021) defined three ME in India with climate and soil data through principal component analysis, followed by a hierarchical clustering based on Euclidean distance with Ward's method. For maize in Africa (CIMMYT's program), Windhausen *et al.* (2012) explored historical (2001-2009) MET data to determine ME according to five subdivision systems (climate, altitude, geographic, country, and yield level), and concluded there was enough genotype by subregion interaction relative to genotypic variance to justify the selection for the low and high-yielding sub-regions separately. Other methodologies such as the additive main effects and multiplicative interaction (AMMI) model (Bustos-Korts 2017; Gauch and Zobel 1997) and factor analytic (FA) models (Smith *et al.* 2021; Bustos-Korts 2017; Smith *et al.* 2015, 2001b; Piepho 1997) also have been used. When ME are included in the model, it is called a zone-based model; therefore, yielding zone-based predictions (Buntaran *et al.* 2019). One of the main advantages of modeling ME in a mixed model framework is the ability to borrow information between zones from the genotype by ME interaction. This is particularly beneficial when fewer testing locations are available creating a sparse representation of genotypes in some ME (Piepho *et al.* 2016; Piepho and Möhring 2005).

The effectiveness of subdividing the TPE into ME was assessed by Atlin *et al.* (2000a) based on the theory of correlated response to selection, first applied to the GEI problem by Falconer (1952). Effective selection occurs when subdivision increases response to selection, which might occur if the genotype by ME interaction variance, i.e. genotype by region (σ^2_{GR}), is large relative to the genotypic variance (σ^2_G). In terms of variance components derived from linear models, the GEI is composed of genotype by location (σ^2_{GL}), genotype by year (σ^2_{GY}), and genotype by location by year (σ^2_{GLY}) interaction variances. Both σ^2_{GY} and σ^2_{GLY} are non-static (unrepeatable) sources of variation. ME can be identified with the static portion of the σ^2_{GL}, which is partially repeatable across years (Yan 2016). When ME are identified and modeled, the σ^2_{GL} can be partitioned into σ^2_{GR} and genotype by location within ME ($\sigma^2_{GL(R)}$). Furthermore, the σ^2_{GLY} can be partitioned into a genotype by ME by year interaction (σ^2_{GRY}), and genotype by location within ME by year ($\sigma^2_{GL(R)Y}$) interaction (Atlin *et al.* 2000a). Consequently, accurate estimates of variance

components are critical to properly identify ME.

Variance components can be estimated with unbalanced historical data to provide information for designing novel breeding strategies and optimizing resource allocation (Aguate *et al.* 2019). Efforts have been made to quantify component variability using historical MET data in wheat, maize, sunflower, sugar beet, potato, rye (Meyer *et al.* 2011; Laidig *et al.* 2008), among other commercial crops. However, proper modeling of historical data can be a significant challenge (Dias *et al.* 2020), and if not done properly can lead to misleading interpretations. In terms of variance component estimators, recent work from Aguate *et al.* (2019) and Hartung and Piepho (2021) considered both the imbalance of data (due to selection) and the properties of the residual maximum likelihood (REML) method (Patterson and Thompson 1971) to shed light onto potential bias of estimated variance components from MET using linear mixed models. Their results served as guidelines for designing the variance estimation portion of this work, which will be discussed later.

To identify and describe ME for soybean seed yield in the primary production area of North America, we obtained historical soybean performance data from Uniform Soybean Cooperative Tests (USDA 2021). We purposely chose this dataset because these trials have been used for decisions on cultivar release by public breeding organizations. The dataset consisted of 39,006 data points for maturity groups (MG) II and III spanning 63 locations between 1989 and 2019. Note that experimental genotypes were not evaluated at all locations within years and most were not evaluated for more than one year. The objectives of this study were to (i) investigate if the observed TPE spanning 31 years of field trial evaluations can be classified into ME, and (ii) estimate probability density functions for the underlying trend of genotypic, genotype by location, genotype by year, genotype by location by year, and residual variance components derived from linear mixed models. This modeling approach allowed us to fit parametric probability distributions to variance components to capture the GEI trend that can be used in future simulation studies, needed for predicting plant breeding outcomes in changing climates. Currently, simulation studies rely on point estimates of variance components (Kleinknecht *et al.* 2016), or set heritability values such

as low or high (Rutkoski 2019). By capturing the GEI trend using historical data, we generate reliable variance estimates that can be used to conduct more realistic simulation studies.

5 Data and Methods

5.1 Phenotypic data

Annual PDF reports from the Northern Region of the Uniform Soybean Tests were obtained from the USDA-ARS (see Section 12). Information from the PDF files was transcribed into CSV format files. The retrieved data for MG II and III contain the empirical best linear unbiased estimated values (eBLUE) of genotypic means for seed yield adjusted to 13% moisture and reported in bushels per acre (bu/ac). Information from the coefficient of variation (CV%), genotypic variance (see Section 5.3.1), and the number of blocks per trial was also obtained. Before data analyses, individual trials with estimates of reliability (i^2, see Section 5.3.1) lower than 0.10, coefficients of variation (CV%) greater than 20%, and individual observations lower than ten bu/ac were removed. A full description of the data cleaning is given in Appendix A.

The trials were divided into MG, Preliminary (PYT), and Uniform Regional (URT) Yield Trials. Because there are large numbers of experimental genotypes created by several public breeding programs within each MZ, the PYT are further split into two groups: PYT-A and PYT-B. Additionally, experimental genotypes with introgressed transgenic alleles were evaluated independently in trials referred to as PYT/URT-RR or PYT/URT-TM, depending on whether the lines included Roundup Ready or other transgenic traits. Preliminary analyses revealed that estimates of variance components in MG II were not significantly different than in MG III. Thus, subsequent analyses were conducted on combined data from both MG (Appendix B).

Operationally, experimental genotypes were first evaluated in PYT, and if not culled, were subsequently evaluated in URT. Most experimental genotypes are culled on an annual basis, thus only 0.47% of all potential combinations of experimental genotypes, locations, and years were evaluated. Field trials were laid out in randomized complete blocks (RCBD) with two or more

blocks. In a given year, a PYT was usually conducted at eight or more locations. Experimental genotypes retained for regional trials were evaluated at 12 or more locations in URT. Some experimental genotypes might be evaluated in two subsequent years of URT. In addition to the experimental genotypes, entries in each field block included common check varieties (~ 3). We noted that check varieties were seldom retained for more than four consecutive years. For more information about the trial field plot design and agronomic practices, please refer to the PDF files.

5.2 Environmental data

In addition to phenotypic (PHE) data from yield trials, environmental data associated with trial locations were obtained. Elevation information was obtained from the "elevatr" package (Hollister *et al.* 2021). Soil characteristics at a depth of 5-15 cm were downloaded from Soilgrids (`https://soilgrids.org/`) with a modified R script available at `https://github.com/zecojls/downloadSoilGridsV2`, and further processed with the package "raster" (Hijmans 2021). The soil characteristics are referred to as soil variables (SV) and include bulk density (SV1), cation exchange capacity (SV2), clay content (SV3), total nitrogen content (SV4), pH (SV5), sand content (SV6), silt content (SV7), and organic carbon content (SV8). Detailed information about SV are available on the Soilgrids website. Latitudes for locations in the USA were downloaded from `https://simplemaps.com/data/us-cities`, and Canadian locations were obtained using Google Maps. Meteorological data, referred herein as MV for each location were obtained from "NASA's Prediction of Worldwide Energy Resources" (NASA POWER, `https://power.larc.nasa.gov/`) with the package "nasapower" (Sparks 2018), and further processed with the "EnvRtype" package (Costa-Neto *et al.* 2021). In total, 19 MV were retrieved daily (averages) from the average planting date until the average check variety maturity date (R8) for each environment (location by year combination). A summary of the environmental variables is provided in Appendix C (Tables C1 and C2), and for more detailed information, please refer to the cited references.

5.3 Data analyses

A stage-wise approach was followed (Piepho *et al.* 2012; Smith *et al.* 2001a; Frensham *et al.* 1997). The first-stage model was applied to individual trials within locations resulting in the vector $\mathbf{y_1}$ of eBLUE values, the second-stage across trials within locations ($\mathbf{y_2}$), and the third-stage across locations and years. All analyses were implemented using `Asreml-R` version 4 (Butler *et al.* 2017) in the R programming environment (R Core Team 2021). Variance components of linear mixed models were estimated with REML followed by estimation/prediction of the fixed and random effects in Henderson's mixed models (Henderson 1949, 1950; Henderson *et al.* 1959). When possible, computation time was sped-up with parallel processing by applying the `doParallel` and `foreach` packages (Microsoft Corporation and Weston 2020a,b).

5.3.1 First-stage analyses

The first-stage analyses were previously performed by the collaborators (public soybean breeders) before the data were submitted to the USDA-ARS for aggregating and reporting as PDF files. Individual trials within locations were analyzed using a linear model in which genotypes and blocks were considered fixed effects, yielding eBLUE values of genotypic means ($\mathbf{y_1}$, entry-mean basis). The standard error (SE) of genotypic means was estimated as $\frac{\sigma_\epsilon}{\sqrt{r}}$, where σ_ϵ is the residual standard deviation calculated from the reported CV%, and r is the reported number of blocks for each trial. Because phenotypic values from each replicate are unavailable, we assumed all genotypes from a given trial had the same SE (i.e. equal replication, so that $\mathbf{y_1}$ was obtained by ordinary least squares and not by generalized least squares). The i^2 was then estimated as $i^2 = \frac{\sigma_G^2}{\sigma_G^2 + \frac{\sigma_\epsilon^2}{r}}$, where σ_G^2 is the genotypic variance and σ_ϵ^2 is the residual variance (Bernardo 2020, p. 173). The σ_G^2 was estimated from the variance of entry means (σ_F^2), where $\sigma_F^2 = \frac{\sigma_\epsilon^2}{r} + \sigma_G^2$.

5.3.2 Second-stage analyses

The second-stage model embraced data from multiple trials with common entries within locations in a given year, as follows:

$$\mathbf{y_1} = \boldsymbol{\mu} + \mathbf{X}_t\mathbf{t} + \mathbf{X}_g\mathbf{g} + \boldsymbol{\epsilon} \tag{1}$$

where $\mathbf{y_1}$ is a vector of eBLUE values from the first-stage analyses, $\boldsymbol{\mu}$ is the intercept, $\mathbf{X}_t$ is the incidence matrix of fixed effects of trials, $\boldsymbol{t}$ is a vector of fixed effects of trials, $\mathbf{X}_g$ is the incidence matrix of fixed effects of genotypes, $\boldsymbol{g}$ is a vector of fixed effects of genotypes, and $\boldsymbol{\epsilon}$ is a vector of residuals with $\boldsymbol{\epsilon} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma_1})$. The residual variance matrix $\boldsymbol{\Sigma_1}$ is a diagonal matrix with elements equal to $\frac{1}{\mathrm{SE}^2}$ (Smith *et al.* 2001a; Frensham *et al.* 1997), where SE was estimated from the first-stage analyses. Estimated values from Model 1 ($\mathbf{y_2}$ and SE) are available in the R package SoyURT.

5.3.3 Multi-location and multi-year analyses

The third-stage analysis used the "baseline" three-way model for multiple locations and years:

$$\mathbf{y_2} = \boldsymbol{\mu} + \mathbf{X}_l\mathrm{l} + \mathbf{Z}_g\mathrm{g} + \mathbf{Z}_{g.l}\mathrm{g.l} + \mathbf{Z}_y\mathrm{y} + \mathbf{Z}_{g.y}\mathrm{g.y} + \mathbf{Z}_{l.y}\mathrm{l.y} + \mathbf{Z}_{g.l.y}\mathrm{g.l.y} + \boldsymbol{\epsilon} \tag{2}$$

where $\mathbf{y_2}$ is a vector of eBLUE values from the second-stage analyses, $\boldsymbol{\mu}$ is the intercept, $\boldsymbol{l}$ is a vector of fixed effects of locations, $\boldsymbol{g}$ is a vector of random effects of genotypes with $\boldsymbol{g} \sim \mathrm{N}(\mathbf{0}, \sigma^2_G\mathbf{I})$, $\boldsymbol{g.l}$ is a vector of random effects of genotype by location interactions with $\boldsymbol{g.l} \sim \mathrm{N}(\mathbf{0}, \sigma^2_{GL}\mathbf{I})$, $\boldsymbol{y}$ is a vector of random effects of years with $\boldsymbol{y} \sim \mathrm{N}(\mathbf{0}, \sigma^2_Y\mathbf{I})$, $\boldsymbol{g.y}$ is a vector of random effects of genotype by year interaction with $\boldsymbol{g.y} \sim \mathrm{N}(\mathbf{0}, \sigma^2_{GY}\mathbf{I})$, $\boldsymbol{l.y}$ is a vector of random effects of location by year interaction with $\boldsymbol{l.y} \sim \mathrm{N}(\mathbf{0}, \sigma^2_{LY}\mathbf{I})$, $\boldsymbol{g.l.y}$ is a vector of random effects of genotype by location by year interaction with $\boldsymbol{g.l.y} \sim \mathrm{N}(\mathbf{0}, \sigma^2_{GLY}\mathbf{I})$, and $\boldsymbol{\epsilon}$ is a vector of residuals with $\boldsymbol{\epsilon} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma_2})$. $\mathbf{X}_l$, $\mathbf{Z}_g$, $\mathbf{Z}_{g.l}$, $\mathbf{Z}_y$, $\mathbf{Z}_{g.y}$, $\mathbf{Z}_{l.y}$, and $\mathbf{Z}_{g.l.y}$ are incidence matrices for their respective effects. The elements of the residual variance matrix $\boldsymbol{\Sigma_2}$ were obtained from Model 1.

5.3.4 Probability distributions of estimated variance components

A modified jackknife resampling approach was used to obtain empirical probability distributions for the variance components σ^2_G, σ^2_{GL}, σ^2_{GY}, and σ^2_{GLY}. Utilizing results from Aguate *et al.* (2019) and Hartung and Piepho (2021), the data were divided into four groups representing consecutive eras of soybean cultivar development: From 1989 to 1995, from 1996 to 2003, from 2004 to 2011, and from 2012 to 2019. For the first group (1989-1995), there were 181 environments (location-year combination); for 1996-2003, 194 environments; for 2004-2011, 100 environments; and for 2012-2019, 116 environments. The modified jackknife approach consisted of leaving-one-environment out (instead of one observation), and then obtaining REML estimates of variance components with a modified version of Model 2 that considered locations as a random effect with variance σ^2_L. Estimates of variance components were combined and evaluated for a best-fit probability distribution with the package "ForestFit" (Teimouri 2021). Given the lack of data from individual plots, trial-based estimates of σ^2_ϵ from the first-stage of analyses were used (i.e., no resampling). Distributional parameters were estimated via the expectation maximization (EM) algorithm (Dempster *et al.* 1977) using the log-likelihood functions of the Gamma, Log-Logistic, Log-Normal, Burr, and F univariate and multivariate distributions. In addition to a visual comparison of the modeled distributions relative to the empirical distributions, Akaike (AIC) and Bayesian (BIC) information criteria (Akaike 1974; Schwarz 1978) as well as the Kolmogorov-Smirnov (KS), Cramer-von Mises (CM), and Anderson-Darling (AD) goodness-of-fit statistics (Stephens 1986) were considered to select the best-fit distribution for each variance component. A classical penalized criteria based on the loglikelihood (AIC, BIC) provided protection from overfitting.

5.3.5 Identification of mega-environments

Herein, "ME" and "cluster" are used interchangeably. We clustered 63 locations using six criteria: (i) phenotypes, i.e. seed yield (PHE), (ii) eight soil variables (SV) plus elevation (SoilE); (iii) latitude, where locations were split into two groups (Lat2), (iv) latitude, where locations were split into three groups (Lat3); (v) 19 meteorological variables (MV) with means across years (WA); and

(vi) MV with means nested within years (WW).

Except for Lat2 and Lat3, the optimal number of clusters was then defined based on the Silhouette and Elbow methods using the package `factoextra` (Kassambara and Mundt 2020), followed by a K-means clustering with the R base function *kmeans()* allowing for a maximum of 1,000 iterations and 100 multiple initial configurations of the K groups.

5.3.5.1 Clustering with PHE data

Several variations of Model 2 were evaluated by modeling the variance-covariance structures (VCOV) for the genotype by location (Σ_{gl}) and genotype by year (Σ_{gy}) interaction terms. The simplest model (M2-1, baseline) assumed independent effects for Σ_{gl} and Σ_{gy}. The next set of models allowed heterogeneous variances for Σ_{gl} (M2-2), Σ_{gy} (M2-3), or both (M2-4). Specific pairwise covariances for both Σ_{gl} and Σ_{gy} were assessed with models M2-5, M2-6, ..., M2-20 (Table 1). The VCOV structures included identity (**I**), diagonal (**D**), and factor-analytic (**FA**(k)) of order k (Piepho 1997; Smith *et al.* 2001b, 2015). Models M2-5, M2-6, ..., M2-20 are variations of the FA for different $k's$.

In addition to the AIC selection criteria to select the best-fit model, the overall percentage of genetic variance accounted by each k factor in FA models was also considered. For each model term (Σ_{gl}, Σ_{gy}), it is defined as 100[tr($\mathbf{\Lambda\Lambda}'$)/tr($\mathbf{\Lambda\Lambda}'$ + $\mathbf{\Psi}$)], where "tr" is the trace of the matrix, $\mathbf{\Lambda}$ ($63 \times k$) is the matrix of loadings, and $\mathbf{\Psi}$ (63×63) is a diagonal matrix of specific variances associated with each location. With the best-fit FA model, locations were clustered based on the estimated Σ_{gl} loadings (Bustos-Korts 2017; Burgueño *et al.* 2008) after Varimax rotation. Genetic correlations between locations ($\mathbf{C}$) were further estimated by $\mathbf{C} = \mathbf{DGD}$, where $\mathbf{G} = (\mathbf{\Lambda\Lambda}' + \mathbf{\Psi})$ is the estimator of genetic variances, and $\mathbf{D}$ is a diagonal matrix composed by the inverse of the square root of the diagonal values of $\mathbf{G}$ (Smith *et al.* 2015).

5.3.5.2 Clustering with SV data

First the SV (including elevation) were centered and scaled to unit variance. Subsequently, a principal component analysis (PCA) by non-linear iterative partial least squares (Wold 1966) was

performed to reduce collinearity with the `pcaMethods` package (Stacklies *et al.* 2007). The number of principal components (PC) was selected with a 90% threshold of cumulative variance explained, followed by a Varimax rotation.

5.3.5.3 Clustering with MV data

Meteorological data was retrieved for each environment for the growing season. Instead of using averaged values for each MV across the season, we used the Critical Environmental Regressor through Informed Search (CERIS) procedure proposed by Li *et al.* (2018) to identify relevant periods during growth and development. The method consists of screening meteorological data across environments to identify a period (window) of days after planting with the highest Pearson correlation between the population average (i.e. environmental average) and the MV. The idea is to identify periods of meteorological data most likely to affect growth and development associated with the phenotypic data (yield, plant height, etc.). Instead of using environmental averages (the average value of $\mathbf{y_2}$ for each environment) to identify the window, we further modified CERIS to account for genotype by environment deviations within years, as follows:

$$\mathbf{y_e} = \boldsymbol{\mu} + \mathbf{X}_e\mathbf{e} + \mathbf{X}_g\mathbf{g} + \boldsymbol{\epsilon} \tag{3}$$

where $\mathbf{y_e}$ is a vector of eBLUE values in a given year (i.e. $\mathbf{y_2}$ in a given year), $\boldsymbol{\mu}$ is the intercept, $\boldsymbol{e}$ is a vector of fixed effects of environments, $\boldsymbol{g}$ is a vector of fixed effects of genotypes, and $\boldsymbol{\epsilon}$ is a vector of residuals with $\boldsymbol{\epsilon} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_\mathbf{e})$. $\mathbf{X}_e$ and $\mathbf{X}_g$ are incidence matrices for their respective effects, and $\boldsymbol{\Sigma}_\mathbf{e}$ is the residual variance matrix obtained from Model 1.

