A Bayesian Shrinkage Approach for AMMI Models

Carlos Pereira da Silva[1☉], Luciano Antonio de Oliveira[2‡], Joel Jorge Nuvunga[1‡], Andrezza Kéllen Alves Pamplona[1‡], Marcio Balestre[1☉]*

1 Department of Exact Science (DEX), Federal University of Lavras, Lavras, Minas Gerais, Brazil, 2 Faculty of Exact Sciences and Technology (FACET), Federal University of Grande Dourados, Grande Dourados, Mato Grosso do Sul, Brazil

☉ These authors contributed equally to this work.
‡ These authors also contributed equally to this work.
* marciobalestre@dex.ufla.br

Abstract

Linear-bilinear models, especially the additive main effects and multiplicative interaction (AMMI) model, are widely applicable to genotype-by-environment interaction (GEI) studies in plant breeding programs. These models allow a parsimonious modeling of GE interactions, retaining a small number of principal components in the analysis. However, one aspect of the AMMI model that is still debated is the selection criteria for determining the number of multiplicative terms required to describe the GE interaction pattern. Shrinkage estimators have been proposed as selection criteria for the GE interaction components. In this study, a Bayesian approach was combined with the AMMI model with shrinkage estimators for the principal components. A total of 55 maize genotypes were evaluated in nine different environments using a complete blocks design with three replicates. The results show that the traditional Bayesian AMMI model produces low shrinkage of singular values but avoids the usual pitfalls in determining the credible intervals in the biplot. On the other hand, Bayesian shrinkage AMMI models have difficulty with the credible interval for model parameters, but produce stronger shrinkage of the principal components, converging to GE matrices that have more shrinkage than those obtained using mixed models. This characteristic allowed more parsimonious models to be chosen, and resulted in models being selected that were similar to those obtained by the Cornelius F-test ($\alpha = 0.05$) in traditional AMMI models and cross validation based on leave-one-out. This characteristic allowed more parsimonious models to be chosen and more GEI pattern retained on the first two components. The resulting model chosen by posterior distribution of singular value was also similar to those produced by the cross-validation approach in traditional AMMI models. Our method enables the estimation of credible interval for AMMI biplot plus the choice of AMMI model based on direct posterior distribution retaining more GEI pattern in the first components and discarding noise without Gaussian assumption as requested in F-based tests or deal with parametric problems as observed in traditional AMMI shrinkage method.

Introduction

Plant breeding programs select and recommend the best genotypes or cultivars based on yield and on their adaptability to and stability in various test environments. However, the breeder's job is hindered by genotype-by-environment interactions *(GEI)*, a phenomenon that can be detected through the responses of the genotypes when evaluated in different environments.

Therefore, studies on *GEI* are of utmost importance, and can be performed using several different methods, including linear-bilinear models [1], which offer an improved description of the effects of interaction among factors. The additive main effects and multiplicative interaction (AMMI) model is currently one of the most popular multiplicative models. The AMMI model was originally proposed by Gollob [2] and Mandel [3,4] in the context of fixed effects, but the actual statistical method itself goes back to work by Pike and Silverberg [5] and Williams [6]. It has been widely used to study *GE* interactions in plant breeding programs and in agronomic experiments in general because it incorporates univariate and multivariate procedures to separate the pattern responsible for the interaction from the noise, which is an unexplained factor. The noise, which is of no agronomic interest, is removed to increase the predictive power of the model [7].

One aspect that is still debated in the literature is the method used to select the number of multiplicative terms necessary to describe the interaction pattern. Bilinear terms of the model have been selected using several different criteria. These include approximate F-tests, notably the F-test from Gollob [2] and the F_R-test from Cornelius et al. [8], and non-parametric methods based on computationally intensive resampling, such as cross validation [7,9,10,11] and parametric bootstrapping, described by Forkman and Piepho [12].

An alternative for selecting the AMMI model multiplicative components has been described by Cornelius and Crossa [10,13]. These authors have proposed that an estimator that promotes shrinkage of the least squares estimates of the scale parameters in multiplicative models, specifically the singular values, be applied. In addition to obtaining more realistic estimates for the parameters, their method eliminates the need for any statistical tests or cross-validation methods to select components that describe the interaction pattern among factors. Additionally, the shrinkage estimators from Cornelius and Crossa [10] produced fitted models that were better than those fitted using approximate F-tests, and they were generally as good as or better than the best linear unbiased predictors (BLUPs) obtained from random models. Their approach was based on the idea that the phenotypic means obtained by the ordinary least squares (OLS) method are not the best at minimizing the mean squared error (MSE) and cannot be used in all situations [14].

Despite the above-mentioned advantages, the method proposed by Cornelius and Crossa [10] does not have a very clear derivation and may provide inconsistent results in the spectral decomposition, for example, negative singular values and magnitude inversion between high-dimensional and low-dimensional singular values. Additionally, the method proposed by these authors can only be used to predict the genotypic value of a genotype present in a specific environment, i.e., it is applied in the context of models with fixed effect parameters and, therefore, requires balanced data and homogeneous variance, which cannot always be guaranteed for data from multi-environment trials (MET). Furthermore, many conclusions about adaptability and stability in multiplicative models are obtained from biplots [15] of the genotype and environment scores that generally do not involve any measure of uncertainty or use inferential procedures prone to criticism, either from the assumptions required for parametric methods or by problematic procedures used in bootstrap resampling [16,17]. Missing data and heterogeneous variance are not large problems for factor analyses equivalent to AMMI models and sites

regression (SREG) models, proposed by Burgueño et al. [18], Piepho [19] and Smith et al.[20], that also allow the GEI to be random. However, as Crossa et al. [21] have shown, it is not clear how the confidence regions used in fixed effect models can be effectively incorporated in BLUPs in multiplicative mixed models for interaction parameters.

An alternative for complete analysis of MET data is to use a Bayesian approach for the AMMI model fit, as proposed by Crossa et al. [21] and Perez—Elisald et al. [22]. Viele and Srinivazan [23] were the first to use an inference method with the AMMI model in a Bayesian context, using Markov chain Monte Carlo (MCMC) methods. These authors showed as sampling bilinear parameters of model when the support for the joint posterior distribution was nontrivial. Liu [24] developed a set of conditional posteriors where sampling for the parameters can be performed directly using a Gibbs sampler, which resulted in a more stable algorithm. The flexibility to include credible regions in the biplot and information in the inference process was demonstrated by Crossa et al. [21], Perez—Elisald et al. [22], Forkman and Piepho [12] and Oliveira et al. [25]. According to these authors, in addition to all of these advantages, Bayesian maximum a posteriori (MAP) estimators for singular values produce shrinkage similar to those observed by Crossa et al. [21], even with uninformative priors for the truncated normal with high variance. The result is a MAP estimator analytically similar to an OLS estimator, but empirically similar to the shrinkage estimator. Therefore, an analysis that effectively explores the analytical properties of shrinkage estimators using Bayesian inference should be developed.

This study proposed singular value shrinkage estimators for AMMI models using the Bayesian shrinkage approach. A shrinkage estimator, similar to the one proposed by Cornelius and Crossa [10], was analytically justified and the results were compared with conventional Bayesian AMMI analysis.

Materials and Methods

To illustrate the proposed method, a dataset using 55 maize hybrids evaluated in nine different environments for grain yield, in $kg.ha^{-1}$, was employed. The experimental design was a complete block design with three replicates and evaluated to rows that were 3m in length during the agricultural year 2005 to 2006 (Table 1).

Table 1. Characteristics of the experimental environments.

Environment	Municipality	Latitude	Longitude	CV[1] (%)	Mean yield (t ha⁻¹)
E1	Lavras, MG	21°13′S	44°58′W	14.1	10.803
E2	Guarda-Mor, MG	17°34′S	47°08′W	13.5	6.212
E3	Barreiras, BA	12°08′S	45°00′W	12.0	4.549
E4	Jussara, GO	23°35′S	52°28′W	12.5	5.152
E5	Lavras, MG	21°12′S	44°58′W	20.8	6.246
E6	São Gotardo, MG	19°18′S	46°03′W	11.3	8.085
E7	Ijaci, MG	21°09′S	44°56′W	10.1	13.192
E8	Ijaci, MG	21°10′S	44°56′W	10.7	8.896
E9	Lavras, MG	21°10′S	45°03′W	14.6	8.737

Source: Machado et al., 2007.

[1]coefficient of variation.

doi:10.1371/journal.pone.0131414.t001

AMMI model and shrinkage effect for singular values in the context of fixed effects

The traditional AMMI model for fixed effects, with y_{ij}, i genotypes ($i = 1, 2,\ldots, g$) and j environments ($j = 1, 2,\ldots, a$), is described as follows:

$$y_{ij} = \mu + \tau_i + \beta_j + \sum_{k=1}^{t} \lambda_k \alpha_{ik} \gamma_{jk} + \in_{ij} \tag{1}$$

where μ is the overall mean of the experiment, τ_i is the effect of the ith genotype, β_j is the effect of the jth environment, λ_k is the kth singular value of the multiplicative component, α_{ik} is the ith element of the kth genotypic singular vector, γ_{jk} is the jth element of the kth environment singular vector, e_{ij} is the error associated with the ith genotype of the jth environment, and $t \leq \min(g\text{-}1, a\text{-}1)$.

This model also has the restrictions of identifiability ($\sum_i \tau_i = \sum_j \beta_j = \sum_i (\tau\beta)_{ij} = \sum_j (\tau\beta)_{ij} = 0$), orthonormality of the singular vectors ($\sum_i \alpha_{ik}^2 = \sum_i \gamma_{jk}^2 = 1$, $\sum_i \alpha_{ik} \alpha_{ik'} = \sum_j \gamma_{jk} \gamma_{jk'} = 0, k' \neq k$) and that λ_k should satisfy the condition $\lambda_k \geq \lambda_{k+1}$, for any $k > 0$.

Shrinkage effect for fixed effect models. Cornelius and Crossa [10] proposed that a shrinkage estimator be used for the bilinear terms to study *GEI* in multiplicative models, notably in AMMI models. The shrinkage estimator assumes that each empirical bilinear least squares component is the "true" value of the effect plus a disturbance (η_{ijk}), i.e., $\hat{\lambda}_k \hat{\alpha}_{ik} \hat{\gamma}_{jk} = \lambda_k \alpha_{ik} \gamma_{jk} + \eta_{ijk}$, such that $\lambda_k \alpha_{ik} \gamma_{jk}$ and η_{ijk} are not correlated, and $\lambda_{k'} \alpha_{ik'} \gamma_{jk'}$ and η_{ijk} are not correlated, for $k \neq k'$. The shrinkage estimator $\sum_k S_k \hat{\lambda}_k \hat{\alpha}_{ik} \hat{\gamma}_{jk}$ estimates the contribution of $\Sigma_k \lambda_k \alpha_{ik} \gamma_{jk}$ to the mean of the observed plot μ_{ij}, and the average quadratic estimation error per plot for the AMMI model is minimized if $S_k = E(\lambda_k^2)/E(\hat{\lambda}_k^2)$. According to these authors the $E(\lambda_k^2)$ means the expectation of estimated lambda on the true value. Defining u_k such that $E(\sum_i \sum_j \eta_{ijk}^2) = u_k \sigma^2/n$, where σ^2 is the residual variance and n the sample size yields $E(\sum_i \sum_j \eta_{ijk}^2) = E(\hat{\lambda}_k^2) - E(\lambda_k^2)$ and, consequently, $S_k = 1 - \frac{u_k \sigma^2/n}{E(\hat{\lambda}_k^2)}$, where $\frac{u_k \sigma^2/n}{E(\hat{\lambda}_k^2)}$ is similar to an approximate F^{-1}-test. Furthermore, $S_k = max(1 - F_k^{-1}, 0)$, the shrinkage factor in this method, can be negative.

The authors showed that, occasionally, $S_k \hat{\lambda}_{k-1} > S_{k-1} \hat{\lambda}_{k-1}$, which violates the model restriction $\lambda_k \leq \lambda_{k\text{-}1}$. To overcome this problem, it is assumed that $\hat{\lambda}_k = \hat{\lambda}_{k-1}$ and a joint estimate was calculated. The rule was applied sequentially; it also yielded joint estimates for sequences of more than two singular values.

Although it violates several critical restrictions of the analysis, models fitted by this method are more parsimonious than those based on the least squares method, and, in general, the shrinkage estimates were as good as or better than the predictions made by random models. This method is based on minimizing the MSE and follows the principle proposed by Stein [26]. Despite all of the advantages of this method, it is naturally subject to the limitations of models conceived in the context of fixed effects.

Bayesian shrinkage AMMI model

The AMMI model described at Eq 1 is related to a two-way table. In order to expand the AMMI effect to plot level at any environment we can to describe the model 1 in matrix form,

with confounded block and environment effects represented as follows:

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\delta} + \sum_{k=1}^{t} \lambda_k diag(\mathbf{Z}\boldsymbol{\alpha}_k)\mathbf{X}_2\boldsymbol{\gamma}_k + \boldsymbol{\epsilon} \qquad (2)$$

where $\mathbf{y}$ is the vector of observations with dimension $n\times1$, $\mathbf{X}_1$ is the design matrix with dimension $n\times a$ associated with the block plus environments effects vector $\boldsymbol{\beta}$, $\mathbf{Z}$ is a design matrix with dimension $n\times g$ associated with the genotypic parameter vector δ with dimension $g\times1$, λ_k is the kth singular value under the conditions mentioned previously, $\boldsymbol{\alpha}_k$ and $\boldsymbol{\gamma}_k$ are the kth eigenvectors for the genotype and the environment, respectively, and ϵ is the vector of the residuals with multivariate normal probability distribution with mean of zero and variance and covariance matrix $\mathbf{I}\sigma_e^2$. The matrix representation of inner product $diag(\mathbf{Z}\boldsymbol{\alpha}_k)\mathbf{X}_2\boldsymbol{\gamma}_k$ is used in order to account for GEI in data level where $\mathbf{X}_2$ account only for environments design. More detail can be obtained in the R program available in S1 Text—line 63).

