RESEARCH ARTICLE

Bayesian factor analytic model: An approach in multiple environment trials

Joel Jorge Nuvunga[1,2‡], **Carlos Pereira da Silva**[1☯], **Luciano Antonio de Oliveira**[1,3☯], **Renato Ribeiro de Lima**[1☯], **Marcio Balestre**[1‡*]

1 Department of Statistics (DES), Federal University of Lavras, Lavras, Minas Gerais, Brazil, **2** Eduardo Mondlane University, Chibuto College of Business and Entrepreneurship, Chibuto, Maputo, Mozambique, **3** 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

Citation: Nuvunga JJ, da Silva CP, de Oliveira LA, de Lima RR, Balestre M (2019) Bayesian factor analytic model: An approach in multiple environment trials. PLoS ONE 14(8): e0220290. https://doi.org/10.1371/journal.pone.0220290

Editor: Lei Shi, Yunnan University of Finance and Economics, CHINA

Received: January 25, 2018

Accepted: July 13, 2019

Published: August 22, 2019

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by Fundação Calouste Gulbenkian, Fundo Nacional de Investigação de Moçambique, Escola Superior de Negócios e Empreendedoriso de Chibuto, Chibuto, Universidade Eduardo Mondlane, CAPES (Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior) and CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico –Grant # 302674/2015-2).

Abstract

One of the main challenges in plant breeding programs is the efficient quantification of the genotype-by-environment interaction (GEI). The presence of significant GEI may create difficulties for breeders in the selection and recommendation of superior genotypes for a wide environmental network. Among the diverse statistical procedures developed for this purpose, we highlight those based on mixed models and factor analysis that are called factor analytic (FA) models. However, some inferential issues are related to the factor analytic model, such as Heywood cases that make the model non-identifiable. Moreover, the representation of the loads and factors in the conventional biplot does not involve any measurement of uncertainty. In this work, we propose dealing with the FA model using the Bayesian framework with direct sampling of factor loadings via spectral decomposition; this guarantees identifiability in the estimation process and eliminates the need for the rotationality of factor loadings or imposition of any *ad hoc* constraints. We used simulated and real data to illustrate the method's application in multi-environment trials (MET) and to compare it with traditional FA mixed models on controlled unbalancing. In general, the Bayesian FA model was robust under different simulated unbalanced levels, presenting the superior predictive ability of missing data when compared to competing models, such as those based on FA mixed models. In addition, for some scenarios, the classical FA mixed model failed in estimating the full FA model, illustrating the parametric problems of convergence in these models. Our results suggest that Bayesian factorial models might be successfully used in plant breeding for MET analysis.

Introduction

The recommendation of new genotypes for commercial use requires confident and accurate estimations of genetic parameters such as marginal genotypic values, stability, adaptability, disease and environmental stress resistance. This information can be obtained by analyzing the

Competing interests: The authors have declared that no competing interests exist.

genotypes across different environments; such an analysis is called multi-environment trials (MET). These trials are required to isolate the effect of the genotype-by-environment interaction (GEI), which means the differential genotypic responses on different environments. In general, the GEI hinders the breeder's work on the selection and recommendation of the best genotypes for a wide class of environments. Thus, it is used to investigate efficient methods that identify stable genotypes (those that do not contribute to GEI) and positive effects of GEI for specific groups of environments aimed at regionalized recommendations.

Several statistical methods have been proposed. One is the additive main effect and multiplicative interaction (AMMI), and another is the genotype plus genotype x environment interaction biplot (GGEbiplot) [1–3]. These methods have been widely applied in plant breeding programs for the identification of mega-environments, which-won-where patterns, ideal genotypes, and specific adaptability, among others. Limitations inherent to these fixed-based parameters models (such as the lack of flexibility to treat unbalanced data and heterogeneity of variances) have motivated the development of more flexible methods, such as those based on mixed models. Piepho [4,5] and Smith et al. [6] proposed the MET analysis based on multivariate mixed models; it uses factor analysis (AF) structures that consider environments/genotypes and interaction as random effects.

In the literature, these models have been frequently referred to as factor analytic (FA) models [6–10]; they have shown great versatility for genotype selection since they combine the stability and adaptability studies into a single approach. The advantage of this approach is related to its ability to address missing data and the heterogeneity of residual and genotypic (co)variances. These models are also notable for allowing the inclusion of heteroscedastic residues of genetic values, which constitute an important aspect to be considered in the analyses [11,12]. It is known that the heterogeneity of variances among genotypes is affected by the heterogeneity of variances among environments and vice versa [13]. Furthermore, this model has been useful for summarizing the covariance pattern in multivariate data [14].

Despite the recognized advantages offered by FA models, the method also has limitations, such as the need to impose some constraints and, under some scenarios, the non-identifiability in parameter estimation. It is worth highlighting the difficulty to construct exact confidence intervals for the components of variance, since they are approximate and require assumptions of asymptotic normality. In addition, there is a great demand for computational resources and efficient algorithms to avoid the occurrence of solutions outside the parametric space (the so-called Heywood cases), among other aspects [15,16].

An interesting alternative to the frequentist or likelihood approaches is the use of Bayesian inference. Bayesian analysis allows greater flexibility for the construction of credible intervals for unknown parameters, since all inference processes are based on the posterior distribution. The flexibility of the Bayesian method for GEI analysis was partly illustrated by Crossa et al. [17], Oliveira et al. [18], Perez-Elizalde et al. [19] and Silva et al. [20], which incorporated credibility regions for the bilinear parameters in the AMMI model. It is known that inferences about genotypic and environmental scores in linear-bilinear models present great difficulties for frequentist methods or non-parametric approaches [17,18,21,22].

Using Bayesian inference in AMMI models, Perez-Elizald et al. [19] have shown how historical information can easily be incorporated into the model. Recently, Jarquin et al. [23] showed how this information could be included in the sites regression (SREG) model using the multilevel (hierarchical) Bayesian approach. Methodological innovations and applications of factorial analysis have been rapidly designed in recent years, partly due to access to computational tools for numeric integration in the Bayesian framework. In particular, it is possible to highlight the use of the Markov chain Monte Carlo (MCMC) in classical factor analysis [24].

The Bayesian analysis for FA model was presented by de Los Campos and Gianola [25]. These authors proposed prior distributions based on the assumptions of the classical factors analysis that avoids the imposition of constraints and reduces the computational requirements that have restricted the use of these models. However, this approach it not founded on the initial FA structure described in Smith et al. [6]. Instead, de Los Campos's and Gianola's [25] model accounts for the decomposition of the genetic variance, ignoring a more general structure for residual variances and performing a two-stage adjustment for the parameter of the model. In addition, only the balanced scenario was considered, since it does not consider situations where missing data are present, which is the great appeal of the FA models for MET data analysis.

Nevertheless, the advantages of the FA models in summarizing the MET network and finding data patterns from breeding programs are remarkable. Several successful applications can be found in Burgueño et al. [7,26], Kelly et al. [9], Meyer [16], Smith, Cullis and Thompson [6,27], Smith et al. [28], Stefanova and Buirchell [29] and Tyrisevä et al. [30].

This study proposes the Bayesian approach for the FA model applied in MET data using spectral decomposition to ensure the model identifiability. Additionally, we sought to evaluate the Bayesian FA predictive ability on unbalanced data with respect to the classic FA models through REML.

Material and methods

Material

Simulated data. We simulated a dataset with 20 genotypes (G1-G20) that were evaluated in five environments (E1-E5) using a randomized complete block design with two replicates. Five genotypes had interactions simulated from the Gaussian distribution with large variances (unstable genotypes) and positive marginal effects. Five more genotypes were marginally negative and contained large Gaussian variances (unstable). 10 genotypes were a standard Gaussian distribution, and the unstable genotypes had variance of 2 (three times the standard Gaussian distribution. Therefore, the stability and instability in this study were considered as a function of the size of the variability across the environments and the genotypes' marginal effects as Gaussian realizations.

Experimental data. The experimental data were described by Melo et al. [31]. These data are related to grain yield traits measured in 50 maize single-cross hybrids (G1-G50). The experiment was conducted using incomplete block designs with 2 replicates, and each plot had 5-m rows with 70-cm spacing between rows. The grain yield was evaluated and adjusted for 13% moisture and converted into t·ha^{-1}. Come from.

These hybrids are originated from crossing among lineages of different backgrounds (tropical–Flint, Lancaster and Stiff Stalk Synthetic sources–see Melo et al. [31]). The crossing design was in a partial diallel system. These hybrids were evaluated during the agricultural years of 2013 and 2014 in 10 environments (E1-E10) representative of the Southeast and Southern regions of Brazil. Details on the environmental characteristics are given in the S1 Table.

Method

Statistical model. The classical multivariate linear mixed model under the unstructured covariance matrix can be described as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon} \tag{1}$$

where $\mathbf{y}_{(n\times1)}$ is the vector of observations for p environments, q blocks and m genotypes.

Vectors $\boldsymbol{\beta}_{(pq\times1)}$, $\mathbf{u}_{(mp\times1)}$ and $\boldsymbol{\varepsilon}_{(n\times1)}$ are fixed, random and residual vector effects, respectively. The matrices $\mathbf{X}_{(n\times pq)}$ and $\mathbf{Z}_{(n\times mp)}$ denote information regarding $\boldsymbol{\beta}$ and $\mathbf{u}$, respectively. For simplicity, it is assumed that $\mathbf{u}$ represents the additive genetic effects. Moreover, $\boldsymbol{\varepsilon}\sim N(\mathbf{0},\mathbf{R})$ and $\mathbf{u}\sim N(\mathbf{0},\Sigma\otimes\mathbf{I}_m)$. In the FA framework, we can approximate the vector related to additive genetic effects by common and specific factors using $\mathbf{u} = (\boldsymbol{\Gamma}_{p\times k}\otimes\mathbf{I}_m)\mathbf{f}+\boldsymbol{\delta}$, where the covariance matrix Σ is represented by a factor analytic structure ($\Sigma = \boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top+\boldsymbol{\Psi}$). Therefore, model (1) can be rewritten as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}(\boldsymbol{\Gamma}_{p\times k}\otimes\mathbf{I}_m)\mathbf{f} + \mathbf{Z}\boldsymbol{\delta} + \boldsymbol{\varepsilon} \tag{2}$$

This has been referred to in the literature as the mixed model factor-analytic structured model or simply the factor analytic (FA) model. $\boldsymbol{\Gamma}_{(p\times k)}$,$\mathbf{f}_{(mk\times1)}$ and $\boldsymbol{\delta}_{(mp\times1)}$ are the loading matrix ($k = 1, \ldots,p$), the factor vector related to additive genetic effects and the vector of specific effects, respectively. Furthermore, $\boldsymbol{\varepsilon}$ is the vector of residuals, $\mathbf{X}$ is the fixed-effect matrix referring to $\boldsymbol{\beta}$, and $\mathbf{Z}$ is the genetic matrix referring to $\mathbf{f}$ and $\boldsymbol{\delta}$, where k represents the number of multiplicative terms. The $\mathbf{Z}_{(n\times mp)}$ is a block diagonal matrix matching the vector $\mathbf{y}_{(n\times1)}$; this is equivalent to assume that $\mathbf{Z} = (\mathbf{Z}_1,\cdots,\mathbf{Z}_p)$ with $\mathbf{Z}_k$ the matrix matching i^{th} genotype at factor k with dimension ($n\times m$), for $k = 1,\ldots,p$, and $\mathbf{f} = (\mathbf{f}_1^\top,\cdots,\mathbf{f}_p^\top)$ with $\mathbf{f}_k$ a m-dimensional vector, then

$$\mathbf{Z}(\boldsymbol{\Gamma}\otimes\mathbf{I}_m)\mathbf{f} = \sum_{k=1}^{p}\mathbf{Z}_k\boldsymbol{\Gamma}\mathbf{f}_k,$$

Given that $\boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top$ is a symmetric matrix, we can rewrite the model (2) observing that the matrix of factor loadings can be obtained by $\boldsymbol{\Gamma} = \mathbf{V}\boldsymbol{\Lambda}^{\frac{1}{2}}$. In this expression, $\mathbf{V}$ represents the matrix of singular vectors and $\boldsymbol{\Lambda}$ a diagonal matrix formed by the eigenvalues obtained by the spectral decomposition of the loading matrix. The spectral decomposition of the unstructured covariance matrix can be approximated by

$$\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top = \mathbf{V}_{\mathrm{p}}\boldsymbol{\Lambda}_{\mathrm{p}}\mathbf{V}_{\mathrm{p}}^\top = \sum_{k=1}^{p}\lambda_k^2\boldsymbol{\alpha}_k\boldsymbol{\alpha}_k^\top \tag{3}$$

where λ_k is the i-th singular value and $\boldsymbol{\alpha}_k$ the i-th singular vector (or eigenvector) of the spectral decomposition, respectively.

A model with $k = p$ (p as the matrix rank) multiplicative terms is called full rank, and the specific effects are assumed as null. However, the advantage of using factor analysis occurs when k is significantly smaller than p. If so, the number of parameters in the factor analysis $k(p+1)$ becomes much smaller than those $p(p+1)/2$ parameters of Σ.