Each environment within years (i.e. a location) was then represented as the average of the squared residuals, which was used by CERIS. Model 3 was exclusively designed to replace environmental means with a metric more useful to identify ME with repeatable GEI. The residuals represent the GL deviations nested within years, which are confounded with GLY effects. An alternative model that contains the genotype by environment interaction is presented in Appendix D.

The CERIS was computed in two ways: across (WA) and within (WW) years, and the best window (i.e. highest correlation) for each MV was selected. In both cases, the final dataset used for clustering consisted of a matrix of 63 rows (locations) by 19 columns (MV). For WA, the algorithm (i) selected the best window across the 591 observed environments, (ii) the average value of each MV was computed with the selected window for each year a given location of observed, and (iii) locations were represented with the averaged value across years. For WW, the only modification is at step (i): each year a given location was observed, the algorithm selected its best window. A minimum window size of seven days was considered in all cases. After identifying the most relevant dataset of MV, the data were centered and scaled to unit variance. Subsequently, clustering was conducted as described for the SV.

5.3.5.4 Effectiveness of clustering

We used the ratio of correlated responses from selection across all environments relative to direct responses to selection within ME (CR/DR) (Atlin *et al.* 2000a; Bustos-Korts 2017) as a metric to assess the relative effectiveness of clustering locations into ME. The zone-based model is defined as follows:

$$\begin{aligned}\mathbf{y_2} =& \mathbf{X}_r\mathbf{r} + \mathbf{Z}_{l(r)}\mathbf{l_{(r)}} + \mathbf{Z}_g\mathbf{g} + \mathbf{Z}_{g.l_r}\mathbf{g.l_{(r)}} + \mathbf{Z}_y\mathbf{y} + \mathbf{Z}_{y.r}\mathbf{y.r} \\ &+ \mathbf{Z}_{g.r}\mathbf{g.r} + \mathbf{Z}_{l_r.y}\mathbf{l_{(r)}.y} + \mathbf{Z}_{g.y}\mathbf{g.y} + \mathbf{Z}_{g.y.r}\mathbf{g.y.r} + \mathbf{Z}_{g.y.l_r}\mathbf{g.y.l_{(r)}} + \boldsymbol{\epsilon}\end{aligned} \tag{4}$$

where $\boldsymbol{r}$ is a vector of fixed effects of clusters, and $\boldsymbol{l_{(r)}}$, $\boldsymbol{g.l_{(r)}}$, $\boldsymbol{y.r}$, $\boldsymbol{g.r}$, $\boldsymbol{l_r.y}$, $\boldsymbol{g.y.r}$, and $\boldsymbol{g.y.l_r}$, are random vectors with specific variances of locations within clusters, genotype by location within clusters interaction, year by clusters interaction, genotype by cluster interaction, locations nested in clusters by year interaction, and genotype by year by location within clusters, respectively. $\mathbf{X}_r$, and $\mathbf{Z}_{l_r}$ up to $\mathbf{Z}_{g.y.l_r}$, are incidence matrices for their respective effects and dimensions. The remaining model terms were previously defined.

Estimates of variance components from Model 4 were used to obtain CR/DR:

$$CR/DR = \rho_g \sqrt{\frac{i_{\mathrm{L}}^2}{i_{\mathrm{SR}}^2}} \tag{5}$$

where ρ_g is the correlation between estimated genotypic effects in the non-clustered and clustered sets of locations and i_{L}^2 and i_{SR}^2 are the estimated reliabilities of genotype means in the non-clustered and clustered sets of locations, respectively. If $CR/DR < 1$, the response to selection will be more effective if selections are made within clusters (Atlin *et al.* 2000a; Bustos-Korts 2017). Note that it is possible for $CR/DR > 1$, indicating that selection will be more effective if the selection is based on eBLUE values obtained from non-clustered locations (Atlin *et al.* 2000a). The terms in equation 5 are defined as follows:

$$\rho_g = \frac{\sigma_G^2}{\sqrt{\sigma_G^2(\sigma_G^2 + \sigma_{GR}^2)}} \tag{6}$$

$$i_{\mathrm{L}}^2 = \frac{\sigma_G^2}{\sigma_G^2 + \frac{\sigma_{GR}^2}{nr} + \frac{\sigma_{GL(R)}^2}{nl \times nr} + \frac{\sigma_{GY}^2}{ny} + \frac{\sigma_{GRY}^2}{ny \times nr} + \frac{\sigma_{GL(R)Y}^2}{nl \times nr \times ny}} \tag{7}$$

$$i_{\mathrm{SR}}^2 = \frac{\sigma_G^2 + \sigma_{GR}^2}{\sigma_G^2 + \sigma_{GR}^2 + \frac{\sigma_{GL(R)}^2}{nl} + \frac{\sigma_{GY}^2}{ny} + \frac{\sigma_{GRY}^2}{ny} + \frac{\sigma_{GL(R)Y}^2}{nl \times ny}} \tag{8}$$

where σ_G^2, σ_{GR}^2, $\sigma_{GL(R)}^2$, σ_{GY}^2, σ_{GRY}^2, and $\sigma_{GL(R)Y}^2$ are the genotypic, genotype by cluster, genotype by location nested in cluster, genotype by year, genotype by clusters by year, and genotype by location nested in cluster by year variance components, respectively. nr and ny are the harmonic means for the number of clusters and years in which genotypes were observed, respectively. nl is the median number of locations, obtained from harmonic means within clusters. The residual terms in both reliability estimators were omitted due to the lack of replicated data within environments.

The Jaccard similarity coefficient was used to compare the coincidence of locations within clusters across clustering types. To explore the dynamic for each clustering type, we reported estimates of σ_{GR}^2 with the jackknife approach presented in section 5.3.4. Lastly, to evaluate if

locations were allocated to clusters by chance, we randomly assigned locations into two, three, and four clusters, and assessed their CR/DR. This process was repeated 100 times.

6 Results

6.1 Single-trial and location analysis

The number of evaluated genotypes by year including checks ranged from 126 to 233, and the number of locations per year ranged from 10 to 32 (Table C3). Estimates of i^2 from single-trial analyses had a median value of 0.55, and the CV% a median value of 7.60% (Figure 1-A). Across years, there has been a positive trend for seed yield with an average increase of 0.49 bu/ac per year (Figure 1-B), although the relative contributions of genetic and non-genetic factors to this trend require further analyses and is the subject of future research.

6.2 Variance components

Most of the estimated phenotypic variance for seed yield has been due to location (σ^2_L) and location by year interaction (σ^2_{LY}) effects (Table 2). Among the estimated GEI variance components (σ^2_{GL} + σ^2_{GY} + σ^2_{GLY}), the static contribution (σ^2_{GL}) represented 26.30% (Table 2, Model M2-1).

Estimated variance components reveal distinct multi-modal distributions over time (Figure 2). For example, the estimated genotypic variance (σ^2_G) more than doubled from $\sim$ 3.54 (bu/ac)^2 for the period 1989-1995 to $\sim$ 7.56 (bu/ac)^2 for the period 1996-2003. While the smallest magnitudes of estimated GEI variance components were usually associated with the period from 1989 to 1995, subsequent changes across years were unique to each estimated variance component. A similar pattern was observed for empirical estimates of location, year, and location by year interaction variances (Figure C1).

The best-fit models for the trend of variance components consisted of a mixture of five Log-Logistic distributions for the empirical distribution of σ^2_G, three Log-Logistic distributions for the empirical distribution of σ^2_{GL}, six Log-Logistic distributions for the empirical distribution

of σ^2_{GY}, and five Gamma distributions for the empirical distribution of σ^2_{GLY} (Table C4, Figure 3). For the empirical distribution of estimated residual variances (σ^2_{ϵ}), obtained directly rather than through jackknife resampling, the best-fit model was a univariate Log-Logistic distribution (Figure 4). Maximum Likelihood estimates for the distributional model parameters are reported for the selected models (Table 3), as well as plots with empirical and fitted cumulative distribution functions (Figures C2 and C3).

6.3 Identification of mega-environments

The six clustering strategies revealed that the observed 63 locations could be divided into at least two or at most three ME (Table 4). The PHE clustering presented the lowest CR/DR value, with an optimal number of clusters of three, followed by two clusters for SoilE, WW, and WA. For the PHE clustering type, the best-fit model (M2-18) accounted for heterogeneous variances and pairwise correlations for Σ_{gl} and Σ_{gy} with FA(5) and FA(2) matrices, explaining 90.2% and 87.2% of the genetic variances, respectively (Table 1 and Figure C5). For all clustering types, both Silhouette and Elbow criteria indicated similar results (Figure C4). The number of locations within clusters ranged from 7 to 54 (Table 4 and Figure 5), with many clusters having common locations. The greatest Jaccard similarity coefficient was between cluster 1 from PHE and 2 from Lat2 (Table C5).

Before K-means clustering, a non-linear PCA was performed with centered and scaled environmental variables (elevation, SV, and MV). The non-linear relationship for most variables was evident from scatterplots (Figures C6, C7, and C8). For SoilE, WA, and WW, the selected number of PC were five (93.06% of variance explained), four (93.42%), and six (91.12%), respectively (results not shown). Furthermore, for MV, just a small proportion of the data was used due to the CERIS biological filter. The selected windows (i.e., with the highest Pearson correlation) were smaller than 20 days, and most of them started at the beginning of the cropping season (Tables C1 and C2, Figure C9).

The effectiveness of clustering was assessed using CR/DR and presented ratios smaller than

one in five out of the six cases. These results suggested an improvement in the response to selection when selecting directly within clusters (i.e., regions, ME) versus selecting across all locations for PHE, SoilE, and Lat2, while a moderate response to selection was suggested for Lat3 and WA. For WW, no improvements were predicted. The PHE presented the lowest ρ_g value with an increase in the reliability (from i^2_L to i^2_{SR}), followed by Lat2 and SoilE. For Lat3 and WA, although effective, the reliabilities remained constant. When locations were randomly assigned to two, three, or four clusters, the mean CR/DR values were always bigger than one, with large decreases in the reliability (Table 4).

REML estimates of variances decreased for σ^2_G when clusters were included in the complete dataset, except for WW. The observed ratios σ^2_{GR}/σ^2_G and $\sigma^2_{GR}/\sigma^2_{GL}$ were greather than 0.50 for PHE, SoilE, and Lat2. In terms of the partitioning of σ^2_{GL}, the $\sigma^2_{GL(R)}$ portion was substantially reduced for PHE, Lat2, and Lat3. The σ^2_{GR} was better captured by PHE, SoilE, and Lat2, being ineffective for WW (estimate bounded at zero due to REML properties). On the other hand, for σ^2_{GLY}, just a small portion of the variation was captured by σ^2_{GRY} (Table 2). Both PHE and Lat2 clustering types were able to greatly capture σ^2_{GR} according to analysis for each group of years (Figure C10). Lastly, large reductions in the variation of years (σ^2_Y) were observed for SoilE and Lat2.

7 Discussion

Analyses of balanced data sets produced by MET using the method of moments and REML estimators provide unbiased estimates of variance components. However, data generated by annual PYT and URT are unbalanced and sparse because every year most field plots (experimental units) are reserved for new experimental genotypes, and most previously evaluated experimental genotypes are culled annually. For example, reports of soybean seed yield in MG II and III evaluated in PYT and URT conducted from 1989 and 2019 included less than one percent of all possible combinations of genotypes, locations, and years. Consequently, estimators of variance compo-

nents might be biased (Rothschild *et al.* 1979). A question to consider is whether the biased estimators produce a large bias in the estimates.

An additional challenge in using soybean seed yield data from PYT and URT is that the data consist of eBLUE values from individual trials within locations for each genotype. These values were transcribed from reports of individual trials formatted as PDF files. To the best of our knowledge, all genotypes were evaluated in replicated field trials organized as RCBD, and were analyzed with a linear model consisting of fixed block and genotypic main effects. To obtain the weights for the stage-wise modeling, the standard errors of genotypic means were assumed to be estimated using equal numbers of replicates per genotype within a field trial. This is equivalent to assuming no missing plots within any field trials, which is highly unlikely. Further, given the lack of replicated data, the estimates of σ^2_{GLY}, σ^2_{GL} and σ^2_{GY} are confounded and biased by σ^2_{ϵ}, i.e. the plot to plot variability.

A consequence of using eBLUE values instead of individual plot (replicated) data is that estimates of variance components needed to be obtained in multiple stages. If the reports of field trials had provided individual plot data, it would have been possible to produce a variance–covariance matrix associated with adjusted means from each trial. Indeed, if individual plot data are provided, the field plot designs need not be restricted to RCBD wherein the covariances among plots may be substantial and assumptions of independence among plots are inappropriate (Möhring and Piepho 2009). Under such field conditions, the use of spatial models and lattice designs where replicates are considered fixed and blocks as random effects (Möhring *et al.* 2015) can be utilized, and the mixed model framework can provide appropriate weights for analyses of data combined across trials, locations and years. Toward this goal, public soybean breeders are working with curators of "SoyBase" (Drs. Rex Nelson and David Grant, personal communication, 2020) to include results from individual plots in future reports of PYT and URT.

If data were obtained from only URT, they would provide very little information for estimating σ^2_{GLY} and σ^2_{GY} because most experimental genotypes are not grown for more than one year of URT. Although check varieties were replicated across multiple years, the checks are a small

sample of genotypes representing mostly commercial cultivars. Thus, inferences about genotypic variability would be limited to the specific sets of checks used in these trials. In addition, by not considering PYT data, the missing data pattern would not be missing at random (MAR). MAR refers to when the missing data pattern in MET is caused by selection (Rothschild *et al.* 1979). If data are not MAR, they can generate biased estimates of variance components (Little and Rubin 2020, Chapter 6). Also, URT can have experimental genotypes not included in a previous year of PYT, a condition known as missing completely at random (MCAR) (Little and Rubin 2020; Rubin 1976). Fortunately, recent work from Hartung and Piepho (2021) demonstrated that both MAR and MCAR conditions for field trials conducted in sequential years result in a minor bias for likelihood-based estimators of variance components, a result previously noted for MAR by Piepho and Mohring (2006). Moreover, data from both trials should be included for variance estimation (Laidig *et al.* 2008; Piepho and Mohring 2006), especially to estimate the genotypic variance (Hartung and Piepho 2021).

Despite the caveats mentioned above, the eBLUE values for genotypes from both PYT and URT provided estimates of variance components with similar relative magnitudes as other studies. For example, we determined that environmental, non-genetic sources were the predominant source of variance for seed yield. Similar results were found for wheat in the USA (George and Lundy 2019), winter wheat field trials in Germany between 1983 to 2014 (Laidig *et al.* 2017a), and winter rye (Laidig *et al.* 2017b). Our analyses also revealed that the interactions of genotype with locations (σ^2_{GL}) were larger than genotype by year interactions (σ^2_{GY}). In addition, the three-way interaction among genotype, location, and years (σ^2_{GLY}), was greater than the two-way interactions involving genotypes. The same pattern was observed when years were combined into four groups of seven to eight years periods. Friesen *et al.* (2016) found similar results for winter wheat evaluated in Canada from 2000 to 2009, where the reported σ^2_{GLY} represented 4.1% of the total variation, and both σ^2_{GL} and σ^2_{GY} together represented less than 2%. Similar trends were also observed for yield in wheat (George and Lundy 2019; Laidig *et al.* 2017a; Arief *et al.* 2015), rye (Laidig *et al.* 2017b), barley, maize, and sunflower (Laidig *et al.* 2008). While seed yields for

multiple crops indicate $\sigma^2_{GL} > \sigma^2_{GY}$, it is not consistent for all traits. For example, Laidig *et al.* (2017b) evaluated the variation in crude protein content, amylogram viscosity, and temperature in winter rye varieties, and reported that $\sigma^2_{GY} > \sigma^2_{GL}$, and that the year-to-year (σ^2_Y) variation was more important than the variation from location to location (σ^2_L). Authors attributed this to the rye seed's susceptibility towards wetness, low temperature, and radiation during harvest time.

To interpret variance components involving genotypes, we recognize that PYT and URT include genotypes from multiple breeding programs. Each breeding program operates independently with distinct breeding objectives and strategies for markets. For example, a couple of breeding programs have objectives that include high seed yields and greater seed protein for food markets. Seed protein is negatively correlated with seed yield. Thus, the estimates of variance components involving genotypes from these trials are likely greater than for any individual breeding program.

Estimates of variance components were also used to fit parametric probability distributions, in which the data was divided into subsets of seven to eight years. Aguate *et al.* (2019) simulated different MET with variable numbers of genotypes, locations, and years to mimic wheat trials. They found that adding years is more beneficial than adding genotypes or locations for obtaining unbiased estimates of genotypic-related variance components. Furthermore, even in highly imbalanced datasets, estimates from at least eight years of trials produced less than 5% bias in the estimates, compared to biases of $\sim$18% for σ^2_G, >40% for σ^2_{GL}, and >15% for σ^2_{GY}, when only two years of MET were considered. Simulation results from Hartung and Piepho (2021) further demonstrated that non-significant bias could be obtained for estimates of σ^2_G, σ^2_{GL}, and σ^2_{GY}, with decreasing dropout rate and increasing the number of years of testing. Both references showed that variance estimates from REML are biased if at least one variance component estimate is bounded at zero, which rarely occurs when using seven or eight years of data. The reader should keep in mind that the goal of minimizing bias in estimators is distinct from the magnitude and proportion of estimates of variance components. The former is a concern for algorithmic es-

timators while the latter is of concern for breeding decisions about whether to use more locations or more years in their crop species and for traits under selection.

In terms of the fitted probability distributions, especially due to the well-known properties of the analysis of variance (ANOVA) among the breeding community, there is a misconception regarding the distribution of estimates of variance components. If the underlying population is normally distributed, the mean squares are distributed as a chi-square (χ^2), whereas normality and independence are requirements to compute valid F tests from ANOVA (Rencher and Schaalje 2007, Chapter 5). The χ^2 distribution is further used for inferences about variance uncertainty, but as mentioned, it does assume that the random variable is normally distributed. The computed empirical distributions from jackknife can be of any form (distribution), likewise, other approaches such as hierarchical Bayesian models can be employed to obtain posterior distributions of variance estimates. It should be emphasized, however, that the obtained empirical density functions might not necessarily be the data-generating functions. When testing the fit of distributions, the systematic dropping of the data for curation and cleaning (6.8% of the data points were removed) was not considered.

The cluster analyses showed that the observed locations could be effectively divided into three clusters using FA models. The obtained clusters can be considered disjoint subsets of environments with minimal genotypic crossover interaction (Burgueño *et al.* 2008; Cooper and DeLacy 1994). Burgueño *et al.* (2008) analyzed a maize MET dataset from CIMMYT and identified five clusters of environments by fitting a FA(2) model. The authors computed the Euclidean distance between pairs of environments from the estimated loading matrix ($\mathbf{\Lambda}$, rotated with singular value decomposition), and environments were clustered based on complete linkage clustering strategy. With a FA(1) model, Bustos-Korts (2017) analyzed data embracing TPE in Denmark, Germany, The Netherlands, and the United Kingdom. The results suggested improvements in response to selection mostly for Denmark, where the CR/DR ratio was 0.93. More recently, Smith *et al.* (2021) proposed a new way to define groups of environments that exhibit minimal crossover interaction based on FA models. The idea is to take advantage of the traditional interpretation of

factor and principal component analysis and classify environments into clusters based on the sign (positive or negative) of the estimated and rotated factor loadings.