Prior distribution for the model parameters. A prior distribution represents the previous knowledge about a parameter of interest. This prior knowledge can be informative or uninformative, depending on the currently available knowledge about the phenomenon or on the amount of information needed for the model, which are set a priori using hyperparameter values.

Therefore, determining the prior distribution for the effects of the singular values defines the Bayesian shrinkage model. In this case, the prior distributions for AMMI model (2) assigned to the parameters were as follows:

$$p(\boldsymbol{\beta}) = 1/\sigma = k(constant)$$

$$\boldsymbol{\delta}|\mu_\delta, \sigma_\delta^2 \sim N(\mathbf{0}, \mathbf{I}\sigma_\delta^2)$$

$$\lambda_k|\mu_{\lambda_k}, \sigma_{\lambda_k}^2 \sim N^+(0, \sigma_{\lambda_k}^2)$$

Where $\boldsymbol{\beta}$ is the fixed effects $\mu_\delta and \sigma_\delta^2$ are mean and variance hypothesis parameters related to effect of genotypes, $\mu_{\lambda_k}, and \sigma_{\lambda_k}^2$ are mean and variance hypothesis parameters related to kth singular value and N^+ means a normal truncated. The non informative priors for singular vector related to genotypes and environments are respectively given by:

$\boldsymbol{\alpha}_k\sim$ uniform spherical distribution in the correct subspace *(SUNI$_d$)*

$\boldsymbol{\gamma}_k\sim$ uniform spherical distribution in the correct subspace *(SUNI$_d$)*.

$\sigma_\delta^2 \sim Inv - scaled - \chi^2(v_\delta, S_\delta^2) \propto \frac{1}{\sigma_\delta^2}$ for $v_\delta = 0$ and $S_\delta^2 = 0$

$\sigma_e^2 \sim Inv - scaled - \chi^2(v_e, S_e^2) \propto \frac{1}{\sigma_e^2}$ for $v_e = 0$ and $S_e^2 = 0$

where *Inv-scaled-χ^2* is the inverted-scaled chi-square distribution with v degree of freedom and S^2 scale parameter. The priors for $\boldsymbol{\beta}$ σ_δ^2 and σ_e^2 are equivalent to Jeffrey priors that can be obtained by the expected Fisher Information of the likelihood distribution (given below). For the singular vectors $\boldsymbol{\alpha}_k$ and $\boldsymbol{\gamma}_k$, a uniform spherical distribution prior, which are uninformative and special cases of the von Mises-Fisher distribution [27], were used. These vectors are distributed in a restricted space in $\Re^p$ orthogonal to t-1 vectors in the space of dimension p ($p = g$ or $p = a$).

Given that $\sigma_{\lambda_k}^2$ is an unknown parameter in $p(\lambda_k|\mu_{\lambda_k}, \sigma_{\lambda_k}^2)$, it must be estimated, the modeling of this effect is included in the joint prior. The prior distribution for this parameter is the same

as that described for σ_e^2 and σ_δ^2, i.e., an uninformative prior was used with the property $p(\sigma_{\lambda_k}^2) = (\sigma_{\lambda_k}^2)^{-1}$.

Analytically, this prior resulted in an improper marginal posterior distribution, also verified by Ter Braak et al. [28] in shrinkage models with specific variances. To correct this problem, the extended prior presented by Ter Braak et al. [28] was adopted, correcting the degrees of freedom to obtain a proper posterior distribution. Therefore, the prior for the specific variance of λ_k is as follows: Assuming $S_{\lambda_k}^2 = 0$ and $\upsilon_{\lambda_k} = n_{\lambda_k} - 1 = -2\Delta$ we have:

$$p(\sigma_{\lambda_k}^2) \propto (\sigma_{\lambda_k}^2)^{-\frac{(n_{\lambda_k}-1)}{2}-1} exp\left\{\frac{S_{\lambda_k}^2}{2\sigma_{\lambda_k}^2}\right\}$$

$$p(\sigma_{\lambda_k}^2) \propto (\sigma_{\lambda_k}^2)^{-\frac{(-2\Delta)}{2}-1} exp\left\{\frac{S_{\lambda_k}^2}{2\sigma_{\lambda_k}^2}\right\} \tag{3}$$

$$p(\sigma_{\lambda_k}^2) \propto (\sigma_{\lambda_k}^2)^{\Delta-1}$$

where $\Delta = -\frac{(n_{\lambda_k}-1)}{2}$, such as $0 < n_{\lambda_k} < 1$ and, consequently, $0 < \Delta < 1/2$. Here n_{λ_k} is a arbitrary value. In this study we assumed $n_{\lambda_k} = 0.95$. Therefore, for $\sigma_{\lambda_k}^2$, $k = 1, 2, \ldots, 8$, the scaled inverse chi-square prior is represented by $\upsilon_{\lambda_k} = -2\Delta$ and $S_{\lambda_k}^2 = 0$ as follows:

$$\sigma_{\lambda_k}^2 \sim Inv - scaled - \chi^2(\upsilon_{\lambda_k}, S_{\lambda_k}^2) \rightarrow (\sigma_{\lambda_k}^2)^{(\Delta-1)}.$$

Therefore, the joint prior distribution is as follows:

$$p(\boldsymbol{\theta}) = p(\boldsymbol{\beta}|\mu_\beta, \sigma_\beta^2)p(\boldsymbol{\delta}|\mu_\delta, \sigma_\delta^2)p(\sigma_\delta^2|\upsilon_\delta, S_\delta^2)p(\sigma_e^2|\upsilon_e, S_e^2)\times$$

$$\prod_k^t [p(\lambda_k|\mu_{\lambda_k}, \sigma_{\lambda_k}^2)p(\sigma_{\lambda_k}^2|\upsilon_{\lambda_k}, S_{\lambda_k}^2)p(\boldsymbol{\alpha}_k)p(\boldsymbol{\gamma}_k)]$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\delta}, \lambda_k, \boldsymbol{\alpha}_k, \boldsymbol{\gamma}_k, \sigma_e^2, \sigma_\delta^2, \sigma_{\lambda_k}^2)$.

Likelihood and full conditional posteriors for the AMMI model parameters. The posterior probability distribution is a combination of the likelihood function (information from the data) and the prior probability distributions. The likelihood function for the AMMI model was implemented as follows:

$$L(\bar{\theta}, \sigma_e^2|\mathbf{y}) = (2\pi|\mathbf{I}\sigma_e^2|)^{-\frac{n}{2}} exp\left\{-\frac{1}{2\sigma_e^2}(\mathbf{y} - \mathbf{X}_1\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{\delta} - \Theta)'(\mathbf{y} - \mathbf{X}_1\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{\delta} - \Theta)\right\} \tag{4}$$

where $\Theta = \sum_{k=1}^t \lambda_k diag(\mathbf{Z}\boldsymbol{\alpha}_k)\mathbf{X}_2\boldsymbol{\gamma}_k$ and $\bar{\theta} = \{\boldsymbol{\beta}, \boldsymbol{\delta}, \lambda_k\boldsymbol{\alpha}_k\boldsymbol{\gamma}_k\}$.

Applying Bayes' theorem gives the full posterior probability distribution as follows:

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto L(\bar{\theta}|\mathbf{y})p(\boldsymbol{\beta})p(\boldsymbol{\delta}|\mu_\delta, \sigma_\delta^2)p(\sigma_\delta^2|\upsilon_\delta, S_\delta^2)p(\sigma_e^2|\upsilon_e, S_e^2)\times$$

$$\prod_k^t [p(\lambda_k|\mu_{\lambda_k}, \sigma_{\lambda_k}^2)p(\sigma_{\lambda_k}^2|\upsilon_{\lambda_k}, S_{\lambda_k}^2)p(\boldsymbol{\alpha}_k)p(\boldsymbol{\gamma}_k)] \tag{5}$$

The full conditional distributions are as follows:

$$p(\boldsymbol{\beta}|\ldots) \propto exp\left\{-\frac{1}{2\sigma_e^2}[\mathbf{A}-(\mathbf{X'}_1\mathbf{X}_1)\mathbf{X'}_1\boldsymbol{\beta}]'(\mathbf{X'}_1\mathbf{X}_1)[\mathbf{A}-(\mathbf{X'}_1\mathbf{X}_1)\mathbf{X'}_1\boldsymbol{\beta}]\right\}$$

$$p(\boldsymbol{\beta}|\ldots) \sim N[(\mathbf{X'}_1\mathbf{X}_1)^{-1}\mathbf{X'}_1\mathbf{A}, (\mathbf{X'}_1\mathbf{X}_1)^{-1}\sigma_e^2] \tag{6}$$

where $\mathbf{A} = \mathbf{y}\text{-}\mathbf{Z}\boldsymbol{\delta}\text{-}\boldsymbol{\Theta}$ and $p(\ |\ldots)$ is the conditional of all of the other parameters of the model.

$$p(\boldsymbol{\delta}|\ldots) \propto exp\left\{-\frac{1}{2\sigma_e^2}\left[\boldsymbol{\delta}-\left(\mathbf{Z'Z}+\mathbf{I}\frac{\sigma_e^2}{\sigma_\delta^2}\right)\mathbf{Z'B}\right]'\left(\mathbf{Z'Z}+\mathbf{I}\frac{\sigma_e^2}{\sigma_\delta^2}\mathbf{Z'B}\right)^{-1}\left[\boldsymbol{\delta}-\left(\mathbf{Z'Z}+\mathbf{I}\frac{\sigma_e^2}{\sigma_\delta^2}\right)\mathbf{Z'B}\right]\right\}$$

$$p(\boldsymbol{\delta}|\ldots) \sim N\left[\left(\mathbf{Z'Z}+\mathbf{I}\frac{\sigma_e^2}{\sigma_\delta^2}\right)^{-1}\mathbf{B}, \left(\mathbf{Z'Z}+I\frac{\sigma_e^2}{\sigma_\delta^2}\right)^{-1}\sigma_e^2\right] \tag{7}$$

where $\mathbf{B} = \mathbf{y}\text{–}\mathbf{X}_1\boldsymbol{\beta}\text{–}\boldsymbol{\Theta}$.

$$p(\sigma_e^2|\ldots) \propto (\sigma_e^2)^{-\left(\frac{n}{2}+1\right)}exp-\frac{1}{2\sigma_e^2}(\mathbf{y}-\boldsymbol{\theta})'(\mathbf{y}-\boldsymbol{\theta})\} \quad p(\sigma_e^2|\ldots) \sim Inv-scaled$$

$$-\chi^2(n,[(\mathbf{y}-\boldsymbol{\theta})'(\mathbf{y}-\boldsymbol{\theta})]/n) \tag{8}$$

$$p(\sigma_\delta^2|\ldots) \propto (\sigma_\delta^2)^{-\left(\frac{n_\delta}{2}+1\right)}exp-\frac{1}{2\sigma_\delta^2}(\boldsymbol{\delta}'\boldsymbol{\delta})\} \quad \sigma_\delta^2|\ldots \sim Inv-scaled-\chi^2(n_\delta,(\boldsymbol{\delta}'\boldsymbol{\delta})/n_\delta) \tag{9}$$

$$p(\lambda_k|\ldots) \propto exp\left\{\frac{1}{2\sigma_e^2}\left[(\mathbf{C}-\phi\lambda_k)'(\mathbf{C}-\phi\lambda_k)+(\lambda_k-\mu\lambda_k)'\frac{\sigma_e^2}{\sigma_{\lambda_k}^2}(\lambda_k-\mu_{\lambda_k})\right]\right\}$$

$$p(\lambda_k|\ldots) \sim N^+\left[\left(\phi'\phi+\frac{\sigma_e^2}{\sigma_{\lambda_k}^2}\right)^{-1}\phi'\mathbf{C}, \left(\phi'\phi+\frac{\sigma_e^2}{\sigma_{\lambda_k}^2}\right)^{-1}\sigma_e^2\right] \tag{10}$$

where $\mathbf{C} = \mathbf{y}\text{-}\mathbf{X}\boldsymbol{\beta}\text{-}\mathbf{Z}\boldsymbol{\delta}\text{-}\mathbf{D}$ and $\phi = diag(\mathbf{Z}\boldsymbol{\alpha}_k)\mathbf{X}_2\gamma_k$, for $\lambda_1 \geq \lambda_2,\ldots,\lambda_t \geq 0$