Using principal components to describe the factor loadings and considering the spectral decomposition properties and appropriate linear transformations, we can rewrite the Eq (2) that involves the loadings as follows:

$$\mathbf{ZL} = \mathbf{Z}(\boldsymbol{\Gamma}_{\mathrm{m}}\otimes\mathbf{I}_{\mathrm{p}}) = \sum_{k=1}^{p}\lambda_k diag(\mathbf{X}_2\boldsymbol{\alpha}_k)\mathbf{Z}_f \tag{4}$$

where $\boldsymbol{\Gamma}_{\mathrm{m}} = \mathbf{V}_{\mathrm{p}}\boldsymbol{\Lambda}_p^{1/2} = \sum_{k=1}^{p}\lambda_k\boldsymbol{\alpha}_k$. Further details on these properties and the theorem's demonstration can be found in Songgui and Suju [32] and in (S1 Text). Replacing the Kronecker product into (2) with the sum in (4), the FA model can be expressed by

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta} + \sum_{k=1}^{p} \lambda_k diag(\mathbf{X}_2\boldsymbol{\alpha}_k)\mathbf{Z}_f\mathbf{f}_k + \mathbf{Z}\boldsymbol{\delta} + \boldsymbol{\varepsilon} \tag{5}$$

The terms λ_k and $\boldsymbol{\alpha}_k$, as already specified in (3), are the k-th singular values and eigenvectors of the spectral decomposition of $\boldsymbol{\Sigma}$, the matrices $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{Z}$ and $\mathbf{Z}_f$ are design matrices destined to distribute the FA effects for experimental unity (see S1 Text). By the summation, we can efficiently distribute the effects by avoiding using the Kronecker product or other artifices in the conditional distributions. This facilitates algebraic manipulations and computational implementations. Thus, the model presented in (5) is more treatable from a Bayesian point of view with respect to its equivalent (2). The conditional likelihood has a multivariate normal density as follows

$$\mathbf{y}|\boldsymbol{\beta}, \lambda, \boldsymbol{\alpha}, \mathbf{f}, \mathbf{R} \sim \mathbf{N}\left[\mathbf{X}_1\boldsymbol{\beta} + \sum_{k=1}^{p} \lambda_k diag(\mathbf{X}_2\boldsymbol{\alpha}_k)\mathbf{Z}_f\mathbf{f}_k + \mathbf{Z}\boldsymbol{\delta}, \mathbf{R}\right].$$

More details about the likelihood in factorial models can be obtained in supplemental material (S2 Text):

Prior distributions. In this study, the prior distributions for the FA parameters were established based on the assumptions of the factor analysis model through the maximum restricted likelihood (REML) [6]. In this sense, equivalent Jeffrey's prior (Gaussian with large variance) was used for the fixed effects and Bayesian AMMI priors for the loading parameters approximated by the eigenvalues and eigenvectors. The prior distributions for each parameter were given by $\boldsymbol{\beta}|\boldsymbol{\mu}_\beta, \sigma^2_\beta \sim N(\boldsymbol{\mu}_\beta, \mathbf{I}_{pq}\sigma^2_\beta)$; $\mathbf{I}_{pq}$ is an identity matrix; $\boldsymbol{\mu}_\beta = \mathbf{0}$ and $\sigma^2_\beta = 10^{12} \Rightarrow \boldsymbol{\beta} \sim$ constant.

$\mathbf{f}|\boldsymbol{\mu}_\mathbf{f}, \sigma^2_\mathbf{f} \sim N(\mathbf{0}, \mathbf{I}_\mathrm{m})$, where: $\mathbf{I}_\mathrm{m}$- is an identity matrix.

$\boldsymbol{\delta}|\boldsymbol{\mu}_{\boldsymbol{\delta}}, \sigma^2_{\boldsymbol{\delta}} \sim N(\mathbf{0}, \boldsymbol{\Psi})$, where: $\boldsymbol{\Psi} = diag(\psi_{11},\ldots,\psi_{kk})$, $k = 1,\ldots,$p is a diagonal matrix in which the elements ψ_{ii} are the specific variances from each environment.

$\boldsymbol{\alpha}_k$~Uniform spherical in the corrected subspace. The uniform spherical distribution is a special case of von Mises-Fisher distribution [15].

$\lambda_k|\mu_{\lambda_k}, \sigma^2_{\lambda_k} \sim N^+(0, \sigma^2_{\lambda_k})$; $\sigma^2_{\lambda_k} = 10^{12} \Rightarrow \lambda_k|\mu_{\lambda_k}, \sigma^2_{\lambda_k} \sim constant$, where N^+ indicates a positive Gaussian distribution truncated on $\lambda_1 \geq \ldots \geq \lambda_p \geq 0$.

Here, to simplify the notation, it will be assumed that $\mathbf{R}$ is a diagonal block matrix composed of $\sigma^2_{e_k}$, with $k = 1, \ldots, p$. For this parameter, we assume an inverse scaled chi-squared prior distribution denoted by

$$\sigma^2_{e_k} \sim Scale - \chi^{-2}(v_e, S^2_k),\; v_e = 0, S^2_k = 0 \;\Rightarrow\; \sigma^2_{e_k} \sim \frac{1}{\sigma^2_{e_k}}$$

Since $\boldsymbol{\Psi}$ is also a diagonal matrix composed of diagonal elements ψ_{kk} with $k = 1 \ldots, p$, we also assumed the inverse scaled chi-squared prior distribution for ψ_{kk} as follows:

$$\psi_{kk} \sim Scale - \chi^{-2}(v_k, S^2_k),\; v_k = 0, S^2_k = 0 \;\Rightarrow\; \psi_{kk} \sim \frac{1}{\psi_{kk}}$$

All previous prior distributions satisfy the model constraints and are conditionally conjugated.

The Jeffrey's prior for the fixed effects is $|I_{(\beta)}|^{1/2}$, where $I_{(\beta)}$ is the expected Fisher information about $\boldsymbol{\beta}$, used for environmental effects is proportional to a constant given that the information about $\boldsymbol{\beta}$ in the likelihood does not depends on $\boldsymbol{\beta}$. Some equivalence is obtained in the posterior distribution by assuming a prior normal distribution with large variance, i.e 10^{12}.

For the singular values λ_k, the truncated normal distributions were used in order to ensure the model identifiability for positive and sorted singular values, i.e., $\lambda_k>0$ and $\lambda_1>\lambda_2>\ldots>\lambda_k$. For the singular vectors $\boldsymbol{\alpha}_k$, the orthonormality restriction request that the coordinate be distributed on the hypersphere space; here an uniform hypersphere was assumed, in which is equivalent to a von Mises-Fisher with concentration parameter equal to zero[15].

For the factor model parameters (**f**, **δ**), the priors were assumed according to the classical factor model assumptions about this effects [6]. The prior hyperparameters for the variance components were chosen in order to impose large entropy and little weight of prior on the posterior distribution.

The data likelihood given above may be simplified as

$$\mathrm{L}(\mathbf{y}|\phi) = \frac{1}{\sqrt{2\pi^n|\mathbf{R}|}}\exp\left\{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\theta})^\top\mathbf{R}^{-1}(\mathbf{y}-\boldsymbol{\theta})\right\} \tag{6}$$

where $\phi = (\boldsymbol{\beta},\boldsymbol{\lambda},\boldsymbol{\delta},\boldsymbol{\alpha},\mathbf{f},\mathbf{R})$ and $\boldsymbol{\theta} = \mathbf{X}_1\boldsymbol{\beta} + \sum_{k=1}^{m}\lambda_k diag(\mathbf{X}_2\boldsymbol{\alpha}_k)\mathbf{Z}_f\mathbf{f}_k + \mathbf{Z}\boldsymbol{\delta}$

Full conditional posterior distributions for the FA parameters. Applying Bayes' theorem on the likelihood and priors for $\boldsymbol{\varphi}= (\boldsymbol{\beta},\boldsymbol{\lambda},\boldsymbol{\delta},\boldsymbol{\alpha},\mathbf{f},\mathbf{R},\boldsymbol{\Psi})$, the joint posterior distribution is given by

$$p(\varphi|\mathbf{y}) \propto \mathrm{L}(\mathrm{y}|\phi)p(\boldsymbol{\beta})p(\boldsymbol{\lambda})p(\boldsymbol{\alpha})p(\boldsymbol{\delta})p(\mathbf{f})p(\mathbf{R})p(\boldsymbol{\Psi}) \tag{7}$$

The full conditional posterior distributions for the Bayesian factorial analytic (BFA) parameters were derived from (7) as follows:

i) Complete conditional a posteriori distribution for **β**

$$\boldsymbol{\beta}|\ldots \sim N[(\mathbf{X}_1^\top\mathbf{R}^{-1}\mathbf{X}_1)^{-1}\mathbf{X}_1^\top\mathbf{R}^{-1}\mathbf{A}_0, (\mathbf{X}_1^\top\mathbf{R}^{-1}\mathbf{X}_1)^{-1}] \tag{8}$$

where: $\mathbf{A}_0 = \mathbf{y} - \sum_{k=1}^{m}\lambda_k diag(\mathbf{X}_2\boldsymbol{\alpha}_k)\mathbf{Z}_f\mathbf{f}_k - \mathbf{Z}\boldsymbol{\delta}$.

ii) Complete conditional a posteriori distribution for λ_k

The full conditional posterior distribution for the singular value is denoted by:

$$\lambda_k|\ldots \propto N^+[(\mathbf{A}_{2k}^\top\mathbf{R}^{-1}\mathbf{A}_{2k})^{-1}\mathbf{A}_{2k}^\top\mathbf{R}^{-1}\mathbf{A}_{1k}, (\mathbf{A}_{2k}^\top\mathbf{R}^{-1}\mathbf{A}_{2k})^{-1}] \tag{9}$$

where $\mathbf{A}_{1k} = \mathbf{y} - \left[\mathbf{X}_1\boldsymbol{\beta} + \sum_{k\neq k'}^{p}\lambda_{k'} diag(\mathbf{X}_2\boldsymbol{\alpha}_{k'})\mathbf{Z}_f\mathbf{f}_{k'} + \mathbf{Z}\boldsymbol{\delta}\right]$ and $\mathbf{A}_{2k} = diag(\mathbf{X}_2\boldsymbol{\alpha}_k)\mathbf{Z}_f\mathbf{f}_k$.

iii) Complete conditional a posteriori distribution for $\boldsymbol{\alpha}_k$

$$\boldsymbol{\alpha}_k|\ldots \sim N[(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\Delta_k)^{-1}\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\mathbf{A}_{3k}, (\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\Delta_k)^{-1}] \tag{10}$$

where $\mathbf{A}_{3k} = \mathbf{y} - \mathbf{X}_1\boldsymbol{\beta} - \sum_{k\neq k'}^{p}\lambda_{k'} diag(\mathbf{X}_2\alpha_{k'})\mathbf{Z}_f\mathbf{f}_{k'} - \mathbf{Z}\boldsymbol{\delta}$ and $\boldsymbol{\Delta}_k = \lambda_k diag(\mathbf{Z}_f\mathbf{f}_k)\mathbf{X}_2$.

Although the prior $\boldsymbol{\alpha}_k$ presented spherical isotropy, it is not possible to obtain a conjugate von Mises-Fisher distribution similar to those obtained in Viele and Srivasan [33], Crossa et al. [17], or Oliveira et al. [18] for the AMMI model or in Crossa et al. [23] and Oliveira et al. [34] for the GGE model. This fact is because the previous-mentioned studies assumed the homogeneity of variances, which is different from the FA proposed here in which different variances were assumed for the environmental network. Instead, the conditional posterior will be a multivariate normal, which can violate the spectral decomposition constraints such as orthonormal eigenvectors.

To overcome these difficulties, the sampling will be performed in the corrected space free of the orthogonality constraint by using auxiliary variables (which will be defined in the further section); the vectors will be returned in the correct subspace in $\Re^m$ through orthogonal linear transformations.

iv) Conditional a posteriori distribution for factor scores $\mathbf{f}$

The complete posterior conditional distribution for $\mathbf{f}$ is given by

$$\mathbf{f}|\ldots \sim \mathbf{N}[(\mathbf{I}+\mathbf{A}_4{}^{\top}\mathbf{R}^{-1}\mathbf{A}_4)^{-1}\mathbf{A}_4{}^{\top}\mathbf{R}^{-1}\mathbf{A}_5;(\mathbf{I}+\mathbf{A}_4{}^{\top}\mathbf{R}^{-1}\mathbf{A}_4)^{-1}] \tag{11}$$

where $\mathbf{A}_4 = \sum_{k=1}^{p} \lambda_k diag(\mathbf{X}_2\boldsymbol{\alpha}_k)\mathbf{Z}_f$ and $\mathbf{A}_5 = \mathbf{y}–\mathbf{X}_1\boldsymbol{\beta}–\mathbf{Z}\boldsymbol{\delta}$.

v) Complete conditional distribution a posteriori for specific variances $\boldsymbol{\delta}$

$$\boldsymbol{\delta}|\ldots \sim N[(\mathbf{Z}^{\top}\mathbf{V}^{-1}\mathbf{Z}+\mathbf{M}^{-1})^{-1}\mathbf{Z}^{\top}\mathbf{R}^{-1}\mathbf{A}_7,(\mathbf{Z}^{\top}\mathbf{V}^{-1}\mathbf{Z}+\mathbf{M}^{-1})^{-1}] \tag{12}$$

where $\mathbf{A}_6 = \sum_{k=1}^{p} \lambda_k diag(\mathbf{X}_2\boldsymbol{\alpha}_k)\mathbf{Z}_f\mathbf{f}_k$, $\mathbf{A}_7 = \mathbf{y}–\mathbf{X}_1\boldsymbol{\beta}–\mathbf{A}_6$ and $\mathbf{M} = \mathbf{I}_m \otimes \boldsymbol{\Psi}$.

vi) Complete conditional a posteriori distribution of $\boldsymbol{\Psi}$

Since $\boldsymbol{\Psi}$ is diagonal matrix with the independent elements $\boldsymbol{\Psi} = diag(\psi_{11},\ldots,\psi_{kk})$ with $k = 1,\ldots,p$, it was considered a scaled inverse chi-squared distribution for each diagonal element. The complete conditional distribution posterior for each ψ_{kk} is

$$\psi_{kk}|\ldots \sim Scale-\chi^{-2}\left(m_k+v_k,\frac{m_k(\boldsymbol{\delta}^{\top}\boldsymbol{\delta}/m_k)+v_kS_k^2}{m_k+v_k}\right) \tag{13}$$

where m_k is the number of genotypes in the environment and $v_k = 1$.

vii) Complete conditional a posteriori distribution for $\sigma^2_{e_k}$

Similar to $\boldsymbol{\Psi}$, the $\mathbf{R}$ is also a diagonal matrix and the independent priors also have scaled inverse chi-squared distributions. The conditional posterior distribution obtained for each $\sigma^2_{e_k}$ is given by

$$\sigma^2_{e_k}|\ldots \sim Scale-\chi^{-2}\left[n_k,\frac{(\mathbf{y}_k-\boldsymbol{\theta}_k)^{\top}(\mathbf{y}_k-\boldsymbol{\theta}_k)}{n_k}\right] \tag{14}$$

where n_k is the number of observations in each environment.