The identification of homogeneous groups of locations was also accomplished by considering soil plus elevation (SoilE) and meteorological variables when CERIS was applied across environments (WA). The rationale is that a portion of the GEI results from static, repeatable variation (Crespo-Herrera *et al.* 2021; Yan 2016). It is well-known that temperature is a key driving force in the rate of seasonal plant growth (Setiyono *et al.* 2007), which is why GDD is commonly a base unit in crop models (Holzworth *et al.* 2014). Photoperiod plays a significant role in soybean plant development, notably the change from vegetative to reproductive growth. Floral induction is essentially day length and temperature-independent (i.e. conversion of shoot apical and nodal meristems from a vegetative to floral mode). In soybeans, this induction occurs as soon as the first unifoliolate leaflets emerge and expand, becoming capable of measuring the night length (from dusk to dawn). Once floral induction occurs at a given apical or axillary node, the few-celled vegetative apical zone is transformed from a vegetative development pathway into a floral inflorescence development pathway. The development pathway is back under thermal control (Setiyono *et al.* 2007). Soybean is a quantitative long-night length sensing (not a short-day length sensing), and hence highly influenced by photoperiod and therefore by the latitude of the growing region/trial (Jackson 2009). This is a major reason that different soybean maturity groups are grown at different latitudes (Mourtzinis and Conley 2017). The estimated clusters (except for the *ad-hoc* Lat2 and Lat3) follow a certain pattern in terms of latitude, which was also confirmed by the Jaccard similarity.

For this work, we focused on identifying ME through reliable estimates of variance components. However, other environmental subsets would have been formed using different clustering strategies (Burgueño *et al.* 2008). Given that our approach is unsupervised learning (i.e. we do not know the truth about ME), the objective is always to discover an interpretable grouping of members. We addressed interpretation using the effectiveness of clustering (Atlin *et al.* 2000a). Other strategies for clustering can also be tested, for example, empirical knowledge of

the TPE. But regardless of the definition/identification of ME, breeders can take advantage of the best linear unbiased prediction (BLUP) that borrows information (strength) between ME from the genotype by ME interaction. This type of modeling can be highly beneficial for ME that rely on a small number of locations (Buntaran *et al.* 2021; Piepho *et al.* 2016; Piepho and Möhring 2005).

A natural continuation for this work would be to (i) evaluate the effectiveness of a combined cluster from soil, elevation, and meteorological variables filtered by CERIS; (ii) evaluate if BLUP based-models will improve the selection response upon regionalization in the Uniform Soybean Cooperative Tests; (iii) select an appropriate model that might account for heterogeneous covariances among ME as well genetic relationships, because we confined attention to the compound symmetry model to facilitate comparison of the different clustering types; and (iv) leverage how far back in the historical data should we go to take maximum advantage of the data in current models. It is also worth investigating if modeling maturity groups (especially when more data is considered) would enhance the ability to find meaningful ME using phenotypic models.

8 Conclusion

We dissected the sources of soybean seed yield variation using reports from Soybean Cooperative Tests for maturity groups II and III. We determined that the sampled locations can be split into mega-environments according to phenotypic, geographic, and meteorological information. Furthermore, it was possible to monitor trends in variance components involving genotypes in terms of parametric probability distributions. Historical field trials also evaluate traits like seed quality and size, iron deficiency chlorosis, green stem, seed oil, and protein content. Therefore, the approach presented herein can be applied to multiple economically important quantitative traits. It is important to emphasize that our inference space is restricted to germplasm adapted to MG II and III and evaluated as nearly homozygous genotypes in public MET. Also, the evaluated environments represent a random sample of the target population of environments. Thus,

inferences about the identified mega-environments (ME) are likewise restricted to the field trial network explored by public soybean breeding programs. Finally, in addition to the practical and theoretical results applied to soybean genetic improvement, the analyses performed in this study may be applied to quantitative traits evaluated in any crop using MET.

9 Acknowledgments

Our sincere thanks to Dr. Aaron Lorenz's group for providing a large portion of the phenotypic data (in particular to Ben Campbell, Liana Nice, and Cleiton Wartha); Dr. Rex Nelson for providing individual plot data from trials conducted in years 2018 and 2019; Dr. Jim Specht for valuable insights on environmental factors that influence soybean growth and development; Dr. Alencar Xavier for careful reading and suggestions on the early draft of the manuscript; and the Iowa State University Research IT team for providing efficient computational power. Most importantly the authors appreciate the long-term commitments of all collaborators (breeders, staff, students, farmers, etc.) that have worked on conducting the cooperative Soybean trials every year since 1983. Funding for this research was provided by the Department of Agronomy - Iowa State University, the North Central Soybean Research Program, an NSF grant (1830478), Baker Center for Plant Breeding, and USDA-ARS CRIS Project IOW04714.

10 Declaration of competing interest

The authors state there is no conflict of interest.

11 Author contributions

MDK and WDB conceived the research; MDK performed the statistical analyses and wrote the first drafts of the manuscript; KOGD provided insights into the methodology and helped in the interpretation of the results; AKS provided interpretation of the results and guidance in scientific

writing; and WDB and AKS were responsible for acquiring funding to support the research. All authors critically revised drafts of the manuscript and approved the final version.

12 Data and code availability

The developed R codes to reproduce most analyses are publicly available on GitHub (https://github.com/mdkrause/VarComp-ME). The data sets (phenotypic, weather, and soil data) are available in the developed R package SoyURT (https://github.com/mdkrause/SoyURT). Raw data can be further downloaded from Soybase (https://soybase.org/ncsrp/queryportal/), and directly from the PDF files from the USDA-ARS webpage (https://www.ars.usda.gov/midwest-area/west-lafayette-in/crop-production-and-pest-control-research/docs/uniform-soybean-tests-northern-region/). All cited web pages here and throughout the manuscript were accessed on 05/04/2023.

13 Tables and Figures

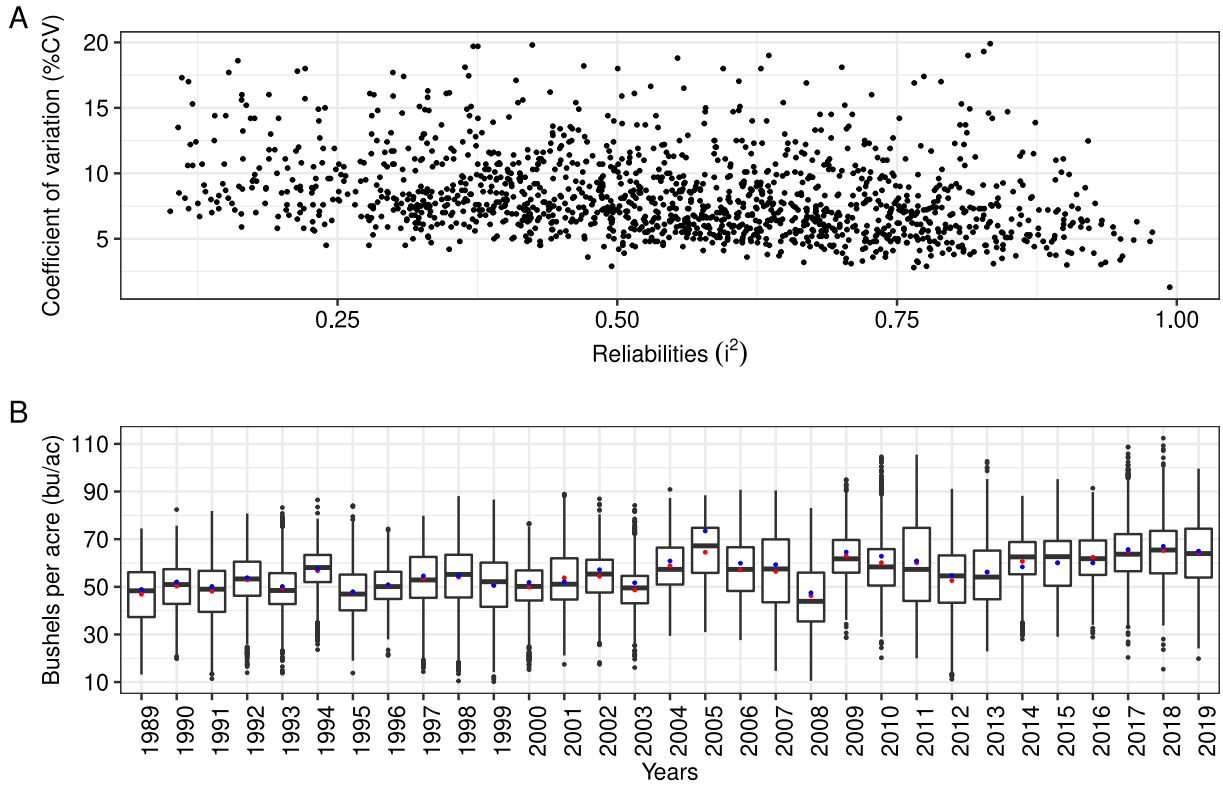

Figure 1: Estimates of reliabilities (i^2) and coefficient of variation (%CV) for 1423 soybean field trials conducted from 1989 to 2019 (A), and boxplots of empirical best linear unbiased estimates (eBLUEs) of seed yield plotted by year (B) from 1989 to 2019. Red dots in B depict the average yield of experimental cultivars excluding checks, whereas blue dots depict the average yield of the check varieties.

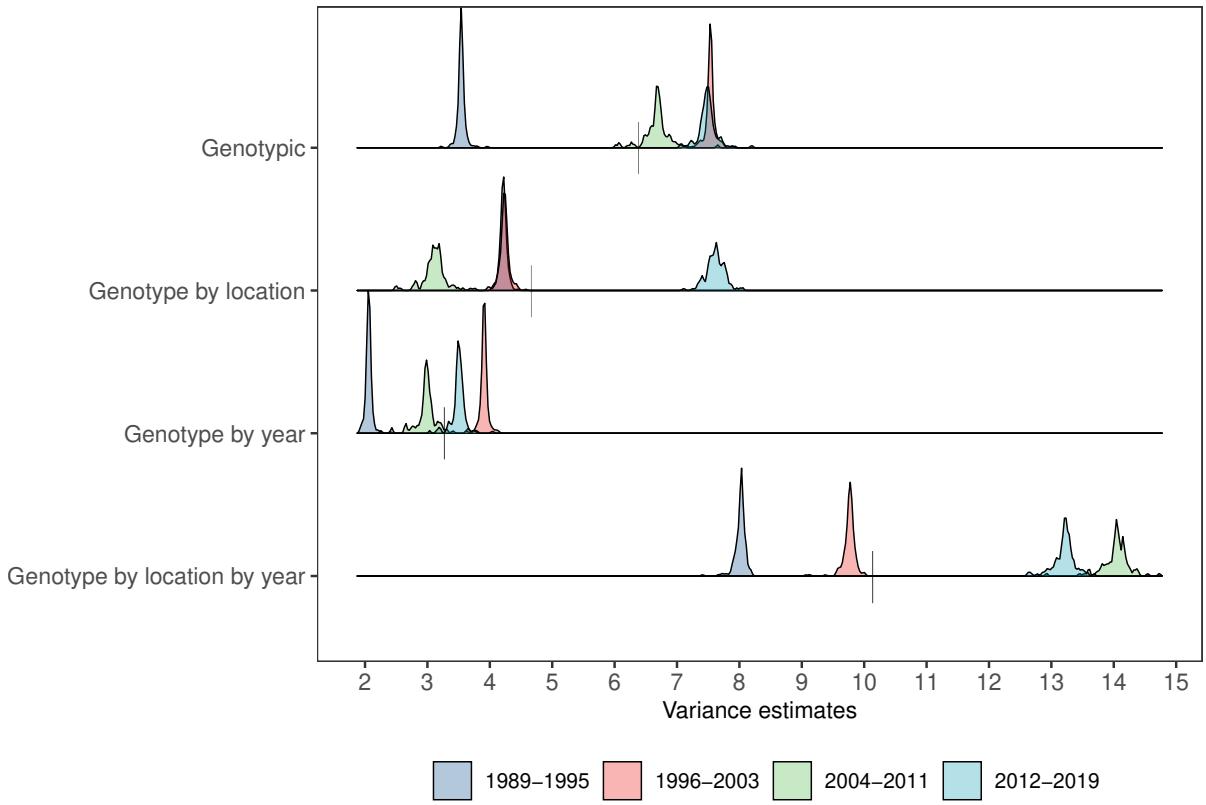

Figure 2: Empirical distributions of estimated variance components consisting of genotypic, genotype by location, genotype by year, and genotype by location by year variances for groups of years 1989-1995, 1996-2003, 2004-2011, and 2012-2019. Empirical estimates were obtained using a jackknife leave-one-location-out method. Vertical bars on the x-axis represent point estimates across all years.

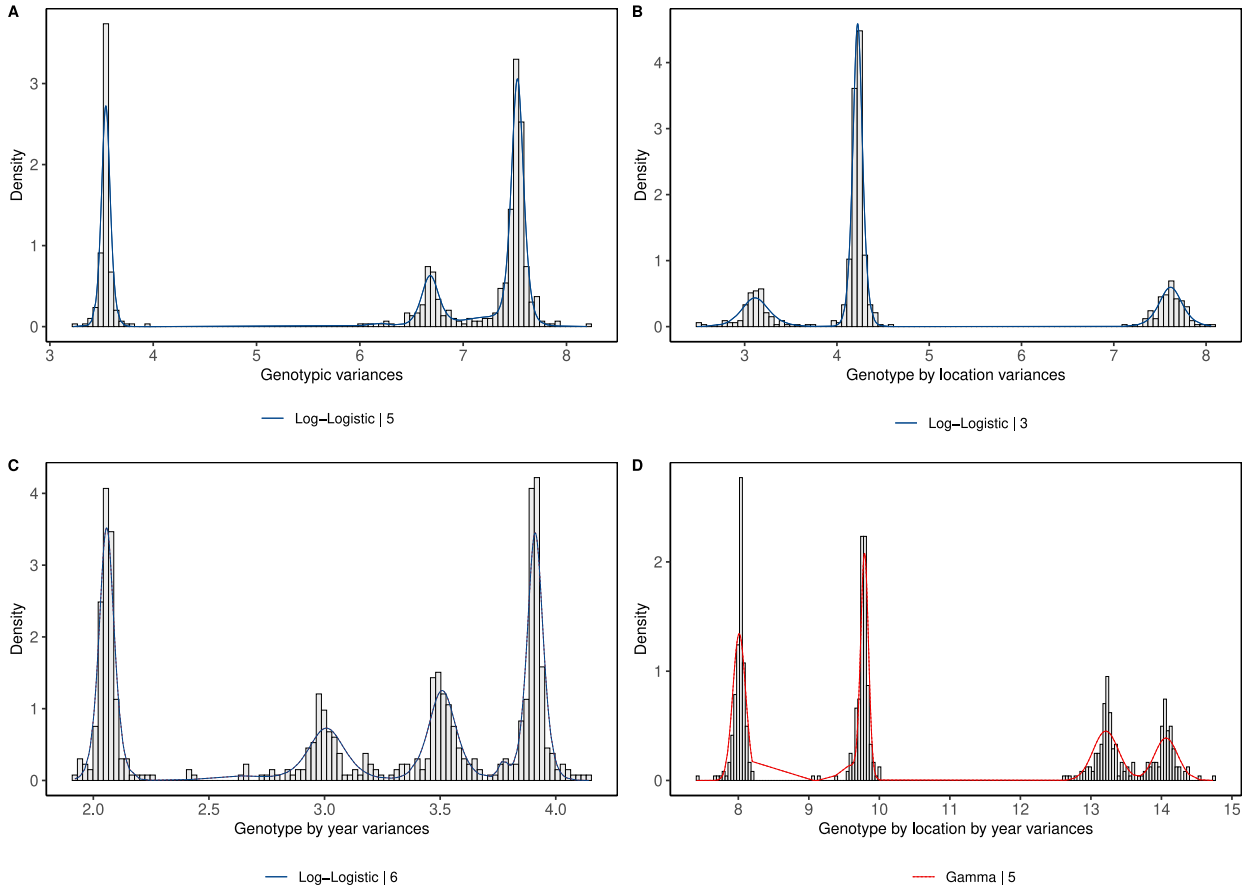

Figure 3: Empirical distributions and probability density function (PDF) of variance components for genotypic (A), genotype by location (B), genotype by year (C), and genotype by location by year (D). Empirical estimates were obtained using a jackknife leave-one-environment-out method. The best-fit PDF models are presented with different colors and include the distribution's name and its mixture number.

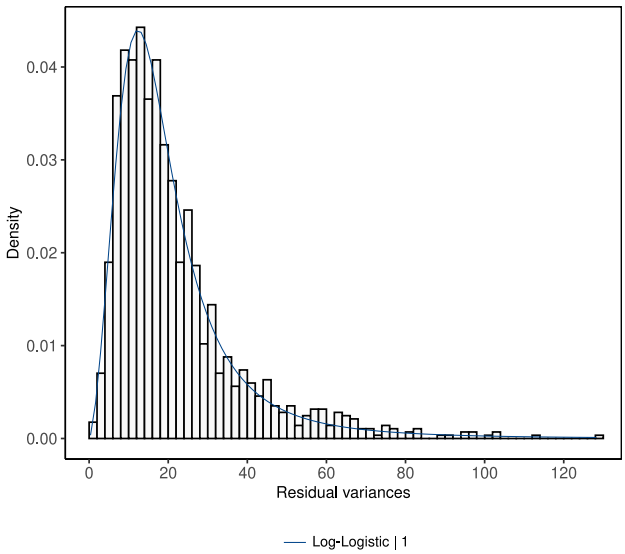

Figure 4: Empirical distribution and probability density function (PDF) for the unimodal Log-Logistic model of residual estimates from individual trials.

Table 1: Models and criteria used to evaluate the models for purposes of clustering locations into groups representing the most likely target population of environments. The best-fit model (M2-18) is highlighted in bold.