In this formulation, the posterior mean of the singular value is similar to a ridge regression estimator. The posterior mean obtained by Crossa et al. [21] and Oliveira et al. [25] is similar to $(\phi'\phi)^{-1}\phi'\mathbf{C}$, referencing the least squares estimator. The distribution of the variance of the singular values is as follows:

$$p(\sigma_{\lambda_k}^2|\ldots) \propto \prod_{k=1}^{t}(\sigma_{\lambda_k}^2)^{\frac{-1}{2}}exp\left\{-\frac{1}{2\sigma_{\lambda_k}^2}(\lambda_k^2)\right\}(\sigma_{\lambda_k}^2)^{(-1+\Delta)}$$

$$p(\sigma_{\lambda_k}^2|\ldots) \propto \prod_{k=1}^{t}(\sigma_{\lambda_k}^2)^{\left[-\frac{(1-2\Delta)}{2}-1\right]}exp\left\{-\frac{(\lambda_k^2)}{2\sigma_{\lambda_k}^2}\frac{(1-2\Delta)}{(1-2\Delta)}\right\} \tag{11}$$

$$p(\sigma_{\lambda_k}^2|\ldots) \propto \prod_{k=1}^{t}(\sigma_{\lambda_k}^2)^{\left[-\frac{(1-2\Delta)}{2}-1\right]}exp\left\{-\frac{(1-2\Delta)}{2\sigma_{\lambda_k}^2}\frac{(\lambda_k^2)}{(1-2\Delta)}\right\}$$

$$\sigma_{\lambda_k}^2|\ldots \sim Inv-scaled-\chi^2(1-2\Delta,[\lambda_k^2/(1-2\Delta)])$$

The posterior distribution for the singular vector are given by:

$$p(\boldsymbol{\alpha}_k|\ldots) \propto exp\left\{\frac{\lambda_k}{2\sigma_e^2}[\boldsymbol{\alpha}_{k'} diag(\mathbf{X}_2\boldsymbol{\gamma}_k)'\mathbf{Z}'](\mathbf{y} - \mathbf{X}_1\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{\delta} - \mathbf{D})\right\}$$

$$p(\boldsymbol{\alpha}_k|\ldots) \sim VMF\left\{\frac{\lambda_k}{2\sigma_e^2}, (diag(\mathbf{X}_2\boldsymbol{\gamma}_k)'\mathbf{Z}')(\mathbf{y} - \mathbf{X}_1\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{\delta} - \mathbf{D})\right\}$$

(12)

$$p(\boldsymbol{\gamma}_k|\ldots) \propto exp\left\{\frac{\lambda_k}{2\sigma_g^2}[\gamma'_k diag(\mathbf{Z}\boldsymbol{\alpha}_k)'\mathbf{X}'_2](\mathbf{y} - \mathbf{X}_1\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{\delta} - \mathbf{D})\right\}$$

$$p(\boldsymbol{\gamma}_k|\ldots) \sim VMF\left\{\frac{\lambda_k}{2\sigma_g^2}, [diag\mathbf{Z}(\boldsymbol{\alpha}_k)'\mathbf{X}'_2](\mathbf{y} - \mathbf{X}_1\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{\delta} - \mathbf{D})\right\}$$

(13)

where VMF is a Von Mises—Fisher distribution

Because the problem is orthogonally restricted, the singular vectors $\boldsymbol{\alpha}_k$ and $\boldsymbol{\gamma}_k$ are sampled by linear transformation [21,25]. The sampling is conducted in the corrected subspace, where there are no restrictions. The values are then returned to the correct space in $\Re^p$, orthogonal to t-1, in the dimension space p ($p = a$ or $p = g$, according to the vector).

Because the full conditional posterior distributions for the model parameters are known, direct sampling can be performed. The Gibbs sampler was used in the MCMC method. Chain convergence was assessed using the methods introduced by Raftery and Lewis [29] and Heidelberger and Welch [30].

To compare the Bayesian shrinkage AMMI method (AMMIBS) with traditional methods, four approaches were used. In the first approach, a classic AMMI analysis was performed using the analysis of variance (ANOVA). The number of components was selected using the F-test by Gollob [2], $F_G = \frac{\lambda_k^2/df_{(G)}}{S^2}$, and the F-test by Cornelius et al.

[31], $F_R = \left(SQ(GE) - \sum_{k=1}^{t} \lambda_k^2\right)/df_R S^2$, where S^2 is the joint mean squared error, $SQ(GE)$ is the sum of the squares of the interaction, and $df_{(G)} = g+a-1-2k$ and $df_R = (g-1-p)(a-1-p)$ are the corresponding degrees of freedom. Both F-based tests were applied using a $\alpha = 0.05$

In the second approach, a singular value decomposition of the BLUPs matrix for the GE interaction was performed using a mixed model with the genotype and interaction effects modeled as random. This is the same model represented by Eq 2, except the singular value decomposition is applied directly in the GE BLUPs using the model as follows:

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\delta} + \mathbf{W}\mathbf{i} + \boldsymbol{\epsilon}$$

(14)

where $\mathbf{w}$ is the matrix of the random effect for GE interaction in classical mixed models. The estimates of the components of variance were obtained using expectation/maximization-restricted maximum likelihood (EM-REML) estimators. Vector $\mathbf{i}$, which corresponds to the interactions, was transformed to the GE matrix, which was then used for the decomposition GE = ULV, where U and V are the singular vectors matrices for the genotypes and the environments, respectively, and L is the diagonal matrix of the singular values.

In the third approach, the AMMI analysis was subjected to the shrinkage estimator by Cornelius and Crossa [10] previously described. In the fourth approach, the Bayesian AMMI analysis (AMMIB) was performed as described in Crossa et al. [21] and Oliveira et al. [25].

Criterion to select Bayesian Shrinkage AMMI model. The use of a additional criterion to selection components could be not an important issue in Bayesian AMMI since all information about the parameter is available on its posterior distribution. Thus, if we use the *ad hoc*

threshold proposed by Cornelius and Crossa [10] for AMMI shrinkage we can to select the AMMI model based on the hypothesis H_0: $\lambda_k \leq 1$. Therefore, if the $P_{\lambda = 1} \leq 0.05$, where $P_{\lambda = 1}$ is the percentile related to $\lambda = 1$ in the $p(\lambda_k | ...)$ posterior distribution, we can retain the kth component or discard otherwise. Since this criterion uses a *ad hoc* value we propose others measures in order to exploit the advantages of shrinkage.

In this study we used an additional test based on Bayes Factor as a comparative measure where the choice of AMMI model (AMMI0, AMMI1...AMMI$_k$) will be depend of the evidence $e \geq 12.79db$ where $e = 10^*\log_{10}(BF)$ and

$$BF = \left[\frac{p(\theta_f | y)}{p(\theta_{f-i} | y)} \right]$$

where θ_f é a full model using p principal components and $p(\theta_{f-i})$ is the reduced model with $p-1$ principal components where $i = 1,2,3....p$. The test was performed initially considering the score for high dimensional model and when $e \geq 12.79db$ the test was stopped [32]. This criterion consider a 0.95 of probability or 19:1 of chance to accept full model over the reduced one.

Additional comparison was accomplished using different cross-validation approach based on leave-one-out method proposed by [11] and [33]. Dias and Krzanowski [11] method is based on Prediction Sum Square (PRESS) criterion corrected by the number of degree of freedom retained on the $AMMI_k$ and the number of degree of freedom remained after adjusting the *kth* component. This is justified since the PRESS in this method is monotonically decreasing and W statistics became necessary. The Gabriel [33] method uses only the PRESS criterion since its cross-validation approach present a "Ockham's profile" where the "pattern" and "noise" can be found by "Ockham's hill" in a plot using a Statistical efficiency criterion given by $SE = PRESS_{full}/PRESS_k$ where full related to raw data and $PRESS_k$ to the prediction sum square related to kth component. The Gabriel method was used in both Bayesian and classical AMMI. However, in cross-validation in AMMI Bayesian methods we prefer to use only the inner product $\lambda_k \alpha_{ik} \gamma_{jk}$ and calculating the PRESS in order to gain computational efficiency and avoiding arbitrary choice of signal in singular vectors.

Model parameter inference

The estimates for σ_e^2, σ_δ^2, $\sigma_{\lambda_k}^2 \lambda_k$, $\boldsymbol{\delta}$, and $\boldsymbol{\beta}$ were obtained using the MCMC samples means. The sample means for the singular vectors do not satisfy the restriction of orthonormality in the model. Therefore, the estimates for $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ were obtained by orthonormalization of the matrices of the means $\bar{\boldsymbol{\alpha}}$ and $\bar{\boldsymbol{\gamma}}$, respectively, performed using the method proposed by Liu [24].

The univariate highest credible posterior density regions for the parameters were constructed using the method proposed by Chen and Shao [34], and implemented in the R statistical software [35].

The bivariate credible regions for the genotypic scores $(\boldsymbol{\alpha}_1 \sqrt{\lambda_1}, \boldsymbol{\alpha}_2 \sqrt{\lambda_2})$ and the environment scores $(\boldsymbol{\gamma}_1 \sqrt{\lambda_1}, \boldsymbol{\gamma}_2 \sqrt{\lambda_2})$ were constructed using the method in Ooms [36] and Oliveira et al. [25].

Results

MCMC chains were simulated using 188,000 iterations for each parameter of the AMMIBS and AMMIB models. For each chain, the first 8,000 observations were discarded (*burn-in*) and values were saved at every 20th observation (*thinning*), resulting in a sample size of 9,000. Convergence of the chains was evaluated using the criteria from Raftery and Lewis [29] and Heidelberger and Welch [30]. The results of the tests showed that all of the parameters had good
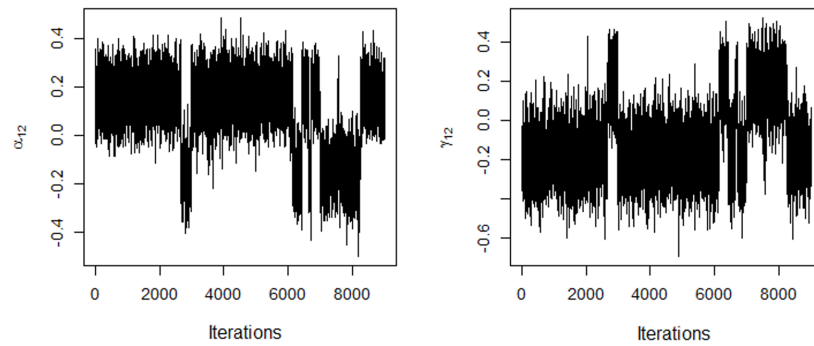
Fig 1. Traces of the chains for the first two coordinates of the genotype and environment singular vectors using the AMMIB model.

convergence properties with a dependence factor that was always less than 5 (I<5) and that they all passed the stationarity test, indicating that convergence was achieved.

In the AMMIB model, the chains for the coordinates of the singular vectors (starting with the second genotype and environment vectors) converged to two solutions, equal in absolute value, similar to the results found by Oliveira et al. [25], with uninformative priors for the singular values. Fig 1 shows the traces for the first coordinates of the singular vectors associated with λ_2 in the AMMIB model. The data show a change of the periodicity in the convergence that separates the chains into sign lags. This behavior was not apparent in the AMMIBS model (Fig 2). In both cases, the choice of sign is arbitrary.

The property of arbitrariness of the singular vectors sign is a characteristic of the principal component analysis. This characteristic ensures that the chains satisfy $E[\alpha_{ik}] = 0$ and $E[\gamma_{jk}] = 0$ for $k \geq 2$. One of the solutions (negative or positive) must be chosen; the choice does not affect the interpretation of the biplot nor the inner product. When the solution is chosen, the other terms adjust automatically. The solution with a positive sign for the first coordinate of each genotype singular vector was chosen. The signs of the other coordinates and the coordinates of the corresponding environment vector are then determined. This choice produced absolute singular values similar to those in Fig 1. Fig 3 shows the traces for the first coordinates of the second genotype and environment singular vectors, respectively, for the AMMIBS analysis. By choosing the positive solution, α_{12} and γ_{12} have opposite signs (see Fig 1).

This figure shows the negative association between genotype 1 and environment 1. Regardless of the sign chosen, α_{12} and γ_{12} have opposite signs (Fig 1).
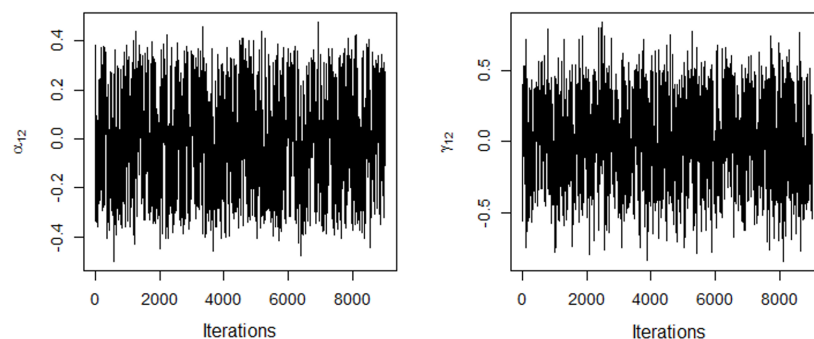

Fig 2. Traces of the chains for the first two coordinates of the genotype and environment singular vectors using the AMMIBS model, without restricting the solutions.
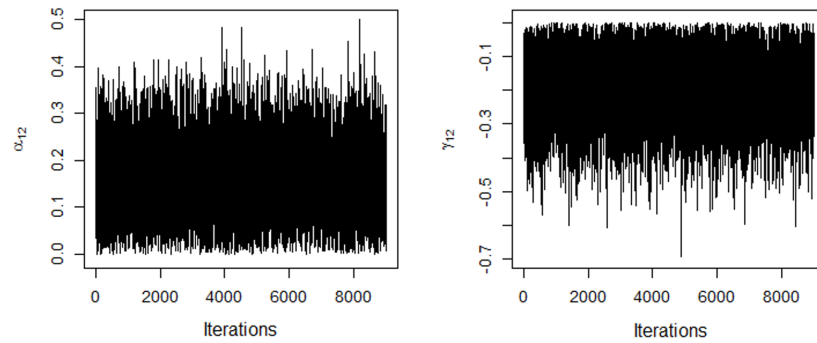
Fig 3. Traces of the chains for the first two coordinates of the genotype and environment singular vectors using the AMMIBS model, with restriction of sign for solutions.