Sampling parameters requesting orthonormal basis. As all conditional distributions were obtained in a closed form, the distributions have known shapes that allow for direct sampling using the Gibbs sampler. However, as was previously highlighted, the conditional distribution for the vector $\boldsymbol{\alpha}_k$ is a multivariate Gaussian distribution instead of a von Mises-Fisher. Nevertheless, this distribution and its parametric space do not ensure the model's constraints since the vectors must have unitary norms and be orthogonal to each other.

However, the sampling can still be performed from the normal multivariate in the corrected subspace by adding two further steps: normalize the $\boldsymbol{\alpha}_k$ vector and return it into the orthogonal basis using the correct subspace through a linear transformation. Viele and Srinivasan [33] have shown how to perform the sampling of the spherical uniform distribution using the standardized Gaussian distribution and how the vectors can be placed in the correct subspace using the Gram–Schmidt orthonormalization process.

Assuming that $\boldsymbol{\alpha}_k$ is an unit vector of *m-s* dimension in space $\Re^m$, ($\boldsymbol{\alpha}_k^{\top}\boldsymbol{\alpha}_k = 1$) is orthogonal to vector $S\ (v_1, v_2, \ldots, v_s)$ in $\Re^m$. Further suppose that the vectors $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\ldots,\boldsymbol{\alpha}_s$ are a set of orthonormal vectors in the subspace generated by $v_1, v_2, \ldots, v_s$. In these terms, given the matrix $\mathbf{H}_s = \boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\ldots,$

$\boldsymbol{\alpha}_s$, there is a matrix $\mathbf{H}_k$ with the dimensions $m{\times}(m{-}s)$ so that $\mathbf{H}_m = (\mathbf{H}_s,\mathbf{H}_k)$ is an orthonormal matrix. This matrix can be obtained by the *Gram-Schmidt* orthonormalization process.

Thereby, we obtained the vector $\tilde{\boldsymbol{\alpha}}_k$ by the linear transformation $\tilde{\boldsymbol{\alpha}}_k = \mathbf{H}_k^\top \boldsymbol{\alpha}_k$ so that $\tilde{\boldsymbol{\alpha}}_k \in \Re^{m-s}$. In other words, the sampling will be performed in the "corrected" subspace without the constraints imposed by the spectral decomposition. We can easily show that $\boldsymbol{\alpha}_k = \mathbf{H}_k\tilde{\boldsymbol{\alpha}}_k$. Therefore, we retrieve the vector in the correct subspace in $\Re^m$ and orthogonal to s vectors by applying the inverse operation, satisfying the model's constraints.

The posterior distribution for the sampling process can be obtained from the kernel of the full conditional posterior distribution of $\boldsymbol{\alpha}_k$ as follows:

$$\exp\left\{-\frac{1}{2}\left(\left[\boldsymbol{\alpha}_k - (\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)^{-1}\boldsymbol{\Delta}^\top\mathbf{R}^{-1}\mathbf{A}_{3k}]^\top(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)[\boldsymbol{\alpha}_k - (\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)^{-1}\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\mathbf{A}_{3k}\right]\right)\right\}$$

By solving within the brackets, we have

$$\begin{aligned}&\boldsymbol{\alpha}_k^\top(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)\boldsymbol{\alpha}_k - \boldsymbol{\alpha}_k^\top(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)(\boldsymbol{\Delta}^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)^{-1}\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\mathbf{A}_{3k}-\\&-[(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\mathbf{A}_{3k}]^\top(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)\boldsymbol{\alpha}_k+\\&+[(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)^{-1}\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\mathbf{A}_{3k}]^\top(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)^{-1}\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\mathbf{A}_{3k}\end{aligned}$$

Using the identity $\mathbf{H}_k\mathbf{H}_k^\top = \mathbf{I}$ and dividing each part of the kernel, we get

i. $\boldsymbol{\alpha}_k^\top\mathbf{H}_k\mathbf{H}_k^\top(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)\mathbf{H}_k\mathbf{H}_r^\top\boldsymbol{\alpha}_k = (\mathbf{H}_k^\top\boldsymbol{\alpha}_k)^\top\mathbf{H}_r^\top(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)\mathbf{H}_k\mathbf{H}_k^\top\boldsymbol{\alpha}_k =$

$= (\tilde{\boldsymbol{\alpha}}_k)^\top\mathbf{H}_k^\top(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)\mathbf{H}_k\tilde{\boldsymbol{\alpha}}_k$

ii. $-\boldsymbol{\alpha}_k^\top(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)^{-1}\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\mathbf{A}_{3k} =$

$= -\boldsymbol{\alpha}_k^\top\mathbf{H}_k\mathbf{H}_k^\top(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)\mathbf{H}_k\mathbf{H}_k^\top(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)^{-1}\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\mathbf{A}_{3k} =$
$= -\tilde{\boldsymbol{\alpha}}_k^\top\mathbf{H}_k^\top(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)\mathbf{H}_k\mathbf{H}_k^\top(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)^{-1}\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\mathbf{A}_{3k}$

iii. $-[(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)^{-1}\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\mathbf{A}_{3k}]^\top(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)\mathbf{H}_k\mathbf{H}_k^\top\boldsymbol{\alpha}_k =$

$= -[(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)^{-1}\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\mathbf{A}_{3k}]^\top(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)\mathbf{H}_k\mathbf{H}_k^\top\boldsymbol{\alpha}_k =$
$= -[(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)^{-1}\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\mathbf{A}_{3k}]^\top\mathbf{H}_k^\top(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)\mathbf{H}_k\tilde{\boldsymbol{\alpha}}_k$

iv. $[(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)^{-1}\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\mathbf{A}_{3k}]^\top(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)^{-1}\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\mathbf{A}_{3k} =$

$= [(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)^{-1}\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\mathbf{A}_{3k}]^\top\mathbf{H}_k\mathbf{H}_k^\top(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)\mathbf{H}_k\mathbf{H}_k^\top(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)^{-1}\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\mathbf{A}_{3k}$

In this way,

$$\begin{aligned}&\boldsymbol{\alpha}_k^\top(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)\boldsymbol{\alpha}_k - \boldsymbol{\alpha}_k^\top(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)^{-1}\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\mathbf{A}_{3k}-\\&-[(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)^{-1}\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\mathbf{A}_{3k}]^\top(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)\boldsymbol{\alpha}_k+\\&+[(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)^{-1}\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\mathbf{A}_{3k}]^\top(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)^{-1}\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\mathbf{A}_{3k} =\\&= (\tilde{\boldsymbol{\alpha}}_k)^\top\mathbf{H}_k^\top(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)\mathbf{H}_k\tilde{\boldsymbol{\alpha}}_k - \tilde{\boldsymbol{\alpha}}_k^\top\mathbf{H}_k^\top(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)\mathbf{H}_k\mathbf{H}_k^\top(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)^{-1}\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\mathbf{A}_{3k}-\\&-[(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)^{-1}\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\mathbf{A}_{3k}]^\top\mathbf{H}_k\mathbf{H}_k^\top(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)\mathbf{H}\tilde{\boldsymbol{\alpha}}_k+\\&[(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)^{-1}\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\mathbf{A}_{3k}]^\top\mathbf{H}_k\mathbf{H}_k^\top(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)\mathbf{H}_k\mathbf{H}_k^\top(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)^{-1}\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\mathbf{A}_{3k}\end{aligned}$$

Thus, the conditional posterior for the auxiliary variable $\tilde{\boldsymbol{\alpha}}_k$ given the other parameters in the corrected subspace is given by

$$\tilde{\boldsymbol{\alpha}}_k|\ldots \sim N[\mathbf{H}_k^\top(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)^{-1}\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\mathbf{A}_{3k}, (\mathbf{H}_k^\top(\boldsymbol{\Delta}_k^\top\mathbf{R}^{-1}\boldsymbol{\Delta}_k)\mathbf{H}_k)^{-1}].$$

Therefore, the sampling of $\tilde{\boldsymbol{\alpha}}_k$ is performed in the corrected subspace through the previously obtained conditional. As presented in Crossa et al. [17] and Oliveira et al. [18], we seek the sample vectors that have the norm 1. Thus, the orthogonal vectors must be normalized as $\boldsymbol{\alpha}_k^* = \frac{\tilde{\boldsymbol{\alpha}}_k}{\sqrt{\tilde{\boldsymbol{\alpha}}_r^\top \tilde{\boldsymbol{\alpha}}_k}}$. Assuming $\bar{\tilde{\boldsymbol{\alpha}}}_k = \mathbf{H}_k^\top (\boldsymbol{\Delta}_k^\top \mathbf{R}^{-1} \boldsymbol{\Delta}_k)^{-1} \boldsymbol{\Delta}_k^\top \mathbf{R}^{-1} \mathbf{A}_{3k}$ and $C_k = \sqrt{\tilde{\boldsymbol{\alpha}}_r^\top \tilde{\boldsymbol{\alpha}}_k}$, through algebraic manipulations, the conditional distribution for the orthonormal $\tilde{\boldsymbol{\alpha}}_k$ is given by

$$\tilde{\boldsymbol{\alpha}}_k | \ldots \sim N[\boldsymbol{\alpha}_k^*, (\mathrm{C}_k^\top \mathbf{H}_k^\top (\boldsymbol{\Delta}_k^\top \mathbf{R}^{-1} \boldsymbol{\Delta}_k) \mathbf{H}_k C_k)^{-1}]$$

where $\boldsymbol{\alpha}_k^* = \frac{\tilde{\boldsymbol{\alpha}}_k}{C_k}$.

To place the singular vector in the correct subspace $\Re^m$ satisfying the orthonormal constraints, we apply the inverse transformation $\boldsymbol{\alpha}_k = \mathbf{H}_k \tilde{\boldsymbol{\alpha}}_k$. The sampled vector is now orthogonal to the other *s* vectors and its transformation preserves the vector norm since

$$(\tilde{\boldsymbol{\alpha}}_k)^\top \tilde{\boldsymbol{\alpha}}_k = (\mathbf{H}_k^\top \boldsymbol{\alpha}_k)^\top \mathbf{H}_k^\top \boldsymbol{\alpha}_k = \boldsymbol{\alpha}_k^\top \mathbf{H}_R \mathbf{H}_R^\top \boldsymbol{\alpha}_k = \boldsymbol{\alpha}_k^\top \mathrm{I}_m \boldsymbol{\alpha}_k^\top = \boldsymbol{\alpha}_k^\top \boldsymbol{\alpha}_k = 1.$$

Thus, the random *m-s* dimensional vector in $\Re^m$ isone-to-one transformation into the same random vector in $\Re^{m-s}$.

From the previously complete conditional distributions, the parameter sampling was performed by the Markov chain Monte Carlo (MCMC) using the Gibbs sampler. The implemented iterative sampling algorithm is illustrated in S3 Text.

After concluded the iterative process and checked the chains' convergence, the samples were considered to have resulted from the marginal densities. The convergence diagnostic was performed using the Raftery and Lewis [35] Heidelberger and Welch [36] criterion. All inference processes were performed using the R statistical software [37].

Inference about the linear and multiplicative parameters of the model. The sampled eigenvectors present orthonormal basis across MCMC process, however, their maximum posterior estimator (MAP) may not be orthogonal. The MAP estimator for the multiplicative terms were constructed according to the method proposed by Chen and Shao [38] implemented in the Bayesian output analysis (BOA) package using the R statistical software (R CORE TEAM, 2016).

Bivariate regions of credibility for factor loadings and scores. The biplot credibility regions for factor loadings ($\lambda_1\boldsymbol{\alpha}_1$,$\lambda_2\boldsymbol{\alpha}_2$) and factor scores ($\mathbf{f}_1$,$\mathbf{f}_2$) were constructed using the Euclidean distances of the sampling points with respect to the distribution center using 5% as the cut off [39].

The FA biplot interpretation was performed similarly to the GGE-biplot, as suggested by Burgueño et al. [7].

Model validation in the prediction of missing data. The cross-validation process was performed considering different levels of missing data in the GEI matrix. The sample was randomly divided into *k*-fold of equal size, with *k* = (10, 3, 2) corresponding respectively to the 10%, 33% and 50% levels of random genotype losses in the environments without replacement. Thus, the simulated missing was performing on GEI table cells where some genotypes (lines) information were totally withdraw from specific environments (columns), but keeping all environments in the dataset. In addition, the cross-validation was performed to allow that all genotypes were evaluated in at least one environment. Therefore, the GEI cells were randomly sampled, but observing the restrictions given above.

The BFA's predictive ability was compared to the two-step FA models using the EM algorithm (FA-EM—expectation maximization) (Nuvunga et al. [40] and the FA via AI algorithm (Average information) (FA-AI) (Smith, Cullis and Thompson [6]) in sparse matrices implemented in Asreml-R [41].

The evaluation of the model's predictive ability was performed using the average PRESS (*predicted residual error sum of squares*) and phenotypic correlation between the predicted

$(\hat{y}_{ij})$ and observed (y_{ij}) values. The PRESS expression is given by

$$PRESS = \frac{1}{n}\sum_{j=1}^{n}(y_{ij} - \hat{y}_{ij})^2 \tag{15}$$

Model selection or choice of the number of factors *k*. The number of latent factors to be retained in the model was selected using the PRESS criterion, which uses a cross-validation approach [42].The statistical efficiency criterion is given by $SE = (PRES_{full})/(PRESS_k)$, which is the ratio between the PRESS of the full model and the PRESS of low-order models.

The model selection for real data was done using the Akaike Information Criterion Monte Carlo (AICM). $\Delta AICM$ corresponds to the difference between the full model and the competing models, as suggested by Raftery et al. [43]. The AICM is calculated as a version based on a posteriori simulation of the AIC [44].

$$AICM = 2(\bar{l} - s_l^2) \tag{16}$$

where $\bar{l}$ is the mean of the marginal log-likelihood and s_l^2 is the posterior variance of the marginal log-likelihood.