Model	Covariance structure[a]			Evaluation criteria		
	Σ_{gl}	Σ_{gy}	Σ_{ϵ}	AIC	% Var(GL)	% Var(GY)
M2-1	$\mathbf{I} \otimes \mathbf{I}$	$\mathbf{I} \otimes \mathbf{I}$	$\mathbf{\Sigma_3}$	175294	-	-
M2-2	$\mathbf{D} \otimes \mathbf{I}$	$\mathbf{I} \otimes \mathbf{I}$	$\mathbf{\Sigma_3}$	174256	-	-
M2-3	$\mathbf{I} \otimes \mathbf{I}$	$\mathbf{D} \otimes \mathbf{I}$	$\mathbf{\Sigma_3}$	174964	-	-
M2-4	$\mathbf{D} \otimes \mathbf{I}$	$\mathbf{D} \otimes \mathbf{I}$	$\mathbf{\Sigma_3}$	173951	-	-
M2-5	$\mathbf{FA}_1 \otimes \mathbf{I}$	$\mathbf{D} \otimes \mathbf{I}$	$\mathbf{\Sigma_3}$	172845	48.5	-
M2-6	$\mathbf{FA}_2 \otimes \mathbf{I}$	$\mathbf{D} \otimes \mathbf{I}$	$\mathbf{\Sigma_3}$	172560	67.3	-
M2-7	$\mathbf{FA}_3 \otimes \mathbf{I}$	$\mathbf{D} \otimes \mathbf{I}$	$\mathbf{\Sigma_3}$	172359	78.2	-
M2-8	$\mathbf{FA}_4 \otimes \mathbf{I}$	$\mathbf{D} \otimes \mathbf{I}$	$\mathbf{\Sigma_3}$	172303	87.6	-
M2-9	$\mathbf{FA}_1 \otimes \mathbf{I}$	$\mathbf{FA}_1 \otimes \mathbf{I}$	$\mathbf{\Sigma_3}$	172754	48.0	63.4
M2-10	$\mathbf{FA}_2 \otimes \mathbf{I}$	$\mathbf{FA}_2 \otimes \mathbf{I}$	$\mathbf{\Sigma_3}$	172472	66.2	82.5
M2-11	$\mathbf{FA}_3 \otimes \mathbf{I}$	$\mathbf{FA}_3 \otimes \mathbf{I}$	$\mathbf{\Sigma_3}$	172288	77.0	96.4
M2-12[b]	$\mathbf{FA}_4 \otimes \mathbf{I}$	$\mathbf{FA}_4 \otimes \mathbf{I}$	$\mathbf{\Sigma_3}$	-	-	-
M2-13	$\mathbf{FA}_2 \otimes \mathbf{I}$	$\mathbf{FA}_1 \otimes \mathbf{I}$	$\mathbf{\Sigma_3}$	172494	66.4	61.0
M2-14	$\mathbf{FA}_3 \otimes \mathbf{I}$	$\mathbf{FA}_1 \otimes \mathbf{I}$	$\mathbf{\Sigma_3}$	172296	77.6	61.0
M2-15	$\mathbf{FA}_3 \otimes \mathbf{I}$	$\mathbf{FA}_2 \otimes \mathbf{I}$	$\mathbf{\Sigma_3}$	172276	77.5	82.9
M2-16	$\mathbf{FA}_4 \otimes \mathbf{I}$	$\mathbf{FA}_2 \otimes \mathbf{I}$	$\mathbf{\Sigma_3}$	172205	84.7	86.3
M2-17	$\mathbf{FA}_4 \otimes \mathbf{I}$	$\mathbf{FA}_3 \otimes \mathbf{I}$	$\mathbf{\Sigma_3}$	172234	87.3	96.7
M2-18	$\mathbf{FA}_5 \otimes \mathbf{I}$	$\mathbf{FA}_2 \otimes \mathbf{I}$	$\mathbf{\Sigma_3}$	**171990**	**90.2**	**87.2**
M2-18	$\mathbf{FA}_5 \otimes \mathbf{I}$	$\mathbf{FA}_3 \otimes \mathbf{I}$	$\mathbf{\Sigma_3}$	172196	88.9	96.9
M2-19[b]	$\mathbf{FA}_6 \otimes \mathbf{I}$	$\mathbf{FA}_2 \otimes \mathbf{I}$	$\mathbf{\Sigma_3}$	-	-	-

[a] The evaluated variance-covariance structures were identity (**I**), diagonal (**D**), and factor analytic ($\mathbf{FA}_k$) from order k = 1, ..., 6. Σ_{gl} and Σ_{gy} are covariance matrices for genotype by location (GL) and genotype by year (GY) interactions, respectively. $\Sigma_{\epsilon} = \mathbf{\Sigma_3}$ is the residual variance matrix assumed to be known from single trial and location analysis.

[b] Singularity in the Average Information matrix.

Table 2: Point estimates and standard error of variance components for seed yield computed from Soybean Cooperative Tests (1989-2019) using Model 2-1 (baseline) and six clustering methods for clustering locations into mega-environments using Model 4.

Variance components[a]	Model 2-1	Clustering criteria[b]					
		PHE	SoilE	Lat2	Lat3	WA	WW
$\hat{\sigma}^2_G$	6.3 (0.3)	4.5 (0.4)	5.0 (0.5)	4.8 (0.4)	5.6 (0.4)	5.7 (0.4)	6.3 (0.4)
$\hat{\sigma}^2_L$	58.9 (13.1)[c]	56.4 (12.6)	58.6 (13.1)	59.5 (13.2)	48.9 (11.4)	59.1 (13.2)	60.5 (13.5)
$\hat{\sigma}^2_Y$	15.3 (5.4)	13.4 (6.6)	10.6 (7.2)	11.5 (6.3)	15.1 (6.1)	16.5 (6.5)	15.0 (6.4)
$\hat{\sigma}^2_{LY}$	81.4 (5.2)	74.6 (5.0)	79.9 (5.2)	76.5 (5.0)	77.7 (5.2)	80.0 (5.1)	79.9 (5.2)
$\hat{\sigma}^2_{GL}$	4.8 (0.3)	-	-	-	-	-	-
$\hat{\sigma}^2_{GY}$	3.2 (0.3)	2.8 (0.3)	2.9 (0.4)	3.2 (0.3)	3.1 (0.3)	2.8 (0.3)	3.0 (0.3)
$\hat{\sigma}^2_{GLY}$	10.1 (0.3)	-	-	-	-	-	-
$\hat{\sigma}^2_{RY}$	-	12.2 (5.7)	7.6 (6.5)	10.2 (5.2)	5.9 (3.9)	2.8 (3.7)	3.6 (4.1)
$\hat{\sigma}^2_{GR}$	-	4.1 (0.3)	2.9 (0.4)	2.5 (0.3)	1.5 (0.2)	1.1 (0.3)	0.0 (0.0)
$\hat{\sigma}^2_{GL(R)}$	-	2.5 (0.3)	4.6 (0.3)	3.6 (0.3)	3.8 (0.3)	4.4 (0.3)	4.8 (0.3)
$\hat{\sigma}^2_{GRY}$	-	0.9 (0.2)	0.5 (0.3)	0.1 (0.2)	0.1 (0.2)	0.7 (0.3)	0.5 (0.2)
$\hat{\sigma}^2_{GL(R)Y}$	-	9.7 (0.3)	9.9 (0.3)	10.3 (0.3)	10.2 (0.3)	9.9 (0.3)	9.9 (0.3)
Clusters	-	3	2	2	3	2	2

[a] Genotypic ($\hat{\sigma}^2_G$), location ($\hat{\sigma}^2_L$), year ($\hat{\sigma}^2_Y$), location by year ($\hat{\sigma}^2_{LY}$), genotype by location ($\hat{\sigma}^2_{GL}$), genotype by year ($\hat{\sigma}^2_{GY}$), genotype by location by year ($\hat{\sigma}^2_{GLY}$), cluster by year ($\hat{\sigma}^2_{RY}$), genotype by cluster ($\hat{\sigma}^2_{GR}$), genotype by location nested in cluster ($\hat{\sigma}^2_{GL(R)}$), genotype by cluster by year ($\hat{\sigma}^2_{GRY}$), and genotype by location nested in cluster by year ($\hat{\sigma}^2_{GL(R)Y}$).

[b] Phenotypic (PHE), soil and elevation (SoilE), latitude split into two groups (Lat2), latitude split into three groups (Lat3), weather from means across years (WA), and weather from means within years (WW).

[c] Locations were modeled as a random effect in Model 2-1 (Table 1).

Table 3: Maximum Likelihood estimates of parameters for the best-fit univariate and multivariate probability distributions for empirical distributions obtained using jackknife resampling. Estimates of residual variance (σ^2_ϵ) were obtained from trials conducted from 1989 to 2019.

Variance[a]	Distribution	Number of distributions	Parameters[b]		
			$\hat{w}_i$	$\hat{\alpha}_i$	$\hat{\beta}_i$
σ^2_G	Log-Logistic	5	0.30	125.98	3.54
			0.14	119.87	6.68
			0.08	44.03	7.23
			0.01	88.03	6.18
			0.47	192.69	7.53
σ^2_{GL}	Log-Logistic	3	0.63	122.27	4.23
			0.17	32.00	3.12
			0.20	92.56	7.62
σ^2_{GY}	Log-Logistic	6	0.02	258.41	3.77
			0.31	172.61	3.91
			0.20	89.05	3.51
			0.30	94.97	2.06
			0.02	32.81	2.64
			0.15	57.79	3.01
σ^2_{GLY}	Gamma	5	0.05	4078.04	0.0024
			0.31	7755.00	0.0011
			0.20	5306.62	0.0025
			0.28	31922.50	0.0003
			0.16	7301.02	0.0020
σ^2_ϵ	Log-Logistic	1	1	2.56	17.03

[a] Genotypic (σ^2_G), genotype by location (σ^2_{GL}), genotype by year (σ^2_{GY}), and genotype by location by year (σ^2_{GLY}) variance components.

[b] Estimates of weight parameters (w_i) sums to one, and both Gamma and Log-Logistic distributions include a shape (α_i) and scale (β_i) parameter.

Table 4: The ratio of correlated responses from selection across all environments relative to direct responses to selection within mega-environments (CR/DR) for each clustering type. ρ_g is the correlation between estimated genotypic effects in the non-clustered and clustered sets of environments, i^2_{L} and i^2_{SR} are the reliabilities of genotype means in the non-clustered and clustered sets of environments, respectively.

Clustering type[a]	Number of		Estimates of			
	Clusters	Locations	$\hat{\rho}_g$	$\hat{i}^2_{\mathrm{L}}$	$\hat{i}^2_{\mathrm{SR}}$	$\widehat{CR/DR}$
PHE	3	36 / 7 / 20	0.72	0.38	0.51	0.62
SoilE	2	9 / 54	0.88	0.39	0.45	0.81
Lat2	2	35 / 28	0.81	0.39	0.48	0.73
Lat3	3	16 / 36 / 11	0.89	0.44	0.44	0.89
WA	2	25 / 38	0.92	0.45	0.45	0.92
WW	2	19 / 44	1.00	0.51	0.43	1.08
	2	-	0.99	0.52	0.44	1.07
At random	3	-	0.99	0.51	0.39	1.14
	4	-	0.99	0.51	0.36	1.18

[a] Phenotypic (PHE), soil and elevation (SoilE), latitude split into two groups (Lat2), latitude split into three groups (Lat3), weather from means across years (WA), and weather from means within years (WW).

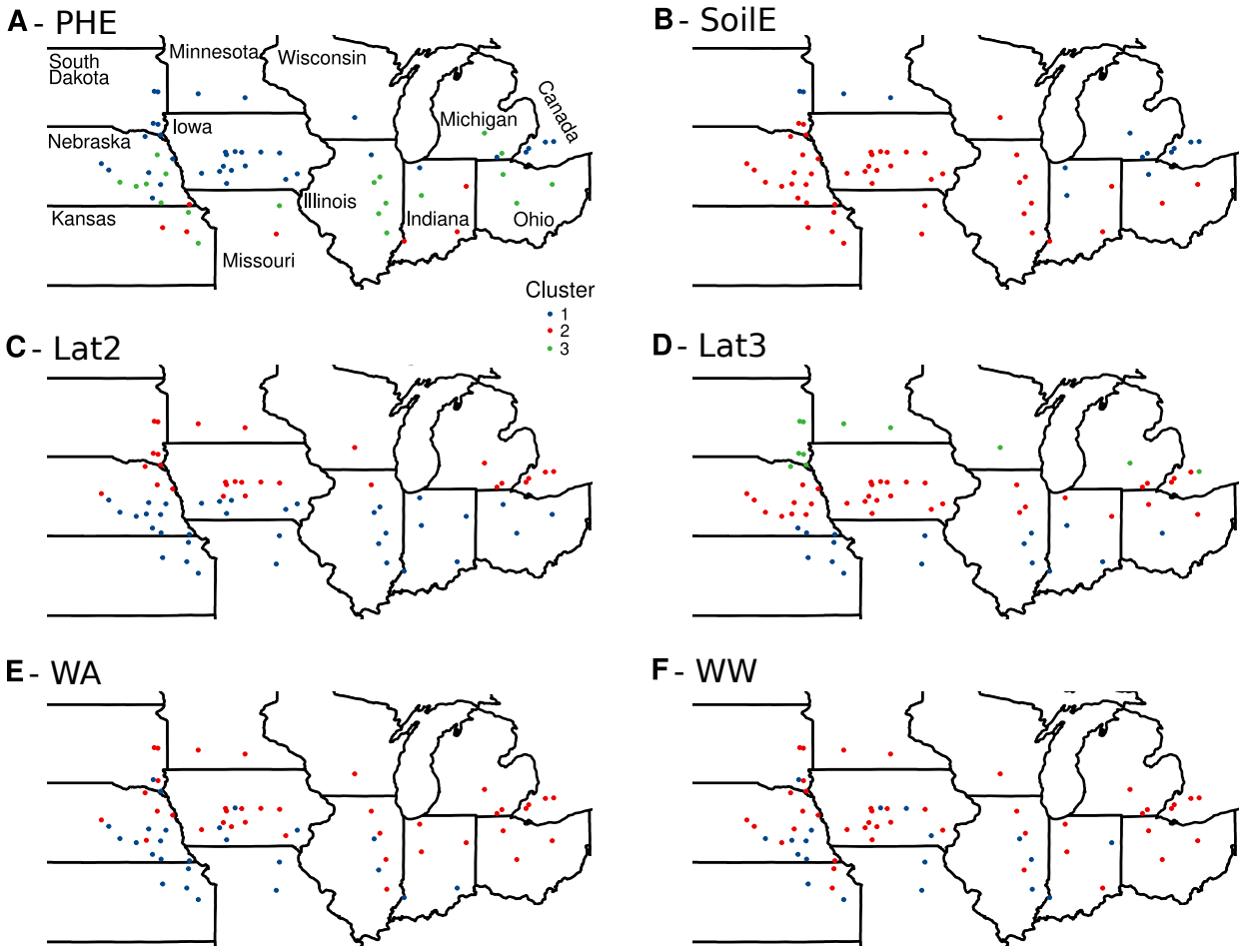

Figure 5: Geographic visualization of the target population of environments divided according to phenotypic (A), soil + elevation (B), latitude split into two groups (C), latitude split into three groups (D), weather across years (E), and weather within years (F) clustering types. In (A), the states' names are provided for geographic orientation.

References

Aguate, F., J. Crossa, and M. Balzarini, 2019 Effect of missing values on variance component estimates in multi-environment trials. Crop Science **59**: 508–517.

Akaike, H., 1974 A New Look at the Statistical Model Identification. IEEE Transactions on Automatic Control **19**: 716–723.

Annicchiarico, P., 2021 Breeding gain from exploitation of regional adaptation: An alfalfa case study. Crop Science **61**: 2254–2271.

Arief, V. N., I. H. DeLacy, J. Crossa, T. Payne, R. Singh, H. J. Braun, T. Tian, K. E. Basford, and M. J. Dieters, 2015 Evaluating Testing Strategies for Plant Breeding Field Trials: Redesigning a CIMMYT International Wheat Nursery. Crop Science **55**: 164–177.

Atlin, G. N., R. J. Baker, K. B. McRae, and X. Lu, 2000a Selection response in subdivided target regions. Crop Science **40**: 7–13.

Atlin, G. N., K. B. McRae, and X. Lu, 2000b Genotype x region interaction for two-row barley yield in Canada. Crop Science **40**: 1–6.

Bernardo, R., 2020 *Breeding for Quantitative Traits in Plants*. Stemma Press, Woodbury, MN, third edition.

Buntaran, H., J. Forkman, and H. P. Piepho, 2021 Projecting results of zoned multi-environment trials to new locations using environmental covariates with random coefficient models: accuracy and precision. Theoretical and Applied Genetics **134**: 1513–1530.

Buntaran, H., H. P. Piepho, J. Hagman, and J. Forkman, 2019 A cross-validation of statistical models for zoned-based prediction in cultivar testing. Crop Science **59**: 1544–1553.

Burgueño, J., J. Crossa, P. L. Cornelius, and R. C. Yang, 2008 Using factor analytic models for joining environments and genotypes without crossover genotype x environment interaction. Crop Science **48**: 1291–1305.

Bustos-Korts, D., 2017 *Modelling of genotype by environment interaction and prediction of complex traits across multiple environments as a synthesis of crop growth modelling, genetics and statistics*. Ph.D. thesis, Wageningen University.

Bustos-Korts, D., M. P. Boer, K. Chenu, B. Zheng, S. Chapman, and F. A. Van Eeuwijk, 2021 Genotype-specific P-spline response surfaces assist interpretation of regional wheat adaptation to climate change. In Silico Plants **3**: 1–23.

Butler, D. G., B. R. Cullis, A. R. Gilmour, B. G. Gogel, and R. Thompson, 2017 ASReml-R Reference Manual Version 4.

CIMMYT, 1989 *Toward the 21st century*. CIMMYT, Mexico, D.F.

Cooper, M. and I. H. DeLacy, 1994 Relationships among analytical methods used to study genotypic variation and genotype-by-environment interaction in plant breeding multi-environment experiments. Theoretical and Applied Genetics **88**: 561–572.

Costa-Neto, G., G. Galli, H. F. Carvalho, J. Crossa, and R. Fritsche-Neto, 2021 EnvRtype: a software to interplay enviromics and quantitative genomics in agriculture. G3 (Bethesda, Md.) **11**.

Crespo-Herrera, L. A., J. Crossa, J. Huerta-Espino, S. Mondal, G. Velu, P. Juliana, M. Vargas, P. Pérez-Rodríguez, A. K. Joshi, H. J. Braun, and R. P. Singh, 2021 Target Population of Environments for Wheat Breeding in India: Definition, Prediction and Genetic Gains. Frontiers in Plant Science **12**: 1–15.

Cullis, B. R., A. Smith, C. Hunt, and A. Gilmour, 2000 An examination of the efficiency of Australian crop variety evaluation programmes. Journal of Agricultural Science **135**: 213–222.

da Silva, K. J., P. E. Teodoro, M. J. da Silva, L. P. R. Teodoro, M. J. Cardoso, V. d. P. C. Godinho, J. H. Mota, G. A. Simon, F. D. Tardin, A. R. da Silva, F. L. Guedes, and C. B. de Menezes, 2021 Identification of mega-environments for grain sorghum in Brazil using GGE biplot methodology. Agronomy Journal **113**: 1–12.

Dempster, A. P., N. M. Laird, and D. B. Rubin, 1977 Maximum Likelihood from Incomplete Data Via the EM Algorithm. Journal of the Royal Statistical Society: Series B (Methodological) **39**: 1–38.

Dias, K. O., H. P. Piepho, L. J. Guimarães, P. E. Guimarães, S. N. Parentoni, M. O. Pinto, R. W. Noda, J. V. Magalhães, C. T. Guimarães, A. A. Garcia, and M. M. Pastina, 2020 Novel strategies for genomic prediction of untested single-cross maize hybrids using unbalanced historical data. Theoretical and Applied Genetics **133**: 443–455.

Falconer, D. S., 1952 The Problem of Environment and Selection. The American Naturalist **86**: 293–298.

Frensham, A., B. Cullis, and A. Verbyla, 1997 Genotype by Environment Variance Heterogeneity in a Two-Stage Analysis. Biometrics **53**: 1373–1383.

Friesen, L. F., A. L. Brûlé-Babel, G. H. Crow, and P. A. Rothenburger, 2016 Mixed model and stability analysis of spring wheat genotype yield evaluation data from Manitoba, Canada. Canadian Journal of Plant Science **96**: 305–320.

Gauch, H. and R. Zobel, 1997 Identifying mega-environments and targeting genotypes. Crop Science **37**: 311–326.

George, N. and M. Lundy, 2019 Quantifying genotype × environment effects in long-term common wheat yield trials from an agroecologically diverse production region. Crop Science **59**: 1960–1972.

Haldane, J., 1947 The interaction of nature and nurture. Annals of Human Genetics **17**: 197–205.