The posterior densities of the singular values for the AMMIBS model are shown in Figs 4 and 5. Starting in the third singular value, the chains converge to near zero, which shows a larger shrinkage when a specific variance is used in the truncated normal. Utilizing the prior for specific variance of the singular values means that the conditional posterior distribution for the second singular value is bimodal. This agrees with Ter Braak's findings [37], when specific variances are assumed that have shrinkage effects on the parameters.

Because of this feature, the posterior mean for the second singular value was slightly shrunk, but the posterior mean was near zero for the others higher-dimensional singular values. On the other hand, the posterior of the second singular value presented values two modes being one of them lower than 1 (Cornelius and Crossa *ad hoc* criterion for classical AMMI shrinkage).

The posterior densities of the components of variance for the singular values are shown in Figs 6 and 7. Only the first two components have high means and variances, while the other components have values near zero with low variance, resulting in shrinkage of the estimates of the singular values (Figs 4 and 5). The shrinkage effect on the estimates of the singular values are directly associated with the components of the variance, as shown in Eq 14. The mean values for these components are shown in Table 2 with their corresponding 95% credible intervals.

Point estimates and credible regions for the other model parameters

The means of the components of variance, σ_δ^2 and σ_e^2, were 0.373 and 1.285 with a 95% highest posterior density (HPD) interval of [0.2262, 0.5426] and [1.1680, 1.406], respectively. The 95% HPD intervals for the primary genotype effects are shown in Fig 8 in increasing order. This ranking allows the genotypes that contribute most to the population mean to be identified, i.e., those that have only positive values in the HPD region (located to the right of genotype G17 in the Fig 8).

To make more accurate conclusions concerning the selection and recommendation of the best genotypes, the effects of the GE interaction must also be considered by analyzing the terms that describe it.

The posterior means for the components of the variance for the singular values that are responsible for shrinkage effects are shown in Table 2. These estimates are calculated using the AMMIBS model. The estimates of the singular values using the AMMIBS model, the AMMIB model, the BLUPs, the OLS and the shrunken estimates (*Shrink.*), proposed by Cornelius and Crossa [10,13], are shown in Table 3.
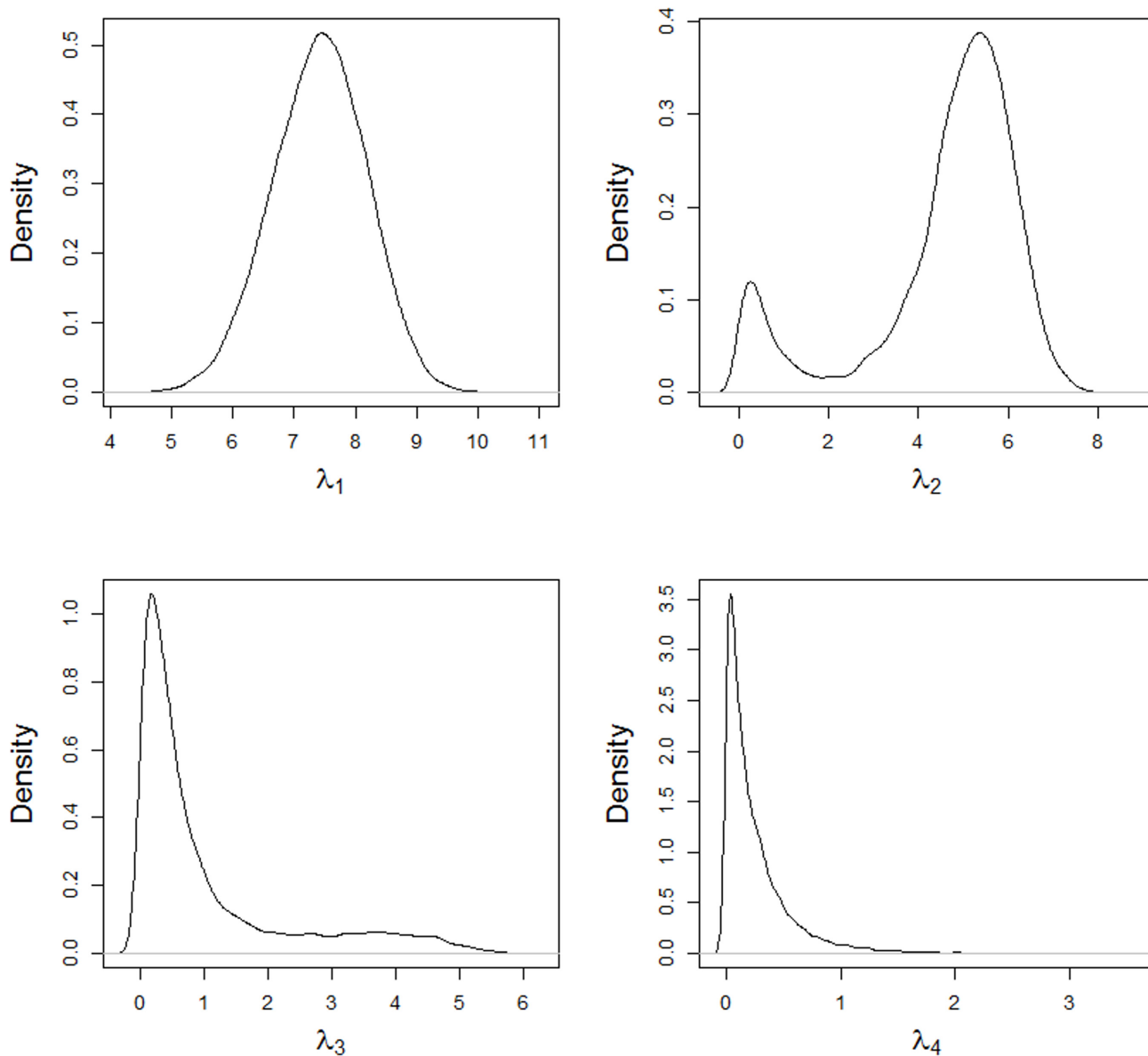
Fig 4. Posterior densities of the singular values λ_1, λ_2, λ_3 and λ_4, obtained using the AMMIBS model.

The data in Table 3 show that the singular values in the AMMIB model from Crossa et al. [21] are slightly shrunken compared to the classical AMMI with GE, estimated by OLS or the mixed models (EBLUP). In the AMMIB model, the first five singular values have means above zero, but, by Bayesian Factor criterion AMMI4 model be chosen. This result is very similar to the values obtained by shrinkage AMMI from Cornelius and Crossa [10], and also by the Gollob F-test, which selected only five principal components. This result suggests that AMMIB tends to select models similar to the Cornelius and Crossa [10] shrinkage estimator, and these results are consistent with the test performed on the traditional AMMI model using the Gollob F-test.

The data in Table 3 also show that the AMMIBS model has a more pronounced shrinkage effect than the AMMIB model or the Cornelius shrinkage model. Specifically, the AMMIBS
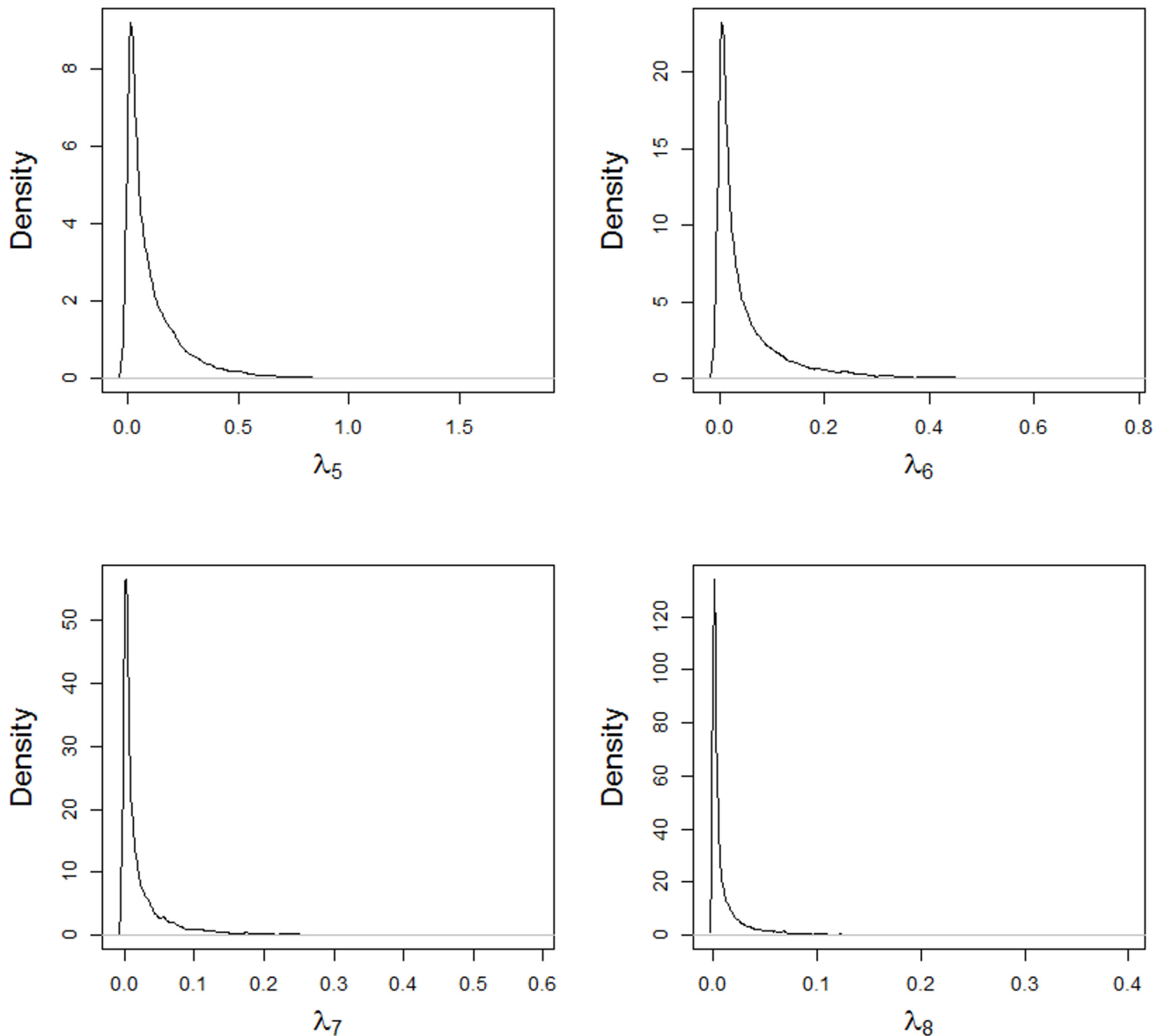
Fig 5. Posterior densities of the singular values λ_5, λ_6, λ_7 and λ_8, obtained using the AMMIBS model.

model captured 89.11% of the interactions explained by the model (not necessarily the raw interactions), while the AMMIB model captured 61.34% in the first two axes. The first two singular values from both models were very similar, suggesting that the AMMIBS model shrinks only the low-magnitude singular values, preserving the estimates of the first ones, capturing most of the interaction pattern than noise. The Bayesian shrinkage method proved to be the most parsimonious of the all methods, capturing most of the interactions with the lowest number of multiplicative terms to explain the interaction pattern.