Thus, the AICM can be seen as the simplified version of the penalized posterior mean of the log-likelihood. The selected model is the one with the highest AICM and lowest $\Delta AICM$.

Results

Simulated data

For this scenario, MCMC chains with 65,560 iterations were simulated for the BFA model. As already pointed out, the convergence of the generated chains was monitored by the criteria Raftery and Lewis and Heidelberger and Welch. The first 8,400 observations were burned-in and a thinning for each four observations was performed to ensure the convergence process. The values of burning and thinning were based on a training sample according to the test of Raftery and Lewis. It was also observed that all parameters had a dependence factor I <5. The final chain length was 14,290 samples for each parameter.

In addition, all parameters passed the stationarity test, indicating that convergence was achieved according to the criterion of Heidelberger and Welch and Geweke. That is, the tests indicated good convergence properties for all model parameters. In S4 Text, the traces for the residual variance chains are shown and its pattern corroborates the convergence test results. Another interesting detail, in the Bayesian context, is the computational time of analysis, which for the simulated data was 22.15 minutes.

The simulated parametric values for the residual variance and FA loadings τ_k (recovered by $\lambda_k\alpha_k$) and the MAP estimates are presented in Table 1. We notice that the estimated values and parametric values do not present large differences, which support the BFA's ability in estimating the loadings of FA models. All values used in the simulations are within the 95% credibility intervals.

Specific estimates and credibility confidence regions for the coordinates related to the first two factor scores can be seen in (S2 Table). The posterior estimates presented values very close to those from FA mixed models obtained by the restricted maximum likelihood (REML) estimates using the average information (FA-AI) algorithm. However, the second axes for both methods present large differences for G1, G2 and G5, but they are within the credibility intervals.

The bivariate credibility regions (highest posterior density—HPD) for the factor loadings $(\lambda_1\boldsymbol{\alpha}_1,\lambda_2\boldsymbol{\alpha}_2)$ and genotypic scores $(\mathbf{f}_1,\mathbf{f}_2)$ that did not included the biplot origin (0,0) are shown

Table 1. **Bayesian maximum a posterior (MAP), simulated parametric value (PV), posterior standard deviation (PSD), credibility intervals (CI 95%, LL: Lower limit, UL: Upper limit).**

				HPD 95%	
Parameter	**PV**	**MAP**	**PSD**	**LL**	**UL**
$\sigma^2_{e_1}$	0.546	0.847	1.260	0.299	1.506
$\sigma^2_{e_2}$	1.209	1.911	1.155	0.846	3.162
$\sigma^2_{e_3}$	4.690	4.381	2.353	2.209	6.826
$\sigma^2_{e_4}$	7.377	10.091	19.936	4.938	15.097
$\sigma^2_{e_5}$	9.026	9.104	3.807	3.503	15.793
σ^2_{e}	2.900	2.848	-	-	-
$\lambda_1\alpha_{11}$	1.468	1.548	0.845	0.500	3.448
$\lambda_1\alpha_{21}$	1.314	1.344	0.761	0.291	3.062
$\lambda_1\alpha_{31}$	2.404	2.353	1.267	0.799	4.916
$\lambda_1\alpha_{31}$	1.771	1.475	1.008	-0.439	3.817
$\lambda_1\alpha_{51}$	2.751	2.191	1.311	0.180	5.022
$\lambda_2\alpha_{12}$	0.579	0.381	0.603	-0.845	1.521
$\lambda_2\alpha_{22}$	0.385	0.436	0.374	0.000	1.154
$\lambda_2\alpha_{32}$	0.257	0.234	0.678	-1.093	1.657
$\lambda_2\alpha_{42}$	1.717	0.307	0.890	-1.466	2.113
$\lambda_2\alpha_{52}$	-1.823	-0.764	1.154	-2.726	1.748

Residual variance mean ($\sigma^2_{e_k}$) simulated factor loadings τ_k and recovered by the BAF model ($\tau_k = \lambda_k\alpha_k$).

https://doi.org/10.1371/journal.pone.0220290.t001

in Fig 1. From these biplots, we see the clustering groups with respect to grain yields and/or GEI patterns. In FA models where G is confounded with GE, the first loading tends to present a positive signal (Fig 1B). According to Burgueño et al. [7], Smith, Cullis and Thompson [6] and Stefanova and Buirchell [29], in these situations, the interpretation for the FA biplot is similar to the GGE-biplot. However, it is necessary to be careful when interpreting the intersections among the credibility regions, given that loadings and scores do not have inner-product properties. If so, the biplots are better justified with separate presentations for loading and factor scores to avoid confusion with the AMMI or GGE models.

However, just to illustrate the usefulness of the BFA model, we plotted both factor loadings and scores in a single biplot (Fig 1A). In this Figure, it is possible to see two distinct groups of genotypes with respect to yields, but they are similar with respect to the GEI pattern since all regions cross the first axis.

Fig 1B shows the HPD's credible regions for factor loadings at 95% confidence. This subgroup of environments has similar effects with respect to the GEI pattern, as indicated by the overlaps between the credibility regions and ellipses that did not encompassed the biplot origin.

Given that the first axis captures much of the main effect of genotypes and the second axis captures the complex part related to the GEI, it was observed that the scale and ranking of the factor scores were equivalent to the marginal E-BLUPS obtained from the mixed model analysis since the regression adjustment between the two estimates was approximately 1 ($r^2 = 0.979$).

Performance evaluation of the unbalanced model. In addition to evaluating the BFA model in terms of parameter estimations, the Bayesian FA model's predictive ability was evaluated and compared to the two step FA model [40], and FA-AI [41]. Although it is difficult to adjust a full FA model in the mixed models framework due to computational costs and

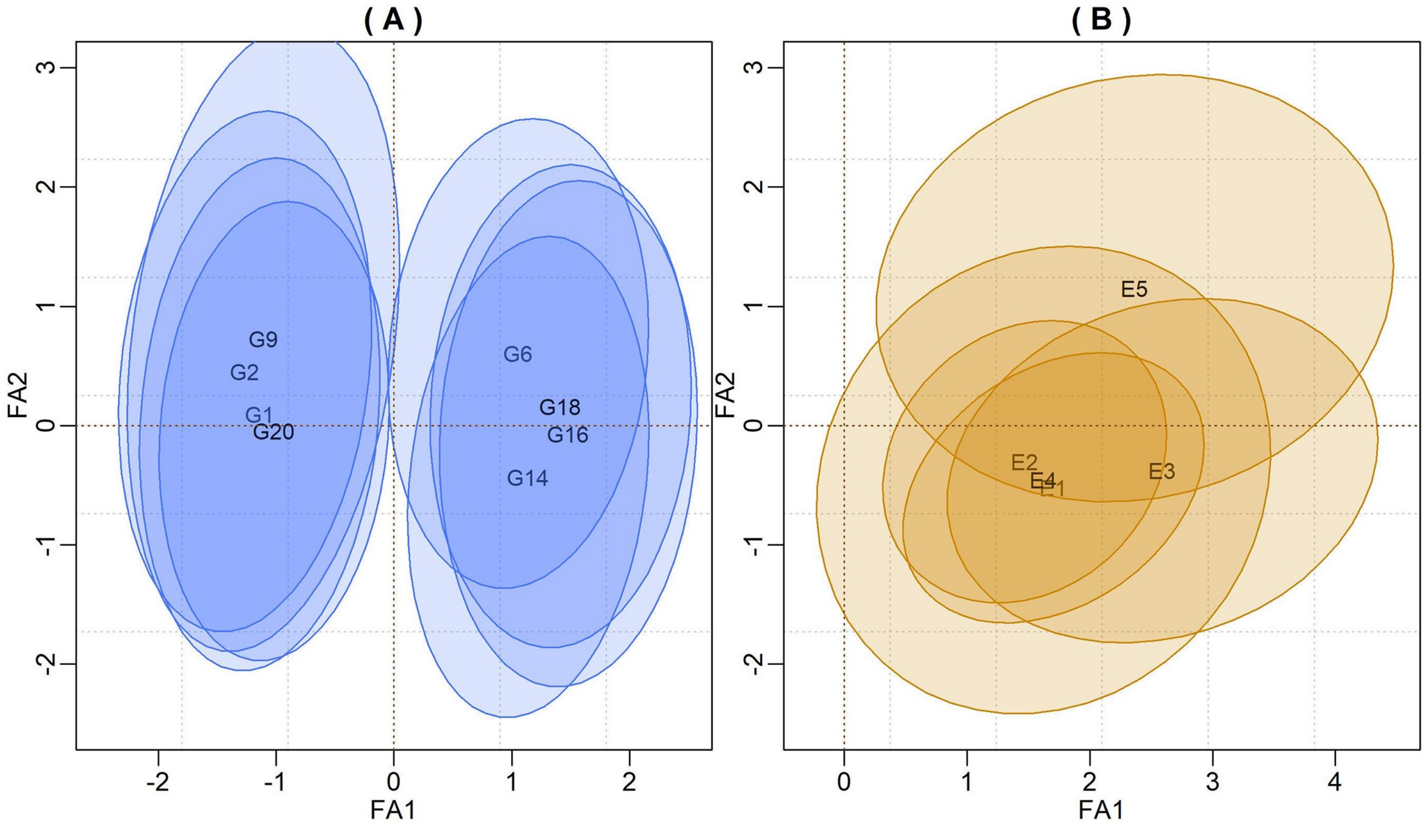

Fig 1. Credibility regions at 95% probability for genotypic factor scores (Fig 1A) and for the factor loadings of environments (Fig 1B) using the first two components. Only the genotypes scores that did not include the biplot origin were represented.

https://doi.org/10.1371/journal.pone.0220290.g001

convergence problems, this problem was not observed for the simulated dataset, given that the set of environments is relatively lower compared to the data set commonly utilized in MET.

The cross-validation results showed that it is possible to predict the performance of hybrids using FA models with high accuracy, reaching up to 0.82 in some folds, as explained in S1–S6 Figs.

Regardless of the unbalanced level applied to the hybrid panel, the magnitude of the correlation values was higher than 0.30 (Fig 2) and the Bayesian model showed the highest prediction ability for all scenarios.

The PRESS in the10-fold scenario (as expected) was lowest when compared to the other folds for the three models (Fig 3). At all considered unbalanced levels, the BAF model had the lowest PRESS. At 10%, the FA-AI and FA-EM models had the same PRESS and alternated precision at the 33% and 50% levels. In other words, the scenarios proposed to PRESS using these two FA approaches were inconclusive.

FA model selection based on cross validation. The numbers of latent factors to be retained in the model were selected using a 10-fold cross-validation. Therefore, the FA5, FA4, FA3, FA2 and FA1 models were adjusted and the model selection was based on the PRESS criterion and statistical efficiency (SE) (defined as the ratio between the full FA model versus the low-dimension ones).

Fig 4B and 4A present the PRESS and the correlation between the simulated phenotypic value and the predicted one using each model. Observing Fig 4B, one can verify that the FA4 model (with $k = 4$) presented the highest predictive accuracy compared to the others candidate

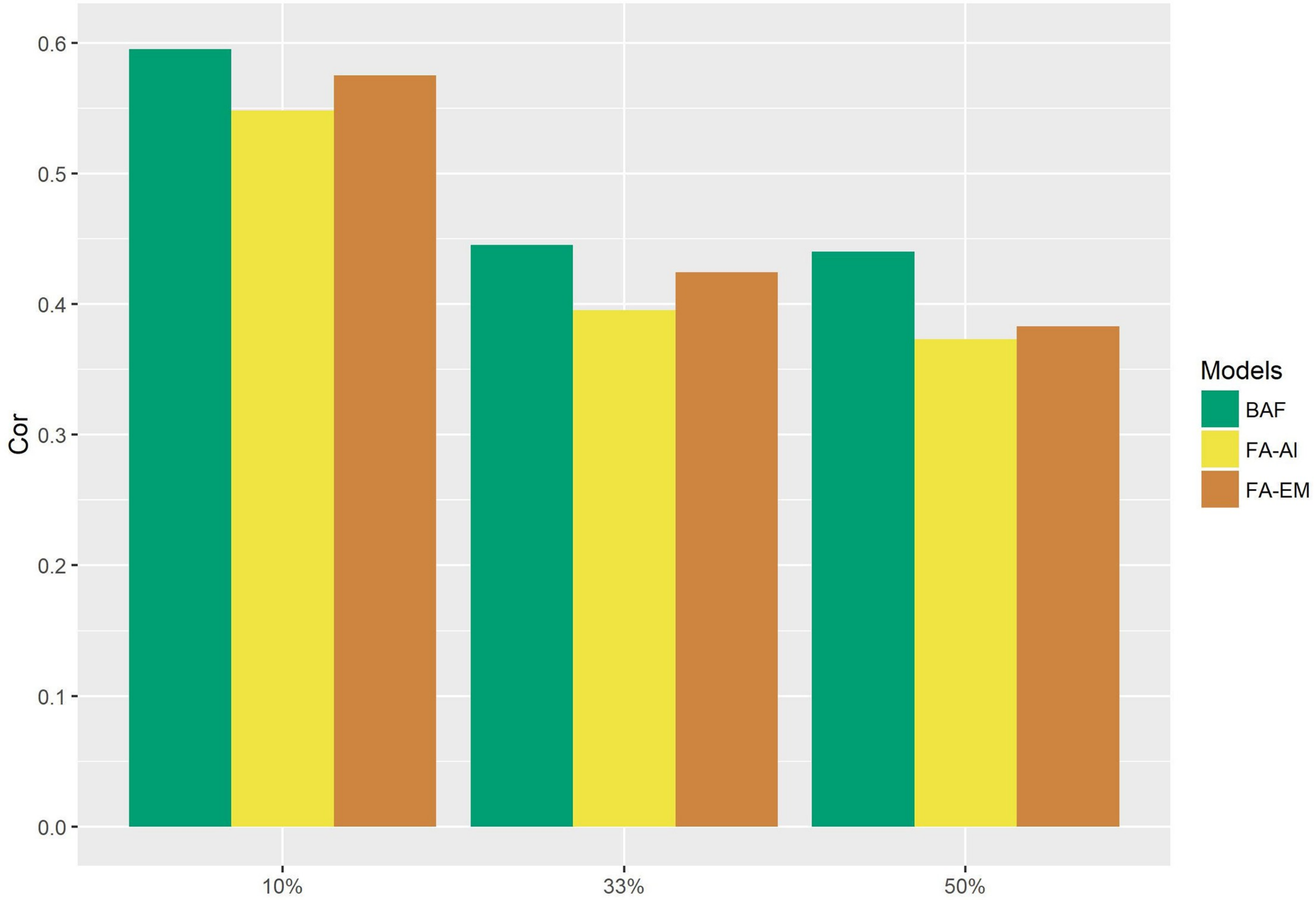

Fig 2. Bar chart related to the correlation for 10-fold, 3-fold and 2-fold scenarios using the Bayesian FA (BAF) and FA models (FA-EM and FA-AI) for simulated data.

https://doi.org/10.1371/journal.pone.0220290.g002

models, indicating that this model is the best one. Additionally, using the PRESS criterion, the best model again was FA4 (Fig 4A).