Hartung, J. and H. Piepho, 2021 Effect of missing values in multi-environmental trials on variance component estimates. Crop Science pp. 1–11.

Henderson, C. R., 1949 Estimates of changes in herd environment. Journal of Dairy Science .

Henderson, C. R., 1950 Estimation of genetic parameters. Annals of Mathematical Statistics **21**: 309–310.

Henderson, C. R., O. Kempthorne, S. R. Searle, and C. M. von Krosigk, 1959 The Estimation of Environmental and Genetic Trends from Records Subject to Culling. Biometrics **15**: 192.

Hijmans, R. J., 2021 raster: Geographic Data Analysis and Modeling.

Hollister, J., T. Shah, A. L. Robitaille, M. W. Beck, and M. Johnson, 2021 elevatr: Access Elevation Data from Various APIs.

Holzworth, D. P., N. I. Huth, P. G. deVoil, E. J. Zurcher, N. I. Herrmann, G. McLean, K. Chenu, E. J. van Oosterom, V. Snow, C. Murphy, A. D. Moore, H. Brown, J. P. Whish, S. Verrall, J. Fainges, L. W. Bell, A. S. Peake, P. L. Poulton, Z. Hochman, and P. J. Thorburn, 2014 APSIM – Evolution towards a new generation of agricultural systems simulation. Environmental modelling & software : with environment data news. **62**: 327–350.

Jackson, S. D., 2009 Plant responses to photoperiod. New Phytologist **181**: 517–531.

Johannsen, W., 1911 The genotype conception of heredity. The American Naturalist **45**: 129–159.

Kang, M., 2020 Genotype-environment interaction and stability analyses: an update. In *Quantitative genetics, genomics and plant breeding*, edited by M. Kang, chapter 9, pp. 140–161, CABI, second edition.

Kassambara, A. and F. Mundt, 2020 factoextra: Extract and Visualize the Results of Multivariate Data Analyses.

Kleinknecht, K., J. Möhring, F. Laidig, U. Meyer, and H. P. Piepho, 2016 A simulation-based approach for evaluating the efficiency of multienvironment trial designs. Crop Science **56**: 2237–2250.

Krishnamurthy, S. L., P. C. Sharma, D. K. Sharma, K. T. Ravikiran, Y. P. Singh, V. K. Mishra, D. Burman, B. Maji, S. Mandal, S. K. Sarangi, R. K. Gautam, P. K. Singh, K. K. Manohara, B. C. Marandi, G. Padmavathi, P. B. Vanve, K. D. Patil, S. Thirumeni, O. P. Verma, A. H. Khan, S. Tiwari, S. Geetha, M. Shakila, R. Gill, V. K. Yadav, S. K. Roy, M. Prakash, J. Bonifacio, A. Ismail, G. B. Gregorio, and R. K. Singh, 2017 Identification of mega-environments and rice genotypes for general and specific adaptation to saline and alkaline stresses in India. Scientific Reports **7**: 1–14.

Laidig, F., T. Drobek, and U. Meyer, 2008 Genotypic and environmental variability of yield for cultivars from 30 different crops in German official variety trials. Plant Breeding **127**: 541–547.

Laidig, F., H. P. Piepho, D. Rentel, T. Drobek, U. Meyer, and A. Huesken, 2017a Breeding progress, environmental variation and correlation of winter wheat yield and quality traits in German official variety trials and on-farm during 1983–2014. Theoretical and Applied Genetics **130**: 223–245.

Laidig, F., H. P. Piepho, D. Rentel, T. Drobek, U. Meyer, and A. Huesken, 2017b Breeding progress, variation, and correlation of grain and quality traits in winter rye hybrid and population varieties and national on-farm progress in Germany over 26 years. Theoretical and Applied Genetics **130**: 981–998.

Li, X. X., T. Guo, Q. Mu, X. X. Li, and J. Yu, 2018 Genomic and environmental determinants and their interplay underlying phenotypic plasticity. Proceedings of the National Academy of Sciences of the United States of America **115**: 6679–6684.

Little, R. J. A. and D. B. Rubin, 2020 *Statistical analysis with missing data*. Wiley series in probability and statistics, John Wiley \& Sons, third edition.

Mackay, I., H. P. Piepho, and A. A. F. Garcia, 2019 Statistical Methods for Plant Breeding. In *Handbook of Statistical Genomics*, edited by D. Balding, I. Moltke, and J. Marioni, chapter 17, pp. 501–530, John Wiley & Sons Ltd, Hoboken, NJ, fourth edition.

Meyer, U., F. Laidig, and T. Drobek, 2011 Optimization of number of trials in official VCU trial series of Germany. Biuletyn Oceny Odmian **33**: 73–82.

Microsoft Corporation and S. Weston, 2020a doParallel: Foreach Parallel Adaptor for the 'parallel' Package.

Microsoft Corporation and S. Weston, 2020b foreach: Provides Foreach Looping Construct.

Möhring, J. and H. P. Piepho, 2009 Comparison of weighting in two-stage analysis of plant breeding trials. Crop Science **49**: 1977–1988.

Möhring, J., E. Williams, and H. P. Piepho, 2015 Inter-block information: to recover or not to recover it? Theoretical and Applied Genetics **128**: 1541–1554.

Mourtzinis, S. and S. P. Conley, 2017 Delineating soybean maturity groups across the United States. Agronomy Journal **109**: 1397–1403.

Oakey, H., B. Cullis, R. Thompson, J. Comadran, C. Halpin, and R. Waugh, 2016 Genomic selection in multi-environment crop trials. G3: Genes, Genomes, Genetics **6**: 1313–1326.

Patterson, H. D. and R. Thompson, 1971 Recovery of inter-block information when block sizes are unequal. Biometrika **58**: 545–554.

Piepho, H.-P., 1997 Analyzing genotype-environment data by mixed models with multiplicative terms. Biometrics **53**: 761–766.

Piepho, H. P. and J. Möhring, 2005 Best linear unbiased prediction of cultivar effects for subdivided target regions. Crop Science **45**: 1151–1159.

Piepho, H. P. and J. Mohring, 2006 Selection in cultivar trials - Is it ignorable? Crop Science **46**: 192–201.

Piepho, H. P., J. Möhring, T. Schulz-Streeck, and J. O. Ogutu, 2012 A stage-wise approach for the analysis of multi-environment trials. Biometrical Journal **54**: 844–860.

Piepho, H. P., M. F. Nazir, M. Qamar, A. U. R. Rattu, Riaz-Ud-Din, M. Hussain, G. Ahmad, Fazal-E-Subhan, J. Ahmad, Abdullah, K. B. Laghari, I. A. Vistro, M. Sharif Kakar, M. A. Sial, and M. Imtiaz, 2016 Stability analysis for a countrywide series of wheat trials in Pakistan. Crop Science **56**: 2465–2475.

Pigliucci, M., 2001 *Phenotypic plasticity: Beyond nature and nurture*. The Johns Hopkins University Press, Baltimore, MD.

R Core Team, 2021 R: A Language and Environment for Statistical Computing.

Rencher, A. C. and G. B. Schaalje, 2007 *Linear Models in Statistics*. John Wiley & Sons, Inc., Hoboken, NJ, USA.

Rothschild, M., C. Henderson, and R. L. Quaas, 1979 Effects of selection on variances and covariances of simulated first and second lactations. Journal of Dairy Science **62**: 996–1002.

Rubin, D., 1976 Inference and missing data. Biometrika **63**: 581–592.

Rutkoski, J. E., 2019 A practical guide to genetic gain. In *Advances in Agronomy*, edited by D. L. Sparks, volume 157 of *Advances in Agronomy*, pp. 217–249, Academic Press.

Schwarz, G., 1978 Estimating the dimension of a model. The Annals of Statistics **6**: 461–464.

Setiyono, T. D., A. Weiss, J. Specht, A. M. Bastidas, K. G. Cassman, and A. Dobermann, 2007 Understanding and modeling the effect of temperature and daylength on soybean phenology under high-yield conditions. Field Crops Research **100**: 257–271.

Singh, D. P., A. K. Singh, and A. Singh, 2021 *Plant Breeding and Cultivar Development*. Academic Press, first edition.

Smith, A., B. Cullis, and A. Gilmour, 2001a The analysis of crop variety evaluation data in Australia. Australian and New Zealand Journal of Statistics **43**: 129–145.

Smith, A., B. R. Cullis, and R. Thompson, 2001b Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. Biometrics **57**: 1138–1147.

Smith, A., A. Norman, H. Kuchel, and B. Cullis, 2021 Plant Variety Selection Using Interaction Classes Derived From Factor Analytic Linear Mixed Models: Models With Independent Variety Effects. Frontiers in Plant Science **12**.

Smith, A. B., A. Ganesalingam, H. Kuchel, and B. R. Cullis, 2015 Factor analytic mixed models for the provision of grower information from national crop variety testing programs. Theoretical and Applied Genetics **128**: 55–72.

Sparks, A., 2018 nasapower: NASA-POWER Data from R.

Sprague, G. and W. Federer, 1951 A comparison of variance components in corn yield trials: II. error, year x variety, location x variety, and variety components. Agronomy Journal **43**: 535–541.

Stacklies, W., H. Redestig, M. Scholz, D. Walther, and J. Selbig, 2007 pcaMethods – a Bioconductor package providing PCA methods for incomplete data. Bioinformatics **23**: 1164–1167.

Stephens, M., 1986 Tests based on edf statistics. In *Goodness-of-fit techniques*, edited by R. D'Agostino and M. Stephens, pp. 97–194, Marcel Dekker, New York.

Tabery, J., 2008 R. A. Fisher, Lancelot Hogben, and the origin(s) of genotype-environment interaction. Journal of the History of Biology **41**: 717–761.

Teimouri, M., 2021 ForestFit: Statistical Modelling for Plant Size Distributions.

USDA, 2021 Uniform Soybean Tests, Northern Region.

Windhausen, V. S., S. Wagener, C. Magorokosho, D. Makumbi, B. Vivek, H. P. Piepho, A. E. Melchinger, and G. N. Atlin, 2012 Strategies to subdivide a target population of environments: Results from the CIMMYT-led maize hybrid testing programs in Africa. Crop Science **52**: 2143–2152.

Wold, H., 1966 Estimation of principal components and related models by iterative least squares. In *Multivariate Analysis*, edited by P. R. Krishnajah, pp. 391–420, Academic Press, New York.

Yan, W., 2016 Analysis and handling of G x E in a practical breeding program. Crop Science **56**: 2106–2118.

Yan, W., J. Frégeau-Reid, D. Pageau, R. Martin, J. Mitchell-Fetch, M. Etienne, J. Rowsell, P. Scott, M. Price, B. de Haan, A. Cummiskey, J. Lajeunesse, J. Durand, and E. Sparry, 2010 Identifying essential test locations for oat breeding in Eastern Canada. Crop Science **50**: 504–515.

Yan, W., L. A. Hunt, Q. Sheng, and Z. Szlavnics, 2000 Cultivar Evaluation and Mega-Environment Investigation Based on the GGE Biplot. Crop Science **40**: 597–605.

Yan, W. and I. Rajcan, 2002 Biplot Analysis of Test Sites and Trait Relations of Soybean in Ontario. Crop Science **42**: 11–20.

Zdziarski, A. D., L. G. Woyann, A. S. Milioli, R. Zanella, L. V. Dallacorte, M. C. Panho, and G. Benin, 2019 Mega-environment identification for soybean (Glycine max) breeding and production in Brazilian Midwest region. Plant Breeding **138**: 336–347.

14 Appendix A - The structure of the data

The complete data consist of eBLUE values for seed yield (bu/ac) of experimental genotypes and check varieties grown in field trials of soybean maturity groups (MG) 00 through IV from 1941 to 2020. For our purposes, we restricted our analyses to data from the most reliable trials of MG II and III from 1989 to 2019. We restricted our interest to MG II and III because these are North America's primary soybean production maturity groups.

We defined unreliable trials as individual trials with estimates of reliability (i^2) less than 0.10 and coefficients of variation (CV%) greater than 20%. Reliability values less than 0.10 and CV values greater than 20% are indicators of multiple experimental units within a replicate that have been severely damaged by unknown sources (such as deer herds, raccoon troops, hail storms, etc.), causing unexpected, inconsistent phenotypic values between replicated genotypes within the trial at an environment. Not only are data from the damaged experimental units suspect, but adjacent undamaged experimental units have unfair advantages from lack of competition. Thus, in keeping with the GIGO (garbage in - garbage out) maxim, plant breeders routinely exclude data from unreliable field trials prior to making decisions about which genotypes should be retained for future trials and breeding purposes. Our exclusion rules are consistent with best practices by public and commercial soybean breeders.

To assure that individual genotypes with unexplained low yields did not affect the variance estimates, 12 individual records within trials with an estimated mean of less than ten bu/ac were removed. In total, only 6.8% of the observed data points were removed. Lastly, locations with less than three years of data were excluded from our analyses. The resulting data comprised 4,257 experimental genotypes evaluated at 63 locations, in 31 years, resulting in 591 location-year combinations (environments) with 39,006 yield values (data points).

15 Appendix B - The analysis within versus across MG

We restricted our analyses to data from the most reliable trials (Appendix A) belonging to MG II and III from 1989 to 2019. Our approach was to make inferences across MG, but here we present the analysis within MG. The results (point estimate followed by the standard error) from the analyses across (presented in Table 2) and within (with individual datasets) MG are the following:

Variance components	Across MG II and II	MG II	MG III
Genotypic (G)	6.3 (0.3)	5.4 (0.4)	4.43 (0.4)
Location (L)	58.9 (13.1)	54.2 (16.1)	48.8 (15.6)
Year (Y)	15.3 (5.4)	13.7 (5.9)	27.9 (10.4)
LY	81.4 (5.2)	85.7 (6.9)	87.8 (7.9)
GL	4.8 (0.3)	4.6 (0.4)	3.8 (0.4)
GY	3.2 (0.3)	3.2 (0.3)	2.8 (0.3)
GLY	10.1 (0.3)	8.3 (0.4)	9.8 (0.4)

The interpretation within MG is the same: GL is more important than GY; and the static portion of the GEI (GL + GY + GLY) due to GL accounted for 26%, 28%, and 23%, respectively. An alternative way to test for homogeneous (pooled) versus heterogeneous variances within MG can be implemented with two different versions of Model 2. The model with heterogeneous variances within MG assumes that there is a unique value for each genotypic-related variance. For example, the estimated genotypic variance using data from both groups is σ^2_G, and within MG is σ^2_{G-MG2} and σ^2_{G-MG3}, for MG II and MG III, respectively. Both models can be fitted in *Asreml-R* (or any other mixed model package) using the *SoyURT* package as follows:

```
library(SoyURT) ; data(pheno)
pheno$location <- as.factor(pheno$location)
pheno$year <- as.factor(pheno$year)
pheno$G <- as.factor(pheno$G)
pheno$maturity_group <- as.factor(pheno$maturity_group)
pheno$weights <- 1/(pheno$SE)^2

library(asreml); library(dplyr)

# heterogeneous variances for G, GL, GY, and GLY
modelF <- asreml(fixed = eBLUE ~ location,
                 random = ~ year + location:year +
                        at(maturity_group):G +
                        location:at(maturity_group):G +
                        year:at(maturity_group):G +
                        location:year:at(maturity_group):G,
                 family = asr_gaussian(dispersion = 1),
                 weights = weights,
                 workspace = '4gb',
                 data = pheno)
```

```r
# homogeneous (pooled) variances, this is Model 2-1 presented in the main text
modelR <- asreml(fixed = eBLUE ~ location,
                 random = ~ year + location:year + G + location:G +
                        year:G + location:year:G,
                 family = asr_gaussian(dispersion = 1),
                 weights = weights,
                 workspace = '2gb',
                 data = pheno)

# variance estimates

## heterogeneous
summary(modelF, coef=TRUE)$varcomp # variance estimates

                                      component std.error   z.ratio bound %ch
year                                   15.464217 5.4901096  2.816741     P   0
location:year                          81.349536 5.2046065 15.630295     P   0
at(maturity_group, 2):G                 6.241550 0.5740897 10.872082     P   0
at(maturity_group, 3):G                 5.847493 0.4346075 13.454653     P   0
at(maturity_group, 2):year:G            4.171794 0.4255065  9.804302     P   0
at(maturity_group, 3):year:G            2.667052 0.3161606  8.435751     P   0
at(maturity_group, 2):location:G        5.553697 0.4002084 13.877015     P   0
at(maturity_group, 3):location:G        4.032175 0.4159511  9.693869     P   0
at(maturity_group, 2):location:year:G   9.017481 0.3962911 22.754687     P   0
at(maturity_group, 3):location:year:G  11.115090 0.4362629 25.477963     P   0
units!R                                 1.000000        NA       NA     F   0

## homogeneous
summary(modelR, coef=TRUE)$varcomp # variance estimates
                  component std.error   z.ratio bound %ch
year              15.272165 5.4408179  2.806961     P   0
location:year     81.356976 5.2052322 15.629846     P   0
G                  6.299500 0.3437735 18.324563     P   0
year:G             3.226249 0.2440203 13.221234     P   0
location:G         4.763123 0.2841663 16.761742     P   0
location:year:G   10.113454 0.2899932 34.874792     P   0
units!R            1.000000        NA       NA     F   0
```

With the fitted models, it is possible to evaluate the AIC/BIC selection criteria to check which model has a better goodness-of-fit:

```r
# selection criteria AIC/BIC
library(asremlPlus)
infoCriteria(list(modelF, modelR))

 fixedDF varDF NBound      AIC      BIC     loglik
1       0    10      1 175082.4 175168.1 -87531.20 # heterogeneous variances
2       0     6      1 175071.4 175122.8 -87529.68 # homogeneous variances
```

There are ten variance components to be estimated in the heterogeneous model (*modelF*), whereas for the homogeneous model (*modelR*), six variance components. Both AIC and BIC values are presented above. Therefore, there is evidence it is reasonable to use the homogeneous variance model and obtain pooled variance estimates for MG II and III. In addition to the statistical modeling, the challenge, of course, is to delineate a continuous characteristic (maturity) that is affected by the environment (heat and drought). For example, what is the difference between a genotype that has been classified as 2.9 and a genotype that has been classified as 3.1? Is the line classified as 2.9 a member of MG II (consisting of lines classified as 2.1) or better classified as MG III? In other words, experimental lines classified as MG II might actually be early MG III lines, whereas early experimental lines classified as MG III might be late MG II lines. Moreover, by combining both MG, the genetics and the locations were better sampled.

Lastly, we fitted the following model considering MG as a fixed effect:

```
modelMG <- asreml(fixed = Value ~ MG + Location,
                  random = ~ year + Location:year + GermplasmId +
                  Location:GermplasmId + year:GermplasmId +
                  Location:year:GermplasmId,
                  family = asr_gaussian(dispersion = 1),
                  weights = weights,
                  workspace = '2gb',
                  data = data)

summary(modelMG)$varcomp

                          component std.error   z.ratio bound %ch
year                      16.956438 5.8644440  2.891397     P   0
Location:year             81.229040 5.1246720 15.850583     P   0
GermplasmId                4.517019 0.3113566 14.507541     P   0
year:GermplasmId           3.558186 0.2543367 13.990062     P   0
Location:GermplasmId       4.528129 0.2764090 16.381989     P   0
Location:year:GermplasmId 10.293894 0.2850479 36.112853     P   0
units!R                    1.000000        NA       NA     F   0
```

In conclusion, the REML estimates of variance are similar to the ones reported in the main manuscript.