The estimates of the first singular value from the AMMIBS, the AMMIB and the EBLUP models were similar to one another, but different from those obtained with the traditional AMMI and Cornelius estimators. Based on this result, the GE interaction matrix, obtained from the inner product of these methods, showed a pronounced shrinkage and, therefore, was
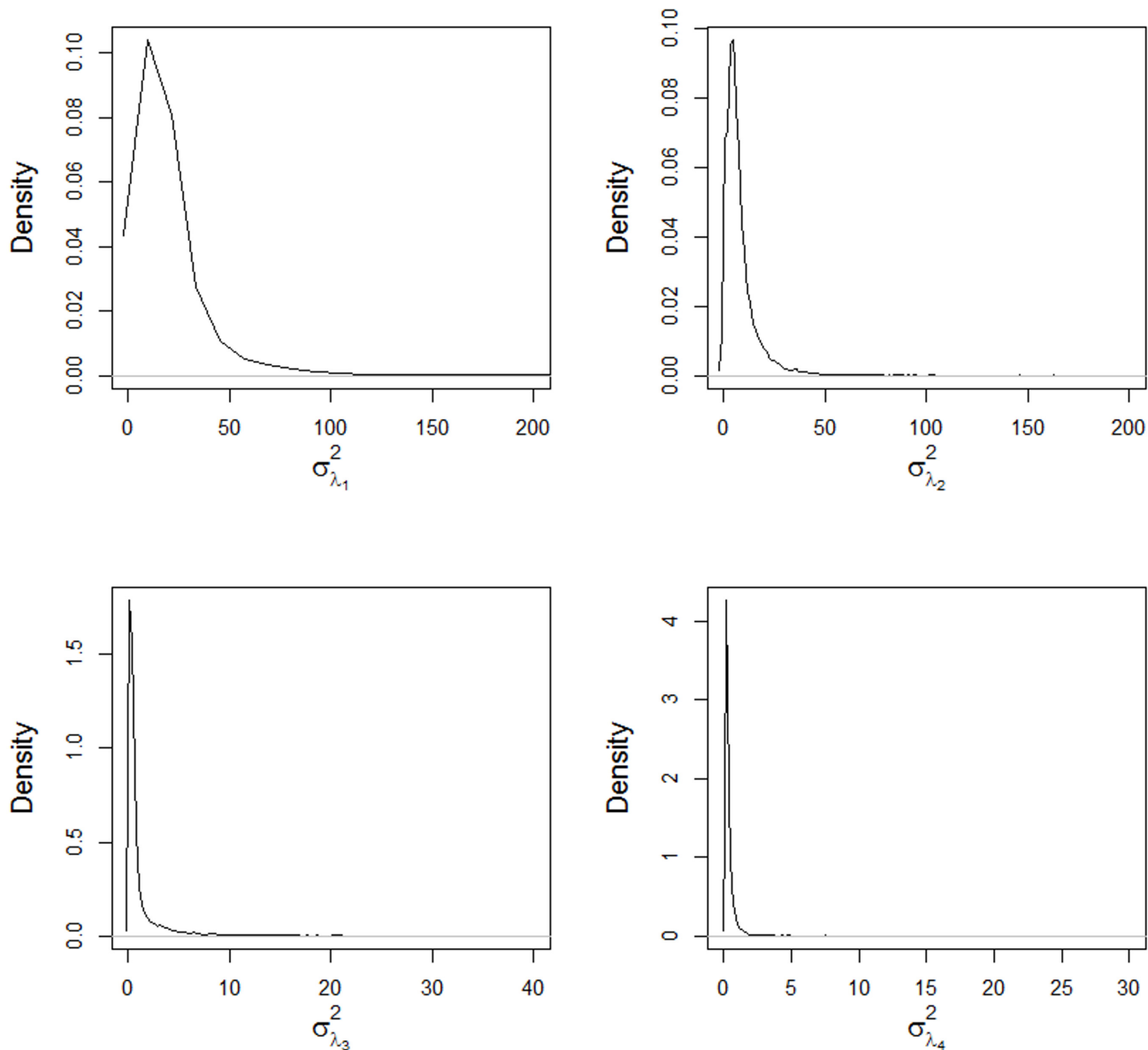
Fig 6. Posterior densities of the variance components for the singular values $\sigma^2_{\lambda_1}$, $\sigma^2_{\lambda_2}$, $\sigma^2_{\lambda_3}$ **and** $\sigma^2_{\lambda_4}$, **obtained using the AMMIBS model.**

doi:10.1371/journal.pone.0131414.g006

similar to modeling the interaction as random in a mixed model. The AMMIB model estimates were also more similar to the AMMI model with EBLUP than to the AMMIBS model.

The model selected by the singular value posterior distribution in AMMBS was AMMI1. It was evident in Fig 3, since the threshold suggested by Cornelius and Crossa presented $p_{\lambda = 1} \leq$ 0.05 To be more exact, all values obtained in MCMC for the first singular value was higher than 1 for the first singular value. This results agree with Dias and Krzanowski [11] criterion where the best model was AMMI1 (Table 3). On the other hand, the BF criterion applied in Bayesian analysis suggests AMMI2 for AMMIBS and AMMI4 for AMMIB. The result for AMMIBS is similar to than obtained using Cornelius F-based test. Using classical AMMI analysis (OLS), the selected models were AMMI1, AMMI2 for Dias and Krzanowski [11] criterion
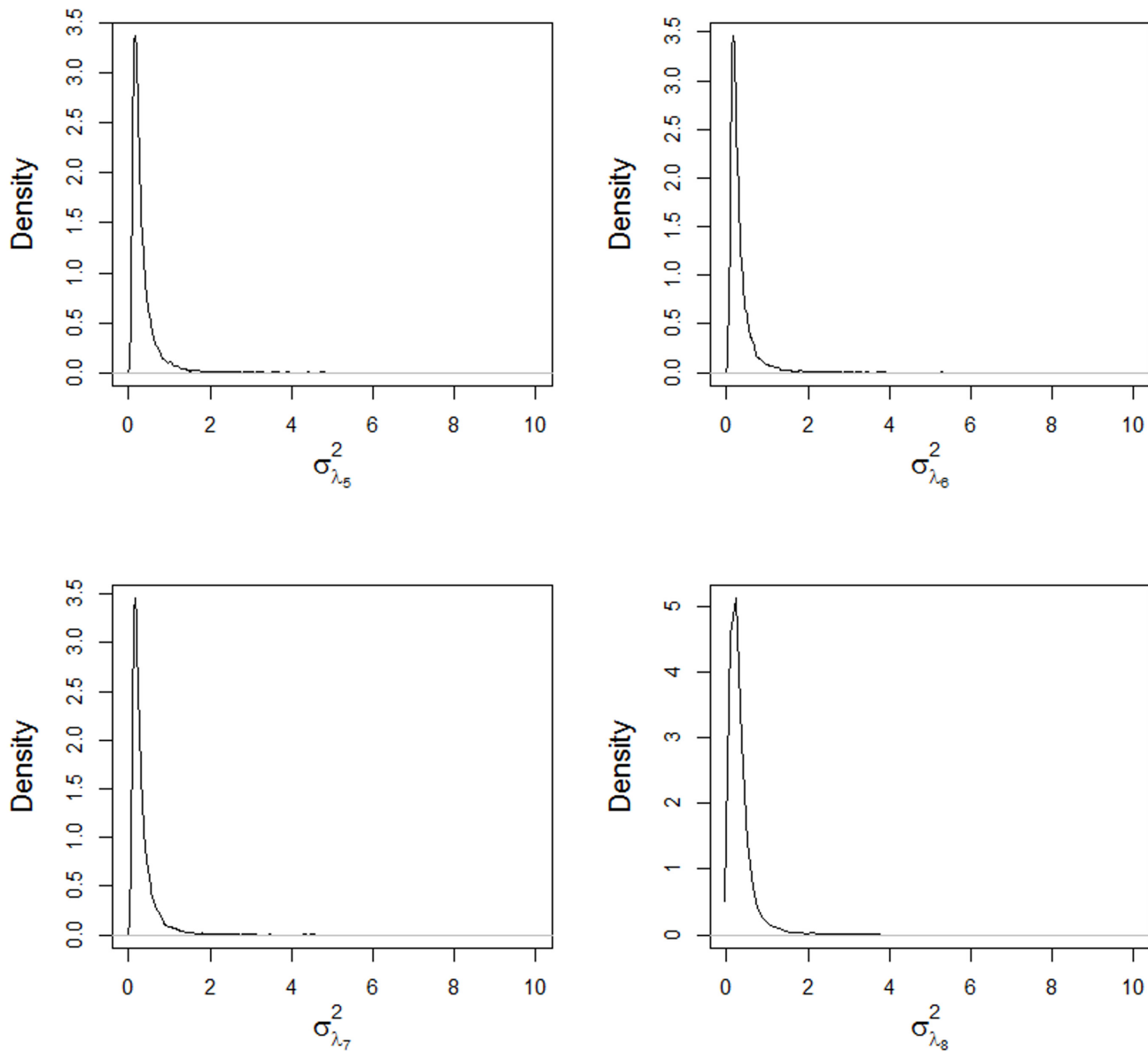
Fig 7. Posterior densities of the variance components for the singular values $\sigma^2_{\lambda_5}$, $\sigma^2_{\lambda_6}$, $\sigma^2_{\lambda_7}$ and $\sigma^2_{\lambda_8}$, obtained using the AMMIBS model.

doi:10.1371/journal.pone.0131414.g007

Table 2. Posterior mean estimates and the 95% credible interval (CI) for the variance components of the singular values.

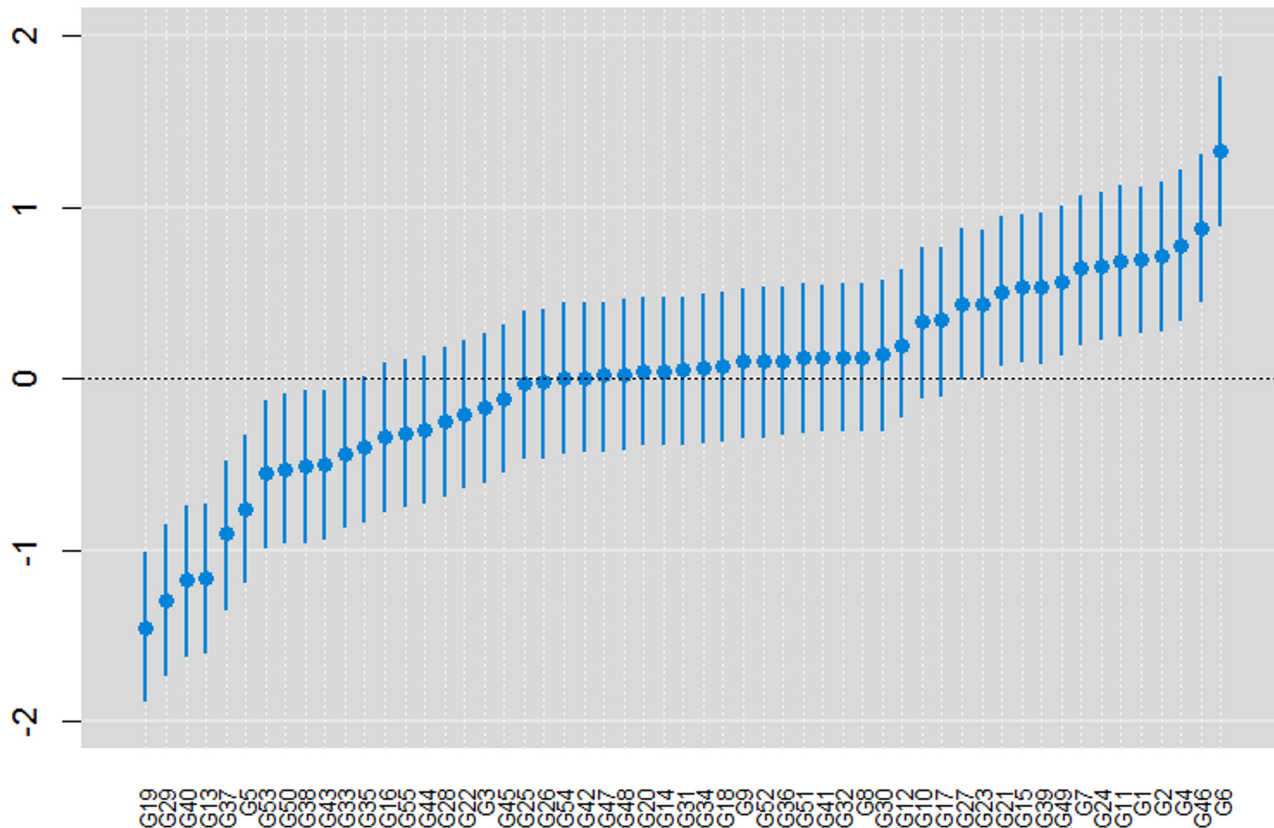| Parameter	Mean	Sd	Lim. Inf.	Lim. Sup.
$\sigma^2_{\lambda_1}$	19.524	68.699	1.977	51.057
$\sigma^2_{\lambda_2}$	8.226	12.467	0.046	24.110
$\sigma^2_{\lambda_3}$	1.232	3.451	0.052	5.062
$\sigma^2_{\lambda_4}$	0.402	1.056	0.050	1.047
$\sigma^2_{\lambda_5}$	0.342	0.392	0.049	0.927
$\sigma^2_{\lambda_6}$	0.332	0.379	0.049	0.879
$\sigma^2_{\lambda_7}$	0.334	0.367	0.049	0.890
$\sigma^2_{\lambda_8}$	0.339	0.822	0.046	0.887

doi:10.1371/journal.pone.0131414.t002

Fig 8. Posterior means and the 95% HPD intervals for the genotype effects.

doi:10.1371/journal.pone.0131414.g008

and Cornelius F-based test respectively and AMMI5 for F-Gollob and Cornelius and Crossa shrinkage analysis.

In general the select models across different analysis were (AMMI1, AMMI2, AMMI4 and AMMI5). These divergent results can be explained when we analyze the Fig 9. In this figure it

Table 3. Estimates of the singular values λ_k for the AMMI model using the AMMIBS model, the AMMIB model and the AMMI model with fixed and random GE (OLS and BLUPs) and the AMMI shrinkage developed by Cornelius.

Parameter	AMMIBS	AMMIB	BLUPs	OLS	Shrink.
λ_1	7.411[§]	7.902[§]	7.027	9.559[*+€]	6.652
λ_2	4.589[§]	5.978[§]	5.887	7.580[*+]	6.442
λ_3	1.015	4.605[§]	5.278	6.156[+]	6.442
λ_4	0.254	2.268[§]	4.309	6.007[+]	6.442
λ_5	0.108	1.058	4.161	5.674[+]	2.783
λ_6	0.051	0.478	3.857	3.920	0
λ_7	0.025	0.213	2.766	3.306	0
λ_8	0.013	0.109	2.379	<0.000	0

*Significant by the Cornelius F-test at a 0.05 significance level

[+]Significant by the Gollob F-test at a 0.05 significance level.

[€]Number of terms retained on the AMMI model based on leave-one-out cross validation using W's score by Dias and Krzanowski.