Fig 4C shows the graphic of the statistical efficiency of the competing models. In this graphic, it is notable that the efficiency increases from FA1 to FA4 where we find the maximum efficiency. This demonstrates that the use of $k = 4$ would be the best choice for these data representation.

The FA2 model is generally considered parsimonious and interpretable for plant breeders, since the first axis can be seen as related to adaptability and the second related to stability, such as in the SREG2/GGE model [7,29]. However, it was verified that the FA4 model showed better results in the three used criteria.

Experimental data

For the real data scenario, 85,000 Markov chains were simulated and (similar to the previous analysis), the first 9,800 observations were burned-in and thinned for every eight samples. A final MCMC chain with 9,400 observations was obtained for each parameter. The chain convergence was verified by Raftery and Lewis [35], Heidelberger and Welch [36] and and

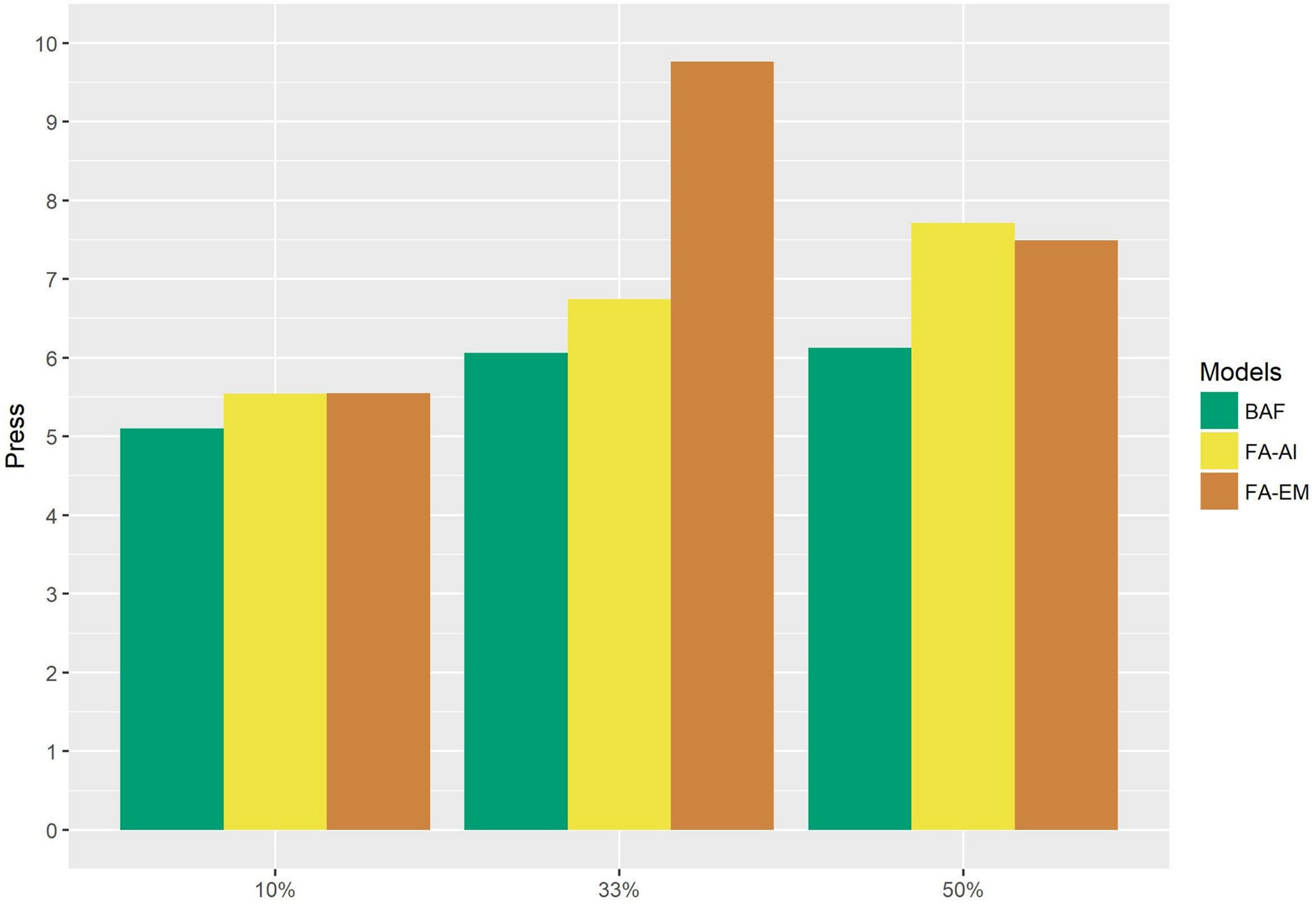

Fig 3. Bar chart related to the PRESS for 10-fold, 3-fold and 2-fold scenarios using the Bayesian FA (BAF) and FA models (FA-EM and FA-AI) for simulated data.

https://doi.org/10.1371/journal.pone.0220290.g003

Geweke criterion [45]. The trace pattern related to the posterior distributions also indicated good convergence. The computational time spent in the MCMC sampling process was 13,938.34 minutes.

The point and interval estimates for the components of variance σ_k^2 are presented in Table 2. It is possible to notice that the environments E8 and E2 presented the highest and lowest residual variances, respectively.

Fig 5A shows the HPD credibility regions at the 95% probability for the genotypic factor scores. For simplicity, only genotypes that did not include the biplot origin (0,0) were represented. A correlation between the first factor scores and the marginal genotypic BLUPs was $r^2 = 0.897$ in the Bayesian FA model and $r^2 = 0.929$ for the FA mixed model, which would also justify the GGE biplot interpretation.

Moreover, for the real data scenario, all factor loadings related to environments were positive and with high overlapping for the credibility intervals (Fig 5B). In this same Figure, it is observed that environments with low residual variances (Table 2) show more concentrated credible regions with respect to those with higher residual variances. Thus, the elliptical range depends on each specific experimental variance. Estimates for factor loadings and factor scores can be found in S3 and S4 Tables.

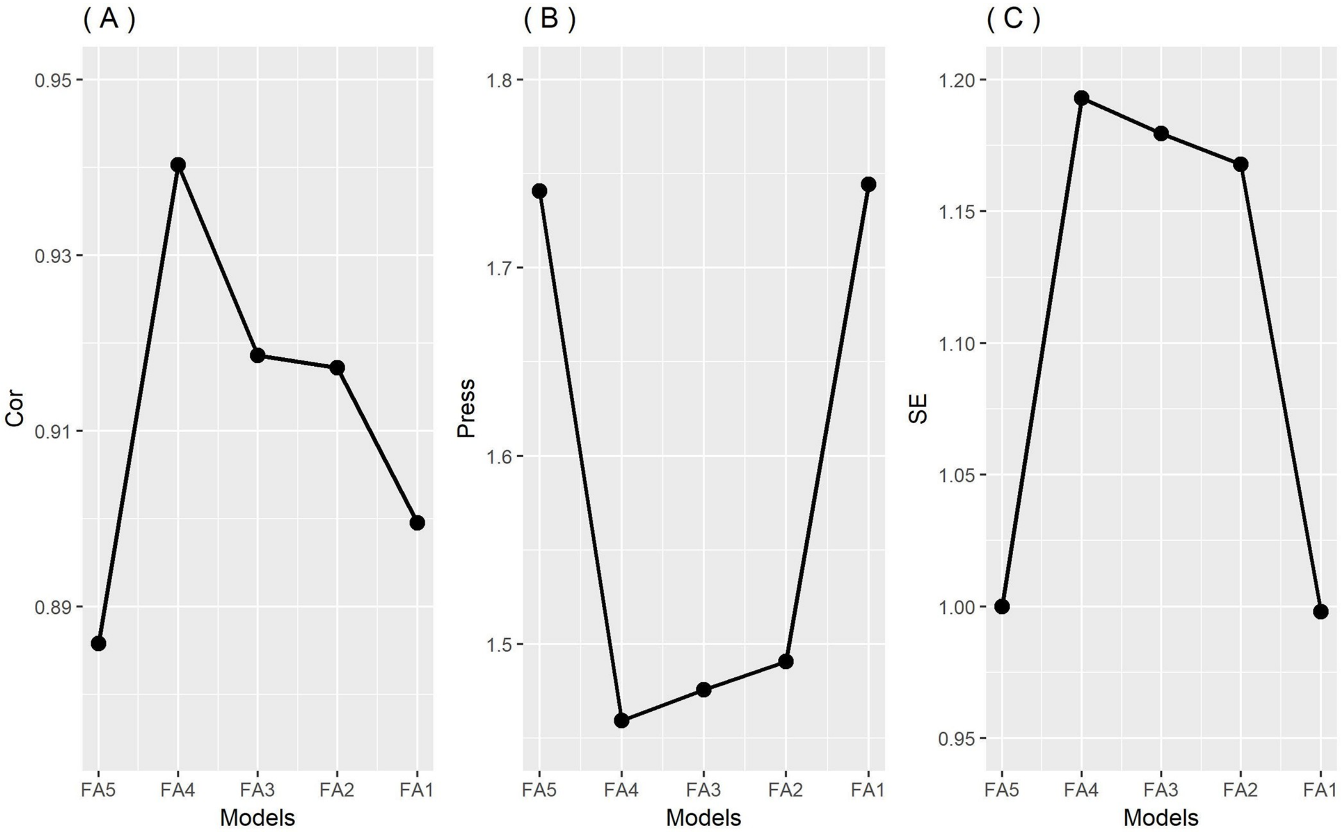

Fig 4. Ockham's plot referring to the BAF model performance using the correlation between the observed and predicted phenotypic values (A) (higher, better), the predicted sum square (PRESS) (B) (lower, better) and the statistical efficiency (SE) (C) (higher, better) for simulated data.

https://doi.org/10.1371/journal.pone.0220290.g004

Figs 6 and 7 show a brief comparison of the biplot results obtained from the mixed model FA(2) method and our BFA model. Both methods produced similar patterns in the biplots separating the genotypes {G32, G35, G36, G37} from the genotypes {G1, G2, G3, G5, G6, G7, G8, G10, G38}, where each group was clustered in opposite biplot quadrants.

In these same Figures, it was verified that the estimates of the two models were coincident for factor scores and factor loadings. Similarities can be observed when comparing the

Table 2. Posterior means (PM), credibility regions (CI 95%, LL: Lower limit, UL: Upper limit) for the residual variance ($\sigma^2_{e_k}$) obtained by the BFA model for real data.

			HPD 95%	
Parameter	PM	SD	LL	UL
$\sigma^2_{e_1}$	3.4642	0.2869	2.9392	3.9960
$\sigma^2_{e_2}$	1.4465	0.2187	1.2076	1.6734
$\sigma^2_{e_3}$	1.9769	0.1655	1.6827	2.2918
$\sigma^2_{e_4}$	1.6672	0.1634	1.3776	1.9581
$\sigma^2_{e_5}$	3.2711	0.3030	2.7526	3.7986
$\sigma^2_{e_6}$	1.8623	0.1538	1.5850	2.1727
$\sigma^2_{e_7}$	2.1392	0.1968	1.8324	2.4674
$\sigma^2_{e_8}$	4.0689	0.2984	3.4846	4.6402
$\sigma^2_{e_9}$	1.6871	0.1472	1.4090	1.9773
$\sigma^2_{e_{10}}$	3.3718	0.2624	2.8713	3.8595

https://doi.org/10.1371/journal.pone.0220290.t002

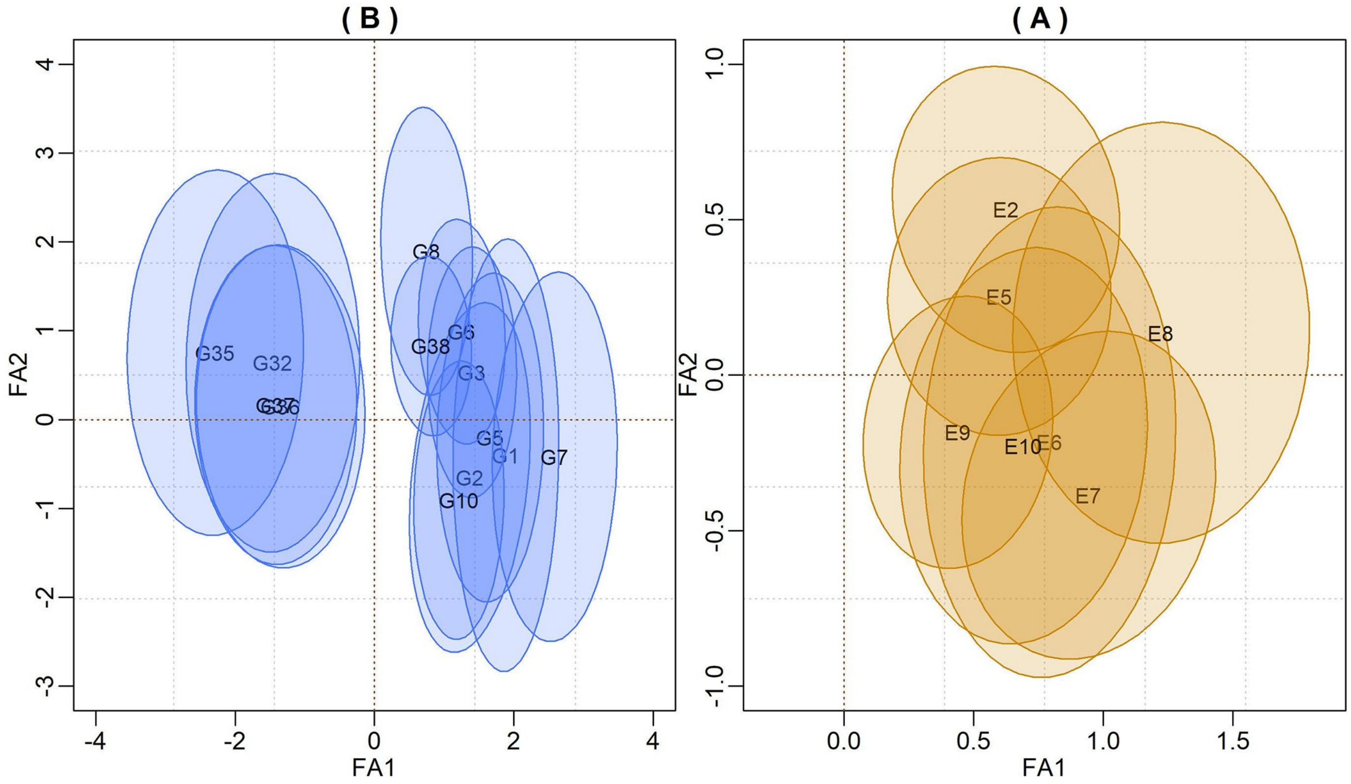

Fig 5. Credibility regions at the 95% probability for genotypic factor scores (a) and for factor loadings of environments (b) using the Bayesian approach for real data. Only the intervals that did not include the biplot origin were represented.