16 Appendix C - Additional figures and tables

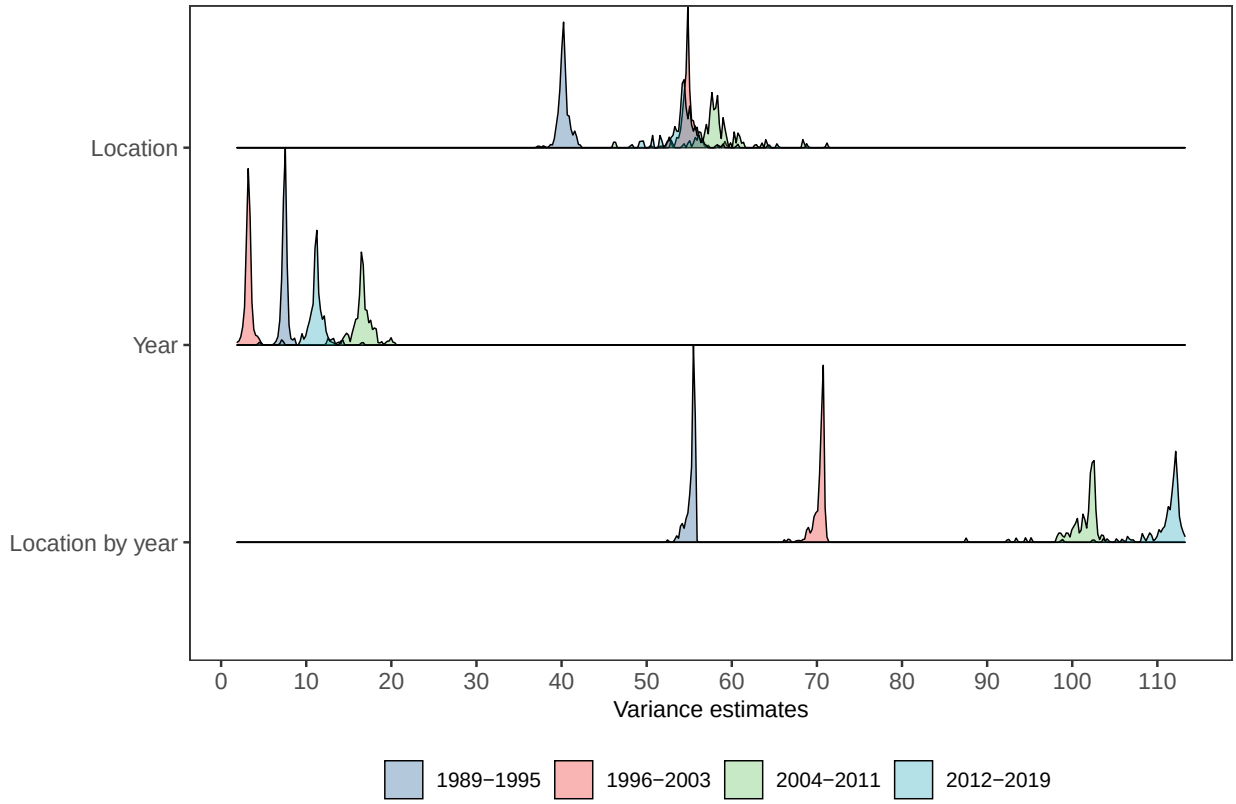

Figure C1: Jackknife estimates of location, year, and location by year variances for the groups of years 1989-1995, 1996-2003, 2004-2011, and 2012-2019.

Table C1: Meteorological variables and selected window across years (WA) with the highest Person correlation ($\hat{\rho}$).

Window			Parameter			$\hat{\rho}$
Beggining	End	Size	Name	Code	Description	
1	8	7	ALLSKY_SFC_LW_DWN	MV1	Downward thermal infrared (longwave) radiative flux	0.1719
9	16	7	ALLSKY_SFC_SW_DWN	MV2	Insolation incident on a horizontal surface	0.1956
9	16	7	ETP	MV3	Evapotranspiration	0.2007
3	10	7	FRUE	MV4	Effect of temperature on radiation use efficiency	0.2021
3	10	7	GDD	MV5	Growing degree-days	0.2012
9	16	7	n	MV6	Duration of sunshine hours	0.1904
52	71	19	PETP	MV7	Deficit of evapotranspiration	-0.1522
52	72	20	PRECTOT	MV8	Rainfall precipitation	-0.1267
3	10	7	PTR	MV9	Photothermal ratio (GDD / daylight in hours)	0.2037
3	10	7	PTT	MV10	Photothermal time (GDD $\times$ daylight in hours)	0.1981
1	8	7	RH2M	MV11	Relative humidity at 2 meters	0.0899
3	10	7	SPV	MV12	The slope of saturation vapor pressure curve	0.1995
3	11	8	T2M	MV13	Daily average temperature at 2 meters	0.2052
3	14	11	T2M_MAX	MV14	Daily minimum temperature at 2 meters	0.1849
1	8	7	T2M_MIN	MV15	Daily maximum temperature at 2 meters	0.2168
1	8	7	T2M_RANGE	MV16	Daily temperature range at 2 meters	-0.0943
1	8	7	T2MDEW	MV17	Dew point at 2 meters	0.2081
7	14	7	VPD	MV18	The deficit of vapor pressure	0.0909
69	77	8	WS2M	MV19	Wind speedy at 2 meters	-0.1397

Table C2: Meteorological variables and selected window within years (WW) with the highest Person correlation ($\hat{\rho}$).

Window				Parameter			$\hat{\rho}$
Beggining	End	Size	Year	Name	Code	Description	
40	47	7	2005	ALLSKY_SFC_LW_DWN	MV1	Downward thermal infrared (longwave) radiative flux	-0.8768
46	55	9	1990	ALLSKY_SFC_SW_DWN	MV2	Insolation incident on a horizontal surface	-0.8845
46	55	9	1990	ETP	MV3	Evapotranspiration	-0.8571
12	44	32	2005	FRUE	MV4	Effect of temperature on radiation use efficiency	-0.9041
12	44	32	2005	GDD	MV5	Growing degree-days	-0.9033
99	124	25	2014	n	MV6	Duration of sunshine hours	0.9126
104	120	16	2005	PETP	MV7	Deficit of evapotranspiration	0.9226
108	120	12	2005	PRECTOT	MV8	Rainfall precipitation	0.9191
17	44	27	2005	PTR	MV9	Photothermal ratio (GDD / daylight in hours)	-0.8858
11	44	33	2005	PTT	MV10	Photothermal time (GDD $\times$ daylight in hours)	-0.9170
31	38	7	2016	RH2M	MV11	Relative humidity at 2 meters	-0.8826
11	44	33	2005	SPV	MV12	The slope of saturation vapor pressure curve	-0.9004
12	44	32	2005	T2M	MV13	Daily average temperature at 2 meters	-0.8996
12	26	14	2005	T2M_MAX	MV14	Daily minimum temperature at 2 meters	-0.8831
108	124	16	2005	T2M_MIN	MV15	Daily maximum temperature at 2 meters	-0.8533
112	121	9	2016	T2M_RANGE	MV16	Daily temperature range at 2 meters	0.8354
5	14	9	2004	T2MDEW	MV17	Dew point at 2 meters	0.7600
22	36	14	2015	VPD	MV18	The deficit of vapor pressure	0.9192
66	80	14	1990	WS2M	MV19	Wind speedy at 2 meters	-0.8760

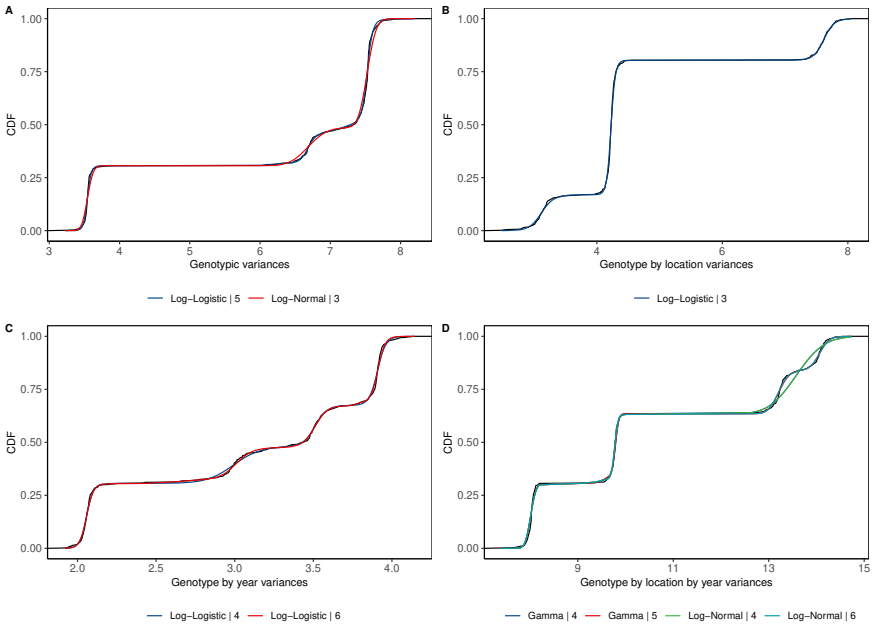

Figure C2: Cumulative distribution function (CDF) for the best-fit models according to the genotypic (A), genotype by location (B), genotype by year (C), and genotype by location by year (D) variances from jackknife. In each plot, the legends states for name of the distribution followed by its mixture number.

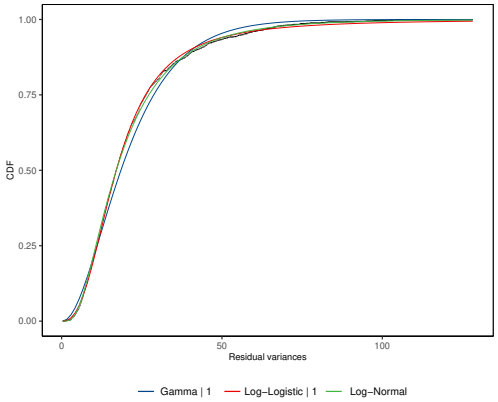

Figure C3: Cumulative distribution function (CDF) for the best-fit models according to the residual variances from individual trial level. The legends state the name of the distribution followed by its mixture number.

Table C3: Number of genotypes (NG) and locations (NL) within each year, the number of common genotypes between years (upper diagonal), and the number of common locations between years (lower diagonal).

Years	89	90	91	92	93	94	95	96	97	98	99	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	NG
89		7	18	9	5	4	2																									192
90	12		28	14	5	4	1																									144
91	22	13		45	13	5	4																									180
92	20	12	21		48	20	8	2	1	1	1																					194
93	16	11	21	24		40	15	2	1	1	1																					179
94	14	11	18	20	25		45	7	1	1	1																					178
95	15	10	18	21	24	21		42	8	4	1																					193
96	14	9	16	20	21	18	21		32	10	4	2	2	2	2	1	1	1	1	1												187
97	16	10	19	21	24	21	21	22		45	17	5	3	2	2	1	1	1	1	1												184
98	17	12	19	22	23	20	20	20	25		43	14	5	3	3	2	1	1	1	1												181
99	15	10	17	17	19	18	18	17	21	25		37	9	5	3	3	2	1	1	1												183
00	11	7	13	15	15	14	12	14	16	17	21		32	10	7	2	1	1	1	1												173
01	7	6	9	10	11	11	9	10	12	14	17	18		44	12	3	1	1	1	1												151
02	11	7	13	15	15	13	13	14	16	17	19	20	18		24	4	2	1	1	1												180
03	5	3	7	6	8	6	7	7	8	8	11	12	11	11		19	4	2	1	1												126
04	6	3	7	8	9	8	8	9	10	9	11	12	12	14	9		46	17	6	3	1	1	1	1	1	1	1	1				171
05	4	1	4	6	5	5	5	5	6	6	8	9	7	10	5	9		42	18	9	5	2	2	3	3	2	3	1				164
06	6	2	6	7	7	6	6	7	8	8	9	9	8	11	6	11	8		50	16	6	2	3	4	4	3	4	2	1	1	1	185
07	4	2	4	5	6	4	4	5	6	6	7	7	7	9	6	10	7	9		49	15	2	3	4	4	3	4	2	1	1	1	229
08	6	2	6	8	7	5	6	7	7	7	8	9	7	11	6	9	8	9	8		41	10	6	4	4	3	4	2	1	1		217
09	8	4	8	10	9	7	8	9	9	9	11	11	8	12	6	10	8	10	8	11		36	13	8	6	4	5	3	1	1		231
10	5	3	5	6	5	4	5	5	5	5	8	8	4	8	4	6	7	6	6	8	11		21	7	6	6	4	3	2	2	1	137
11	4	3	4	6	5	5	4	5	5	5	7	8	5	8	2	6	6	7	5	7	11	9		27	13	6	5	4	3	3	2	139
12	4	2	4	6	5	4	4	5	5	6	8	8	6	8	3	7	6	8	6	7	10	7	11		43	10	8	8	5	5	4	179
13	7	3	6	8	7	5	6	7	7	7	7	7	4	8	3	7	6	8	6	7	11	9	11	13		34	16	8	8	6	4	161
14	6	3	6	6	5	4	5	5	5	6	7	7	3	7	3	5	5	5	4	5	8	7	6	8	9		41	17	6	5	3	140
15	4	3	4	5	4	4	4	4	4	4	5	6	2	5	2	4	4	4	3	4	7	7	6	8	9	9		48	19	10	7	154
16	5	3	5	6	4	4	4	4	4	5	6	6	3	6	2	4	4	5	2	4	6	4	5	7	7	9	7		56	26	13	169
17	9	6	8	11	8	7	8	8	8	11	12	10	8	11	6	5	4	5	3	6	9	7	6	8	8	8	9	8		60	28	175
18	6	4	6	7	7	5	6	6	6	6	8	8	6	8	7	5	3	3	3	5	7	6	4	5	7	5	7	5	14		63	183
19	5	3	3	6	3	3	3	3	3	4	4	5	3	5	2	2	4	2	2	3	4	5	3	4	6	5	7	4	11	13		233
NL	24	15	27	30	32	27	26	23	27	31	30	25	21	23	14	15	12	12	10	11	15	12	13	16	17	12	11	11	20	16	13	

Table C4: Goodness-of-fit (GOF) statistics and selection criteria for the fit of univariate and multivariate probability distributions for genotypic, genotype by location, genotype by year, and genotype by location by year variance components estimated from the jackknife analysis, and residuals variances from trial-level. The best-fit model is highlighted in bold. KS, CM, AD, AIC, and BIC stand for Kolmogorov-Smirnov, Cramer-von Mises, Anderson-Darling, Akaike's and Bayesian Information Criterion, respectively.

Variance	Distribution	Mixture of	GOF statistics			GOF criteria		
			KS	CM	AD	AIC	BIC	Log-likelihood
Genotypic	Gamma	2	0.19	4.39	25.23	672.09	694.00	-331.05
	Log-Logistic	2	0.15	3.32	19.97	598.57	620.48	-294.29
	Log-Normal	2	0.19	4.48	25.66	680.36	702.27	-335.18
	Burr	2	0.50	40.79	193.24	3978.69	4000.60	-1984.35
	F	2	0.67	77.08	356.80	4338.62	4360.52	-2164.31
	Gamma	3	0.07	0.81	5.56	182.11	217.16	-83.06
	Log-Logistic	3	0.15	3.26	19.45	554.84	589.90	-269.42
	Log-Normal	3	0.08	0.66	4.90	170.74	205.79	-77.37
	Burr	3	0.52	44.39	207.91	3978.03	4013.09	-1981.02
	F	3	0.67	77.08	356.80	4338.62	4360.52	-2164.31
	Gamma	4	0.09	1.03	6.20	177.38	225.58	-77.69
	Log-Logistic	4	0.15	3.26	19.45	565.62	613.82	-271.81
	Log-Normal	4	0.08	0.52	3.61	74.67	122.87	-26.33
	Burr	4	0.52	44.39	207.91	3984.04	4032.24	-1981.02
	F	4	0.67	75.30	347.50	4351.57	4399.77	-2164.79
	Gamma	5	0.10	0.74	5.85	155.12	216.47	-63.56
	Log-Logistic	**5**	0.05	0.21	1.67	-9.63	51.71	18.82
	Log-Normal	5	0.06	0.36	2.53	89.40	150.75	-30.70
	Burr	5	0.53	45.19	211.25	3989.54	4050.89	-1980.77
	F	5	0.67	75.93	350.80	4357.11	4418.45	-2164.55
	Gamma	6	0.07	0.46	3.15	94.30	168.79	-30.15
	Log-Logistic	6	0.05	0.20	1.89	32.97	107.47	0.51
	Log-Normal	6	0.06	0.37	2.68	103.17	177.66	-34.59
	Burr	6	0.52	43.68	205.00	3996.91	4071.40	-1981.45
	F	6	0.66	74.44	343.11	4364.38	4438.87	-2165.19
Genotype by Location	Gamma	2	0.30	10.51	48.78	1164.59	1186.50	-577.29
	Log-Logistic	2	0.24	8.35	49.77	1139.99	1161.90	-565.00
	Log-Normal	2	0.30	10.80	50.82	1194.97	1216.87	-592.48
	Burr	2	0.50	42.63	197.90	3466.48	3488.39	-1728.24
	F	2	0.63	67.82	305.85	3830.34	3852.24	-1910.17
	Gamma	3	0.30	10.52	48.87	1185.88	1220.94	-584.94
	Log-Logistic	**3**	0.04	0.14	0.72	-80.42	-45.36	48.21
	Log-Normal	3	0.06	0.49	2.38	-5.10	29.96	10.55
	Burr	3	0.50	42.78	198.50	3477.58	3512.63	-1730.79
	F	3	0.63	67.82	305.85	3830.34	3852.24	-1910.17
	Gamma	4	0.30	10.51	48.83	1185.69	1233.89	-581.85
	Log-Logistic	4	0.04	0.13	0.62	-80.72	-32.52	51.36
	Log-Normal	4	0.06	0.48	2.20	-10.68	37.52	16.34
	Burr	4	0.50	41.96	195.32	3482.35	3530.55	-1730.18
	F	4	0.65	73.19	331.38	3854.78	3902.98	-1916.39
	Gamma	5	0.02	0.07	0.56	-34.66	26.69	31.33
	Log-Logistic	5	0.03	0.10	0.59	-73.99	-12.64	50.99
	Log-Normal	5	0.02	0.06	0.49	-66.28	-4.93	47.14