[§] Model selection based on Bayes Factor.
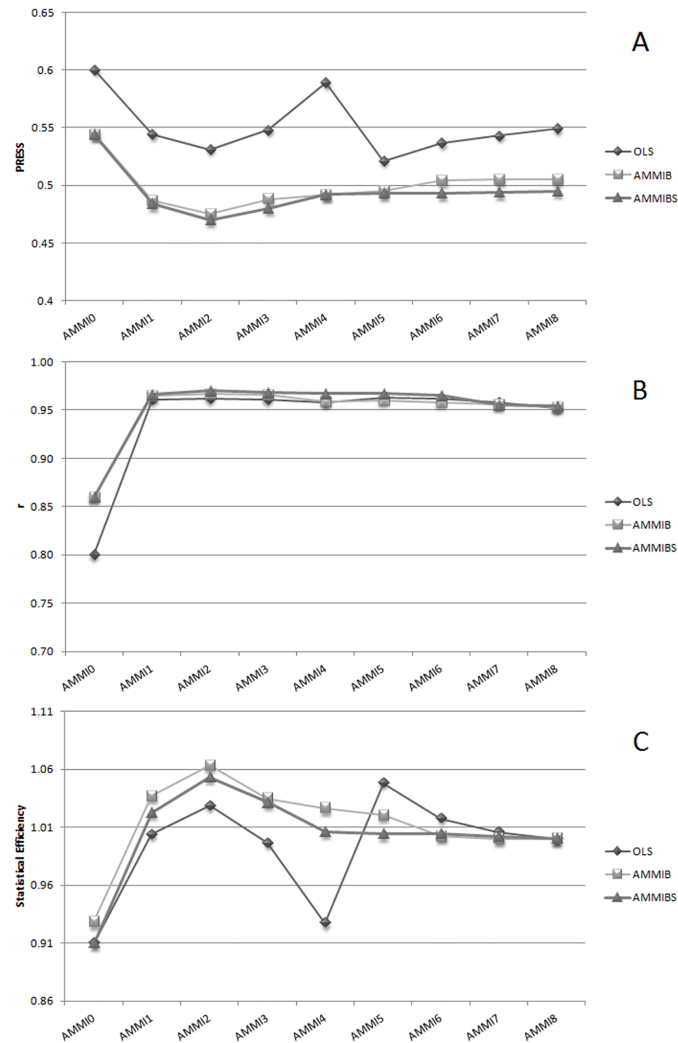
doi:10.1371/journal.pone.0131414.t003

Fig 9. Ockham's plot related to classical AMMI model (OLS), Bayesian AMMI (AMMIB) and Bayesian Shrinkage AMMI (AMMIBS). (A) AMMI vs prediction sum square- PRESS (lowest is best—Ockham's valley); (B) AMMI vs correlation (r) between "true" and predicted cell values in GEI matrix; (C) AMMI vs Statistical Efficiency (higher is best—Ockham's hill).

doi:10.1371/journal.pone.0131414.g009

is evident that AMMI2 and AMMI5 presented two peak in the plot for statistical efficiency; where in OLS analysis the "Ockham's hill" was observed in AMMI5. But this model retain two times more degree of freedom than AMMI2. In both Bayesian analysis AMMI5 was marginal and the "Ockham's hill" was allocated at AMMI2 suggesting Bayesian methods the best approach to retain pattern and discarding noise using lesser degree of freedom. The predictive accuracy based on correlation for all models were equivalent showing marginal gain starting from AMMI1.

Bivariate credible regions for the genotype and the environment scores

Bivariate credible regions were constructed for the AMMI2 biplot, similar to those presented by Oliveira et al. [25]. Genotypes or environments whose credible regions for the scores in the biplot include the origin are considered stable. In addition, large overlaps between the credible regions indicate that the respective genotypes or environments have similar GE interactions.
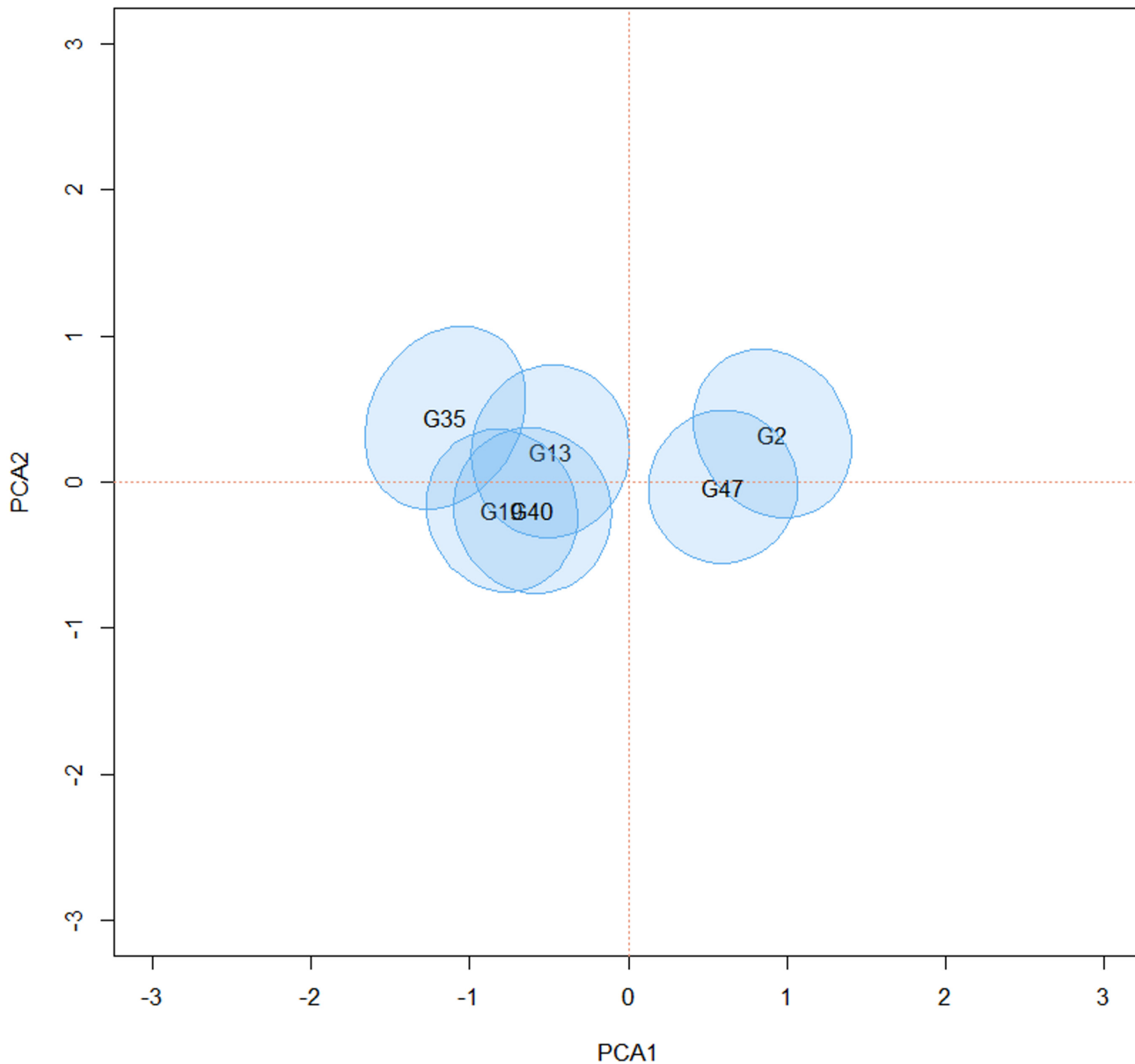
Fig 10. 95% credible regions that do not include the origin for the first two principal axes of the genotype scores using the AMMIB analysis.

The Figs 10 and 11 show the same groups of genotypes for the AMMIB and the AMMIBS analyses, respectively. The figures show that the grouping pattern is retained, even with the truncation of solutions of the second singular vectors in the AMMIBS analysis.

The genotypes shown in Fig 11 have nonzero values for the interactions (unstable genotypes), and two homogeneous groups can be formed, i.e., one group comprising G2 and G47 and the other group comprising G13, G19, G35 and G40. These two groups are responsible for many of the significant contribution to the GE interaction. Only the regions that do not include the origin are shown in the figure for simplicity and ease of interpretation. The bivariate credible regions in this biplot were not regular in the second axis. This distortion was from the truncation in the choice of the sign that, together with the shrinkage model, prevented the correct
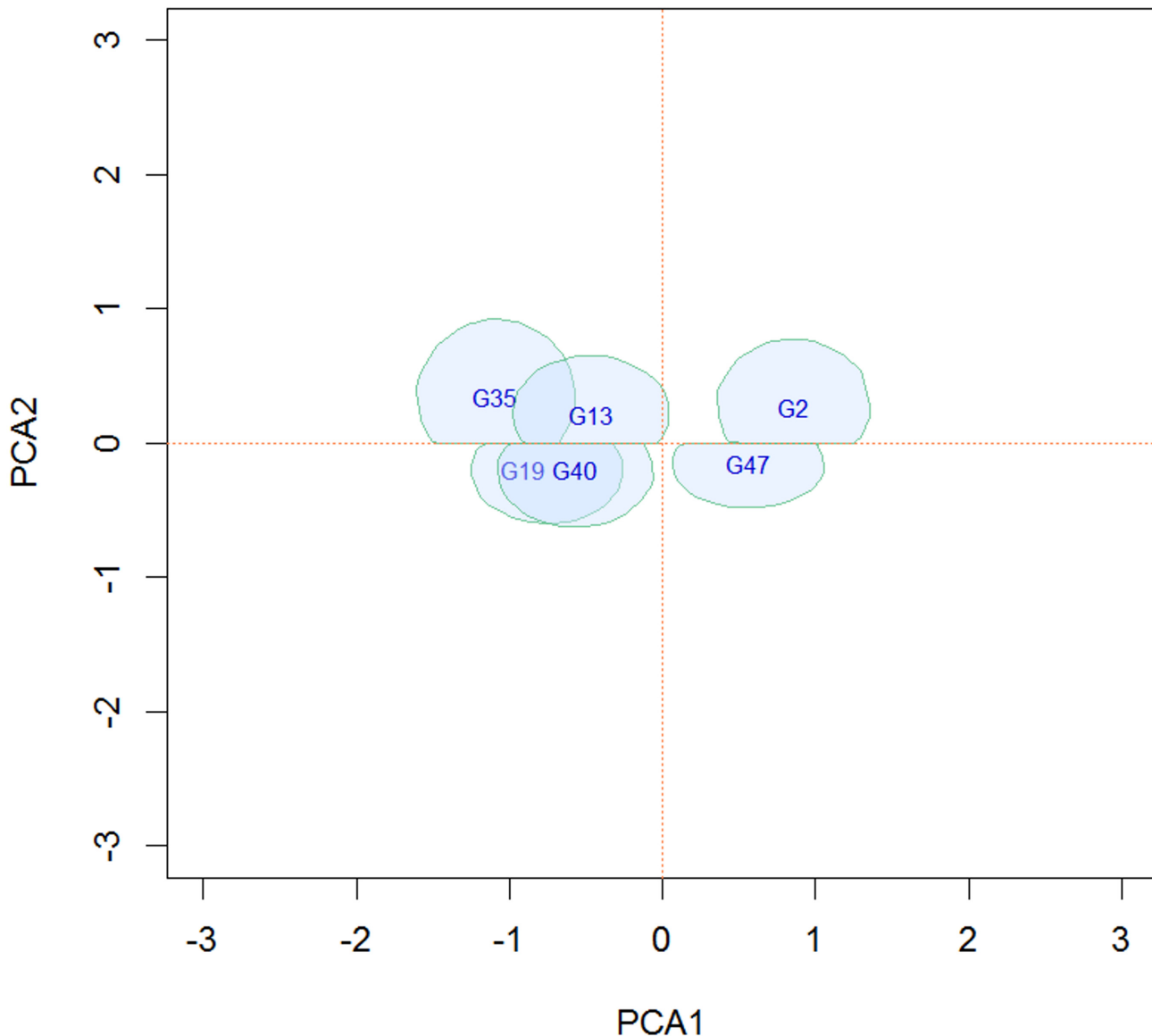
Fig 11. 95% credible regions that do not include the origin for the first two principal axes of the genotype scores using the AMMIBS analysis.

selection of the points in the chain of singular vectors (see Fig 2). This problem did not occur in the AMMIB model and the separation of the chain into lags and the choice of sign were more pronounced (Fig 1). In Fig 11, the change in the signs during the MCMC is in evidence, but the data show that the absolute value of the two estimates and the position of the bivariate credible regions in the biplot for the two methods did not differ.