https://doi.org/10.1371/journal.pone.0220290.g005

distribution of environmental scores in standard biplot analyses (Fig 6) compared to those obtained from the Bayesian model. It is possible to verify the change in the E8 position from the FA-mixed model (fourth quadrant) to the Bayesian FA (second quadrant) that may result in small differences in the mega-environments formation and the specific adaptability.

Model selection. In the Bayesian factors analysis, one of the most important issues to be addressed is the choice of the appropriate number of factors to be retained in the model.

In Table 3, we present the log-likelihood AICM values of the ten competing models for the real data set. It was verified that the AICM criterion was unable to select the best FA model since the AICM were practically equal.

In this same table, it is possible to verify the ΔAICM for each model. Similarly, there is not a fair criterion to select the number of factor loadings using this information. The results presented in Table 3 indicate the difficulty of selecting the best model since the criterion differentiation occurs only in the third decimal place, making it necessary to add an additional measure for model selection.

In FA models, the genetic covariance is estimated by $[\Sigma = (\boldsymbol{\Gamma}\boldsymbol{\Gamma}^{\mathrm{T}})^{k-1}+\boldsymbol{\Psi}]$. To ensure the FA model's identifiability, the loading matrix must be conditioned to the following equality $(\boldsymbol{\Gamma}\boldsymbol{\Gamma}^{\mathrm{T}} = \Sigma-\boldsymbol{\Psi})$.

In Fig 8, we can verify that when the full model-FA k is adjusted, the proportion of marginal genetic variance given by the geometric mean $[\sigma_g^2 = \sqrt[10]{(diag|\boldsymbol{\Sigma}|)}]$ is fully recovered by the loadings $[\sqrt[10]{diag(\boldsymbol{\Gamma}\boldsymbol{\Gamma}^{\mathrm{T}})}]$ with a geometric mean of the residual variance given by $[\sigma_e^2 = \sqrt[10]{|\mathbf{R}|}]$. This situation corresponds to an unstructured model. When the one axis is removed from the

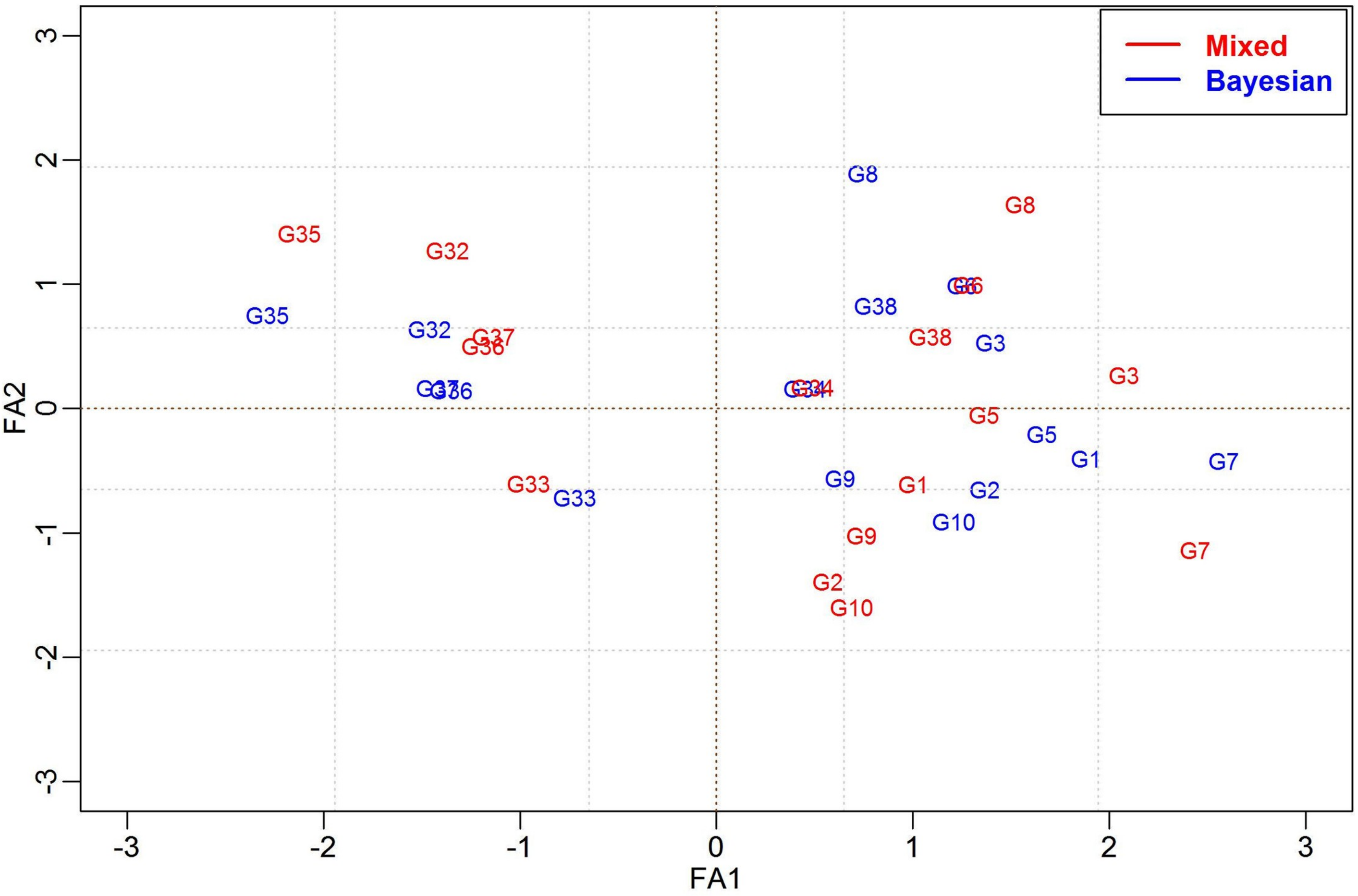

Fig 6. Biplot analysis of genotype scores using the FA-AI model (Red) and the BAF (Blue), considering 50 genotypes evaluated in 10 environments for real data.

https://doi.org/10.1371/journal.pone.0220290.g006

full model, one can verify the decreasing residual variance and the fast increase of the specific variance that merges the genetic and residual (noise) variances.

It can be verified that starting from FA *(k-1)* model to model FA1, the variance σ_e^2 remains constant, but the harmonic mean related to the specific variance $[\psi = \sqrt[10]{diag|\mathbf{\Psi}|}]$ increases as the FA model becomes more parsimonious. In addition, (as expected) the genotypic variance recovered by the loadings $\sqrt[10]{diag(\mathbf{\Gamma}\mathbf{\Gamma}^{\mathrm{T}})}$ decreases. It is possible to note that the use of complex models (high *k*-dimensional FA) has no advantage or gain in the calculation of σ_g^2 or σ_e^2. Thus, under the identifiability imposed in this study, the performance of the model selection tests using the likelihood can be complicated, since the specific variances (or noise) are estimated separately from the experimental error and the missing variance unrecovered by the loading in low-dimension FA structure recovered by the specific variances given in $\mathbf{\Psi}$.

Discussion

The development of models able to describe the response of genotypes in environmental networks has become a great challenge to quantitative breeders, mainly in the genome selection context [46] The use of factorial analytical structures in MET analyses has contributed greatly to the analysis and meta-analysis of phenotypic data in trial networks, presenting unbalanced

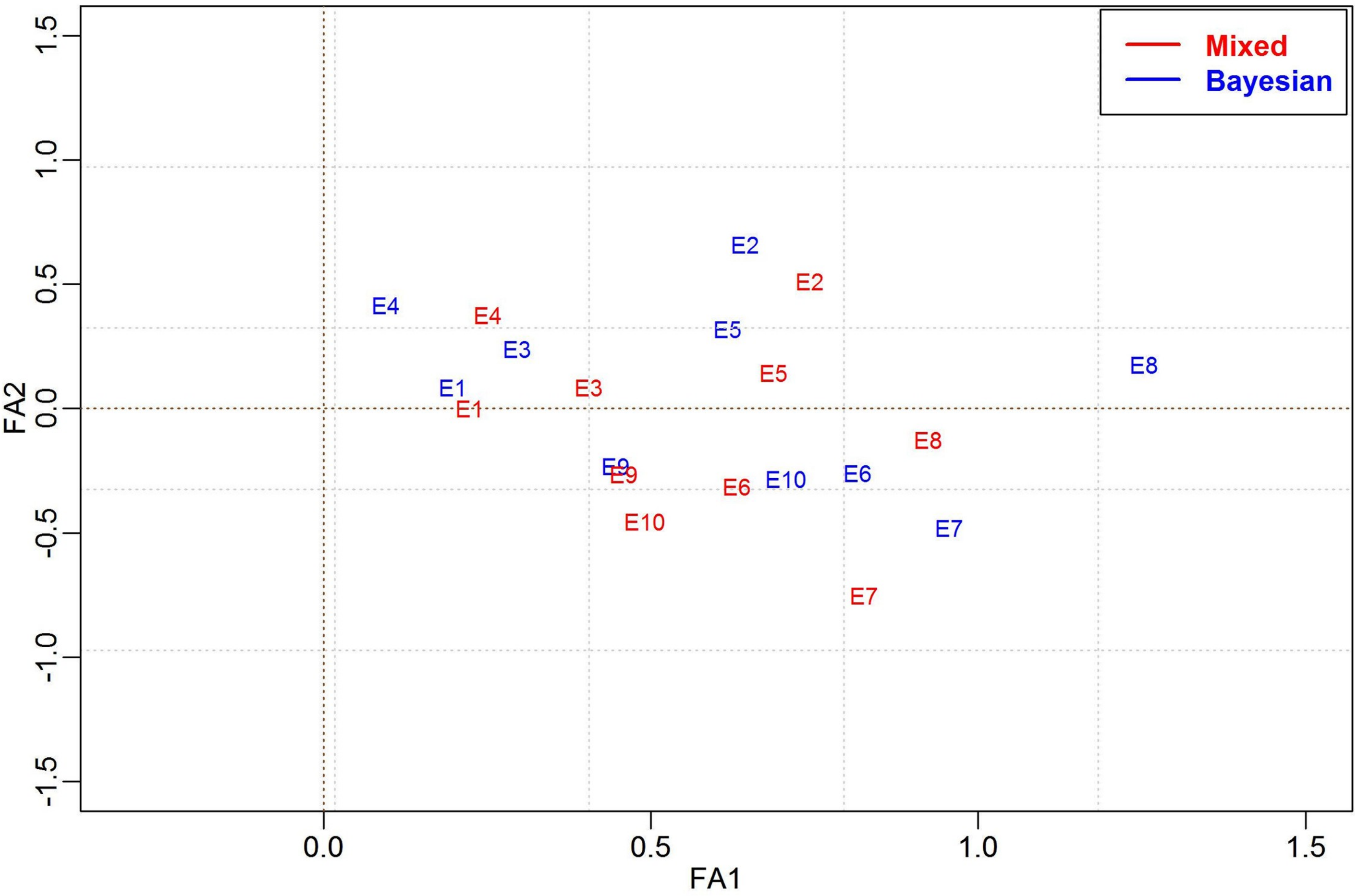

Fig 7. Biplot analysis of environmental scores using the FA-AI model (Red) and the BAF (Blue), considering 50 genotypes evaluated in 10 environments for real data.

https://doi.org/10.1371/journal.pone.0220290.g007

data and different experimental accuracies. The main difference between the BFA method and the classical mixed model FA lies in the BFA model assumptions that are founded on factor analysis via spectral decomposition of the genetic covariance matrix. This approach allows one to incorporate inference in the biplot of the loadings and rotationally, which is not directly performed in classical FA analysis.

Table 3. AICM values and ΔAICM (difference between the AICM of the complete model and the others) for the selection of FAk models, (k = 1, . . ., 10) and the ranking of models for real data.