Table C4 continued from previous page

Variance	Distribution	Mixture of	GOF statistics			GOF criteria		
			KS	CM	AD	AIC	BIC	Log-likelihood
	Burr	5	0.49	41.22	192.48	3488.02	3549.36	-1730.01
	F	5	0.65	72.56	328.33	3860.42	3921.76	-1916.21
	Gamma	6	0.02	0.06	0.40	-30.91	43.58	32.46
	Log-Logistic	6	0.03	0.09	0.46	-74.96	-0.47	54.48
	Log-Normal	6	0.03	0.04	0.25	-71.57	2.92	52.78
	Burr	6	0.49	41.61	194.00	3494.23	3568.72	-1730.11
	F	6	0.65	72.49	328.02	3866.55	3941.05	-1916.28
Genotype by Year	Gamma	2	0.13	2.33	17.77	805.22	827.13	-397.61
	Log-Logistic	2	0.12	1.39	10.73	522.98	544.89	-256.49
	Log-Normal	2	0.13	1.80	12.89	546.28	568.19	-268.14
	Burr	2	0.46	34.26	165.40	2641.18	2663.09	-1315.59
	F	2	0.67	80.39	375.82	3042.72	3064.63	-1516.36
	Gamma	3	0.12	1.48	10.61	438.20	473.26	-211.10
	Log-Logistic	3	0.08	0.29	1.48	69.69	104.75	-26.85
	Log-Normal	3	0.13	1.36	10.01	402.06	437.11	-193.03
	Burr	3	0.49	38.99	184.26	2636.35	2671.40	-1310.17
	F	3	0.67	80.39	375.82	3042.72	3064.63	-1516.36
	Gamma	4	0.12	1.34	9.44	391.85	440.05	-184.92
	Log-Logistic	4	0.04	0.11	1.14	-21.98	26.22	21.99
	Log-Normal	4	0.05	0.26	2.48	69.53	117.73	-23.77
	Burr	4	0.48	36.59	174.58	2642.16	2690.36	-1310.08
	F	4	0.65	73.44	338.37	3062.06	3110.26	-1520.03
	Gamma	5	0.07	0.45	4.18	142.71	204.05	-57.35
	Log-Logistic	5	0.03	0.09	1.07	-25.40	35.94	26.70
	Log-Normal	5	0.05	0.27	2.52	73.76	135.10	-22.88
	Burr	5	0.46	33.87	163.84	2652.24	2713.58	-1312.12
	F	5	0.65	73.45	338.42	3068.16	3129.51	-1520.08
	Gamma	6	0.05	0.29	2.90	78.16	152.65	-22.08
	Log-Logistic	**6**	0.03	0.08	1.04	-35.86	38.63	34.93
	Log-Normal	6	0.04	0.13	1.11	2.00	76.49	16.00
	Burr	6	0.45	32.15	157.07	2645.95	2720.44	-1305.98
	F	6	0.65	73.41	338.24	3073.76	3148.25	-1519.88
Genotype by Location by Year	Gamma	2	0.18	3.72	23.32	2027.94	2049.84	-1008.97
	Log-Logistic	2	0.17	3.40	21.40	2104.98	2126.88	-1047.49
	Log-Normal	2	0.18	3.73	23.33	2028.46	2050.36	-1009.23
	Burr	2	0.57	46.63	216.45	4962.90	4984.81	-2476.45
	F	2	0.68	68.15	308.55	5298.60	5320.51	-2644.30
	Gamma	3	0.06	0.52	3.94	905.88	940.93	-444.94
	Log-Logistic	3	0.17	3.22	19.52	1938.73	1973.78	-961.36
	Log-Normal	3	0.06	0.52	3.96	909.36	944.42	-446.68
	Burr	3	0.60	51.37	235.46	4969.12	5004.18	-2476.56
	F	3	0.68	68.15	308.57	5304.63	5339.68	-2644.31
	Gamma	4	0.06	0.41	3.50	830.66	878.86	-404.33
	Log-Normal	4	0.06	0.36	3.30	810.14	858.34	-394.07
	Burr	4	0.60	51.37	235.46	4975.11	5023.31	-2476.55
	F	4	0.69	71.27	323.14	5313.82	5362.02	-2645.91
	Gamma	**5**	0.05	0.21	1.74	709.98	771.33	-340.99
	Log-Normal	5	0.06	0.22	1.15	715.12	776.47	-343.56
	Burr	5	0.60	51.37	235.46	4981.10	5042.45	-2476.55

Table C4 continued from previous page

Variance	Distribution	Mixture of	GOF statistics			GOF criteria		
			KS	CM	AD	AIC	BIC	Log-likelihood
	F	5	0.70	75.75	344.66	5317.16	5378.51	-2644.58
	Gamma	6	0.06	0.32	2.58	781.22	855.71	-373.61
	Log-Normal	6	0.05	0.16	1.53	691.83	766.32	-328.91
	Burr	6	0.58	48.52	223.96	4989.49	5063.98	-2477.75
	F	6	0.68	68.92	312.13	5322.81	5397.30	-2644.40
Residual (Trial-level)	Gamma	1	0.06	1.60	9.57	11139.50	11150.02	-5567.80
	Log-Logistic	**1**	0.02	0.07	0.72	11063.82	11074.34	-5529.90
	Log-Normal	1	0.02	0.09	0.67	11066.60	11077.12	-5531.10
	Burr	1	0.41	79.00	377.80	13896.60	13907.11	-6946.30
	F	1	0.54	131.37	594.40	14677.45	14687.97	-7336.73

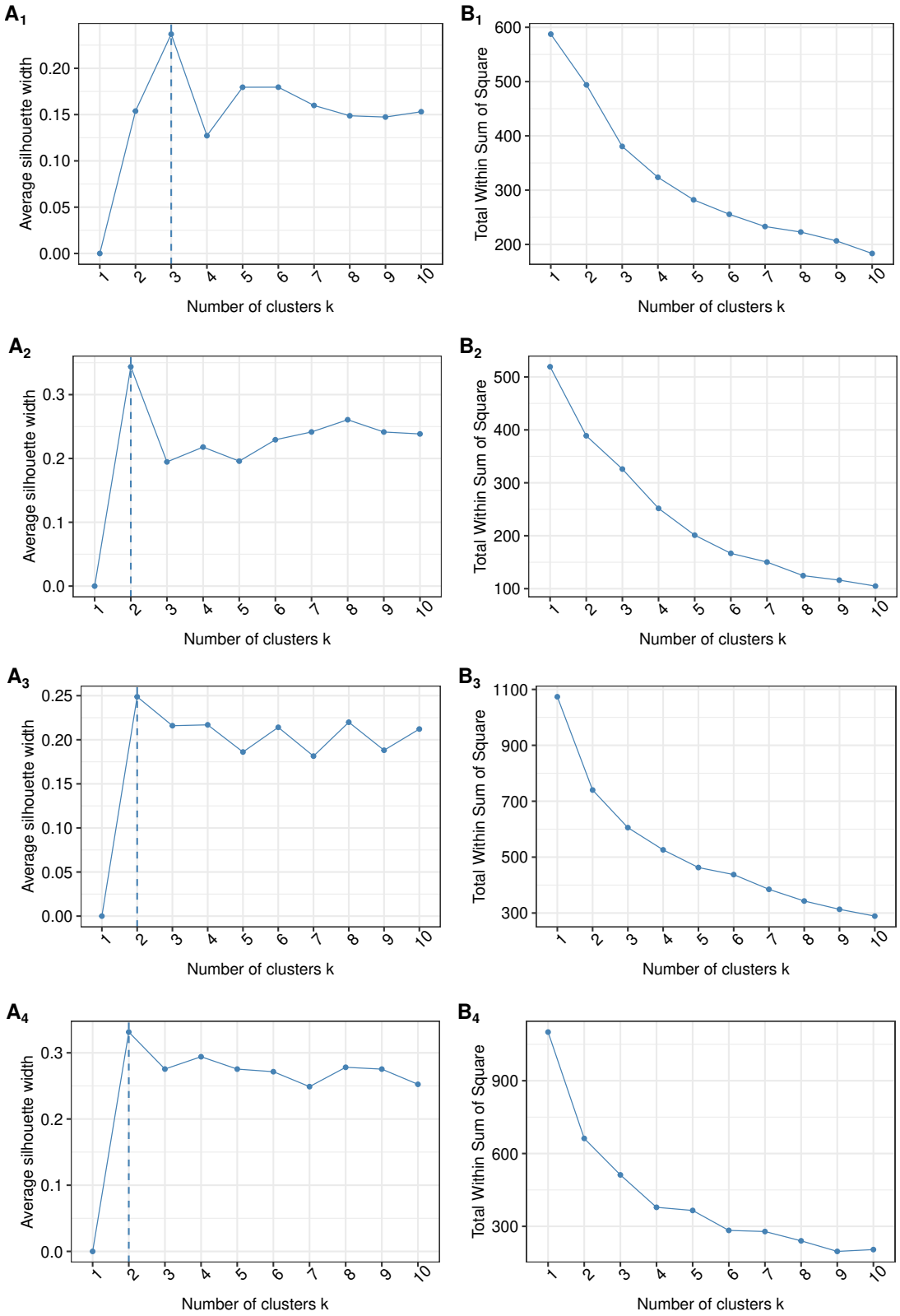

Figure C4: Graphical display of the optimal number of clusters based on the Silhouette (A) and Elbow (B) methods for the phenotypic clustering type (PHE, A_1 and B_1), soil + elevation variables (SoilE, A_2 and B_2), weather within year variables (WW, A_3 and B_3), and weather across year variables (WA, A_4 and B_4).

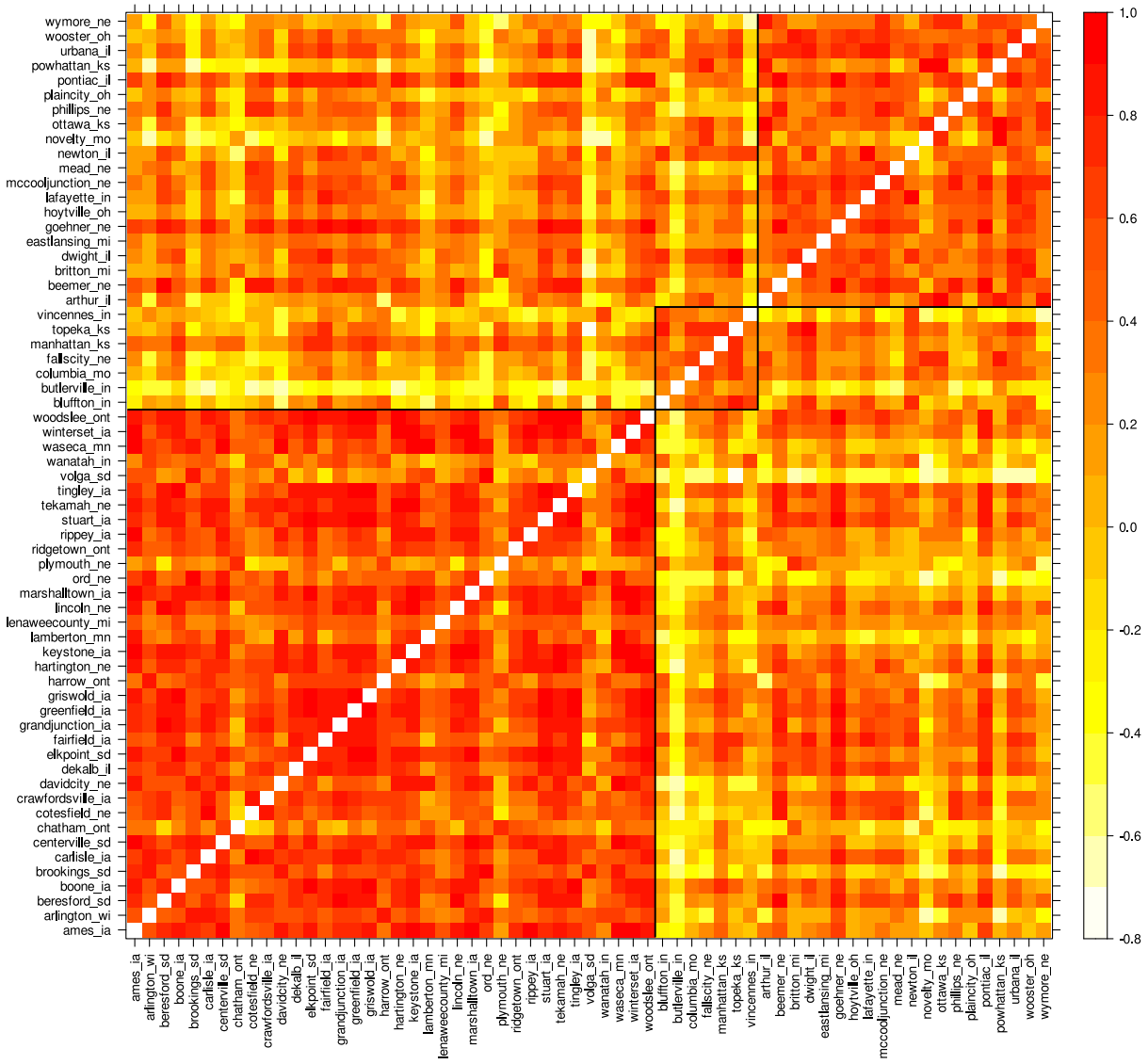

Figure C5: Heatmap of genetic correlations between the 63 observed locations from 1989 to 2019, estimated from the factor analytic (FA) model M2-18. Fitted K-means clusters are ordered from 1 to 3.

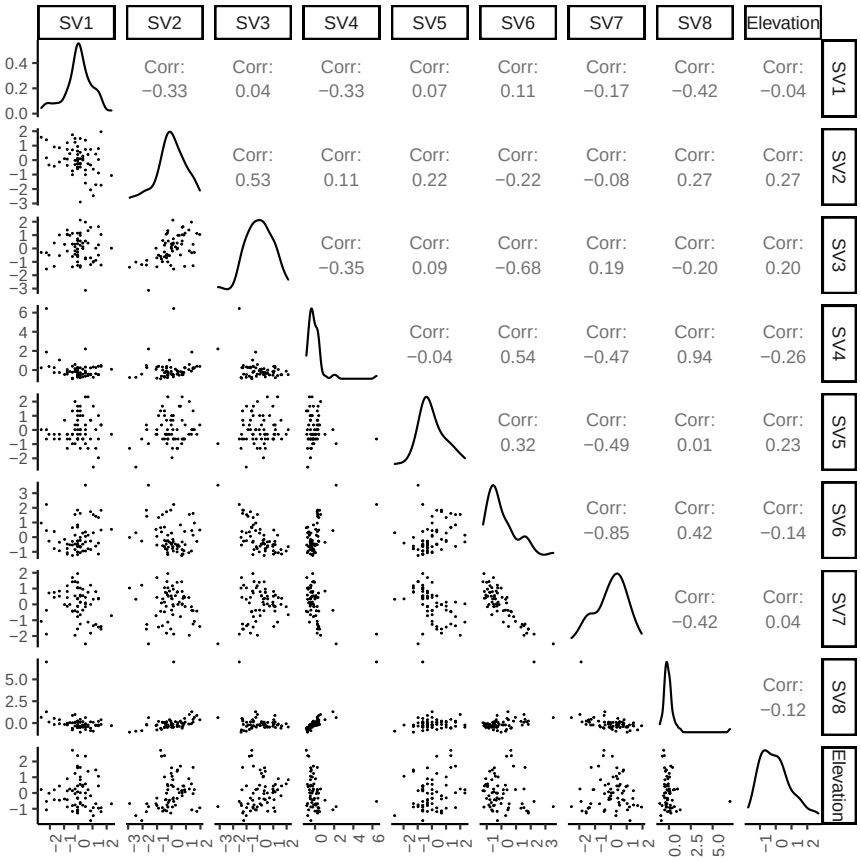

Figure C6: Scatterplots of the scaled and centered soil variables and elevation. Pearson correlation is displayed on the right. Variable distribution is available on the diagonal. The labels SV1, SV2, to SV8 are the soil variables described in section 5.3.5.

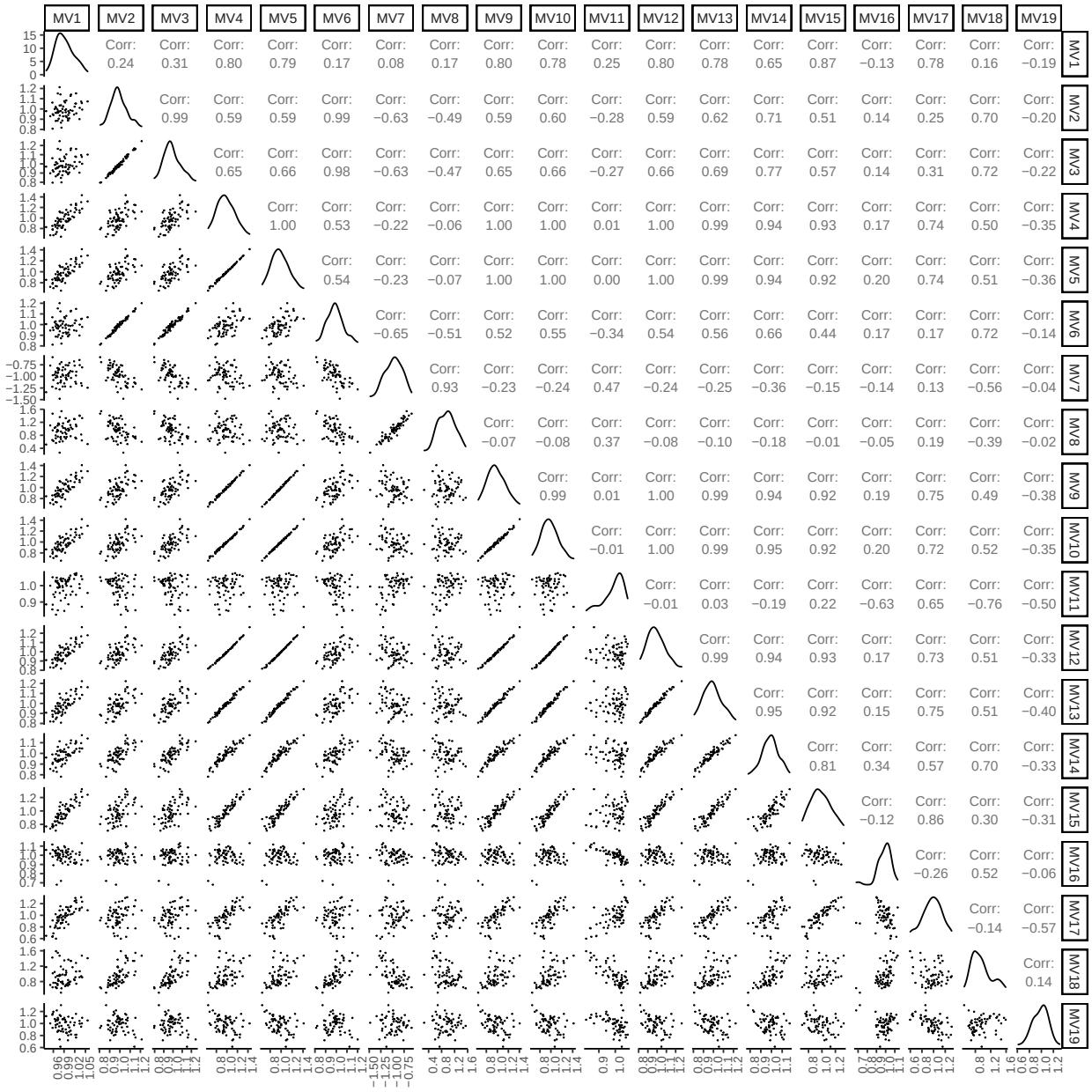

Figure C7: Scatterplots of the scaled and centered weather variables computed across years. Pearson correlation is displayed on the right. Variable distribution is available on the diagonal.The labels MV1, MV2, ..., to MV19, are the weather variables described in Table C1.