The same procedure was performed to create homogeneous groups of environments with respect to the interaction effect. Similar to the genotypes, the results for environments A3 and A8 are not shown in the biplot for simplicity because they do not contribute significantly to the interaction effect. The others were classified by similarity among groups, resulting in the three groups shown in Fig 12, i.e., (A6), (A1 and A7) and (A2, A4, A5 and A9). In this case, the same restriction regarding the truncation when choosing the sign was applied.
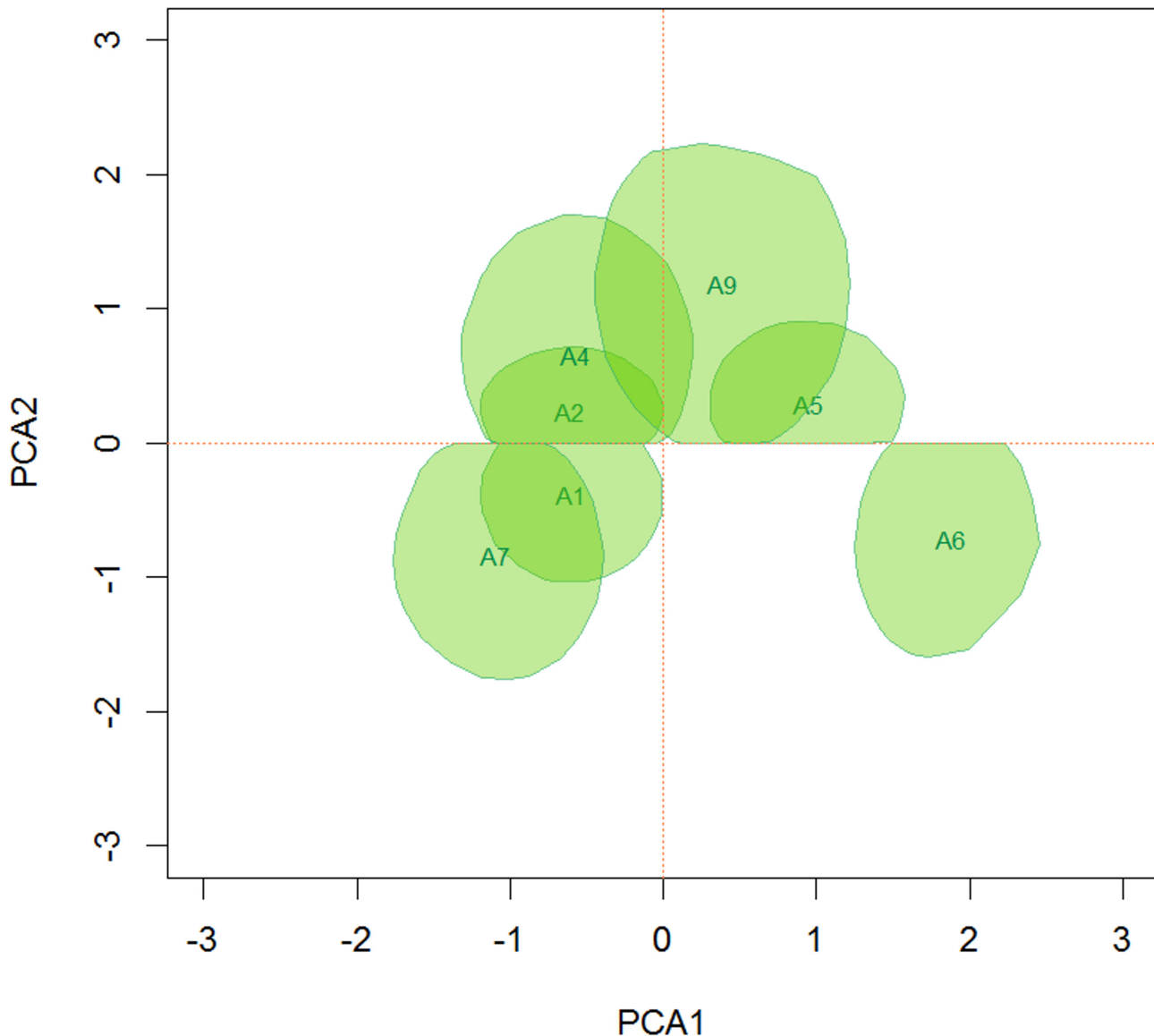
Fig 12. 95% credible regions that do not include the origin for the first two principal axis of the environmental scores.

Using the biplot, the adaptability between genotypes and environments was also analyzed (Fig 13). The superposition of the credible regions showed that the group consisting of G13, G19, G35 and G40 was specifically adapted to the environment group consisting of A1, A2, A4 and A7, while genotype G2, based on the overlap of the credible regions, appears to be specifically adapted to environments A2 and A9. Genotype G47 does not overlap the environments because of the restriction of the solution of the second singular vector; however, the proximity of the credible regions in the truncated area indicates that G47 would be adapted to environments A2 and A9. Apparently, none of the genotypes would be adapted to environment A6.

These results, the univariate HPD regions for the BLUPs of the effects and the bivariate credible regions for the genotype and environment scores in the biplot can help breeders to select the best genotypes, providing statistical support for their choices.
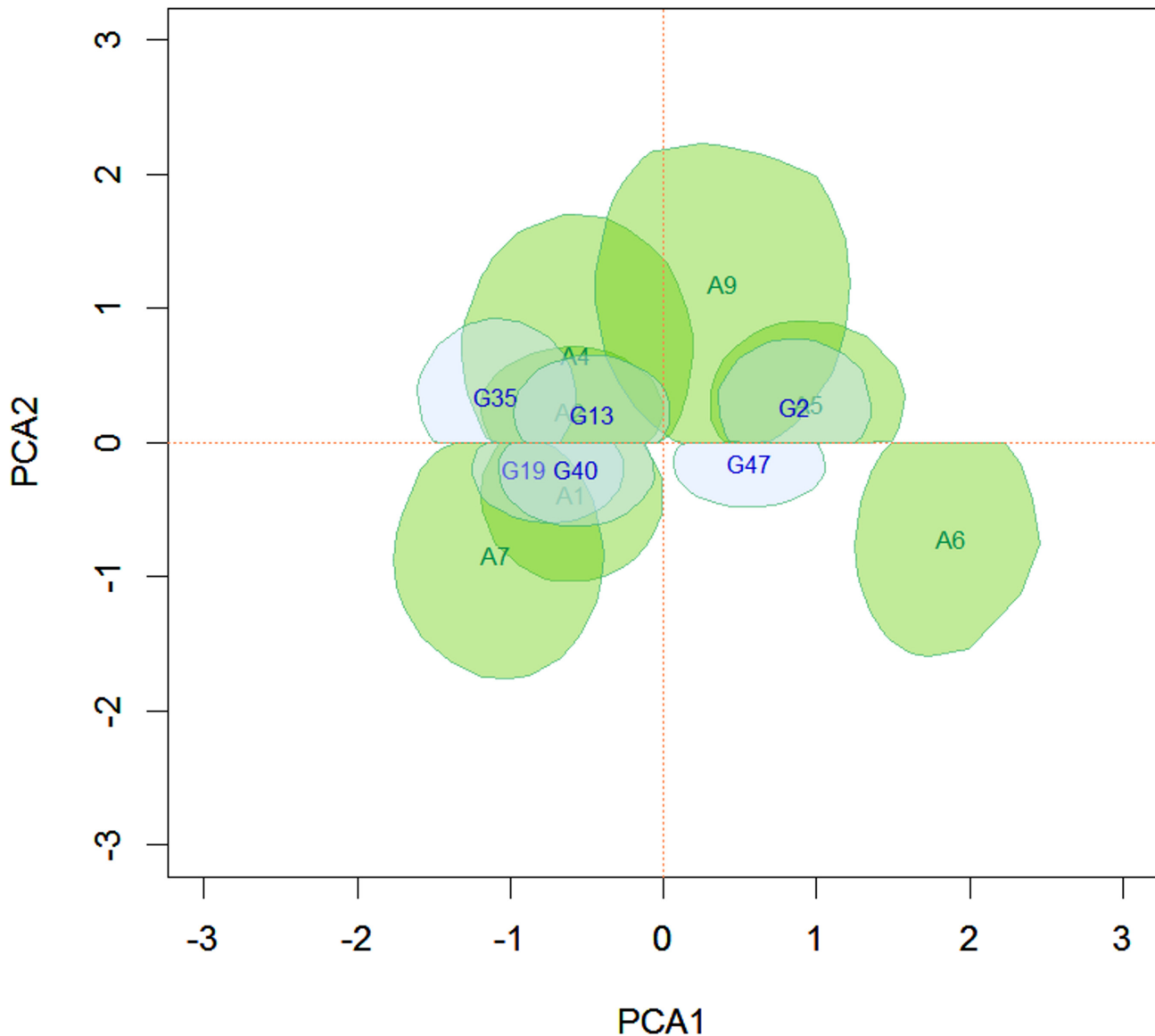
Fig 13. 95% credible regions that do not include the origin for the first two principal axes of the genotype and the environmental scores.

Discussion

In Bayesian inference, the number of bilinear components to be retained in AMMI model is selected using the Bayesian information criterion (BIC) and the Bayes factor (BF) [22, 24]. However, Cornelius and Crossa [10,13] showed that when shrinkage estimators are used, the zero singular values do not provide any information on interaction effects and can thus be excluded from the model. Therefore, by assigning specific variance to the singular values, the selection of the interaction terms is similar to a shrinkage estimator of fixed effects, i.e., singular values that contribute little or no information on the interaction pattern have estimates close to zero and are not included in the model, and those with true importance for the bilinear model have little or no shrinkage effect and are thus retained in the model.

The results of Cornelius and Crossa [10,13] showed that shrunken estimates for linear-bilinear models with fixed effects are superior to OLS and as good as or better than estimates produced by BLUPs for models with random effects. Their method, however, was subject to the limitations of models with fixed effects and, therefore, could only be used with balanced data and homogeneous variances. Another disadvantage is the need for extra steps in the iterative estimation process because the estimates obtained by the iterative method could violate the restrictions imposed on singular values, as observed in this study, where the restriction $\lambda_k \geq \lambda_{k+1}$ was generally not obeyed. To circumvent this violation of the model, it is assumed that $\lambda_k = \lambda_{k+1}$ and a joint estimate for the terms violating this restriction is calculated, as followed in the present analyses (Table 3).

Our approach showed that shrinkage estimators can be obtained for singular values without violating the restrictions of the analysis allowing to obtain credible intervals for the principal components and, simultaneously, to select the best model, i.e., the one with the fewest components. Our approach is a Bayesian interpretation of the estimators developed by Cornelius and Crossa [10], although in practical terms, the selected model was more parsimonious (AMMI2 vs. AMMI5 and AMMI1 vs. AMMI5). Despite this difference, our method based on Bayes Factor or evidence criterion [32] converged to the same model indicated by the Cornelius F-test and purely Bayesian inference selected values similar to Dias and Krzanowski cross validation criterion. The first principal component showed no shrinkage compared with those obtained by the AMMIB and the AMMI models using a random GE matrix. This result showed that the model successfully retains axis with greater ability to capture the pattern and eliminates axis with high noise as confirmed by cross validation result.

In AMMIBS, the use of any test on the number of components might not be necessary since all information we need is present at posterior distribution. In this context, the model to be selected by the AMMIBS was AMMI1 instead AMMI 2 as by BF measure. This can be observed by the use of the posterior percentile in the Fig 1. However, this approach might be ambiguous since the second singular value present a high posterior average and hold the "Ockham hill" in the plot and, therefore, might be important to explain GEI. In this sense, we adopted a extra criterion by used of Bayes Factor. The Bayes Factor reveled the best model in a very parsimonious models (AMMI2) and in this context what is the best approach? A general and definitive answer is very difficult but in the Fig 9 we can obtain a clarification for the range of models observed. The AMMI OLS gives us a ambiguous response about the best model since AMMI5 and AMMI2 presenting two peaks being the Ockham´s hill localized at AMMI5. By the classical interpretation this could be the best model since it retain more pattern than noise. But the degree of freedom requested to adjust AMMI5 is two times larger than requested by AMMI2 and thus the best model could be AMMI2 in OLS context [11]. This reason was used by [11] in their approach to find the more parsimonious model by account for the number of degree of freedom used in PRESS adjustment. On the other hand, in Bayesian cross validation context, this problem was not observed since higher dimensional components were shrunk and no more than one peak was observed in Ockham's plot. Thus, the Bayes Factor was a good criterion to select model since all pattern was retained in the two first components—presenting lower PRESS than AMMI5-OLS. However, the Bayes Factor analysis presented a high computational cost because the number of rounds requested to obtain the criterion was almost equal to the number of components in AMMI model. In addition, in AMMIB, the BF criterion selected the worst AMMI model considering the predictive ability (AMMI4).

As alternative we applied a direct cutoff criterion using the threshold given by Cornelius and Crossa [10] for shrinkage AMMI on the posterior distribution. In this sense, the AMMI1 model was selected. It is obvious that AMMI 1 is not the best model as showed in Ockham's plot, but it is the second one. Therefore, we might claim the parsimonious principle and

verifying that the number of replications requested by full model to obtain the same root mean square predictive difference (RMSPD) of AMMI1 and AMMI2 is 7.4 and 8 respectably. Thus, the loss of GEI pattern between choosing AMMI1 instead AMMI 2 could be marginal, but computably more efficient and therefore we suggested the AMMI1 model as the best model for our dataset. In addition, the direct posterior analysis as proposed here do not request several rounds of tests as in BF test and do not claim for Gaussian assumption about principal components as in F test—about this last claim, our result showed that Gaussian assumption, in general, is very strong.

It worth to highlight also that our BF algorithm was not optimized since we started by full model instead lower dimensional models. Thus, we suggest the development of a more efficient algorithms starting from more parsimonious models first in order to avoid more cycles of Bayes Factor test.

A good discussion about parsimony and predictive ability in AMMI models was given by Dias and Krzanowski [11]. These authors observed that cross-validation based on leave-one-out methods usually select more parsimonious models than F-test or cross-validation based on randomized methods. In other words, they found that leave-one-out selected AMMI1 while AMMI3 and AMMI2 were selected by F-Gollob and Cornelius F-test respectively. In addition these authors observed very larger range of different models selected by F-based tests and cross-validation as observed in this study, pointing the difficult task of selecting models in AMMI methods. In general, the authors concluded that cross-validation methods are more stable across several trials, more parsimonious than F-based tests and are free from Gaussian assumption.

Although it is very difficult to relate analytical proof for our empirical result, the relation between AMMIBS and F-based test is very similar. It is worth highlighting that different breeder's aims could be related to the number of components retained in the AMMI model and that predictive ability sometimes may be useful to breeder, but sometimes not [38]. In this work, we prefer to emphasize the advantage of applying models selection plus biplot inference using just a model and further discussion could be raised in a more practical study.