Model	AICM	Rank	ΔAICM	Rank
FA1	-4.351	1st-FA10	-0.007	1st -FA2
FA2	-4.347	2nd -FA2	-0.003	2nd FA4
FA3	-4.351	3rd- FA4	-0.007	3rd -FA1
FA4	-4.35	4th -FA1	-0.006	4° -FA3
FA5	-4.351	5th -FA3	-0.007	5th -FA5
FA6	-4.351	6th-FA5	-0.007	6th -FA6
FA7	-4.354	7th -FA6	-0.01	7th -FA9
FA8	-4.354	8th-FA9	-0.01	8th -FA7
FA9	-4.353	9th-FA7	-0.009	9th -FA8
FA10	-4.344	10th -FA8	-	-

https://doi.org/10.1371/journal.pone.0220290.t003

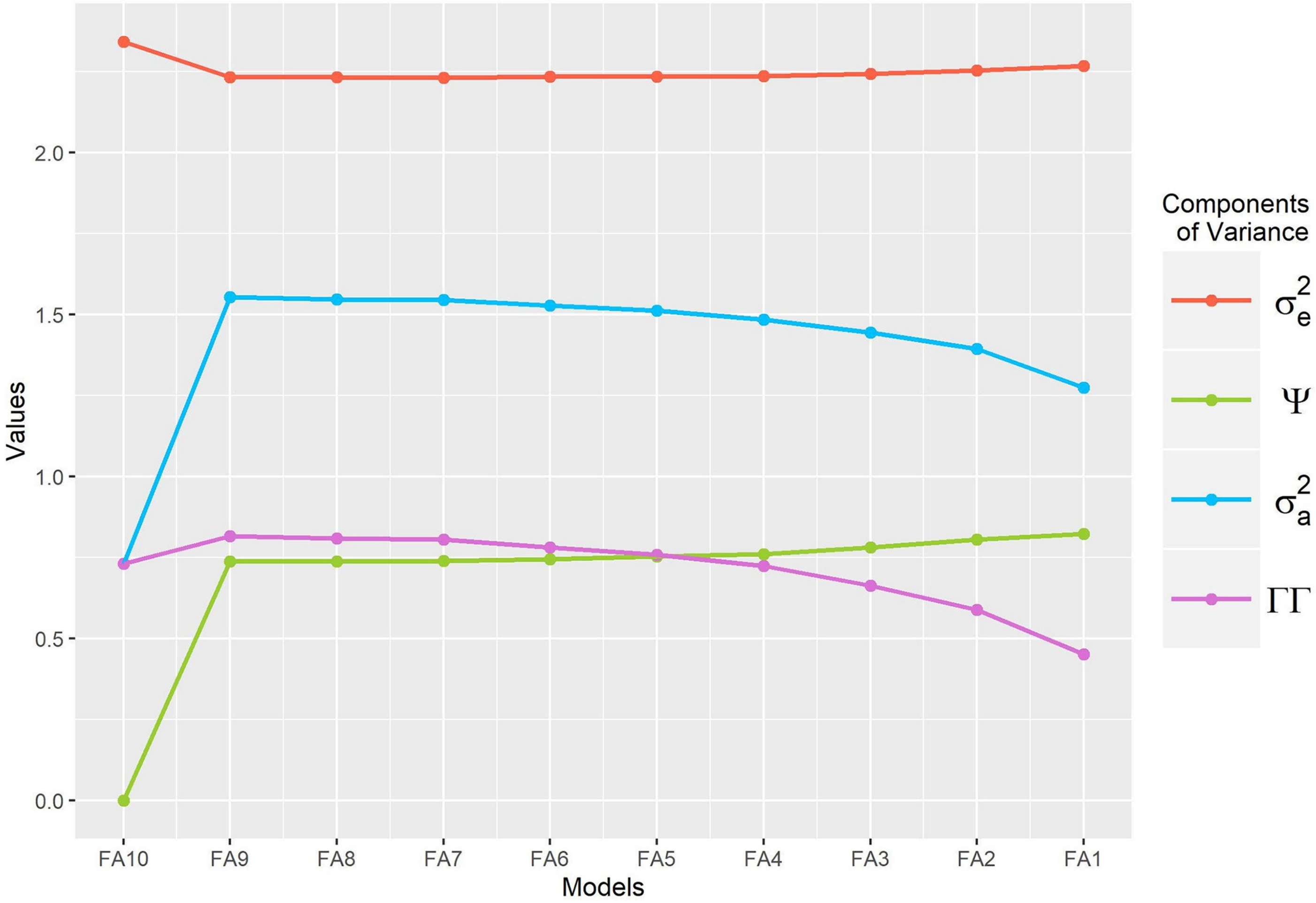

Fig 8. Residual variance (σ_e^2), specific (ψ) variance, loading matrix ($\mathbf{\Gamma\Gamma}$) and recovered genetic variance [$\sigma_g^2 = \sqrt[10]{(diag|(\mathbf{\Gamma\Gamma}^{\mathrm{T}})^{k-1} + \mathbf{\Psi}|)10}$] on different FA(k) structures for real data.

https://doi.org/10.1371/journal.pone.0220290.g008

This particularity ensures the model's identifiability and avoids the occurrence of estimates outside of the parametric space (Heywood cases), as observed in Smith, Cullis and Thompson [6]. In addition, it eliminates the need for loadings' rotationality, providing robust estimates for covariance parameters through loadings and factor scores. Thus, the BFA method guarantees the ability to test all latent factors that cannot be ensured by classical FA model. For example, it was observed that the FA analysis conducted in Asreml-R on real data did not converge on more complex FA models (FA$\geq$3) and a fair scanning for the best model was not possible using the real data. The selection of the number of *k* factors is still a non-trivial issue and inadequate choices can result in biased and unstable estimates of Σ and **R** [47,48].

In this study, we assumed independence among environments assigning the scaled-inverse chi-squared distributions for each diagonal element of the residual variance. However, if this assumption is relaxed, we can use an inverse Wishart prior for the residual covariance matrix (as in de Los Campos and Gianola [25]) or the sparse prior matrix for the loading matrices (as proposed by Runcie and Mukherjee [48]).

It is important to highlight that some problems related to improper posterior may emerge in hierarchical models when non informative priors are used. This issue was discussed in

Hobert and Casella [49] and Gelman [50]. While it is not ease checking all marginal posterior in complex models some tips can be observed during MCMC sampling such as non convergence or bimodal posterior presenting high mass on zero value.

In this study, we did not observed such problems as those verified in our previous work using Bayesian hierarchical AMMI model [20]. On the scenario presented by these authors, some *ad hoc* procedures were applied to obtain proper posteriors in family of inverted Gamma-Gaussian hierarchical models when improper Jeffrey's prior replaces the inverted Gamma distribution. Silva et al. [20] proposed the correction of degree of freedom based on Ter Braak [51] approach and suggested such correction when there is evidence of improper posterior.

The rotationally issue is not discussed in the study of de Los Campos and Gianola [25]. Runcie and Mukherjee [48] argue that the rotationality may be guaranteed by imposing constraints on the loading matrixin the prior specification (hierarchical modeling). As it is known, the rotationality of the loading matrix does not influence the estimation of model parameters. However, it may blur some biological interpretations [6] and hamper the MCMC's convergence [52].

Moreover, the parametric advantages of the BFA model in its efficiency in representing the FA pattern can be seen in Figs 6 and 7, where the BFA is compared with the FA-AI. The same FA pattern was expected, given the BFA'spriors based on the FA model's assumptions reported in Smith et al. [6]. Furthermore, in the study involving the simulated data, the estimates obtained for residual variances were close to the true values and the recovery of the first two factor loadings by the eigenvalues and eigenvectors (Table 1).

Other proposals for multi-environment data analysis can be found in the literature; some highlight the use of Bayesian AMMI or SREG models [17–19,23,34]. One limitation of these approaches with respect to BAF is to assume homogeneity of variances and other model assumptions based on ANOVA.

When the homogeneity of variances is assumed, the sampling of the eigenvectors is performed in hyperspherical space through the von Mises-Fisher distributions, as proposed by Liu [52] for the AMMI models. Here, given that different residual variances were assumed across the environments, it was not possible to approximate a von Mises-Fisher distribution as the posterior distribution; rather, the eigenvectors are sampled from a multivariate Gaussian distribution that provides hyper-ellipses instead of hyper-sphere over multivariate spaces. The eigenvectors were further placed in the correct subspace by orthogonal transformation, thus meeting the restrictions imposed by spectral decomposition.

Although the aim of our study is not to provide an intensive comparison of the predictive ability between the BAF and FA-based mixed models, the results showed that even using different likelihood approaches for FA mixed models (AI and EM), the BFA outperformed these models in most of the missing data scenarios (S1–S6 Figs). Given the intra-class correlation (or the average heritability) obtained for the simulated data, one could expect a maximum correlation between the observed and predicted values of 0.82; however, this threshold was exceeded in some folds (folds 3 and 5 in S1 Fig), showing that the model's accuracy may be higher than expected in some scenarios, especially when the group to be predicted is composed of stable genotypes.

The BFA was not superior in all *k*-fold scenarios. For instance, in some *k*-folds, we can verify a marginal loss of BFA with respect to FA-EM. This result was observed under low levels of missing data (10%). Additionally, the predictive ability of the model was similar for some folds, which were 33%, and 50% of data were missing. This result suggests that BFA might be inferior to FA-based mixed models under some data scenarios. As emphasized by Wolpert and Macready [53], the high performance of some algorithms in a class of problems is compensated by the low performance in another class (No Free Lunch Theorem).

The main drawback of the Bayesian approach is the high demand for computational studies of for the MCMC's convergence. However, this disadvantage can be offset by the greater model flexibility and guaranteed convergence in the parameter space. However, high-technology computers associated with optimized codes and parallel processing can alleviate this disadvantage of MCMCs.

The FA model proposed by Smith et al. [6] and improved by Thompson et al. [54] uses one-stage analysis. Meyer [16] describes this same model using a factor analysis approach that was implemented using a two-stage FA analysis (FA-EM) by Nuvunga et al. [40]. In this study, it was observed that the FA-EM was slightly superior to the classic FA. The relative gain of the FA-EM with respect to the FA-AI can be explained by the structure of the simulated data. According to Nuvunga et al. [40], the FA-EM is estimated in two stages. In the first stage, an unstructured covariance matrix (UN) is estimated by EM. In the second stage, the factor analysis is used to estimate the loading and FA scores using the varimax rotationally. In turn, the FA-AI estimates the covariance matrix approximating the FA loadings and scores using a single-stage analysis. Since few environments were simulated and given that the FA-EM is based on previous UN analysis (which tends to be the best choice for low-dimensional data), under this scenario, the FA based on two-stage analysis may have advantages for low-dimensional FA.

Crossa [55] note that the FA model can be interpreted in a similar way to the SREG model when the G effect is confounded with GEI or as in the AMMI model if the G effects are marginalized from GEI. Burgueño et al. [56] notes that there is no clear difference in the predictions between these two approaches. However, it should be noted that the FA models using the confounding of G+GE (as presented here) are more parsimonious than those FA models that marginalize the genotype effect from the GEI.

The bivariate credibility regions (at 95% probability) were incorporated into the FA biplots (Figs 1A, 1B, 5A and 5B), identifying homogeneous subgroups of genotypes and environments for adaptability and stability. Through the credibility regions, it is possible to identify which environments show greater variability or contribute to GEI. The interpretation of the BFA biplot is similar to that in classical factor analysis, although the uncertainty is considered in the BFA biplot. The joint decomposition of G+GE becomes the model's interpretation, similar to the GGE-biplot. However, it is noteworthy to highlight that this view must be assumed with caution, since the graphical representation must be performed separately for genotypes and environments, since their responses do not have the same scale or inner-product proprieties.

The selection of models for simulated data was performed by three complementary criteria: the PRESS criterion, the correlation between the observed and the predicted phenotypic value, and the statistical efficiency (SE). It was observed that the model chosen from the simulated data was FA(4). For the real data, we adopted a parametric criterion to select the best model, including the AICM and the ΔAICM information criteria [43]. As observed, the AICM was not informative enough to select the number of factors to be retained in the model, although it presented a tendency using the full model, which in theory recovers the UN matrix. The widely used FA(2) model did not present the best information criterion, even with it being very parsimonious, presenting graphical justifications and genetic interpretations [26,29,40]. The use of information criteria as a selection method for the number of factors has several ambiguities and is always subject to criticism since different criteria tend to select different models, as shown in Table 3 and in Smith et al. [28]. In addition, it can also be observed that the differentiation of models was small, which could be interpreted as a non-selection of models (Table 3).

In addition to information criteria, Smith et al. [28] emphasized the need to use additional measures in the model selection, such as the proportion of variance explained by the components. It was observed that as FA(*k*-1*)* model is adjusted, the residual variance decreases and the genetic variance increases, which is now explained by the loadings and specific variance (Fig 8). From the FA(*k*-2), (*k*-3), . . ., FA(1) model, it was verified that the specific and residual variances remain constant, showing that the application of information criteria as a model selection method is not easy t under the restrictions imposed in the BFA analysis.

Some other information criterion based on expectation of likelihood could be applied such us the averaged BIC, DIC, AEBIC and so on [57]. While these methods may present divergent behavior of AICM and the ΔAICM information criteria, we understand that, on the present scenario, the best Bayesian FA model selected by these information criterions could differ qualitatively from AICM (i.e. selecting the FA(1) model) but presenting low quantitative differences since the average marginal likelihood will be the same and the penalty criterion could not be large enough to efficiently separate the FA models within each information criterion. It worth to highlight that this occurrence is not a problem related to the information criterions, but a characteristic of our model since the restriction ($\mathbf{\Gamma\Gamma}^{T} = \mathbf{\Sigma}-\mathbf{\Psi}$) ensure equivalence among the marginal likelihoods across the FA models.

In general, the choice of the FA2 model has become consensual among researchers. It is argued that increasing the number of components (such as FA3, FA4 and FA5) does not guarantee better prediction ability, but it will certainly increase the model's complexity; therefore, it is doubtful that a better adjustment will be produced, as observed in our study with real and simulated data. The results obtained by Burgueño et al. [26] Burgueño et al. [7], Kelly et al. [9] and Nuvunga et al. [40] show that FA models with more than two components improved variance-covariance estimates, but this was not reflected in the genotypic predicted values (EBLUPs).