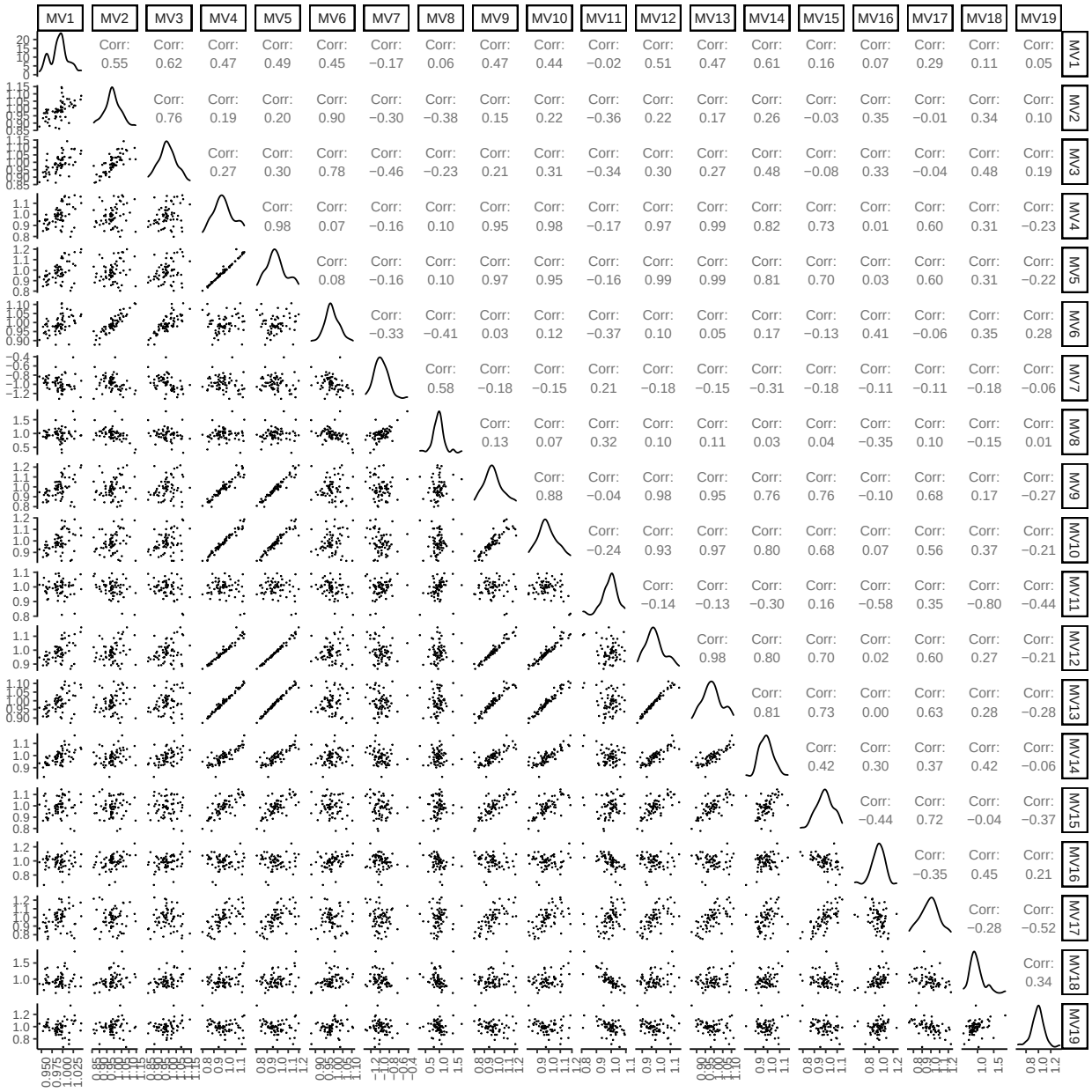

Figure C8: Scatterplots of the scaled and centered weather variables computed within years. Pearson correlation is displayed on the right. Variable distribution is available on the diagonal. The labels MV1, MV2, ..., to MV19, are the weather variables described in Table C2.

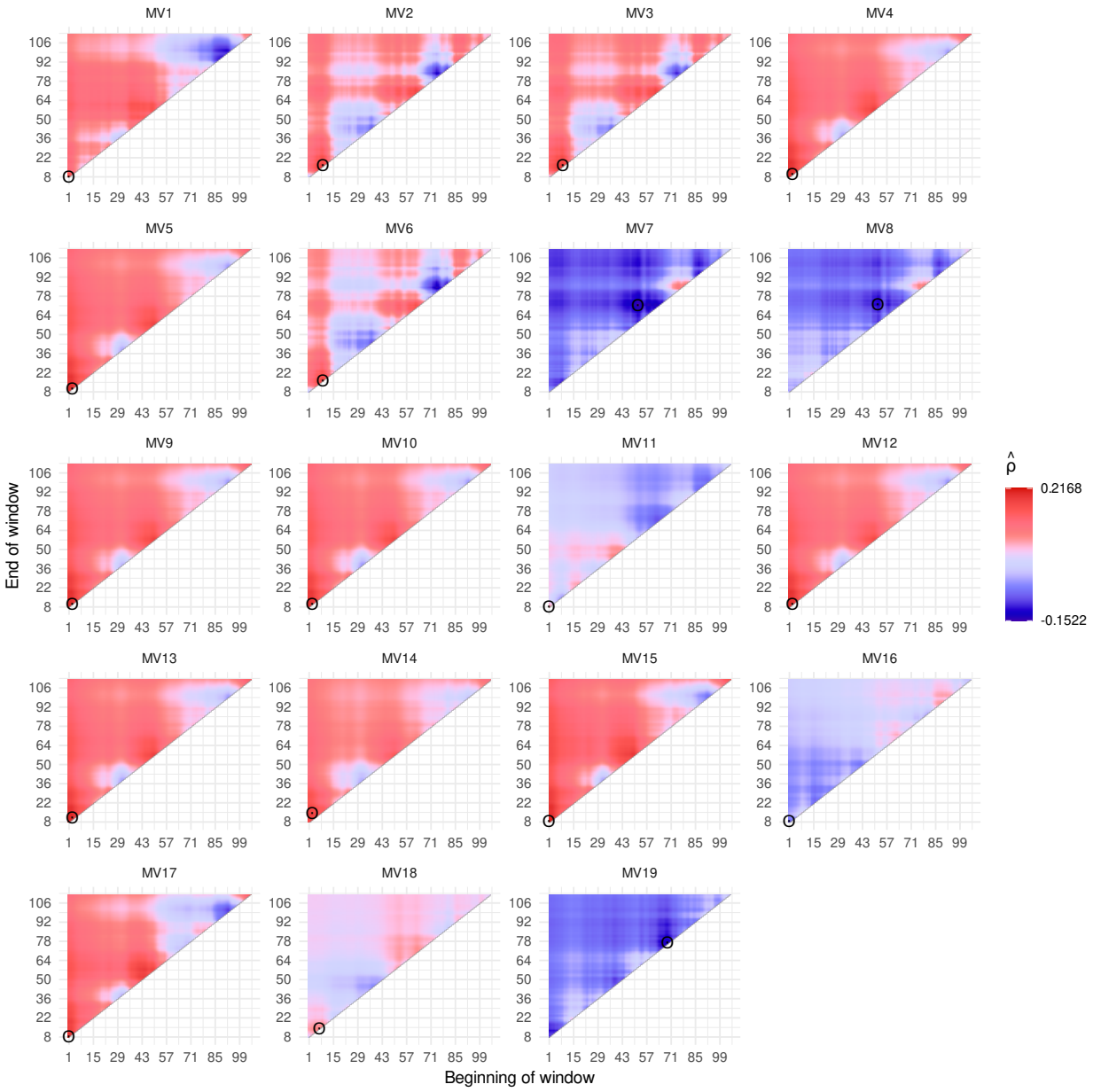

Figure C9: Exhaustive search from the critical environmental window computed from the Pearson correlation between the weather variables across years and the genotype by location deviations of each environment (location-year combination). The dots depict the highest window correlation. The labels MV1, MV2, to MV19 are the weather variables described in Table C1.

Table C5: Jaccard similarity matrix between clustering-types (phenotypic - PHE, soil and elevation - SoilE, latitude split into 2 clusters - Lat2, latitude split into 3 clusters - Lat3, weather across years - WA, and weather within years - WW).

Clusters		PHE			SoilE		Lat2		Lat3			WA		WW	
		1	2	3	1	2	1	2	1	2	3	1	2	1	2
PHE	1	1	-	-	0.25	0.44	0.18	0.64	0.02	0.53	0.27	0.20	0.54	0.12	0.60
	2		1	-	0.00	0.14	0.20	0.00	0.35	0.02	0.00	0.23	0.02	0.18	0.06
	3			1	0.13	0.30	0.45	0.07	0.33	0.22	0.03	0.25	0.23	0.30	0.21
SoilE	1				1	-	0.07	0.35	0.03	0.16	0.32	0.00	0.37	0.00	0.32
	2					1	0.62	0.28	0.30	0.52	0.09	0.51	0.38	0.39	0.48
Lat2	1						1	-	0.46	0.37	0.00	0.58	0.22	0.42	0.32
	2							1	0.00	0.36	0.39	0.06	0.61	0.07	0.53
Lat3	1								1	-	-	0.41	0.08	0.35	0.13
	2									1	-	0.22	0.51	0.20	0.51
	3										1	0.06	0.23	0.03	0.22
WA	1											1	-	0.42	0.21
	2												1	0.12	0.64

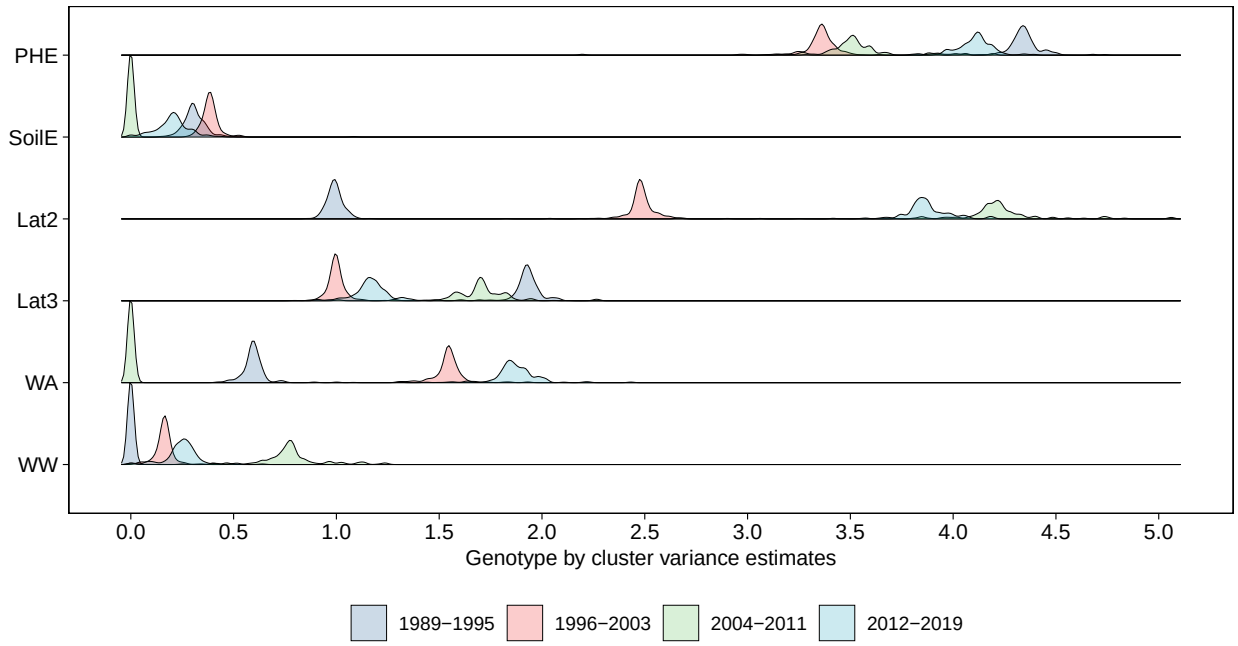

Figure C10: Jackknife estimates of genotype by cluster variances for the groups of years 1989-1995, 1996-2003, 2004-2011, and 2012-2019, for phenotypic (PHE), soil + elevation (SoilE), latitude split into two groups (Lat2), latitude split into three groups (Lat3), weather across years (WA), and weather within years (WW) clustering types.

17 Appendix D - A variation of Model 3

Here we present a variation of Model 3 in which the genotype by environment interaction is included. This model was designed to be used with CERIS, where locations within years (i.e. environments) were represented as the average of the squared residuals (i.e., environmental means). The model is the following:

$$\mathbf{y_e} = \boldsymbol{\mu} + \mathbf{X}_e\mathbf{e} + \mathbf{X}_g\mathbf{g} + \mathbf{Z}_{g.e}\mathbf{g.e} + \boldsymbol{\epsilon} \tag{3.1}$$

where $\mathbf{y_e}$ is a vector of eBLUE values in a given year (i.e. $\mathbf{y_2}$ in a given year), $\boldsymbol{\mu}$ is the intercept, $\boldsymbol{e}$ is a vector of fixed effects of environments, $\boldsymbol{g}$ is a vector of fixed effects of genotypes, $\boldsymbol{g.e}$ is a vector of random effects of genotype by environment interaction with $\boldsymbol{g.e} \sim \mathrm{N}(\mathbf{0}, \sigma^2_{GL}\mathbf{I})$, and $\boldsymbol{\epsilon}$ is a vector of residuals with $\boldsymbol{\epsilon} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_\mathbf{e})$. $\mathbf{X}_e$, $\mathbf{X}_g$, and $\mathbf{Z}_{g.e}$ are incidence matrices for their respective effects, and $\boldsymbol{\Sigma}_\mathbf{e}$ was previously defined. With Model 3.1, locations within years were represented as the average of the observed predicted genotype by location values (i.e., eBLUP values). Then, the same procedure described in Sections 5.3.5 and 5.3.5.3 was applied.

The results are presented below, and in summary, are similar to the results from Table 4, Figures 5, and C4. The Jaccard similarity (data not shown) was also similar to Table C5. The $\widehat{CR/DR}$ for WA using Model 3 was 0.92 (Table 4), which is slightly superior to the value obtained using Model 3.1 (Table D1).

Table D1: The ratio of correlated responses from selection across all environments relative to direct responses to selection within mega-environments (CR/DR) for weather variables using Model 3.1. ρ_g is the correlation between estimated genotypic effects in the non-clustered and clustered sets of environments, i^2_{L} and i^2_{SR} are the reliabilities of genotype means in the non-clustered and clustered sets of environments, respectively. Weather from means across years (WA), and weather from means within years (WW).

Clustering type[a]	Number of		Estimates of			
	Clusters	Locations	$\hat{\rho}_g$	$\hat{i}^2_{\mathrm{L}}$	$\hat{i}^2_{\mathrm{SR}}$	$\widehat{CR/DR}$
WA	2	23 / 40	0.94	0.48	0.45	0.98
WW	2	30 / 33	0.97	0.49	0.46	1.00

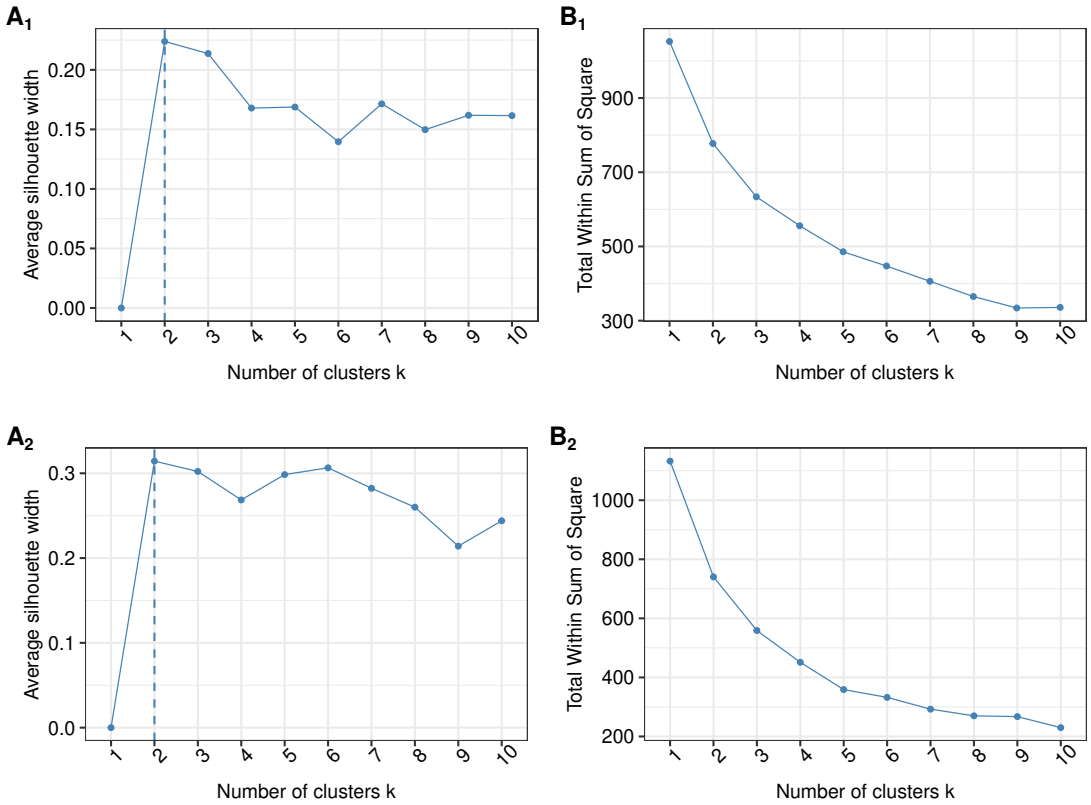

Figure D1: Graphical display of the optimal number of clusters based on the Silhouette (A) and Elbow (B) methods for weather within year variables (WW, A_1 and B_1) and weather across year variables (WA, A_2 and B_2). CERIS was implemented with Model 3.1.

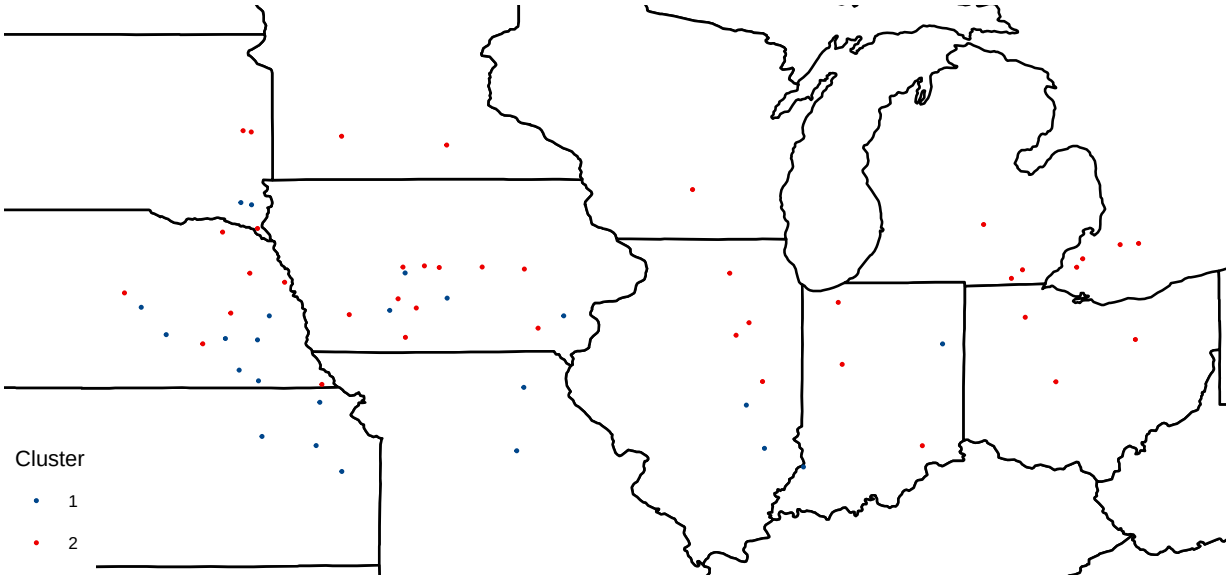

Figure D2: Geographic visualization of the target population of environments divided according to weather across years (E), computed with Model 3.1.