The primary difficulty encountered in the shrinkage estimators was in the separation of the sign of the singular vector obtained by the AMMIB method arising from the precision parameter of the von Mises-Fisher conditional distribution. Because the ratio of singular value and residual variance represents the concentration parameter, in high-dimensional coordinates, when the singular value is strongly shrunken, these vectors are sampled with high dispersion distribution or nearly uniformly over the hypersphere. Therefore, this effect combined with the arbitrary change in the sign showed that the separation of the chain into lags in high dimensional parameter is impractical.

Another cause of the lack of periodicity in the chain is from the Jeffreys priors for the specific variances, resulting in scaled inverse chi-square conditionals with only one degree of freedom of high-variance and improper posteriors for the singular values. To solve this situation, an extended prior, from Ter Braak et al. [28], was implemented to obtain a proper posterior distribution. However, as shown by Ter Braak et al. [28] and verified by the data in this study, the conditional posterior density can be bimodal, with one mode peaking near zero. According to Ter Braak [37], this is caused by the shrinkage effect provided by the prior. The posterior mean for the second singular value, also shown by Ter Braak [37], is the least squares estimate of this value times its shrinkage factor.

It was empirically shown that the absolute value of each coordinate is approximately equal to that obtained from the conventional analysis, and, therefore, this pattern was used to choose the sign (Figs 1 and 2).

The decision to use the absolute value of the chain resulted in point estimates of the singular values similar in the AMMI, the AMMIB, and the AMMIBS methods, but hindered the construction of the credible intervals. However, when there was little shrinkage of the singular value, the bivariate credible regions were not truncated, as observed for the first principal component. This implies that in models with a large amount of shrinkage, the inferences for the intervals can be hindered, even though the estimates of the posterior means are similar to conventional estimators.

Despite this limitation, overall, the AMMIBS model successfully chose the terms for the AMMI model and parameter estimates, and offers several benefits compared with traditional methods. For example, although not addressed in this study, the model can have heterogeneous variance in the data, which, as previously mentioned, would be a limitation of frequentist approaches; although some works have tried to address heterogeneity by weighted models [39]. Moreover, additional information, such as relatedness and historical experimental data can be incorporated with relative ease into the model. Although these advantages can also be obtained from mixed models in MET in versions equivalent to AMMI models and genotype main effect plus genotype environment effect (GGE) biplots [18,19,20], parametric confidence regions for the scores in the biplot are not easily incorporated [17,21]. Some suboptimal solutions have been proposed by Yang et al. [17], suggesting that confidence regions in the biplot be constructed using bootstrap resampling techniques. This method has been criticized by destroying the interaction pattern [16]. The critics emphasized that the position of the scores in a biplot is important for the analysis and mutually defined, so that the signs and values lose meaning when genotype scores are randomly separated from environment scores in the resampling process. But, in general, our results showed that Yang et al. [17] claim about biplot interpretation must be taking into account since the uncertain about relative position of environments and genotypes in the biplot may conduct to erroneous classification of stability and adaptability plus mega-environments discovery.

When estimating the singular values for the AMMIB model, the estimates were shrunk relative to the least squares solutions, which is consistent with the results in Crossa et al. [21] and Oliveira et al. [25]. However, as shown in Table 3, AMMIBS produced even more shrunken estimates, i.e., the results suggest that shrinkage using specific variance for the singular values was stronger for the high-dimensional terms and had little effect on the first two singular values. This implies that the AMMIBS model was more parsimonious because it retained only one non zero singular value and two principal components to describe the GE interaction. As described above, the inner product of these components produces shrunken GE estimators, showing that the AMMIBS model produces more shrunken interaction estimates than BLUPs from a mixed model, suggesting that the interaction is a random. The AMMIB model produced the same result, with an inner product much closer to the BLUP, suggesting that this model restricted by priors with a fixed estimator produces shrunken GE estimates that are similar to the BLUPs. It is worth to highlight that the BLUP method instead of shrinking the singular value directly as in the AMMIBS, it shrunk the cell means related to GEI at ratio of σ_e^2/σ_I^2 where σ_I^2 is the GEI variance. The ratio of shrinkage obtained in our method is defined directly on the singular value on the ratio $\sigma_e^2/\sigma_{\lambda_k}^2$ (S1 Appendix). Thus, in mixed models, the EBLUPs for GEI are function of a constant ratio of shrinkage, but in AMMIBS, different ratios are applied directly on singular value posterior showing the difference among the methods.

In the last decade, several methods for analyzing GE interactions in MET data have been proposed. Models that incorporate genotypes as random effects are being widely published and defended in the literature. The application of Bayesian methods, although not new [23,24], brings new prospects for MET data analysis. Using relatedness matrices replacing

$\boldsymbol{\delta}|\mu_\delta, \sigma_\delta^2 \sim N(\mathbf{0}, \mathbf{I}\sigma_\delta^2)$ by $\boldsymbol{\delta}|\mu_\delta, \sigma_\delta^2 \sim N(\mathbf{0}, G)$ and heterogeneity of variance by replacing $I\sigma_e^2$ by V can be easily implemented and more informative priors can helps to generate more realistic estimates for the parameters of interest and should be a part of new research approaches.

This study illustrated that Bayesian shrinkage can be applied to AMMI models with relative success to select models and estimate parameters. However, the derivation of reference priors or maximum entropy for the parameters of the terms of the principal components may solve the problems of improper or even bimodal posteriors.

The results demonstrated that the traditional Bayesian AMMI model produces a small amount of shrinkage in the singular values and avoids problems determining credible intervals in the biplot. Bayesian shrinkage AMMI models have problems with the credible interval, but produce stronger shrinkage of the principal components, converging to more shrunken GE matrices than those obtained using mixed models. This characteristic allowed more parsimonious models to be chosen and more GEI pattern retained on the first two components. The resulting model chosen by posterior distribution of singular value was also similar to those produced by the cross-validation approach in traditional AMMI models. Our method enables the estimation of credible interval for AMMI biplot plus the choice of AMMI model based on direct posterior distribution retaining more GEI pattern in the first components and discarding noise without Gaussian assumption as requested in F-based tests or deal with parametric problems as observed in traditional AMMI shrinkage method.

Supporting Information

S1 Appendix. The AMMIBS model and the effect of shrinkage on the bilinear parameters.
(DOCX)

S1 Table. Cross validation based on leave-one-out approach.
(DOCX)

S2 Table. ANOVA for grain yield in 55 genotypes evaluated at 8 environments.
(DOCX)

S1 Dataset. Dataset.
(CSV)

S1 Text. R-code.
(R)

Acknowledgments

Author Contributions

Conceived and designed the experiments: MB. Analyzed the data: CPdS LAdO AKAP. Wrote the paper: CPdS. Performed the simulation data: JJN. Wrote the R Program: MB.

References

1. Cornelius PL, Seyedsadr MS (1997) Estimation of general linear-bilinear models for two-way tables. Journal of Statistical Computation and Simulation 58: 287–322.

2. Gollob H F (1968) A statistical model which combines features of factor analytic and analysis of variance techniques. Psychometrika 33: 73–115. PMID: 5239571

3. Mandel J (1969) The partitioning of interactions in analysis of variance. Journal of Research of the National Bureau of Standards, Series B 73: 309–328.

4. Mandel J (1971) A new analysis of variance model for non-additive data, Technometrics 13: 1–18.

5. Pike EW, Silverberg T R., Pike E. and Silverberg T. 1952. Designing Mechanical Computers. *Machine Design*, 24:131–137.

6. Williams EJ (1952) The Interpretation of Interactions in Factorial Experiments Biometrika, 39:65–81.

7. Gauch HG (2006) Statistical analysis of yield trials by AMMI and GGE. Crop Science 46:1488–1500.

8. Cornelius PL, Seyedsadr M, Crossa J (1992) Using the shifted multiplicative model to search for "separability" in crop cultivar trials. Theoretical and Applied Genetics 84:161–172. doi: 10.1007/BF00223996 PMID: 24203043

9. Balestre M, Souza JC, Von Pinho RG, Oliveira RL, Paes JMV (2009) Yield stability and adaptability of maize hybrids based on GGE biplot analysis characteristics. Crop Breeding and Applied Biotechnology 9: 219–228.

10. Cornelius PL, Crossa J (1999) Prediction assessment of shrinkage estimators of multiplicative models for multi-environment cultivar trials. Crop Science 39: 998–1009.

11. Dias CTDS, Krzanowski WJ (2003) Model selection and cross validation in additive main effect and multiplicative interaction models. Crop Science 43: 865–873.

12. Forkman J, Piepho H-P (2014) A parametric bootstrap procedure for selecting the number of multiplicative terms in GGE and AMMI models. Biometrics (01111/biom12162).

13. Cornelius P, Crossa J, Seyedsadr M (1996) Statistical tests and estimators of multiplicative models for genotype-by-environment interaction. In:, Kang M S, and Gauch H G (ed) Genotype-by-environment interaction. Boca Raton: CRC Press pp 199–234

14. James W, Stein C (1961) Estimation with quadratic loss. In: Proceedings of the fourth Berkeley symposium on mathematical statistics and probability. University of California Press 1: 361–379.

15. Gabriel KR (1971) The biplot graphic display of matrices with application to principal component analysis. Biometrika 58: 453–467.

16. Yan W, Glover KD, Kang MS (2010) Comment on "Biplot Analysis of Genotype× Environment Interaction: Proceed with Caution," by Yang R-C, Crossa J, Cornelius PL, Burgueño J in 2009 49: 1564–1576. Crop Science 50: 1121–1123.

17. Yang R-C, Crossa J, Cornelius PL, Burgueño J (2009) Biplot analysis of genotype× environment interaction: Proceed with caution. Crop Science 49: 1564–1576.

18. Burgueño J, Crossa J, Cornelius PL, Yang R-C (2008) Using factor analytic models for joining environments and genotypes without crossover genotype × environment interaction Crop Science 48:1291–1305.

19. Piepho H-P (1997) Analyzing genotype-environment data by mixed models with multiplicative terms. Biometrics 90: 761–766.

20. Smith A, Cullis B, Gilmour A (2001) Applications: the analysis of crop variety evaluation data in Australia. Australian & New Zealand Journal of Statistics 43: 129–145.

21. Crossa J, Perez-Elizalde S, Jarquin D, Cotes JM, Viele K, Liu G, et al. (2011) Bayesian estimation of the additive main effects and multiplicative interaction model. Crop Science 51:1458–1469.

22. Perez-Elizalde S, Jarquin D, Crossa J (2012) A General Bayesian Estimation Method of Linear—Bilinear Models Applied to Plant Breeding Trials With Genotype×Environment Interaction. Journal of agricultural, biological, and environmental statistics 17: 15–37.

23. Viele K, Srinivasan C (2000) Parsimonious estimation of multiplicative interaction in analysis of variance using Kullback—Leibler Information. Journal of statistical planning and inference 84: 201–219.

24. Liu G (2001) Bayesian computations for general linear-bilinear models. M.Sc. Thesis, University of Kentucky.

25. Oliveira LA, Silva CP, Nuvunga JJ, Silva AQ, Balestre M (2014) Credible intervals for genotypic and environmental scores in the AMMI model with random effects for genotype. Crop Science doi: 10.2135/cropsci2014.05.0369

26. Stein C (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In: Proceedings of the Third Berkeley symposium on mathematical statistics and probability. Los Angeles: University of California 1:197–206.

27. Mardia KV, Kent JT, Bibby JM (1979) Multivariate analysis. London: Academic Press. 521 p.

28. Ter Braak CJ, Boer MP, Bink MC (2005) Extending Xu's Bayesian model for estimating polygenic effects using markers of the entire genome. Genetics 170: 1435–1438. PMID: 15911578

29. Raftery AE, Lewis SM (1992) [Practical Markov Chain Monte Carlo]: Comment: One Long Run with Diagnostics: Implementation Strategies for Markov Chain Monte Carlo. Statistical Science 7: 493–497.

30. Heidelberger P, Welch PD (1983) Simulation run length control in the presence of an initial transient. Operations Research 31: 1109–1144.

31. Cornelius PL, Seyedsadr M, Crossa J (1992) Using the shifted multiplicative model to search for "separability" in crop cultivar trials. Theoretical and Applied Genetics 84: 161–172. doi: 10.1007/BF00223996 PMID: 24203043

32. Jaynes ET (2003) Probability theory: the logic of science. Cambridge university press, 2003.

33. Gabriel KR (2002). Le biplot—outil d'exploration de données multidimensionelles. Journal de la Societe Francaise de Statistique 143:5–55

34. Chen MH, Shao QM (1999) Monte Carlo estimation of Bayesian credible and HPD intervals. Journal of Computational and Graphical Statistics 8: 69–92.

35. R Core Team (2013) R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

36. Ooms JCL (2009) The highest posterior density posterior prior for bayesian model selection. MSc Thesis, Utrecht University. Available: http://dspace.library.uu.nl/handle/1874/34234. Accessed 12 Nov 2013.

37. Ter Braak CJ (2006) Bayesian sigmoid shrinkage with improper variance priors and an application to wavelet denoising. Computational statistics & data analysis 51: 1232–1242.

38. Gauch HG (2013) A simple protocol for AMMI analysis of yield trials. Crop Science 53: 1860–1869.

39. Rodrigues PC, Malosetti M, Gauch HG, van Eeuwijk FA (2014) A Weighted AMMI Algorithm to Study Genotype-by-Environment Interaction and QTL-by-Environment Interaction. Crop Science 54: 1555–1570.