In this study, it was not our intention to give a fine perspective of model selection in FA bayesian framework; instead, our aim was to provide a bayesian perspective of FA models in MET analysis. While the model selection is an open issue in FA models, the cross-validation approaches used here, in general, were more informative to select FA scores than AIC criterion. Others methods based on transdimentional models such as reversible jump could be proposed in order to select FA scores [58]; but, until now, results in this area are scarce in Bayesian context.

Although both data sets used in this study may be relatively small when compared to the data available from experimental trials in breeding programs, it is sufficient to show the strength of the BFA method in MET analysis, presenting a better predictive ability than the classical FA based on mixed models. Given that in the MET analysis some environments are not present in the trial network and further years are unpredictable, some functional information may be included in the FA analysis to predict the genotypic values for coming years; for example, we can take the covariance matrix as a functional response surface related to some distance measures in the Hilbert space.

Our results demonstrate that the Bayesian FA model can be effectively implemented to study GEI patterns in MET networks and predict missing data with high levels of imbalance.

Supporting information

S1 Text. Justification of bayesian factor analytical model.
(DOCX)

S2 Text. Likelihood justification.
(DOCX)

S3 Text. Gibbs sampler algorithm.
(DOCX)

S4 Text. Traces and densities for the MCMC chains for some parameters of model.
(DOCX)

S1 Table. Description of the environments where the experiments were conducted.
(DOCX)

S2 Table. Posterior means (PM), regions of credibility (95%. LL: Lower limit. UL: Upper limit) and estimates of restricted maximum likelihood (REML) of FA-AI and genotypic scores (f_{i1}–f_{i2}), simulated data.
(DOCX)

S3 Table. Posterior means (PM), regions of credibility (95%. LL: Lower limit. UL: Upper limit) for the first two factor loadings ($\lambda_1\alpha_1$– $\lambda_2\alpha_2$), real data.
(DOCX)

S4 Table. Posterior means (PM), regions of credibility (95%. LL: Lower limit. UL: Upper limit) for the first two factor scores (f_{i1}–f_{i2}), real data.
(DOCX)

S1 Fig. Bar chart representing the correlation for the 10% loss level using the Bayesian FA (BAF) and FA (FA-EM and FA-AI) models for simulated data.
(JPEG)

S2 Fig. Bar chart representing the correlation for the 33% loss level using the Bayesian FA (BAF) and FA (FA-EM and FA-AI) models for simulated data.
(JPEG)

S3 Fig. Bar chart representing the correlation for the 50% loss level using the Bayesian FA (BAF) and FA (FA-EM and FA-AI) models for simulated data.
(JPEG)

S4 Fig. Bar chart representing the PRESS for the 10% loss level using the Bayesian FA (BAF) and FA (FA-EM and FA-AI) models for simulated data.
(JPEG)

S5 Fig. Bar chart representing the PRESS for the 33% loss level using the Bayesian FA (BAF) and FA (FA-EM and FA-AI) models for simulated data.
(JPEG)

S6 Fig. Bar chart representing the PRESS for the 50% loss level using the Bayesian FA (BAF) and FA (FA-EM and FA-AI) models for simulated data.
(JPEG)

S1 Data. Simulated data.
(RAR)

S2 Data. Real data.
(RAR)

S3 Data. R BFA package.
(ZIP)

Acknowledgments

The authors would like to thank the Fundação Calouste Gulbenkian, Fundo Nacional de Investigação de Moçambique, Escola Superior de Negócios e Empreדedoriso de Chibuto, Chibuto, Universidade Eduardo Mondlane, CAPES (Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior) and CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico–Grant # 302674/2015-2) for supporting this research.

Author Contributions

Conceptualization: Renato Ribeiro de Lima, Marcio Balestre.

Data curation: Marcio Balestre.

Formal analysis: Joel Jorge Nuvunga, Carlos Pereira da Silva, Marcio Balestre.

Methodology: Luciano Antonio de Oliveira, Marcio Balestre.

Software: Joel Jorge Nuvunga, Marcio Balestre.

Supervision: Renato Ribeiro de Lima, Marcio Balestre.

Writing – original draft: Joel Jorge Nuvunga, Carlos Pereira da Silva, Luciano Antonio de Oliveira.

Writing – review & editing: Marcio Balestre.

References

1. Cornelius PL, Seyedsadr MS. Estimation of general linear-bilinear models for two-way tables. J Stat Comput Simul. 1997;
2. Crossa J. Statistical Analyses of Multilocation Trials. Adv Agron. 1990;
3. Yan W, Hunt L a., Sheng Q, Szlavnics Z. Cultivar Evaluation and Mega-Environment Investigation Based on the GGE Biplot. Crop Sci. 2000;
4. Piepho HP. Analyzing genotype-environment data by mixed models with multiplicative terms. Biometrics. 1997;
5. Piepho HP. Empirical best linear unbiased prediction in cultivar trials using factor-analytic variance-covariance structures. Theor Appl Genet. 1998;
6. Smith A, Cullis B, Thompson R. Analyzing Variety by Environment Data Using Multiplicative Mixed Models and Adjustments for Spatial Field Trend. Biometrics. 2001;
7. Burgueño J, Crossa J, Cornelius PL, Yang R-C. Using Factor Analytic Models for Joining Environments and Genotypes without Crossover Genotype × Environment Interaction. Crop Sci [Internet]. 2008; 48 (4):1291. Available from: https://www.crops.org/publications/cs/abstracts/48/4/1291
8. Crossa J, Burgueño J, Cornelius PL, McLaren G, Trethowan R, Anitha K. Modeling genotype x environment interaction using additive genetic covariances of relatives for predicting breeding values of wheat genotypes. Crop Sci. 2006;
9. Kelly AM, Smith AB, Eccleston JA, Cullis BR. The accuracy of varietal selection using factor analytic models for multi-environment plant breeding trials. Crop Sci. 2007;
10. Smith AB, Cullis BR, Thompson R. Exploring Variety–Environment Data Using Random Effects AMMI Models with Adjustments for Spatial Field Trend: Part 1: Theory. Quant Genet genomics plant breeding. 2002;
11. Hill WG. On selection among groups with heterogeneous variance. Anim Prod. 1984;
12. Rönnegård L, Felleki M, Fikse F, Mulder HA, Strandberg E. Genetic heterogeneity of residual variance—estimation of variance components using double hierarchical generalized linear models. Genet Sel Evol. 2010;
13. Edwards JW, Jannink J-L. Bayesian Modeling of Heterogeneous Error and Genotype × Environment Interaction Variances. Crop Sci. 2006;
14. Hogan JW, Tchernis R. Bayesian Factor Analysis for Spatially Correlated Data, With Application to Summarizing Area-Level Material Deprivation From Census Data. J Am Stat Assoc. 2004;

15. Mardia K V, Kent JT, Bibby JM. Multivariate Analysis. Multivar Anal London Acad Press. 1979;

16. Meyer K. Factor-analytic models for genotype × environment type problems and structured covariance matrices. Genet Sel Evol. 2009;

17. Crossa J, Perez-Elizalde S, Jarquin D, Cotes JM, Viele K, Liu G, et al. Bayesian estimation of the additive main effects and multiplicative interaction model. Crop Sci. 2011;

18. De Oliveira LA, Da Silva CP, Nuvunga JJ, Da Silva AQ, Balestre M. Credible intervals for scores in the AMMI with random effects for genotype. Crop Sci. 2015;

19. Perez-Elizalde S, Jarquin D, Crossa J. A General Bayesian Estimation Method of Linear-Bilinear Models Applied to Plant Breeding Trials With Genotype × Environment Interaction. J Agric Biol Environ Stat. 2012;

20. Da Silva CP, De Oliveira LA, Nuvunga JJ, Pamplona AKA, Balestre M. A Bayesian Shrinkage approach for AMMI models. PLoS One. 2015;

21. Yan W, Glover KD, Kang MS. Comment on biplot analysis of genotype x environment interaction: Proceed with caution, by r.-c. yang, j. crossa, p.l. cornelius, and j. burgueno in crop science 2009 49:1564–1576. Crop Science. 2010.

22. Yang RC, Crossa J, Cornelius PL, Burgueño J. Biplot analysis of genotype × environment interaction: Proceed with caution. Crop Sci. 2009;

23. Jarquín D, Pérez-Elizalde S, Burgueño J, Crossa J. A hierarchical Bayesian estimation model for multienvironment plant breeding trials in successive years. Crop Sci. 2016; https://doi.org/10.2135/cropsci2015.07.0410 PMID: 27814405

24. Geweke J, Zhou G. Measuring the pricing error of the arbitrage pricing theory. Review of Financial Studies. 1996.

25. de los Campos G, Gianola D. Factor analysis models for structuring covariance matrices of additive genetic effects: a Bayesian implementation. Genet Sel Evol. 2007;

26. Burgueño J, Crossa J, Cornelius PL, Trethowan R, McLaren G, Krishnamachari A. Modeling additive ?? environment and additive ?? additive ?? environment using genetic covariances of relatives of wheat genotypes. Crop Sci. 2007;

27. SMITH AB, CULLIS BR, THOMPSON R. The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. J Agric Sci. 2005;

28. Smith AB, Ganesalingam A, Kuchel H, Cullis BR. Factor analytic mixed models for the provision of grower information from national crop variety testing programs. Theor Appl Genet. 2015;

29. Stefanova KT, Buirchell B. Multiplicative mixed models for genetic gain assessment in lupin breeding. Crop Sci. 2010;

30. Tyrisevä A-M, Meyer K, Fikse WF, Ducrocq V, Jakobsen J, Lidauer MH, et al. Principal component approach in variance component estimation for international sire evaluation. Genet Sel Evol. 2011;

31. Melo WMC, Pinho RG Von, Balestre M. Prediction of maize single cross hybrids using the total effects of associated markers approach assessed by cross-validation and regional trials. Sci World J. 2014;

32. Songgui W, Suju Y. A new estimate of the parameters in linear mixed models. Sci China Ser A Math [Internet]. 2002; 45(10):1301–11. Available from: https://doi.org/10.1360/02ys9140

33. Viele K, Srinivasan C. Parsimonious estimation of multiplicative interaction in analysis of variance using Kullback–Leibler Information. J Stat Plan Inference [Internet]. 2000 Mar; 84(1–2):201–19. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0378375899001512

34. de Oliveira LA, da Silva CP, Nuvunga JJ, da Silva AQ, Balestre M. Bayesian GGE biplot models applied to maize multi-environments trials. Genet Mol Res. 2016;

35. Raftery AE, Lewis S. How many iterations in the Gibbs sampler? Bayesian Stat. 1992;

36. Heidelberger P, Welch PD. Simulation Run Length Control in the Presence of an Initial Transient. Oper Res. 1983;

37. Team RC. R Core Team R. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org. 2016.

38. Chen MH, Shao QM. Monte carlo estimation of bayesian credible and hpd intervals? J Comput Graph Stat. 1999;

39. Ooms JCL. The highest posterior density posterior prior for bayesian model selection. 2009.

40. Nuvunga JJ, Oliveira LA, Pamplona AKA, Silva CP, Lima RR, Balestre M. Factor analysis using mixed models of multi-environment trials with different levels of unbalancing. Genet Mol Res. 2015;

41. Butler DG, Cullis BR, Gilmour AR, Gogel BJ. Mixed models for S language environments: ASReml-R reference manual (version 3). Queensl Dep Prim Ind Fish. 2009;

42. Gabriel KR. Le biplot-outil d'exploration de données multidimensionelles. J la Soc Fr Stat. 2002; 143(3–4):5–55.

43. Raftery AE, Newton MA, Satagopan JM, Krivitsky PN. Estimating the Integrated Likelihood via Posterior Simulation Using the Harmonic Mean Identity. Bayesian Stat. 2007;

44. Akaike H. Information theory and an extensión of the maximum likelihood principle. Int Symp Inf theory. 1973;

45. Lado B, Barrios PG, Quincke M, Silva P, Gutiérrez L. Modeling Genotype × Environment Interaction for Genomic Selection with Unbalanced Data from a Wheat Breeding Program. Crop Sci [Internet]. 2016; 56(5):2165. Available from: https://dl.sciencesocieties.org/publications/cs/abstracts/56/5/2165

46. Meyer K, Kirkpatrick M. Perils of parsimony: Properties of reduced-rank estimates of genetic covariance matrices. Genetics. 2008.

47. Runcie DE, Mukherjee S. Dissecting high-dimensional phenotypes with Bayesian sparse factor analysis of genetic covariance matrices. Genetics. 2013;

48. Huang E, Cheng SH, Dressman H, Pittman J, Tsou MH, Horng CF, et al. Gene expression predictors of breast cancer outcomes. Lance. 2003;

49. Hobert JP, Casella G. The Effect of Improper Priors on Gibbs Sampling in Hierarchical Linear Mixed Models. J Am Stat Assoc. 1996;

50. Gelman A. Prior distributions for variance parameters in hierarchical models (Comment on Article by Browne and Draper). Bayesian Analysis. 2006.

51. Ter Braak CJF, Boer MP, Bink MCAM. Extending Xu's Bayesian model for estimating polygenic effects using markers of the entire genome. Genetics. 2005;

52. Liu G. Bayesian computations for general linear-bilinear models. University of Kentucky; 2001.

53. Wolpert DH, Macready WG. No free lunch theorems for optimization. IEEE Trans Evol Comput. 1997;

54. Thompson R, Cullis B, Smith A, Gilmour A. A Sparse Implementation of the Average Information Algorithm for Factor Analytic and Reduced Rank Variance Models. Aust <html_ent glyph = "@amp;" ascii = "&"/> New Zeal J Stat. 2003;

55. Crossa J. From genotype × environment interaction to gene × environment interaction. Curr Genomics. 2012;

56. Burgueno J, de los Campos G, Weigel K, Crossa J. Genomic prediction of breeding values when modeling genotype × environment interaction using pedigree and dense molecular markers. Crop Sci. 2012; PMID: 26924849

57. Xue J, Luo Y, Liang F. Average (E)BIC-like Criteria for Bayesian Model Selection. Work Pap [Internet]. 2017;(Mcmc). Available from: https://people.clas.ufl.edu/yeluo/files/ave4.pdf

58. Green PJ. Reversible jump Markov chain monte carlo computation and Bayesian model determination. Biometrika. 1995;